



# The Palgrave Handbook of Digital Russia Studies

*Edited by* Daria Gritsenko  
Mariëlle Wijermars · Mikhail Kopotev

OPEN ACCESS

palgrave  
macmillan

# The Palgrave Handbook of Digital Russia Studies

Daria Gritsenko  
Mariëlle Wijermars • Mikhail Kopotev  
Editors

# The Palgrave Handbook of Digital Russia Studies

palgrave  
macmillan

*Editors*

Daria Gritsenko  
University of Helsinki  
Helsinki, Finland

Mariëlle Wijermars  
Maastricht University  
Maastricht, The Netherlands

Mikhail Kopotev  
Higher School of Economics  
(HSE University)  
Saint Petersburg, Russia



ISBN 978-3-030-42854-9      ISBN 978-3-030-42855-6 (eBook)  
<https://doi.org/10.1007/978-3-030-42855-6>

© The Editor(s) (if applicable) and The Author(s) 2021, corrected publication 2021. This book is an open access publication.

**Open Access** This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Cover illustration: FrankRamspott / gettyimages  
Cover design: eStudioCalamar

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG.

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland



The original version of this book was revised. The affiliation for the authors Ekaterina Artemova, Anastasia Bonch-Osmolovskaya, Frank Fischer, Arto Mustajoki, Polina Kolozaridi and Daniil Skorinkin have now been updated. The correction to the book is available at [https://doi.org/10.1007/978-3-030-42855-6\\_33](https://doi.org/10.1007/978-3-030-42855-6_33).

## PREFACE

This Handbook emerged out of the Digital Russia Studies (DRS) initiative,<sup>1</sup> launched by Daria Gritsenko and Mariëlle Wijermars at the University of Helsinki's Aleksanteri Institute and Helsinki Center for Digital Humanities (HELDIG) in January 2018. The aim of the DRS initiative was to unite scholars of the humanities and the social and computer sciences working at the intersection of “digital” and “social” in the Russian context. By providing a regular meeting place and networking opportunities, we sought to establish open discussion and knowledge sharing among those who study the various aspects of digitalization processes in Russia and those studying Russia with the use of (innovative) digital methods. The many positive responses to our interdisciplinary approach and the exciting research that is currently conducted in this area of study inspired us to join forces with Mikhail Kopotev to compile this Handbook.

The editors would like to thank Lucy Batrouney and Mala Sanghera-Warren, our commissioning editors at Palgrave Macmillan, for their enthusiasm for the project as well as the anonymous reviewers for their critical eye. We thank the Faculty of Arts of the University of Helsinki for making it possible to publish the Handbook in Open Access. We are particularly grateful to Aleksandr Klimov, our research assistant, who was of great help in preparing the manuscript for publication.

The interdisciplinarity of the Handbook has affected our choice concerning the transliteration of Russian. While it is customary for scholars working in the humanities and social sciences to apply the Library of Congress system of transliteration, for scholars in linguistics and computer science a different system, ISO 9, is more appropriate. For consistency, we have chosen to follow the

Library of Congress system for references (authors' names) and ISO 9 for all other Russian terms and names throughout the book. Where appropriate, customary English spellings are maintained for familiar terms, places, and personal names.

Helsinki, Finland  
Maastricht, The Netherlands  
Saint Petersburg, Russia

Daria Gritsenko  
Mariëlle Wijermars  
Mikhail Kopotev

#### NOTE

1. <https://blogs.helsinki.fi/digital-russia-studies/>.

# CONTENTS

<b>1</b>	<b>Digital Russia Studies: An Introduction</b>	<b>1</b>
	Daria Gritsenko, Mikhail Kopotev, and Mariëlle Wijermars	
<b>Part I</b>	<b>Studying Digital Russia</b>	<b>13</b>
<b>2</b>	<b>The Digitalization of Russian Politics and Political Participation</b>	<b>15</b>
	Mariëlle Wijermars	
<b>3</b>	<b>E-Government in Russia: Plans, Reality, and Future Outlook</b>	<b>33</b>
	Daria Gritsenko and Mikhail Zherebtsov	
<b>4</b>	<b>Russia's Digital Economy Program: An Effective Strategy for Digital Transformation?</b>	<b>53</b>
	Anna Lowry	
<b>5</b>	<b>Law and Digitization in Russia</b>	<b>77</b>
	Marianna Muravyeva and Alexander Gurkov	
<b>6</b>	<b>Personal Data Protection in Russia</b>	<b>95</b>
	Alexander Gurkov	
<b>7</b>	<b>Cybercrime and Punishment: Security, Information War, and the Future of Runet</b>	<b>115</b>
	Elizaveta Gaufman	

<b>8</b>	<b>Digital Activism in Russia: The Evolution and Forms of Online Participation in an Authoritarian State</b>	135
	Markku Lonkila, Larisa Shpakovskaya, and Philip Torchinsky	
<b>9</b>	<b>Digital Journalism: Toward a Theory of Journalistic Practice in the Twenty-First Century</b>	155
	Vlad Strukov	
<b>10</b>	<b>Digitalization of Russian Education: Changing Actors and Spaces of Governance</b>	171
	Nelli Piattoeva and Galina Gurova	
<b>11</b>	<b>Digitalization of Religion in Russia: Adjusting Preaching to New Formats, Channels and Platforms</b>	187
	Victor Khroul	
<b>12</b>	<b>Doing Gender Online: Digital Spaces for Identity Politics</b>	205
	Olga Andreevskikh and Marianna Muravyeva	
<b>13</b>	<b>Digitalization of Consumption in Russia: Online Platforms, Regulations and Consumer Behavior</b>	221
	Olga Gurova and Daria Morozova	
<b>14</b>	<b>Digital Art: A Sourcebook of Ideas for Conceptualizing New Practices, Networks and Modes of Self-Expression</b>	241
	Vlad Strukov	
<b>15</b>	<b>From Samizdat to New Sincerity. Digital Literature on the Russian-Language Internet</b>	255
	Henrike Schmidt	
<b>16</b>	<b>Run Runet Runaway: The Transformation of the Russian Internet as a Cultural-Historical Object</b>	277
	Gregory Asmolov and Polina Kolozaridi	

<b>Part II Digital Sources and Methods</b>	297
<b>17 Corpora in Text-Based Russian Studies</b> Mikhail Kopotey, Arto Mustajoki, and Anastasia Bonch-Osmolovskaya	299
<b>18 RuThes Thesaurus for Natural Language Processing</b> Natalia Loukachevitch and Boris Dobrov	319
<b>19 Social Media-based Research of Interpersonal and Group Communication in Russia</b> Olessia Koltsova, Alexander Porshnev, and Yadviga Sinyavskaya	335
<b>20 Digitizing Archives in Russia: Epistemic Sovereignty and Its Challenges in the Digital Age</b> Alexey Golubev	353
<b>21 Affordances of Digital Archives: The Case of the <i>Prozbito</i> Archive of Personal Diaries</b> Ekaterina Kalinina	371
<b>22 Open Government Data in Russia</b> Olga Parkhimovich and Daria Gritsenko	389
<b>23 Topic Modeling in Russia: Current Approaches and Issues in Methodology</b> Svetlana S. Bodrunova	409
<b>24 Topic Modeling Russian History</b> Mila Oiva	427
<b>25 Studying Ideational Change in Russian Politics with Topic Models and Word Embeddings</b> Andrey Indukaev	443

<b>26</b>	<b>Deep Learning for the Russian Language</b>	<b>465</b>
	Ekaterina Artemova	
<b>27</b>	<b>Shifting the Norm: The Case of Academic Plagiarism Detection</b>	<b>483</b>
	Mikhail Kopotev, Andrey Rostovtsev, and Mikhail Sokolov	
<b>28</b>	<b>Automatic Sentiment Analysis of Texts: The Case of Russian</b>	<b>501</b>
	Natalia Loukachevitch	
<b>29</b>	<b>Social Network Analysis in Russian Literary Studies</b>	<b>517</b>
	Frank Fischer and Daniil Skorinkin	
<b>30</b>	<b>Tweeting Russian Politics: Studying Online Political Dynamics</b>	<b>537</b>
	Mikhail Zherebtsov and Sergei Goussev	
<b>31</b>	<b>The State of the Art: Surveying Digital Russian Art History</b>	<b>569</b>
	Reeta E. Kangas	
<b>32</b>	<b>Geospatial Data Analysis in Russia's Geoweb</b>	<b>585</b>
	Mykola Makhortykh	
	<b>Correction to: The Palgrave Handbook of Digital Russia Studies</b>	<b>C1</b>
	Daria Gritsenko, Mariëlle Wijermars, and Mikhail Kopotev	
	<b>Index</b>	<b>605</b>

## NOTES ON CONTRIBUTORS

**Olga Andreevskikh** is a PhD candidate at the University of Leeds, UK. Her PhD thesis focuses on the verbal and visual communication of non-heteronormative masculinities in contemporary Russian media. Her other research projects focus on offline and online activism for women's and LGBTQ rights in contemporary Russia.

**Ekaterina Artemova** is a researcher at the National Researcher University Higher School of Economics (HSE University), Moscow, Russia. Her research interests include natural language processing in general as well as applications of deep learning to information extraction and question answering.

**Gregory Asmolov** is Leverhulme Early Career Fellow at the Russia Institute, King's College London. His research focuses on how ICTs constitute the role of crowds in crisis situations. He has published in *Journalism Studies*, *Policy & Internet*, *Russian Politics*, MIT's *Journal of Design and Science*, and others on vertical crowdsourcing, digital propaganda, and Internet regulation.

**Svetlana S. Bodrunova** is a professor at the School of Journalism and Mass Communications, St. Petersburg State University, Russia, where she also leads the Center for International Media Research. She has published two books, several chapters, and over 80 research papers in Russian and English, including in *Journalism*, *International Journal of Communication*, *Media and Communication*, and *Digital Journalism*. Her research interests include Russian and European journalism, media and politics, social media, and ethnicity in communication.

**Anastasia Bonch-Osmolovskaya** is an associate professor at the School of Linguistics, National Research University Higher School of Economics (HSE University), Moscow, Russia. She is also an academic supervisor of the master's program in computational linguistics and the founder of the Digital Humanities Center at HSE. She is a member of the Russian Association for Digital Humanities (DH-Russia). Her main research interests concern digital archives, linguistic and



other textual corpora, and the implementation of quantitative methods in the studies of language, literature, and culture.

**Boris Dobrov** is the head of the Laboratory in Research Computing Center of Lomonosov Moscow State University in Moscow, Russia. He leads the development of software tools for processing and analyzing large text collections, including RuThes-like large linguistic ontologies, ALOT (Automated Linguistic Text Processing) tools, and NearIdx corporate information-analytical system.

**Frank Fischer** is Associate Professor for Digital Humanities at the Higher School of Economics (HSE University), Moscow, Russia, and director of DARIAH-EU, the pan-European digital infrastructure for the arts and humanities. He studied computer science, German literature, and Spanish philology in Leipzig and London and is an Ancien Pensionnaire de l'École Normale Supérieure in Paris. He holds a PhD from the University of Jena for a study of revenge in Enlightenment drama.

**Elizaveta Gaufrman** is Assistant Professor of Russian Discourse and Politics at the University of Groningen, the Netherlands. She is the author of *Security Threats and Public Perception: Digital Russia and the Ukraine Crisis* (Palgrave Macmillan, 2017). Her other publications include peer-reviewed articles on nationalism, sexuality, and social networks, as well as regular blog posts at *The Duck of Minerva*.

**Alexey Golubev** is Assistant Professor of Russian History and Digital Humanities at the University of Houston. He holds a PhD degree in history from the University of British Columbia (2016). Before starting his current position, he was Banting Postdoctoral Fellow at the University of Toronto.

**Sergei Goussev** is an independent scholar whose research focuses on computational propaganda, social media analytics, and political sociology of digital social networks. He holds a master's degree in financial economics from Carleton University, Canada.

**Daria Gritsenko** is an assistant professor at the University of Helsinki, Finland, affiliated with the Aleksanteri Institute and the Helsinki Center for Digital Humanities (HELDIG). She is a co-founder of Digital Russia Studies, an interdisciplinary network of scholars working at the intersection of “digital” and “social” in Russia and beyond.

**Alexander Gurkov** is a postdoctoral researcher at the University of Helsinki, Faculty of Law. He researches international cooperative law, alternative dispute resolution, and blockchain economy. Before working in academia, Gurkov practiced law as an attorney in St. Petersburg.

**Galina Gurova** is a researcher at the SKOLKOVO Education Development Centre with expertise in the field of school education. She currently pursues a PhD degree at the Doctoral Programme of Education and Society at the University of Tampere, Finland, and is a member of the EduKnow—Knowledge, Power, and Politics in Education research group at the same university.

**Olga Gurova** is Assistant Professor of Consumption Studies at Aalborg University, Denmark. Her research interests include consumption studies, social theory, and qualitative methods of social research. She is the author of *Fashion and the Consumer Revolution in Contemporary Russia* (2015) and several articles, published in the *Journal of Consumer Culture*, *Consumption, Markets and Culture*, and *Cultural Studies*, among other journals.

**Andrey Indukaev** is a postdoctoral researcher at the Aleksanteri Institute of the University of Helsinki. He is a member of the Digital Russia Studies research group, working on the politics of digitalization in Russia and developing digital methods of textual analysis. His doctoral dissertation, defended in 2018 (ENS Paris-Saclay, France), focused on innovation policy and academic entrepreneurship in Russia.

**Ekaterina Kalinina** is a senior lecturer at the Department of Media and Communication Studies of Jönköping University, Sweden. She worked as a research fellow at Swedish National Defense University, researching questions of Russian patriotism and biopolitics. Her recent research project investigated the role of affective mnemonic experiences in triggering social mobilization. Kalinina also runs the Swedish NGO Nordkonst, where she manages cultural projects.

**Reeta E. Kangas** is a postdoctoral researcher in the School of Art History at the University of Turku. Her research focuses on Soviet and Russian visual art and propaganda and the use of animal symbols for political purposes. In 2017, she completed a PhD on animal symbolism in Soviet political cartoons.

**Victor Khroul** is an associate professor at the Department of Sociology of Mass Communications, Faculty of Journalism, Lomonosov Moscow State University. He is the author of *Media and Religion in Russia* (2012) and over 90 publications in Russian and English. He is the founding editor of the “Media and Religion” book series (published since 2011).

**Polina Kolozaridi** is a social researcher focusing on the Internet. Her key research interests are around the issues of how we know, feel, anticipate, and imagine technologies. In Moscow, Kolozaridi coordinates a grassroots community of researchers (academic and independent) called the Club for Internet and Society Enthusiasts (<http://clubforinternet.net/>). Together with other club members, she organizes research projects and acts as a knowledge activist. She teaches courses about Internet studies, amateur online media, and critical data studies at the Higher School of Economics (HSE University), Moscow, Russia.

**Olessia Koltsova** is an associate professor at the Department of Sociology of the St. Petersburg School of Social Sciences and Humanities, Higher School of Economics (HSE University), Saint Petersburg, Russia. She is the director of the Laboratory for Internet Studies (LINIS) at HSE. Her research interests include sociology of the Internet and of mass communication, political communication, online interpersonal communication, online social networks, text mining, big data analysis, and modeling.

**Mikhail Kopotev** is the academic director of the MA program in Language Technology at Higher School of Economics (HSE University), Saint Petersburg, Russia, and an associate professor at the University of Helsinki, Finland. His research interests include corpus linguistics, quantitative analysis of big textual data, plagiarism detection, and computer-assisted language learning. He is the author of *Introduction to Corpus Linguistics* (Praha, 2014) and a co-editor of *Quantitative Approaches to the Russian Language* (Routledge, 2018).

**Markku Lonkila** works as Professor of Sociology at the University of Jyväskylä. His research concerns the study of social movements, social media, social networks, and Russian society. Since the early 1990s, Lonkila has carried out several empirical research projects in Russia, investigating the role of social networks in civil society and economy and politics in Russia, often in a comparative perspective. Of late, he has focused on new forms of Russian civic and political activism enabled by social media.

**Natalia Loukachevitch** is a leading researcher at the Research Computing Center of Lomonosov Moscow State University, Russia. She is the main author of several large Russian resources for natural language processing including RuThes thesaurus, Russian wordnet RuWordNet, the ontology on natural science and technologies OENT, and Russian sentiment lexicon RuSentiLex. She is the co-organizer of several Russian language evaluations such as sentiment analysis evaluations SentiRuEval and semantic similarity evaluations RUSSE.

**Anna Lowry** is a postdoctoral researcher at the Aleksanteri Institute of the University of Helsinki. She has published on a range of topics dealing with the political economy of Russia and Eurasia. Recent publications have focused on Russia's development strategy and industrial policy.

**Mykola Makhortykh** is a postdoctoral researcher at the University of Bern. He holds a PhD from the University of Amsterdam, where he studied digital remembrance of WWII in Eastern Europe. Recently, he published on (counter)memory, discursive construction of (in)security, and conflict framing in articles published in journals such as *Media, War and Conflict*, *Memory Studies*, and *Visual Communication*.

**Daria Morozova** is a PhD student at Aalborg University, Denmark. She holds a master's degree in social sciences from the University of Helsinki. Her research focuses on sociology of consumption, cultural entrepreneurship, and

sustainability. Her current project concerns wearable technology, particularly how it changes daily practices of consumers.

**Marianna Muravyeva** is Professor of Russian Law and Administration at the University of Helsinki. Her research is interdisciplinary, bringing together history, social sciences, and law to examine long-term trends and patterns in social development with a special focus on normativity, gender, and violence.

**Arto Mustajoki** is professor emeritus at the University of Helsinki. He is a leading research fellow at the National Research University Higher School of Economics (HSE University), Moscow, Russia.

**Mila Oiva** is a postdoctoral researcher at the University of Turku, Finland. She is an expert on Russian and Polish history with a particular interest in the transnational transfer of information and computational research methods. Her work has been published by *Media History*, among others.

**Olga Parkhimovich** is affiliated with the Faculty of Software Engineering and Computer Systems at ITMO University (ITMO: Information Technologies, Mechanics and Optics). She is a project manager at the NGO Information Culture, the team leader of [spending.gov.ru](http://spending.gov.ru), and a member of the Public Council at the Federal Treasury of the Russian Federation.

**Nelli Piattoeva** is University Lecturer in Education Sciences at the University of Tampere, Finland, and Adjunct Professor of International and Comparative Education Policy Research at the University of Oulu, Finland. Her research interest in the institutions of formal education is mainly focused on the complexities of the relationship between education policy-making and the wider society.

**Alexander Porshnev** is a senior research fellow at the Laboratory of Internet Studies of the National Research University Higher School of Economics (HSE University), Saint Petersburg, Russia. His research interests include mathematical models and data mining in psychology, marketing and management (including social network analysis, cluster, factor, regression analysis, and machine learning methods), and natural language processing.

**Andrey Rostovtsev** is a professor at the Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute) in Moscow, Russia. He is a co-founder of the Dissernet network project.

**Henrike Schmidt** is a private lecturer at the Szondi-Institute for Comparative Literature, Freie Universität Berlin. Her research interests include digital culture in Eastern and Central Europe. Her publications include the monograph *Russian Literature on the Internet. Between Digital Folklore and Political*

*Propaganda* (Transcript Verlag, 2011, in German). She is a co-editor of the online journal *Studies in Russian, Eurasian and Central European New Media* ([digitalicons.org](http://digitalicons.org)).

**Larisa Shpakovskaya** holds a PhD in sociology and works on several research projects at the University of Helsinki, Finland. Her research interests include Internet communities on the Russian Internet, hate speech in the Russian-language segment of the Internet, state regulation of Runet, and processes of accumulating social capital in virtual communities and its conversion into a resource for mobilizing collective actions.

**Yadviga Sinyavskaya** is a junior research fellow at the Laboratory for Internet Studies (LINIS) and a PhD student at the Department of Sociology, St. Petersburg School of Social Sciences and Humanities, Higher School of Economics (HSE University), Saint Petersburg, Russia. Her research interests include Internet studies, social psychology, online communication and online behavior, online privacy, social network analysis, and online experiments.

**Daniil Skorinkin** is the head of the MA in digital humanities program at the Higher School of Economics (HSE University), Moscow, Russia. He teaches digital humanities and digital literacy courses and conducts research at the HSE Centre for Digital Humanities. His research interests include stylometry, network analysis of literary texts, and the use of computational methods for literary research.

**Mikhail Sokolov** is Professor of Sociology at the European University at St. Petersburg, Russia. His principal research interests are in the fields of micro-sociological theory, social stratification, cultural consumption, sociology, and history of the social sciences.

**Vlad Strukov** is an associate professor at the University of Leeds, specializing in Russian and Russophone media and communication with special focus on visual culture and digital culture. He is the founding and principal editor of a research journal called *Studies in Russian, Eurasian and Central European New Media* ([www.digitalicons.org](http://www.digitalicons.org)).

**Philip Torchinsky** is an independent scholar who has specialized in computer networks from 1993 onward. His current interests in social media include technology marketing and messenger bots. He has co-authored a booklet on hate speech on Runet while working as the head of the Computer Science Center at the European University at St. Petersburg.

**Mariëlle Wijermars** is Assistant Professor of Cyber-Security and Politics at Maastricht University, The Netherlands. She is a co-founder of the Digital Russia Studies network and an editor of the journal *Studies in Russian, Eurasian*

*and Central European New Media* (digitalicons.org). She is the author of *Memory Politics in Contemporary Russia: Television, Cinema and the State* (2019) and editor (with Katja Lehtisaari) of *Freedom of Expression in Russia's New Mediasphere* (2020).

**Mikhail Zhrebtsov** is an adjunct research professor at the Institute of European, Russian and Eurasian Studies at Carleton University. His research interests are focused on contemporary issues of governance and public policy in Russia and other post-Soviet countries as well as computational political studies, particularly social media analytics.

## LIST OF FIGURES

Fig. 17.1	Frequency of adjective decade constructions for each decade	307
Fig. 17.2	Distribution of <i>rannie</i> (early) and <i>pozdnie</i> (late) in decade constructions	308
Fig. 17.3	Frequencies of <i>libie devânostye</i> (wild nineties) compared to all adjectives attested in the construction (2001–2013, the Newspaper subcorpus)	309
Fig. 17.4	The relative frequency (%) of <i>modernizaciâ</i> (modernization) occurring in texts from Russian national newspapers (Source: Integrum, Dec. 31, 2000—Dec. 31, 2012)	311
Fig. 17.5	The usage of <i>modernizaciâ Rossii</i> (modernization of Russia) in comparison to <i>importozamešenie</i> (import substitution) (Source: Integrum, Russian National Media, 2013–2015)	313
Fig. 18.1	Representation of the current international sanctions situation in the thesaurus form	329
Fig. 21.1	<i>Prozbito</i> . User interface	376
Fig. 21.2	<i>Prozbito</i> . Author page	377
Fig. 21.3	<i>Prozbito</i> . The page <i>Pomoš'</i> (Help), where volunteers can learn how they can contribute to the project	378
Fig. 21.4	<i>Prozbito</i> . A suggested page of a diary	379
Fig. 24.1	The ten topics covered in newspaper articles that discussed Yves Montand's 1956 tour of the Soviet Union	435
Fig. 24.2	The top eight topics in the Polish <i>Życie Gospodarcze</i> newspaper, 1950–1980	436
Fig. 25.1	“Modernization” topic prevalence dynamic	456
Fig. 25.2	Projection of keywords on a two-dimensional space	459
Fig. 26.1	Neural network layers. (a) feed forward layer, (b) convolutional building layer, (c) recurrent layer, (d) transformer layer	469
Fig. 26.2	Word2vec configurations. (a) continuous bag of words, (b) skip-gram	471
Fig. 27.1	A network in the MGPU producing large-scale plagiarism (A. Abalkina, Dissnetnet.org). The full interactive graph is available at: <a href="https://www.dissnetnet.org/publications/mpgu_graf.htm">https://www.dissnetnet.org/publications/mpgu_graf.htm</a>	491

Fig. 27.2	The overall distribution of small-scale plagiarism across disciplines	495
Fig. 29.1	The seven bridges of Königsberg. Wikimedia Commons, <a href="https://commons.wikimedia.org/wiki/File:7_bridges.svg">https://commons.wikimedia.org/wiki/File:7_bridges.svg</a> , licence: CC BY-SA 3.0	518
Fig. 29.2	Extracted social networks of 20 Russian plays. Excerpt (left-upper corner) from a larger poster displaying 144 plays in chronological order (1747–1940s). Version in full resolution: <a href="https://doi.org/10.6084/m9.figshare.12058179">https://doi.org/10.6084/m9.figshare.12058179</a>	521
Fig. 29.3	Network graph for Ostrovsky’s <i>Groza</i>	522
Fig. 29.4	Network sizes of 144 Russian plays in chronological order, x-axis: (normalized) year of publication, y-axis: number of speaking entities per play. Arrow indicates Pushkin’s “Boris Godunov.” Russian Drama Corpus ( <a href="https://dracor.org/rus">https://dracor.org/rus</a> )	526
Fig. 29.5	Network visualization of A. Pushkin’s <i>Boris Godunov</i> . Russian Drama Corpus ( <a href="https://dracor.org/rus">https://dracor.org/rus</a> )	527
Fig. 29.6	Network visualization of L. Tolstoy’s <i>War and Peace</i>	530
Fig. 29.7	Network visualization of L. Tolstoy’s <i>War and Peace</i> , book 1 (first part of the first volume)	531
Fig. 29.8	Network visualization of L. Tolstoy’s <i>War and Peace</i> , book 10 (second part of the third volume)	531
Fig. 29.9	Network visualization of L. Tolstoy’s <i>War and Peace</i> , epilogue	532
Fig. 29.10	Network densities of separate books (parts) of <i>War and Peace</i>	532
Fig. 30.1	The structure of political communities on Twitter by event	543
Fig. 30.2	Clique size frequency distribution by community—Crimea sample	545
Fig. 30.3	Clique size frequency distribution by community—Medvedev sample	545
Fig. 30.4	Influence of leaders’ content, distribution across quantiles	549
Fig. 30.5	Pro-government community reaction to Medvedev’s comment to pensioners in Crimea	555
Fig. 30.6	Opposition community reaction of disbelief to Belykh’s guilt	556
Fig. 30.7	First and second quantile breakdown of account types by political communities	561
Fig. 30.8	Change in influence of leaders’ accounts when centrality compliments the impact of their distributed content	561
Fig. 30.9	Tweet patterns for the two main communities, tweets per hour by event	562



## LIST OF TABLES

Table 4.1	Main indicators of the national program “Digital Economy of the Russian Federation” (2018)	58
Table 8.1	European Social Survey questions on various forms of activism (European Social Survey Round 8 Data 2016)	141
Table 11.1	Religions and digital media normative expectations	192
Table 13.1	Classification of online retail platforms in Russia	226
Table 17.1	Russian National Corpus: texts by subcorpora	304
Table 17.2	Russian National Corpus: texts by creation date (the main subcorpus only)	305
Table 17.3	Frequency of adjective decade constructions for each decade	307
Table 18.1	Types and examples of part-whole relations in RuThes	326
Table 18.2	Examples of conceptual dependence relations denoted as nonsymmetrical associations in RuThes	328
Table 19.1	SNS use in Russia according to media research (October 2018, in mil)	337
Table 22.1	Cumulative effect of using applications based on open data in Moscow’s public transport system	404
Table 25.1	Top-30 semantic neighbors of the word <i>modernizaciã</i> (modernization) before and after January 1, 2012. Neighbors appearing only in one list are highlighted	458
Table 26.1	Word embeddings for Russian	472
Table 26.2	Tools to train word embeddings	472
Table 26.3	Two examples of sequence labeling tasks	475
Table 26.4	Transfer learning models for Russian	477
Table 27.1	A source text (left) and the copy-pasted text after OCR (right)	486
Table 27.2	A source text (left) and the paraphrased text (right)	487
Table 27.3	A source text (left) and the translated text (right)	488
Table 27.4	Percentage of borrowing in dissertations defended at various Russian institutions	496
Table 27.5	Differences in publication and citation performance among authors demonstrating the highest and the lowest amount of borrowing	497

Table 29.1	Number of co-occurrences of characters in A. Ostrovsky's <i>Groza</i> (abbreviated)	521
Table 29.2	Graph-related values for five selected Russian plays	525
Table 29.3	Selected network metrics for five characters in A. Pushkin's <i>Boris Godunov</i>	528
Table 29.4	Central characters in L. Tolstoy's <i>War and Peace</i> ranked according to three different centrality measures	530
Table 30.1	Six events and size of captured networks	541
Table 30.2	Density and transitivity of the network in its entirety and within its main communities	545
Table 30.3	Leaders' impact metrics using all data collected or focused on the top 10k popular tweets	548
Table 30.4	Relative community sizes and standardized homophily indicators	554
Table 30.5	Duplication and similarity of content in two main communities (% identical; % similar)	558
Table 30.6	Sample of "bot" communities detected (% identical; % similar)	558
Table 30.7	Botometer (Bot or Not) results (average universal probability of automation; proportion of accounts no longer present two years after samples were originally collected)	559
Table 30.8	Polarization of communities: retweets	560
Table 30.9	Polarization of communities: mentions	560
Table 30.10	Polarization of communities: replies	561
Table 32.1	Open datasets in Russian geoweb	589



# Digital Russia Studies: An Introduction

*Daria Gritsenko, Mikhail Kopotev, and Mariëlle Wijermars*

## 1.1 AREA STUDIES GO DIGITAL

The “digital” is profoundly changing Russia today. While in the mid-1990s less than 1 per cent of the Russian population had Internet access, today Russia ranks sixth globally with approximately 110 million Internet users, or three quarters of the population (The World Factbook 2019). The proliferation of affordable smartphones in the 2010s has made Internet access a common place by 2020, with over 60 per cent of users connecting through mobile devices, and Russia’s Internet market is the largest in Europe (GfK 2019). According to the Russian Ministry of Digital Development, Communications and Mass Media, the Russian Internet industry amounted to an estimated value of five trillion rubles in 2019, or 5 per cent of the country’s gross domestic product (GDP) (TASS 2019). Taking into account the additional 25 million Russians who live outside of Russia, it is no surprise that Russian is the second most

---

The names of the authors are given in alphabetical order.

---

D. Gritsenko  
University of Helsinki, Helsinki, Finland  
e-mail: [daria.gritsenko@helsinki.fi](mailto:daria.gritsenko@helsinki.fi)

M. Kopotev  
Higher School of Economics (HSE University), Saint Petersburg, Russia  
e-mail: [mkopotev@hse.ru](mailto:mkopotev@hse.ru)

M. Wijermars (✉)  
Maastricht University, Maastricht, The Netherlands  
e-mail: [m.wijermars@maastrichtuniversity.nl](mailto:m.wijermars@maastrichtuniversity.nl)

popular language on the Net after English (Historical trends 2019). These figures alone make Russia an attractive object for researchers interested in the development of today's digital society. The Russian information technologies (IT) industry, moreover, is an ample provider of highly sophisticated digital tools and well-organized software solutions: Nginx's popular web server that is used by, for instance, Netflix; Kaspersky antivirus software; optical character recognition application ABBYY FineReader, to mention just a few. In Russian-speaking markets, tech conglomerate Yandex furthermore successfully rivals with Google, while social networking sites VK (formerly known as VKontakte) and Odnoklassniki outperform their international competitor Facebook.

The global digitalization trend and the major societal shifts that accompany the process of converting ever more information and communications into digital form, challenge and transform existing practices across all spheres of life. In many ways, the digital transformation Russia is undergoing is far from unique. For example, the Russian government, similar to governments elsewhere, actively develops digital strategies, looking to reform education, finances and telecommunications and to increase governmental efficiency. Russian businesses seek to reap the benefits afforded by information and communication technologies (ICTs) and big data as they operate in and expand into domestic and global markets. Russian citizens, meanwhile, actively engage in the production and consumption of web-based content, while their dealings with state authorities increasingly occur through online e-government portals. New trends and practices emerge in the arts, where literary authors experiment with virtual personae and hyperlinked narration, while visual artists explore collaborative and cooperative online work and digital forms of expression.

At the same time, the impact of and responses to these digitalization practices in Russia are evidently context driven. The conservative and authoritarian turn in Russian politics (Smyth 2016) during Vladimir Putin's third presidential term (2012–2018), for example, has influenced not only the political, but also the technological landscape. State attempts to control the online sphere have materialized in various forms, including the regulation of data flows and the blocking of access to unfavorable online content and unruly platforms. Russia also exerts pressure on major domestic and international Internet companies, for example to transfer personal data of Russian citizens to servers located in Russia, and seeks to shape global Internet governance to reflect its favored terms. At the same time, digital communications have created new opportunities for the facilitation of civic resistance, as is evidenced by the success of oppositional leader Alexei Navalny in rallying support and mobilizing political resistance through his online activities.

For researchers investigating Russia, digitalization has resulted in the emergence of a wealth of new (big) data sources, including social media and other kinds of digital-born content that allow us to investigate Russian society in novel ways. The accelerating speed at which Russian archives are being digitized means that collections of research materials have become more easily available, while simultaneously new methodological possibilities open up for examining Russian historical sources with the help of digital tools. The

abundance of computational methods, ranging from simple automated keyword sorting to complex machine learning algorithms, allow us to tap into the opportunities offered by combining different types of data that have not previously been used together, or to explore patterns in large datasets that are difficult to grasp with a “manual” approach.

Given the intricate combination of the universal and the particular in how Russia is influenced by the digital as well as gives shape to digitalization trends, and the specificities involved in the availability and use of digital sources and methods, we argue that an area studies approach is both timely and productive. Area studies, as we know them today, developed in American and Western European universities in the second half of the twentieth century, when departments studying non-Western cultures welcomed sociology, economics, and political science specialists to, together with language and literature scholars, explore the contemporary social life of the regions they studied (Colonomos 2016, 65). The value of area studies, essentially a Cold War project striving to provide a general framework to describe and explain what was going on in different parts of the non-Western world, be it the Soviet Block, the Middle East, Africa, Latin America, or China (Rafael 1994), was increasingly questioned after “the end of history” (Fukuyama 1989). The forces of globalization, the third wave of democratization, and the worldwide triumph of the market economy were expected to diminish the value and necessity of studying an area, with its emphasis on contexts; disciplinary knowledge was thought to be central and contextualized “place knowledge” secondary. This volume asserts that area studies, as a geographically and geopolitically motivated interdisciplinary research domain, is of particular value to and can provide a general framework for describing the variety of responses to digitalization and explaining the mechanisms that assist or obstruct the domestication of global trends. In this respect, we can build upon earlier efforts in this direction, such as the volume *Digital Russia: The Language, Culture and Politics of New Media Communication* (2014) edited by Michael Gorham, Ingunn Lunde, and Martin Paulsen and the journal *Studies in Russian, Eurasian and Central European New Media* ([digitalicons.org](http://digitalicons.org)). Other area studies fields have similarly turned their attention to digitalization. Consider, for example, the launch of the *Digital America* journal in 2012 and publications such as *The Other Digital China* by J. Wang (2019). All such emerging digital area studies initiatives, in turn, draw upon and contribute to the by-now-established field of Internet Studies, exemplified by, for example, *The Oxford Handbook of Internet Studies* (2013).

The fact that digitalization started making major headline appearances around the same time the post–Cold War end of history was declared is instructive for understanding how it came to be viewed (even though the process of converting traditional forms of information storage and processing into the binary code of computer storage can be traced back to the advent of computing after the Second World War). The ideals closely connected to the early development of the Internet, such as freedom, decentralized control, the claim of universality of technological development and so on, fitted well with the overall narrative of global modernity (Dirlik 2003). Yet, during the past decade

we have witnessed backlashes on all “global fronts”—including democratic backsliding, the rise of populism, the return of economic protectionism and borders, first off- and then online—allowing area studies to make a comeback. More than half a century of area studies scholarship has brought forward important methodological accomplishments that turn out to be extremely useful in approaching these global backlashes. First, the idea that context matters, a staple in the disciplines of geography and anthropology, has been explicitly brought into studies on economics, politics, and society through in-depth field research. Area studies have routinely challenged the US- and Euro-centric assumptions of many disciplines, while Szanton (2004) even argued that mainstream disciplines are in fact special cases of area studies, American and European Studies, respectively. Practices of place-based research that produce contextually and culturally rooted explanations are useful if we seek to fully understand questions of digital transformation.

Second, the multi- and interdisciplinary approaches that are inherent to research projects in area studies have led to extensive conceptual borrowing, cross-fertilization among disciplinary fields, and an emphasis on comparative methodologies (Katzenstein 2001). Practical circumstances—colleagues working at centers for area studies are likely to have various disciplinary backgrounds and area studies conferences bring together scholars working across the humanities and social sciences—not only push individual scholars out of their (disciplinary) comfort zone, but also provide ideas and nourish creative conceptual development. This feature, we want to suggest, is invaluable for studying digitalization across societies. Finally, language, which has been at the center of area studies from its very inception, has been recognized “as productive and powerful in its own right” (Gibson-Graham 2004) and capable of shaping social practices. Accentuating the performativity of language and the power of discourse as a method for critical deconstruction, area studies have been at the forefront of the so-called interpretative turn in the social sciences. By the same token, language-based approaches—in particular computational approaches—are among the backbones of digital studies.

Therefore, it makes sense to talk about Digital Russia Studies. Yet, a comprehensive volume that offers novice-friendly guidance for navigating the full breadth of this new territory is currently lacking. To grasp the simultaneous transformation of research object and research practices, this Handbook brings together world-leading experts and emerging scholars to lead the way in the emerging field of Digital Russia Studies. That being said, we are moving away from the conventional label of Russian Studies to highlight that we aim to contribute to and consolidate a methodological broadening in area studies: *Digital Russia* studies focuses on the digital transformation of the (geographical) area of study, while digital *Russia Studies* indicates the use of digital sources and methods in studying it and that is only partially captured by the term “digital humanities.” Together, *Digital Russia Studies* emphasizes how these two research lines are intertwined, interdependent, and mutually reinforcing.

Drawing the borders of Digital Russia is no easy feat, even though it is clear that it cannot be reduced to the digital projection of the state within its physical borders. For one, many political and economic digital actors of significance are located outside Russia, for example online media outlet Meduza that operates from Latvia and Yandex N.V. that is registered in the Netherlands. Russian services also operate in languages other than Russian and are not merely hosted on the Russian .ru domain, but also on international domains (such as .com or .edu) and the still functional Soviet .su domain. Russian Studies for the digital era therefore deals with opaque, negotiable, and constantly moving borders—material and virtual—that cannot be set once and for all, but rather require careful consideration depending on the case-study, level of analysis, or specific research application.

Aiming to present a multidisciplinary and multifaceted perspective on the issues outlined above, the objective of this Handbook is twofold, as reflected in its two-part structure. The first part of the book, *Studying Digital Russia*, provides a critical and conceptual update on how Russian society, politics, economy, and culture are reconfigured in the context of digitalization, datafication, and the—by now—widespread use of algorithmic systems. Reviewing the state of the art in scholarship on a broad range of policy sectors and issues, the chapters investigate the transformative power of the digital and the particularities of how these transformations manifest themselves in the context of Russia. The chapters also reflect on societal responses to these ongoing transformation processes.

The second part of the Handbook, *Digital Sources and Methods*, combines two subsections that aim to answer practical and methodological questions in dealing with Russian data. *Digital Sources* describes the main resources that are available to investigate the multifaceted Digital Russia sketched above: textual, visual, and numeric. In addition, the vulnerabilities, uncertainties, legal and ethical controversies involved in working with Russian digital materials are addressed. The second subsection, *Digital Methods*, showcases examples of cutting-edge digital methods applied in different fields of research. The chapters provide a concise overview of the manifold opportunities for studying society, politics, and culture in novel ways. The chapters also address the particular methodological issues that researchers will encounter when working with Russian data, such as working with Russian social media platforms and processing sources written in Cyrillic rather than Latin script. The chapters in this section demonstrate how the area studies tradition of invoking context as an essential element of scientific explanation can leverage some of the criticism that is being directed to the use of digital methodologies and big data in humanities and social sciences research. In the remainder of this introduction, we provide an overview of the topics, questions, and methods covered by the contributions in this Handbook and briefly sketch the emergence of digital technologies and networks in the region.

## 1.2 STUDYING DIGITAL RUSSIA

The first attempts to establish a national digital network in Russia can be traced back to the late Soviet period and the never realized project called OGAS (*Obşegosudarstvennaâ avtomatizirovannaâ sistema*, All-State Automated System). As is recounted by Benjamin Peters (2016), the story of OGAS is a troubled one that ended in total failure due to the forces of Soviet bureaucracy, effectively resisting innovations capable of jeopardizing state power, or the positions of those in power. In the 1990s, local- and national-level networks were overtaken by the expansion of the global Internet, emerging out of the efforts of, for example, research institutions Conseil Européen pour la Recherche Nucléaire (CERN) in Switzerland and the Institute for High Energy Physics in Russia (Abbate 1999; Gerovitch 2002). Since then, global technological developments have followed similar trends, albeit at different paces; for example, the transitions from low- to high-speed Internet, from wired to wireless access, and from expensive to affordable services offered by Internet service providers. While the Internet was becoming more user-friendly, functional and attractive all around, its social and political domestication in Russia had its specificities: whereas many Western publicly available online services were developed by IT geeks in garages, the Russian Internet, as legend has it, was born in the kitchens of the intelligentsia. This local feature is sealed in the term “Runet” coined from the words “Russian” and “Internet.”

The concept of Runet has evolved over the course of the past decades, along with the object it describes, as Asmolov and Kolozaridi explain in their chapter. Yet, in any circumstance it cannot be reduced to the .ru-domain or to online content in the Russian language. In the late 1990s–early 2000s, when the concept gained a foothold, Runet was defined as having two fundamental features: it was logocentric and free. The first feature refers to the fact that many of Runet’s forerunners had an interest in the arts and humanities:

The RuNet is specific with regard to the topic of literature: the myth of ‘literature-centrism’ of Russian culture (almost dead, as it seems) has been resurrected on the RuNet’s literary sites, which have no analogues in the other (national) segments of the Internet. (Konradova et al. 2006)

The first Runet websites, for example [lib.ru](http://lib.ru), while technologically and economically amateurish, were oriented toward the free distribution of information and deeply rooted into the domestic cultural context. Many of these features are still preserved in Runet today (see Chaps. 15 and 9), even though it has become technologically advanced and market oriented, as the chapter by O. Gurova and Morozova on digital consumption shows. Runet preserves some of the spirit of freedom, although the legality of some of these activities can be questioned (see Chaps. 7, 8 and Chap. 6). Digital technologies have



given a great impetus to innovation of the arts, as Strukov demonstrates in his chapter. The Internet has also been instrumental in facilitating the expression and negotiation of gender identities, analyzed by Andreevskikh and Muravyeva, and is leaving its mark on religious practices, as Khroul's chapter demonstrates.

In its early days, the Internet in Russia developed practically free from state interference. As many sources testify (e.g. Babaeva 2015), soon-to-be president Vladimir Putin hosted a meeting with representatives of the IT (information technologies) industry on December 28, 1999, during which the sector was promised a decade of free development. Limited state regulation and meddling indeed was among the defining features of Rунet for a considerable period of time (e.g. the lack of effective online copyright protection), but the screws have been steadily tightening, most rapidly from 2012 onwards, as is addressed in multiple chapters in this volume (e.g. Chaps. 16, 8 and 2). The Russian Internet has come under ever more direct and indirect control of the state, among others in terms of extensive surveillance capabilities and prerogatives concerning digital communications and the economic dependence of IT businesses. In 2019 alone, there have been several milestone decisions that illustrate the extend of state control over how the Internet develops in Russia. For example, the expansion of 5G network technology has been significantly delayed because of continued resistance, among others by the Security Council of the Russian Federation, against making the preferred frequency band available for civilian uses (the 3.4–3.8 GHz range earmarked for 5G use by, e.g. European Union [EU] countries, is currently used by the Russian military and security services), while Yandex changed its corporate governance structure to accommodate governmental pressure and avert the introduction of legislation limiting foreign ownership of major Internet companies (Yandex N.V. is registered in the Netherlands).

While the Russian government has sought to counteract the freedoms previously afforded to the Internet through regulation and other control strategies, the analyses in the first part of the Handbook make clear how it at the same time recognizes the enormous potential of digital technologies. Indeed, the Russian government frequently points toward digitalization as a cornerstone of the country's development. At the 2017 Saint Petersburg Economic Forum, for example, Putin highlighted Russia's place among the forefront of research into artificial intelligence (AI):

Just like other leading nations, Russia has drafted a national strategy for developing AI technologies. It was designed by the Government along with domestic hi-tech companies. (<http://en.kremlin.ru/events/president/news/60707>, official translation)

The federal government runs various programs to support digitalization across sectors, such as government (analyzed by Gritsenko and Zherebtsov), politics (discussed by Wijermars), law and justice (addressed by Muravyeva and Gurkov), economy (examined by Lowry), and education (analyzed by Piattoeva

and G. Gurova). Billions of rubles from the federal budget have been invested into infrastructure, making available many e-services, as well as an abundance of administrative, legislative, archival, textual, geospatial data (explored in Part II of this volume). As the chapters in this Handbook discuss in more detail, the success of these federal programs is ambivalent. It is however undeniable that the massive amount of data that is produced by various agencies as a result is now available to experts and citizen scientists alike, enabling them to conduct in-depth big data analyses, among others to reveal breakdowns in governance (as is argued by Parkhimovich and Gritsenko, and Kopotev, Rostovstev, and Sokolov in their respective chapters).

The ostensibly clear-cut image of Russia's Internet status changing from free to not free over the course of the past two decades, as is evidenced by annual rankings of Internet freedom, therefore fails to tell the full story and its inherent paradoxes. Manifold examples demonstrate how the Internet continues to be instrumental for facilitating civic resistance, as Lonkila et al. recount in their analysis. From this perspective, today's digital dissidents can be seen as acting in the vein of the Soviet intelligentsia, even though the two groups represent different generations and values.

### 1.3 DIGITAL SOURCES AND METHODS

The second part of the Handbook is diverse and of a more applied nature. It starts with chapters discussing the most widely used digital sources, mainly those for text-based studies that depart from the assumption that language can be studied as a reflection of society. Collections of texts, or textual corpora, are a key resource for linguistic studies as well as for a wide variety of applications within the humanities and social sciences. Kopotev, Mustajoki, and Bonch-Osmolovskaya describe these sources with a focus on the *Russian National Corpus* (RNC), a deeply annotated and well-designed resource on the Russian language, and the *Integrum* database, which comprises most newspapers, journals, and online media published in Russia or in Russian, as well as TV and radio transcripts. Thesauri, for example the Russian RuThes thesaurus that is discussed by Loukachevitch and Dobrov, are more sophisticated linguistic and terminological resources for automatic text processing that can be used to explore concepts, changes in word meaning, text categorization, and so forth. More recently, social media have established themselves as a new channel of communication and novel resource for studying a wide set of societal questions. In a chapter that focuses on assessing the applicability of existing models of social media research in the Russian context, Koltsova et al. present the limitations of existing approaches and suggest best practices for social media research that uses Russian sources.

Two chapters are devoted to digital archives and digitized archival materials. While all standard text-analytical techniques, both qualitative and quantitative, can be applied to these materials, the contributions draw attention to questions

regarding their provenance, objectivity, and affordances, and the complex political economy of historical knowledge production. Providing an overview of digitization practices in Russia, Golubev reveals an underlying political agenda to restore epistemic sovereignty over Russian history. Kalinina, in turn, raises a series of techno-methodological questions concerning the composition and affordances of a digital archive platform created by a community of volunteers. The final digital source covered is open government data, which is presented by Parkhimovich and Gritsenko from an infrastructural, legal, and technical viewpoint. Illustrating their argument with examples of projects and applications utilizing open government data, especially open financial data, the authors provide concrete use cases that show the perceived benefits for government agencies and citizens.

The final collection of chapters is methodological in orientation, presenting a variety of digital and computational techniques and providing concrete examples of their use in Russian Studies. First, topic modeling, a method of probabilistic text clustering, is explored. Bodrunova looks at how topic modeling techniques have been developed and employed by Russian scholars—applied both to Russian and other languages—paying special attention to questions of validity and assessment of model quality. Oiva shows how topic modeling can be applied to Russian historical sources—such as Soviet newspapers—and offers an accessible step-by-step walk through of the basics of topic modeling. Indukaev then applies topic modeling to a contemporary media collection obtained from the *Integrum* database and showcases how the analysis can be enriched by incorporating the word embedding technique. He argues that the latter is capable of providing more accurate observations of the data. Artemova dives even deeper into Natural Language Processing (NLP). She focuses on deep-learning applications for processing Russian, presenting state-of-the-art methods in the field. The chapter written by Kopotev, Rostovtsev, and Sokolov investigates the issue of academic plagiarism and how its detection posits a challenge for computational linguistics. Another popular NLP application—sentiment analysis—is discussed by Loukachevitch, who explains the main contemporary applications of the method focusing on Russian-specific components of automatic sentiment analysis.

While computational text-analytical techniques constitute the backbone of Digital Russia Studies, other methods provide equally exiting opportunities for future research. The first of these is network analysis, a method for exploring relationships and structures based on graph theory. To show the versatility of its application, we have included two chapters. Fischer and Skorinkin apply network analysis in the field of literary studies. They demonstrate how texts can be formalized into a set of nodes and edges, where nodes represent characters and edges describe interactions between these characters, based on a selection of Russian plays and the classic novel *War and Peace* by Leo Tolstoy. The second application concerns a study of Russian politics and society on microblogging platform Twitter. Zherebtsov and Goussev analyze six resonant political

events to demonstrate how network analysis enables an alternative approach to answer classic questions within political science, such as designating political communities, tracing group reactions to informational events, and detecting opinion leaders and influencers.

The Handbook concludes with two methods that operate with nontextual data. The field of art history, Kangas argues, has lagged behind in joining the digital humanities trend; yet, digital image analysis opens up various new avenues for research. Drawing upon the example of Soviet political cartoons, she advocates the use of mixed methods to best utilize computational and human interpretative strengths. The final chapter is devoted to the analytical use of geospatial data, their attributes in Russia's online ecosystem, and the methodologies best suited for their analysis. Makhortykh discusses novel techniques for extracting geolocations from various data formats and demonstrates different ways of using these data, from mapping the spatial distribution of social and political phenomena to the use of the geoweb for narrating individual and collective identities online.

#### 1.4 CONCLUDING REMARKS

With this Handbook we have aimed to lay down the foundations for the emerging research direction of Digital Russia Studies. Through its 32 chapters, the book makes a timely intervention in our understanding of the changing field of Russian Studies at the intersection of the societal and the digital in order to become a first comprehensive review and guide for scholars as well as graduate and advanced undergraduate students studying Russia today.

As is true for any work that seeks to carve out the contours of an emerging field of study, the range of topics, approaches, and methods covered in this Handbook is necessarily incomplete. However, by compiling analyses of the impact of digitalization on various spheres of Russian politics, society, and culture in a single volume together with chapters exemplifying best practices in using digital sources and methods in Russian Studies, we hope to have demonstrated the value of an area studies approach in studying the digital domain. At the same time, it has to be acknowledged that this Handbook is itself a product and expression of the shifts we are currently witnessing: while most analyses included here are still predicated to some extent on the opposition between, coexistence, and interwovenness of digital and analogue, such distinctions may rapidly become obsolete as digital becomes the new norm in ever more domains. In this regard, the Handbook also functions as an important landmark, documenting these transitional pathways as they take shape across various spheres of society and human activity.

## REFERENCES

- Abbate, Janet. 1999. *Inventing the Internet*. Cambridge, MA: MIT Press.
- Babaeva, Svetlana. 2015. Interv'ù s Antonom Nosikom [Interview with Anton Nosik]. Accessed December 10, 2019. <https://www.gazeta.ru/social/2017/07/09/10779572.shtml>.
- Colonomos, Ariel. 2016. *Selling the Future: The Perils of Predicting Global Politics*. London: Hurst.
- Dirlík, Arif. 2003. Global Modernity? Modernity in an Age of Global Capitalism. *European Journal of Social Theory* 6 (3): 275–292.
- Dutton, W.H., ed. 2013. *The Oxford Handbook of Internet Studies*. Oxford: Oxford University Press.
- Fukuyama, Francis. 1989. The End of History? *The National Interest* 16: 3–18.
- Gerovitch, Slava. 2002. *From Newspeak to Cyberspeak: A History of Soviet Cybernetics*. Cambridge, MA: MIT Press.
- GfK. 2019. *Issledovanie GfK: Proniknovenie interneta v Rossii* [GfK Research: Internet Penetration in Russia]. Accessed December 16, 2019. <https://www.gfk.com/ru/insaity/press-release/issledovanie-gfk-proniknovenie-interneta-v-rossii>.
- Gibson-Graham, Julie. 2004. Area Studies After Post-structuralism. *Environment and Planning A* 36 (3): 405–419.
- Gorham, Michael, Ingunn Lunde, and Martin Paulsen. 2014. *Digital Russia: The Language, Culture and Politics of New Media Communication*. Abingdon and New York: Routledge.
- Historical trends. 2019. *Historical Trends in the Usage of Content Languages for Websites*. Accessed December 16, 2019. [https://w3techs.com/technologies/history\\_overview/content\\_language](https://w3techs.com/technologies/history_overview/content_language).
- Katzenstein, Peter. 2001. Area and Regional Studies in the United States. *PS: Political Science and Politics* 34 (4): 789–791.
- Konradova, Natalja, Henrike Schmidt, and Katy Teubener, eds. 2006. *Control+Shift: Public and Private Usages of the Russian Internet*. Norderstedt: BOD Verlag.
- Peters, Benjamin. 2016. *How Not to Network a Nation: The Uneasy History of the Soviet Internet*. Cambridge, MA: MIT Press.
- Rafael, Vicente L. 1994. The Cultures of Area Studies in the United States. *Social Text* 41: 91–111.
- Smyth, Regina. 2016. Studying Russia's Authoritarian Turn: New Directions in Political Research on Russia. *Russian Politics* 1 (4): 337–346.
- Szanton, David L. 2004. Introduction: The Origin, Nature, and Challenges of Area Studies in the United States. In *The Politics of Knowledge: Area Studies and the Disciplines*, ed. D. Szanton. Berkeley: University of California Press.
- TASS. 2019. *V Minkomsvázi ocenili ob'em ekonomiki Runeta* [The Ministry of Communications Estimated the Volume of the Runic Economy]. Accessed January 11, 2020. <https://tass.ru/ekonomika/7016909>.
- The World Factbook*. 2019. Washington, DC: Central Intelligence Agency. Accessed December 10, 2019. <https://www.cia.gov/library/publications/resources/the-world-factbook/index.html>.
- Wang, Jing. 2019. *The Other Digital China: Non-confrontational Activism on the Social Web*. Cambridge: Harvard University Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



PART I

---

# Studying Digital Russia



# The Digitalization of Russian Politics and Political Participation

*Mariëlle Wijermars*

## 2.1 INTRODUCTION

Digitalization has affected politics in manifold ways, which result from the broad variety of technologies the term comprises. The Internet and, more recently, social media have, for instance, transformed political campaigning. The publication of public policy documents on government websites has created new expectations for political transparency. And, the introduction of voting computers and other e-voting solutions has made it possible to fundamentally rethink the voting process (e.g. online voting) while raising novel security concerns. In its strictest sense, digital politics can be defined as “how politicians employ the Internet to reach, court, and mobilize citizens and about how citizens rely on the web to inform themselves and engage with others politically” (Vaccari 2013, 4). Yet, as is pointed out by Stephen Coleman and Deen Freelon, “[t]o speak of digital politics is not simply to tell a story about how political routines are replicated online,” rather it is about the (unforeseen) transformations of political practices that result from digitalization:

One feature of all technologies is that they are constitutive: they do not simply support predetermined courses of action, but open up new spaces of action, often contrary to the original intentions of inventors and sponsors. (Coleman and Freelon 2015, 1)

---

M. Wijermars (✉)  
Maastricht University, Maastricht, The Netherlands  
e-mail: [m.wijermars@maastrichtuniversity.nl](mailto:m.wijermars@maastrichtuniversity.nl)



In this chapter, I discuss the impact of digitalization on politics in Russia and the extent to which such unforeseen transformations in the political process have taken place. My discussion highlights four areas: political communication; political campaigning; voting; and, political participation and civic engagement.

While the digitalization of politics is a global trend, the characteristics and constraints of the national political context, such as the uptake speed of particular technologies, condition the shape digital politics takes. In the case of Russia, the proliferation of digital technologies unfolded in parallel with the “authoritarian turn” under President Vladimir Putin (Smyth 2016). As the examples discussed in this chapter will illustrate, digitalization has in fact been a deliberate politics on the part of the Russian state. While it is therefore necessary to consider to what extent the impact of digital technologies on politics unfolds differently in democracies as compared to hybrid regimes or non-democracies, the opposite poles of the scholarly debate are similar: they either highlight the democratizing potential of digital tools or focus on their unintended or negative consequences. Given the different starting points—for example, the extent to which the object of study can be classified as a functioning democracy—this nonetheless results in different questions being asked.

Regarding Western liberal democracies, the democratizing potential is thought to lie in the opportunities digitalization provides for remedying the democratic deficit, for example through increased citizen participation, more direct communication channels between politicians and citizens through social media and the facilitation of forms of direct democracy. On the flipside, concerns have emerged about how online communications, in particular social media, may have polarizing effects that negatively affect societal stability and may be used to manipulate public opinion and election outcomes, as well as concerns about expanding state surveillance. In a similar vein, in the context of hybrid or non-democratic states, scholarly debate placed high hopes on the democratizing potential of the Internet. It was assumed that, among other factors, increased access to information online and the facilitation of political mobilization through the use of social media would empower citizens to challenge state power and demand a greater say in political decision-making (e.g. Castells 2012). Departing from the same assumption, many studies have examined states’ efforts to control online communications and protect the political status quo in response (e.g. Deibert et al. 2010). The extent to which the Internet indeed functions as a “liberation technology” is increasingly questioned (e.g. Diamond 2010). Rather, it appears that the proliferation of Internet access has given rise to “networked authoritarianism” (MacKinnon 2011, 33), a condition in which:

the single ruling party remains in control while a wide range of conversations about the country’s problems nonetheless occurs on websites and social-networking services. The government follows this online chatter, and sometimes people are able to use the Internet to call attention to social problems or injustices and even manage to have an impact on government policies. As a result, the

average person with Internet or mobile access has a much greater sense of freedom—and may feel that he has the ability to speak and be heard—in ways that were not possible under classic authoritarianism. At the same time, in the networked authoritarian state, there is no guarantee of individual rights and freedoms. (MacKinnon 2011, 33)

Notwithstanding the challenges that online communications raise for maintaining political control by increasing citizens' access to information and opportunities for free speech, it appears many authoritarian states are disinclined to (fully) limit access to the Internet. The seeming paradox—often referred to as the digital “dictator’s dilemma”—may be explained by the potential economic consequences of such a decision, fear of popular unrest or the undermining a regime’s democratic image or other sources of regime legitimacy. Yet, scholars have also noted that digitalization may, in fact, strengthen rather than weaken authoritarian regimes since the Internet can be used to effectuate political control, and information and opinions shared by citizens online can be a valuable resource to gauge public opinion on policy issues (e.g. Gunitsky 2015).

In this chapter, I first examine how the activities of political actors in Russia have changed as a result of digitalization, focusing on political communication and election campaigning, before turning my attention towards changes in voting and other forms of political participation. Many of these changes result from or developed against the backdrop of the introduction of open government ideas. Therefore, I open with an overview of actions in this domain. I argue that, while some of the changes described can be categorized as mere digital reproductions of existing political practices, several spheres of Russian politics have been transformed as a result of digitalization, in particular the opportunities for political opposition and civic engagement.

## 2.2 OPEN GOVERNMENT

The concept of open government promotes the ideal of transparency and accountability in governance: citizens should be able to access governmental documents and proceedings in order to establish an effective climate of checks and balances. In the past two decades, the concept has been inseparably intertwined with the notion of “e-government”: the spread of Internet access and information technology (IT) infrastructures have made the Internet the perfect solution for achieving the aims of “open” government. Combined, the overall goals of open and e-government are to increase efficiency and transparency, as well as to simplify and improve the provision of governmental services to civilians and government-to-citizen communication. In Russia, the government initiated the expansion of information technologies, digitization, provision of online services, increased governmental transparency and so forth in earnest in

the early 2000s (see also Chaps. 3 and 5). The Federal Program “*Ėlektronnaâ Rossiâ* (2002–2010)” (Electronic Russia)

called for the ‘widespread integration’ of information technology in government operations for such tasks as document management, registrations and declarations, and procurement tenders. To accomplish this mission, E-Russia’s goals also included building up the nation’s IT hardware and telecommunications infrastructure and developing a supportive legal and regulatory environment. Of note, the program’s mission statement also called for ‘significantly increasing the volume of information [that] government institutions provide to citizens, including via the Internet,’ such as draft laws and decrees, government revenues, and budgets; performance reports by public enterprises; and assessments by auditing agencies. In the process, information technologies were seen as ‘cardinally changing the basis of the government’s relationship with citizens and businesses’. (Peterson 2005, 51)

While some government bodies were early adopters, from 2003 onwards all federal agencies were required to make a broad range of information accessible online, such as regulations and legislation, and information on the activities of their officials (Peterson 2005, 58).

With modernization and innovation as the buzzwords to define his “liberal” presidency, Dmitry Medvedev (2008–2012) launched a federal program aiming towards turning Russia into an “Information Society” (2011–2020) (Toepfl 2012, for more, see also Chap. 25) and a Minister of Open Government was appointed in 2012. In 2018, the ministerial position was discontinued, signaling the topic had lost priority with the authorities. The push towards open government has resulted in a significant increase in the availability of open government data. For example, information concerning government tenders can be accessed on the *Goszakupki* (Government procurement) portal, [zakupki.gov.ru](http://zakupki.gov.ru), while various open data sources are collected on the open data portal [data.gov.ru](http://data.gov.ru). Through the creation of dedicated online platforms, the transparency of the legislative process has been enhanced; for example, the video recording of the Russian State Duma can be viewed on the platform [video.duma.gov.ru](http://video.duma.gov.ru) and draft laws are made available for public discussion on the platform [regulation.gov.ru](http://regulation.gov.ru) (for more, see also Chap. 5). Yet, many issues remain, including a tendency to reintroduce restrictions on publicly available information. For example, in response to investigations by Alexei Navalny’s FBK (*Fond bor’by s korrupciej*, Anti-Corruption Foundation), examples of which will be discussed below, the FSB (*Federal’naâ služba bezopasnosti*, Federal Security Service) proposed a law in 2015 that would severely restrict access to information about property ownership contained in *Rosreestr* (Federal Register). While the law was not passed, the Supreme Court determined in 2017 that *Rosreestr* is permitted to limit third-party access to ownership data, invoking the protection of personal data, thus setting a precedent (Kornia 2017).

### 2.3 POLITICAL COMMUNICATION

Parallel to the emphasis on adopting digital technologies in the policy sphere, significant changes were implemented in the authorities' communication strategies that, to an extent, resemble trends in political communication elsewhere. As a public advocate for technological innovation, Medvedev can be credited with pushing forward both the open government agenda and expanding Russian political communication from traditional media to online platforms. Through videos posted on the Kremlin website and, from 2009 onwards, his blog on LiveJournal (at the time the most popular blogging platform, see Podshibyakin 2010), Medvedev set an example for novel ways of communicating and engaging with citizens, and he pushed other government officials to start blogging as well (Gorham 2014). In 2010, some 35 per cent of Russian regional governors had a blog, a third of which emulated the videoblog format exemplified by the president (Toepfl 2012).

Medvedev's blogging activities were criticized for being "a blog without a blogger" (Yagodin 2012, 1422): his page featured videos posted by the presidential administration and functioned rather as a one-way channel for communication, lacking signs of Medvedev's direct contribution or his interaction with the online community, for example, with those commenting on his posts. Notwithstanding Medvedev's initial statements about aspiring towards a form of direct democracy through digital means, in practice most Russian politicians used their online communications "in ways that minimize the perils of truly direct online interaction and opting, instead, for a more hierarchical model of communication grounded in the discourse of 'e-government'" (Gorham 2014, 235). Rather than entering into conversations with engaged citizens, the online communication strategies they chose opted for "the carefully structured, monitored, and filtered interfaces such as the online opinion polling, 'online reception area,' or the sound-bite sized Twitter scroll" (Gorham 2014, 246). In a similar vein, Florian Toepfl (2012, 1454) argues that, when it concerns the leaders of Russia's federal subjects, "most Russian governors did not set up their blog primarily with the intention of gaining electoral support." Instead, blogging was predominantly "a symbolic action that showcased their allegiance and loyalty to the president, who was widely known for his Internet enthusiasm" (Toepfl 2012, 1454).

In his capacity as prime minister, following Vladimir Putin's return to the presidential office in 2012, Medvedev moved his most visible online presence to Twitter and Instagram, following the shifts in the platforms' popularity. Compared to his earlier presence on LiveJournal, the Instagram account is administered as a personal account, alternating between press photographs and pictures taken by Medvedev himself, accompanied with brief captions. Contrary to the LiveJournal blog, there is some interaction between the prime minister's account and other users on the platform, with Medvedev now and then commenting and responding. The increased personal dimension of Medvedev's social media presence may be explained by changing public relations (PR)

needs—aimed to remedy the previous lack of connection with citizens and following the more general trend of increased personalization of politics. The fact that Instagram is predominantly image-based—Medvedev is known to have an interest in photography—and allows one to post, edit and comment quickly through the application on one’s smartphone may also have been factors.

Yet, the decision to switch to Instagram also created vulnerabilities. Indeed, it was Medvedev’s Instagram that provided opposition leader Alexei Navalny’s Anti-Corruption Foundation with crucial visual evidence to tie together various publicly available sources of information indicating the prime minister’s involvement in large-scale corruption (including hacked emails leaked by Russian hacker collective *Šaltaj Boltaj* [Humpty Dumpty], Global Positioning System [GPS] tracking of naval movements and various official registries). The results of the investigation were published in a video entitled “*On vam ne Dimon*” (“He is not Dimon to you”) shared through FBK’s YouTube channel and website. While this was not the first video FBK published that exposes corrupt practices by Russian state officials—indeed, there are many—the Dimon video gained particular traction online (by December 2019: 32.8 million views). More importantly, it served as the occasion for mass anti-corruption protests on March 26, 2017, that mobilized thousands of protesters across Russia<sup>1</sup>; the largest demonstrations to take place since the protest movement of 2011–2012. FBK’s investigations demonstrate how open source data—some of which became available as part of the implementation of open government ideas—can be effectively used to scrutinize and challenge government practices.

On the sub-federal level, Ramzan Kadyrov, the head of the Republic of Chechnya, is one of the Russian political actors who has most successfully used social media to increase his popularity, both in Chechnya and (far) beyond. His Instagram account, with posts that blended “discussion of politics with photos of himself hugging cats, posing in a knight’s outfit, working out in a gym, and throwing snowballs with friends” (Rodina and Dligach 2019, 95) collected some three million followers, before the platform decided to shut down his account. Kadyrov’s posts merged public, political and private spheres to the extent that “all of the personal topics contain elements of political framing, and most of the public/political topics include terminology that refers to personal topics such as friendship and family” (Rodina and Dligach 2019, 106). Kadyrov’s success exemplifies how social media “can be used to normalize despotism, giving a modern-day dictator ‘a human face’” (Rodina and Dligach 2019, 96). The increasing use of social media in political communication is visibly changing the communication strategies used by the Russian Ministry of Foreign Affairs as well, whose official Twitter account incorporates vernacular language and actively partakes in online debates (Zvereva 2020). The Ministry’s spokesperson, Maria Zakharova, in particular, has adopted a style of communication that blends formal and informal statements, expressed through multiple (and at times parallel) accounts on, for example, Facebook and Twitter.

Digitalization has also changed the rules of the game when it comes to political contestation by citizens. The rise of the Russian “blogosphere” and,

subsequently, the popularity of bloggers, citizen journalists and vloggers on social media and YouTube, brought about novel opportunities for sharing political criticism with a wide audience, and for creating communities around a political cause (of which Navalny is but one example). Over time, the Russian government has responded to this perceived threat in multiple ways. Most notably, with the so-called “Bloggers’ Law” (Federal Law No. 97-FZ) it introduced a special register for bloggers with a daily audience of >3000 visitors in 2014. For bloggers, some of whom published under a pseudonym, the registration involved, among other requirements, the disclosure of their real identities to the Russian authorities. The impact of the measure on the expression of political criticism online is difficult to ascertain, yet it is known that its introduction did not lead to any blogs being blocked or fines imposed (Soldatov 2019). Nonetheless, as Oleg Soldatov points out, “the mere existence of the public list of popular Internet personalities, administered by and conceived in the interests of a governmental body, should have led to a certain number of such personalities thinking twice before making public their criticism of the government” (Soldatov 2019, 70–71).

The law was repealed in 2017, which can be explained by a combination of factors: the ineffectiveness of the register and difficulties in enforcing the law (e.g. poor definition of who counts as a blogger, estimation of daily audience); a change of policy towards other control strategies (expanding restrictions on the publication of particular types of content); as well as the recognition that the practice of blogging was rapidly losing ground to other forms of online expression, most notably the shift to social media and video sharing platforms. Around the same time, the government attempted to co-opt some of these online “influencers.” Popular vlogger Sasha Spilberg was invited to address the State Duma in May 2017, and soon after a special “bloggers council”—in full, *Sovet po razvitiu informacionnogo obščestva i sredstv massovoj informacii* (Council on the Development of Information Society and Mass Media)—was convened on the initiative of Vladimir Vlasov, the youngest member of parliament. The council got off to a bad start since only a third of the invited bloggers took part, and the most popular Russian vloggers publicly distanced themselves from the initiative, including oppositional vloggers such as Kamikadzedead (Makutina 2017). The council has since convened incidentally, yet appears to be of limited influence and predominantly speaks out in support of governmental restrictions on online speech.

## 2.4 POLITICAL CAMPAIGNING

Political campaigns in Russia tend to be candidate-centered, rather than focusing on policy issues or political parties, a feature resulting from the constitutionally strong president and other characteristics of the electoral system (Ishiyama 2019). As an “electoral authoritarian regime” (Gel’man 2015), election outcomes in Russia are deemed important, even if the elections themselves are unfair. By extension, political campaigns are a significant feature of Russian

politics.<sup>2</sup> As is noted by Sergei Samoilenko and Elina Erzikova, “[t]he traditional boundaries between news and political advertising have eroded in Russia” and unfair practices, such as “[h]idden advertising, black PR and biased news reporting” have been a common feature since the 1990s (2017, 265). Television and print media have played an important role in political campaigning and media ownership is generally seen as an important factor in explaining election outcomes, most notably Boris Yeltsin’s victory in the 1996 presidential elections.<sup>3</sup>

The parliamentary elections of 2011 were the first in which the Internet played a role of significance in how election campaigns were run, resulting from both the increase of Internet access and the expansion of online party presence in the years preceding it (Roberts 2015; Samoilenko and Erzikova 2017). While party websites appeared already at the time of the 1999 parliamentary elections, by 2011 political campaigning via social networking sites had become a common feature. *Edinaâ Rossiâ* (United Russia), as the “party of power,” was particularly prolific and was active on multiple platforms in large measure because it had access to the resources needed to finance investing in the online dimension of its campaign. For example, the party’s Twitter account (er\_2011) “issued an average of over 360 tweets per day during the intensive campaign period—more in a single day than the LDPR [Liberal Democratic Party of Russia, led by conservative nationalist Vladimir Zhirinovskiy, M.W.] and Yabloko [party of social-liberal orientation, M.W.] managed in the whole of the campaign, literally swamping the tweets from other parties,” while amassing 600 videos on its YouTube channel by December 2011 (Roberts 2015, 1235).

On the candidate level, however, a different picture emerges. Analyzing the online campaigns of 910 candidates representing the seven political parties that were successfully registered for the elections, Sean Roberts found that only 111 of them (12%) maintained either a website, a Twitter account or a LiveJournal blog, while this percentage was markedly higher among United Russia candidates (43%) (Roberts 2015, 1236, 1238). However, a significant number of these accounts were dormant during the campaign period, suggesting “that United Russia candidates were being forced to use social networks by the party leadership making them at best reluctant web users, at worst ‘dissenters’ by deliberately failing to maintain their accounts” (Roberts 2015, 1245). Notwithstanding United Russia’s more extensive online activities, Roberts found “evidence of equalization [a relative leveling of the political playing field in favor of opposition parties, M.W.], as the online message of the remaining party candidates converged on an anti-United Russia theme” (Roberts 2015, 1229).

The availability of resources appears to be a key determinant in whether a party decides to invest in developing online campaigning strategies. In this respect, a clear difference has emerged between the campaigning style of United Russia, whose “campaigns have become increasingly professionalized and digitized, with expansive media campaigns funded by administrative



resources” while its main competitor, the communist party KPRF (*Kommunističeskáá partiá Rossijskoj Federacii*, Communist Party of the Russian Federation), still “relies heavily on traditional methods of local party organization, voter mobilization (particularly older voters), newspaper advertisements, short television spots, public appearances by Zyuganov and other KPRF leaders, and campaign flyers and posters” (Ishiyama 2019). Of the remaining parties represented in parliament, the LDPR operates more similar to United Russia, but without the same access to large budgets, while the campaigning of *Spravedlivaá Rossiá* (A Just Russia) is more alike to the KPRF (Ishiyama 2019).

The significance of the availability of digital technologies appears to have been the greatest for opposition groups who are not represented in the Russian parliament (sometimes referred to as the “non-systemic” opposition) and who lack access to traditional media. A closer look at two campaigns run by Alexei Navalny—for the 2013 Moscow mayoral elections and 2018 presidential elections—demonstrates this well. As is argued by Renira Gambarato and Sergei Medvedev (2015), Navalny’s mayoral campaign (which build upon the 2011–2012 protest movement; see Lonkila et al. 2020) introduced a new form of political campaigning in Russia that was more grassroots (e.g. through online fundraising) and characterized by the use of transmedia strategies.<sup>4</sup> Online tools were essential for spreading information regarding his political program—as the opposition candidate, Navalny was and continues to be barred access to mainstream media, in particular federal television—and to recruit campaign volunteers (Gambarato and Medvedev, 2015). These volunteers, in turn, campaigned both on- and offline, while social media played an important facilitating role in attracting people to these offline events. While Sergey Sobyenin won the elections in the first round by garnering some 51 percent of the votes, Navalny’s 27 per cent showed the success of the campaigning strategies employed. Navalny’s 2018 presidential campaign, which built upon the momentum generated following the anti-corruption protests discussed earlier, optimized many of these strategies, incorporating sophisticated big data analysis techniques. At the same time, it invested heavily in the creation of a network of local headquarters and volunteer groups. Navalny’s campaign activities therefore show the continued mutual interdependence of online and offline campaigning, and the need to coordinate between and integrate both approaches. Contrary to the mayoral elections, the success of Navalny’s presidential campaign cannot be substantiated by election results: in December 2017, the Central Election Commission of the Russian Federation decided Navalny was not eligible to run for president because of his previous conviction in a (much contested) fraud case.<sup>5</sup>

Notwithstanding the novel opportunities for political opposition, mobilization and campaigning provided by digital technologies, it remains difficult for those acting outside of the political establishment to be elected to a post of political importance or to otherwise effectuate significant political change. Gunitsky (2015) furthermore argues that the co-optation of social media by



authoritarian regimes in fact serves as a way out of the limitations contained in the “dictator’s dilemma” that was introduced above. Social media co-optation, he argues, can serve the resilience of authoritarian regimes by enabling, among other things, the introduction of alternative frames—for example, counter to those formulated by opposition groups—to shape public discourse online.

## 2.5 VOTING

The digitalization of various aspects of the voting process was made possible by the adoption of the law “O gosudarstvennoj avtomatizirovannoj sisteme Rossijskoj Federacii ‘Vybory’” (On the State Automated System of the Russian Federation [called] ‘Elections’, no. 20-FZ, 20 January 2003). “Electronic urns,” that is, ballot boxes equipped with a special lid that scans the ballot paper when it is entered, counts the votes that have been cast and prints out the results, were first introduced in 2004 (*kompleks obrabotki izbiratel’nyh búlletenej*, referred to in Russian by the abbreviation KOIB). The systems were introduced with the stated aim to prevent miscalculations and speed up the voting process, while also preventing ballot box stuffing since only one paper can be passed through the scanner at a time. E-voting machines (*kompleks dlá èlektronnogo golosovaniá*, or KEG) were introduced on a small scale during the 2007 elections, after having been successfully tested in 2006 in an election in Veliky Novgorod. By 2018, most Russian federal districts used KOIB and/or KEG systems, albeit on greatly diverging scales; in total 11.1% of votes were counted automatically (RIA 2018).

Russia only recently trialed *remote* electronic voting, and on a modest scale: during the 2019 Moscow City Duma elections the voters of three electoral districts were given the option to vote online. The experiment did not run flawlessly. Already during the preparatory phase, the security of the system was questioned; moreover, the fact that it was run by the city of Moscow and voters’ identity and right to vote were verified by the Moscow Mayor’s portal, rather than the Multifunctional Centers for Governmental and Municipal Services normally endowed with this task, was criticized (Vasil’chuk 2019). In May 2019, the Communist Party filed a case with the Supreme Court in an attempt to prohibit the use of online voting in the Moscow elections, citing concerns about the violation of voting secrecy and the risk of manipulation and coercion of voters (Garmonenko 2019); the Supreme Court found the experiment not to be in violation of the Constitution. On the day of voting, September 8, 2019, the online voting system experienced multiple interruptions, which caused the service to be offline for periods of up to one hour (Kommersant 2019).

In the three districts where it was introduced, online voting appears to have worked in favor of pro-regime candidates who received a higher percentage of the online votes as compared to the paper votes, while the opposite was the case for opposition candidates (Uspenskiy 2019). In one of the districts that participated in the trial, the independent candidate would have won on the basis of paper votes only, yet lost the election by a mere 84 votes with the

addition of votes cast online (Vasil'chuk 2019). The explanation for the fact that pro-regime candidates fared comparatively well among those voters who voted online has yet to be determined. One thinkable scenario is that the introduction of online voting, and thereby the removal of the controlled conditions of the polling station that aim to ensure voter secrecy and freedom of choice, makes, for example, civil servants even more vulnerable to coercion. While it commonly understood state employees are placed under pressure to vote (to increase voter turnout) and support a given candidate, online voting creates the opportunity for superiors to directly supervise how their employees vote (e.g. by having them vote at the workplace). Whether and to what extent this is indeed the case, and to what extent other factors may be able to explain this difference, requires further investigation. Moreover, to be able to draw definitive conclusions on how the introduction of online voting may affect political outcomes, the empirical base needs to be extended as further trials with online voting are conducted.

Apart from the automation of voting and the gradual introduction of voting machines, the conditions under which Russians vote has changed through the placement of webcams. In response to (proven) accusations of electoral fraud committed during the December 2011 parliamentary elections, that gave cause to a series of mass protests, the government installed webcams at nearly all polling stations for the 2012 presidential elections to allow for real-time monitoring via a special website ([webvybory2012.ru](http://webvybory2012.ru)). In total, 91,000 of the 95,000 polling stations had a total of 180,000 cameras installed; of these, 80,000 were streamed online and with sound (Asmolov 2014). Webcams had been in use earlier, but only on a small scale. According to Gregory Asmolov, the actual impact of this massive infrastructural investment on increasing the transparency and, in particular, the accountability of the voting process was limited by the lack of an integrated mechanism for reporting fraudulent behavior, the impossibility of recording live-streamed footage (requiring one to file an official request to gain access to centrally stored footage from the webcams) and the ill-defined legal status of the recordings. As a result, no “criminal conviction of electoral fraud or revision of election results” were made on the basis of the videos (Asmolov 2014). Moreover, for volunteer monitors, the sheer number of available live streams made it difficult to monitor effectively. Beyond polling stations, webcams had earlier been used on smaller scale to monitor the progress on national projects in 2007, and in 2010 to monitor the reconstruction process following the wildfires. According to Asmolov, however, these initiatives symbolized rather than truly increased government transparency and accountability, as was their supposed aim (Asmolov 2014).

The 2012 presidential elections also saw the first use of a specially developed app for election observers called *Web-nablúdatel'* (web-observer) (Ermoshina 2016). The app, developed with the involvement of NGO (non-governmental organization) *Golos* (Voice), provided observers with guidance on how to conduct their activities, as well as giving them the ability to report any violations. The app was connected to a website hosting a collaborative map and statistics,

which provided novel insight into the extent and distribution of suspect and fraudulent behaviors. The aggregation of information, as well as the support the app provided for individuals volunteering to act as election observers, are important for consolidating proper election observation practices and enabling follow-up political actions.

In addition to the government, opposition forces have also turned to online voting as a means for creating legitimacy. As the 2011–2012 protest movement sought to transition from street protests into a sustained political opposition movement, an online vote was organized to elect the *Koordinacionnyj sovet rossijskoj opozicii* (Coordination Council of the Opposition) (Toepfl 2017). It was believed that this strategy would help remedy the lack of internal coherence and coordination (and as a result, credibility and legitimacy) that has undermined the success of earlier protests and opposition movements. The council was short lived, however, as the legitimacy provided by the voting process proved insufficient to remedy the fault lines within the opposition it sought to unite and was dissolved in 2013.

## 2.6 CIVIC TECH AND CIVIC ENGAGEMENT

In addition to the changes described above, digitalization has enabled new forms of political participation, among others, through the introduction of online consultation platforms. Florian Toepfl (2018, 960) proposes to categorize such digital participatory tools into four groups: tools that allow citizens to monitor policy implementation; tools enabling the public discussion of policies, measures or draft laws; tools that collect citizen preferences; and, forms of Internet voting outside of the electoral system. Above, we have already come across examples of the first—webcams used to monitor the progress of national projects—and second groups—the [regulation.gov.ru](http://regulation.gov.ru) portal for the public discussion of draft laws. The third group Toepfl identifies comprises tools that collect citizen preferences and thereby allow the government to “gauge the intensity of support for, or resistance to, planned measures or policy changes” (Toepfl 2018, 960). For example, the *Rossijskaâ obšestvennaâ iniciativa* (Russian public initiative) portal ([roi.ru](http://roi.ru)) that was introduced in 2013 allows citizens to submit an initiative to the government and cast their vote for proposals posted by others. If the initiative receives a sufficient number of votes, it will be discussed by expert working groups of the relevant federal, regional or municipal authorities (at least 100,000 signatures for proposals at the federal level or in regions with a population of over two million; or over five percent of the registered population for proposals aimed at regional and municipal governments). According to data published by the portal on the occasion of its sixth anniversary in April 2019, a total of 50,531 initiatives were submitted since its introduction, that received 17,970,021 votes in favor and 2,615,479 against (*Rossijskaâ obšestvennaâ iniciativa* 2019). The number of initiatives that led to government action, however, is limited: 33 initiatives resulted in a

decision, while 19 proposals succeeded in gathering over 100,000 votes in support.

The final group outlined by Toepfl concerns forms of Internet voting outside of the electoral system. The Active Citizen Platform of the city of Moscow, for example, allows inhabitants to decide on questions put before them by the city council; from naming metro stations and trains to school vacation dates. Citizen budgets—where online platforms are used as a tool for increasing budgetary transparency or to facilitate participatory budgeting, in which citizens have a say in the spending of state resources—are another example of this category. The city of Yakutsk, for instance, provides extensive insight into its sources of income and spending, while providing core information concerning the budgetary process ([openbudget.yakadm.ru](http://openbudget.yakadm.ru)). While, in the case of such citizen budget portals, opportunities for citizen participation are limited, participatory budgeting initiatives are more ambitious. In 2016 the city of St. Petersburg, for example, launched the *Tvoj Búdžet* project (Your Budget, [tvojbudget.spb.ru](http://tvojbudget.spb.ru)) in collaboration with the European University of St. Petersburg. Through its online portal, citizens can propose how resources should be spent in their neighborhood. Based on the total number of submitted proposals, a small number of districts (both inner city and suburbs) is then selected and allocated an earmarked budget of up to 15 million rubles for the realization of between one and three initiatives. In a special meeting, a budget committee is formed from among the initiators (by draw). The members of the committee then take part in a series of lectures to learn about, for example, urban planning and budgeting, in order to further develop their ideas. The final plans need to secure support from the district administration and be voted upon by the members of the budgeting committee in order to receive funding (Antonov 2018). One of the most visible citizen initiatives realized through Your Budget is a stretch of cycling lanes along one of the city's central canals.

Analyzing another example of the last category—the online voting to elect members for the President's Council on the Development of Civil Society and Human Rights in 2012—Toepfl argues such tools serve to strengthen, rather than weaken authoritarian rule, while simultaneously “convey[ing] to the mass public the image of transparent, accountable, and responsive government” (Toepfl 2018, 958). Studies of the use of online participatory tools by autocratic regimes elsewhere indicate that we, indeed, should not expect too much of a democratizing effect from civic tech. In China, for example, the authorities do appear to incorporate citizen input received through online consultation platforms, where a higher number of comments demanding a revision is found to increase the likelihood of the policy being revised (Kornreich 2019). In a similar vein, Jiang et al. (2019, 532) find that “cities that receive a larger number of online petitions in a year tend to devote significantly higher proportions of government reports in the following year to a topic on social welfare,” which reflects the majority of concerns expressed in the petitions. Yet, this type of citizen influence remains limited, at best, and does not necessarily translate into sustained political change or the upscaling of political participation to other/

higher levels of politics. As Yoel Kornreich explains, a certain degree of “authoritarian responsiveness” is to be expected since “[f]ailure to address citizen feedback will undermine the regime’s credibility,” while simultaneously undermining “citizens’ motivation to participate in consultation, thus depriving the authorities of an important information gathering channel” (Kornreich 2019, 549). Since legitimacy and information gathering are the main incentives for implementing civic tech, minimal functionality and effectiveness are insufficient indicators of democratization.

Whereas Toepfl’s categorization captures the governmental side of civic tech, digitalization has also enabled novel forms of civic engagement. On the local and regional levels, in particular, manifold civic initiatives (portals) have been successfully launched aimed at e-participation (e.g., urban improvement), at times acting in direct competition with government-initiated e-participation portals. Analyzing such “civic apps” in Russia, Ksenia Ermoshina argues that, while “a civic application can become a means to overcome the existing dysfunctions in communication between citizens and official institutions,” they are still best suited to solving “problems that can be easily classified and are regulated by a definite legal basis” (Ermoshina 2016, 128, 137). Successful examples include RosYama (Russian pit), an app developed by Alexei Navalny’s Anti-Corruption Foundation to map and draw attention to potholes in Russian roads or RosZKH (Russian housing and communal services) that “help[ed] individuals write petitions to the Housing Inspection Committees responsible for oversight of their particular block of flats” (Ermoshina 2014).

The two types—civic tech and civic apps—are not always perfectly separated, nor do civic apps always empower citizens vis-à-vis the state. In his study of emergency response volunteering platforms, Gregory Asmolov demonstrates the different shapes the power relations between authorities and/or platform administrators and volunteers can take. Rather than enabling more horizontal, peer-to-peer forms of (self-)organization, the way platforms for citizen engagement operate risk taking on the characteristics of “vertical crowdsourcing,” in which,

the structure of activity is defined by the institutional actor, with no space for the influence of agency on the system’s structure. In this case the purpose of the system, the boundaries, the rules, the right to participate in the community, and the division of labor are dictated by the agent who created the platform. In many cases the purpose of this type of activity system is primarily to control the activity of the crowd and to neutralize the potential for independent forms of activity. (Asmolov 2015, 311)

Instead of empowering citizens in their capacity to address societal issues, vertical crowdsourcing of resources impedes action independent of state or state-affiliated structures, who may view such citizen initiatives as threatening their position.

## 2.7 CONCLUSION

In this chapter, I have examined the impact of digitalization on Russian politics, covering the spheres of political communication, campaigning, voting, civic tech and civic engagement. From blogging politicians to online political campaigning, open government data and participatory budgeting—digital technologies evidently are shaping how politics is conducted in Russia and who can participate in and influence political decision-making. Some of the changes and initiatives I have examined are best categorized as digital replications of existing political practices or have only limited impact on political practices. The introduction of voting computers, for example, is a slow process that, thus far, does not appear to affect election outcomes. Most online participatory tools lack bite. Yet, it appears that several spheres of Russian politics have indeed been transformed as a result of digitalization. This concerns, in particular, the novel opportunities that have emerged for conducting and organizing political opposition, including political campaigning by opposition candidates, and civic engagement. At the same time, these transformations do not necessarily result in the strengthening of the democratic degree of political practices. Rather, the cases and studies reviewed in this chapter support the claim that in many cases digital tools for political participation serve to strengthen, rather than weaken, state control.

## NOTES

1. According to police estimates, some 7000 persons took part in the Moscow protest and 5000 in St. Petersburg. Several hundreds of protesters were arrested, including Navalny.
2. For an overview of campaign characteristics from 1993–2016, see Ishiyama (2019).
3. On election campaigning and changes in political advertisement, including the use of compromising materials (*kompromat*) in the period 1993–2014, see Samoilenko and Erzikova (2017).
4. Gambarato and Medvedev (2015, 176) identify a total of 32 different elements to the campaign, ranging from political advertisements, banners and stickers to distributing campaign materials on public transport and an online couch surfing service for volunteers.
5. Unlike Navalny, former socialite Ksenia Sobchak did succeed in being registered as a candidate and ran an oppositional campaign with the motto “Sobchak against all.” Boasting a massive following on Instagram, social media were at the center of her campaign. Contrary to Navalny, though, Sobchak did receive coverage on federal television and participated in the televised debates of presidential candidates (president Putin was conspicuous by his absence).

## REFERENCES

- Antonov, E. 2018. Lúboj peterburžec možet postroit' velodorožki ili otkryt' punkt obogreva za sčet gorodskogo búdžeta [Any Petersburgers Can Build Bike Paths or Open a Heating Point at the Expense of the City Budget]. *Bumaga*, February 27. <https://paperpaper.ru/tvoi-budzhet/>.
- Asmolov, Gregory. 2014. The Kremlin's Cameras and Virtual Potemkin Villages: ICT and the Construction of Statehood. In *Bits and Atoms. Information and Communication Technology in Areas of Limited Statehood*, ed. Steven Livingston and G. Walter-Drop, 30–46. New York: Oxford University Press.
- . 2015. Vertical Crowdsourcing in Russia: Balancing Governance of Crowds and State-Citizen Partnership in Emergency Situations. *Policy and Internet* 7 (3): 292–318.
- Castells, Manuel. 2012. *Networks of Outrage and Hope: Social Movements in the Internet Age*. Cambridge and Malden: Polity Press.
- Coleman, S., and D. Freelon. 2015. Introduction: Conceptualizing Digital Politics. In *Handbook of Digital Politics*, ed. Stephen Coleman and Deen Freelon, 1–13. Cheltenham and Northampton: Edward Elgar Publishing Limited.
- Deibert, Ronald, John Palfrey, Rafal Rohozinski, and Jonathan Zittrain. 2010. *Access Controlled. The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, MA: MIT Press.
- Diamond, L. 2010. Liberation Technology. *Journal of Democracy* 21 (3): 69–83.
- Ermoshina, Ksenia. 2014. Democracy as Pothole Repair: Civic Applications and Cyberempowerment in Russia. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 8 (3). <https://doi.org/10.5817/CP2014-3-4>.
- . 2016. Is There an App for Everything? Potentials and Limits of Civic Hacking. *Observatorio Journal* 10: 116–140. <https://doi.org/10.15847/obsOBS0020161088>.
- Gambarato, Renira, and Sergei Medvedev. 2015. Grassroots Political Campaign in Russia: Alexey Navalny and Transmedia Strategies for Democratic Development. In *Promoting Social Change and Democracy through Information Technology*, ed. Vikas Kumar and Jakob Svensson, 165–192. Hershey: IGI Global.
- Garmonenko, Daria. 2019. KPRF trebuet priznat' èlektronnoe golosovanie ne sootvetstvuušim Konstitucii [KPRF Demands to Recognize Electronic Voting as Not Complying with the Constitution]. *Nezavisimáá gazeta*, May 13. [http://www.ng.ru/politics/2019-05-13/3\\_7571\\_kprf.html](http://www.ng.ru/politics/2019-05-13/3_7571_kprf.html).
- Gel'man, V. 2015. *Authoritarian Russia: Analyzing Post-Soviet Regime Changes*. Pittsburgh: Pittsburgh University Press.
- Gorham, Michael. 2014. Politicians Online: Prospects and Perils of 'Direct Internet Democracy'. In *Digital Russia*, ed. Michael S. Gorham, Ingunn Lunde, and Martin Paulsen, 233–250. Abingdon and New York: Routledge.
- Gunitsky, Seva. 2015. Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability. *Perspectives on Politics* 13 (1): 42–54.
- Ishiyama, John. 2019. Russia. In *Thirty Years of Political Campaigning in Central and Eastern Europe*, ed. Otto Eibl and Milos Gregor, 391–408. Palgrave Macmillan.
- Jiang, Junyan, Tianguang Meng, and Qing Zhang. 2019. From Internet to Social Safety Net: The Policy Consequences of Online Participation in China. *Governance* 32: 531–546.



- Kommersant. 2019. Onlajn-golosovanie v Moskve dvaždy priostanavlivali iz-za sboá [Online Voting in Moscow Twice Suspended Due to a Failure]. *Kommersant*, September 8. <https://www.kommersant.ru/doc/4086901>.
- Kornia, Anastasia. 2017. Verhovnyj sud razrešil Rosreestru zasekretit' svedeniá o vladel'cah nedvizimosti [The Supreme Court Allowed Rosreestr to Classify Information about Property Owners]. *Vedomosti*, October 18. <https://www.vedomosti.ru/realty/articles/2017/10/19/738417-razreshil-zasekretit-svedeniya-vladeltsah-nedvizhimosti>.
- Kornreich, Yoel. 2019. Authoritarian Responsiveness: Online Consultation with 'Issue Publics' in China. *Governance* 32: 547–564.
- Lonkila, Markku, Larisa Shpakovskaya, and Philip Torchinsky. 2020. The Occupation of Runet? The Tightening State Regulation of the Russian-Language Section of the Internet. In *Freedom of Expression in Russia's New Mediasphere*, ed. Mariëlle Wijermars and Katja Lehtisaari, 17–38. Abingdon and New York: Routledge.
- MacKinnon, R. 2011. Liberation Technology: China's 'Networked Authoritarianism'. *Journal of Democracy* 22 (2): 32–46.
- Makutina, Mariia. 2017. Sovet bez blogerov [Council Without Bloggers]. *RBC*, June 19. <https://www.rbc.ru/newspaper/2017/06/20/5947e6469a794764344b7a27>.
- Peterson, D. J. 2005. *Russia and the Information Revolution*. Santa Monica, CA: RAND Corporation.
- Podshibyakin, A. 2010. *Po živomu. 1999–2009: LiveJournal v Rossii* [Live Cut. 1999–2009: LiveJournal in Russia]. Moscow: KoLibri.
- RIA Novosti. 2018. Demokraciá v bezopasnosti [Democracy Is Safe]. February 21. <https://ria.ru/20180221/1514773317.html>.
- Roberts, Sean P. 2015. Online Campaigning in Russia: Evidence from the 2011 State Duma Election. *Europe-Asia Studies* 67 (8): 1228–1250.
- Rodina, E., and D. Dligach. 2019. Dictator's Instagram: Personal and Political Narratives in a Chechen Leader's Social Network. *Caucasus Survey* 7 (2): 95–109.
- Rossijskaá obšestvennaá iniciatiiva [Russian Public Initiative]. 2019. Portalu Rossijskaá obšestvennaá iniciatiiva ispolnáetsá 6 let! [Russian Public Initiative Portal is 6 Years Old!], April 2. <https://www.roi.ru/news/79/>.
- Samoilenko, Sergei A., and Elina Erzikova. 2017. Media, Political Advertising and Election Campaigning in Russia. In *Routledge Handbook of Political Advertising*, ed. Christina Holz-Bacha and Marion R. Just, 253–268. New York: Routledge.
- Smyth, Regina. 2016. Studying Russia's Authoritarian Turn. *Russian Politics* 1 (4): 337–346.
- Soldatov, Oleg. 2019. Half-hearted Inception, Miserable Existence, and the Untimely Death of the Bloggers' Register in Russia. *Israel Law Review* 52 (1): 61–75.
- Toepfl, Florian. 2012. Blogging for the Sake of the President: The Online Diaries of Russian Governors. *Europe-Asia Studies* 64 (8): 1435–1459.
- . 2017. From Connective to Collective Action: Internet Elections as a Digital Tool to Centralize and Formalize Protest in Russia. *Information, Communication & Society* 21 (4): 531–547.
- . 2018. Innovative Consultative Authoritarianism: Internet Votes as a Novel Tool to Stabilize Non-democratic Rule in Russia. *New Media & Society* 20 (3): 956–972.
- Uspenskiy, Alexander. 2019. Èlektronnoe golosovanie v Moskve vyigrali kandidaty ot vlasti. V oflajn-golosovanii u nih net takogo perevesa [Candidates from the



- Authorities Won the Electronic Voting in Moscow. They Have No Such Advantage in Offline Voting]. *TJ*, September 9. <https://tjournal.ru/analysis/115536-elektronnoe-golosovanie-v-moskve-vyigrali-kandidaty-ot-vlasti-v-oflajn-golosovanii-u-nih-net-takogo-perevesa>.
- Vaccari, Cristian. 2013. *Digital Politics in Western Democracies: A Comparative Study*. Baltimore: Johns Hopkins University Press.
- Vasil'chuk, Tatjana. 2019. 'Miški' na servere. Kak onlajn-golosovanie privelo v Mosgordumu kandidatov, podderžannyh 'Edinoj Rossiej' ['Bears' on the Server. How Online Voting Brought Candidates Supported by 'United Russia' to the Moscow City Duma]. *Novaya gazeta*, September 12. <https://novyagazeta.ru/articles/2019/09/12/81950-mishki-na-servere>.
- Yagodin, Dmitry. 2012. Blog Medvedev: Aiming for Public Consent. *Europe-Asia Studies* 64 (8): 1415–1434.
- Zvereva, Vera. 2020. State Propaganda and Popular Culture in the Russian-Speaking Internet. In *Freedom of Expression in Russia's New Mediasphere*, ed. Mariëlle Wijermars and Katja Lehtisaari, 225–247. Abingdon and New York: Routledge.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# E-Government in Russia: Plans, Reality, and Future Outlook

*Daria Gritsenko and Mikhail Zherebtsov*

## 3.1 INTRODUCTION

Digitalization is a “Faustian bargain” for the state (Owen 2015, 15). On the one hand, it lends a promise to raise the efficiency of public administration by increasing the speed of bureaucratic processes and decreasing their cost. On the other hand, it poses a challenge to preserving the power, authority and control, threatening the public system to be “disrupted” by the new actors who previously had a limited opportunity to participate in public policy. For the government, exploring new forms of governance that relies on new Information and Communication Technologies (ICT) is arguably a way to navigate this bargain. As a result, in late-1990s, a new concept—electronic government or simply *e-Government*—became prominent on the agenda of government reformers (Heeks, R., and S. Bailur. 2007).

According to Layne and Lee (2001, 123), e-Government is “government’s use of technology, particularly web-based Internet applications to enhance the access to and delivery of government information and service to citizens, business partners, employees, other agencies, and government entities.” E-government borrowed heavily from applications and managerial approaches that originated in the private sector (Systems, Applications, and Products

---

D. Gritsenko (✉)

University of Helsinki, Helsinki, Finland

e-mail: [daria.gritsenko@helsinki.fi](mailto:daria.gritsenko@helsinki.fi)

M. Zherebtsov

Carleton University, Ottawa, ON, Canada

e-mail: [mikhail.zherebtsov@carleton.ca](mailto:mikhail.zherebtsov@carleton.ca)

[SAP], enterprise resource planning, portfolio analysis, and the like). This reliance on private-sector, market-based techniques provoked conversations that e-Government is a digitally enhanced version of the “new public management” (NPM), an ideology and a number of more and less successful reforms that were implemented across the world in pursuit of greater government efficiency, the reduction of cost of public administration and improvement of public services by making the public sector more businesslike (Homburg 2004). Other scholars considered “digital era governance” as a reaction (“course-correction”) to the new public management through the re-integration of processes and functions disintegrated in the course of NPM reforms (Dunleavy et al. 2006). ICT is, in short, an option for the government to remain in control while lowering the cost of bureaucratic government.

This chapter traces the development of e-Government in Russia from 2002 to 2020 through the lens of public administration reform. Whereas in many countries digitization of the public sphere was implemented on an already developed and properly functional government apparatus, in Russia both reform projects coexisted for quite some time. The public administration reform (2003–2013), led by the Ministry for Economic Development (then the Ministry of Economic Development and Trade), at its early stages was primarily focused on developing a new vertically integrated government infrastructure, reducing the burden of administrative redtape and over-regulation as well as streamlining the bureaucratic *modus operandi*. At the early stages, it was more intertwined with the civil service reform, controlled by the Presidential Administration, than with the initiatives in the sphere of Information and Communication Technologies (ICTs) that were championed by the Ministry for Communication (then Ministry for Communication and Mass Media). The overlap between the two major reforms created internal tensions that affected e-Government development trajectory. As a result, in the context of digitization and global e-Government development, the Russian case appears to be a peculiar instance.

Since its inception, the dynamics in e-Government development in Russia has been fluctuating (Zherebtsov 2019, 603). Only in 2012 the outcomes of activities pursued by the government became detectable, with Russia improving its United Nations (UN) e-Government ranking to place 27, having started at 58th place in 2003, and improving its eParticipation index for the same period from 0.05 to 0.65 (<https://publicadministration.un.org/egovkb>). After 2012, the development stagnated again. By 2016, the progress of e-Government included user-facing advancements (such as implementation of the Multi-Function Centers and a Unified Portal for public services ([www.gosuslugi.ru](http://www.gosuslugi.ru)), as well as introducing common services online such as identification, authentication, and payments systems) and “back office” solutions (setting up infrastructure to link different government institutions and establishing national databases) (Petrov et al. 2016, 5). Yet, the citizen uptake of many electronic services remained slow, some legislative changes were missing, and a significant part of the “back office” remained analogue (Petrov et al. 2016).

Despite the ambitious plans and strategies, the implementation of e-Government in Russia still lags behind most of the European countries. We argue that the level of implementation fell short of projected goals because of the resistance of the incumbent public administration system, but also due to the discrepancy of e-Government ideas and ideals between the members of the governing elite.

The chapter proceeds as follows. First, we introduce general considerations on the digital transformation of government. Next, we discuss the stages of e-Government unfolding in Russia, paying attention to both progress and problems. We mainly discuss the federal reforms, allocating only brief remarks to the regional and local dimension of the process. The conclusion provides an assessment of the past e-Government reforms and an outlook for the near future.

## 3.2 DIGITALIZATION AND GOVERNMENT—WHY AND HOW?

### 3.2.1 *Motivations for e-Government Uptake*

Garson (2006) put forward four theories to analyze the uptake of digital technology by the governments. First, technological determinism postulates that technology is a way (or even *the* way) of achieving change. It sees technology as an unstoppable force to which everyone, including governments, has to adapt. Second, the reinforcement theory suggests that technology tends to reinforce the existing power structure. ICT has no “magic powers,” but it is a tool of control and domination that governments can deploy to maintain their authoritative position. Third, the systems theory assumes that while technology does not prescribe change, it is the main force that enables change. ICTs can be used to integrate organizations, to achieve higher levels of efficiency, to improve performance, and this motivates governments to deploy them. Finally, the sociotechnical theory suggests that human factors determine the outcomes of technological change. ICT can be developed to support centralization or decentralization, democracy or autocracy, hierarchy or networks, depending on the design choices made by whoever develops and implements the system. Following the recent advances in sociotechnical change theorizing, we suggest that any technology is not implemented in a vacuum but rather embedded in a sociopolitical context and that individual practices and perceptions are indicative of the contextual sociotechnical change. Thus, the uptake and functioning of digital technology in public administration will depend on the (political) context upon which this technology is superimposed.

In the context of non-democratic political regimes, a further theory of government digitalization has been proposed. Maerz (2016), bridging the reinforcement and the sociotechnical theories, has argued that e-Government is used by competitive authoritarian regimes, such as Russia or Kazakhstan, as a tool for gaining internal legitimacy. She suggested that e-Government allows to “simulate” transparency and participation, offering the citizens a number of services and engagement opportunities which, nevertheless, remain a façade covering the authoritarian core. The study concluded that e-Government

facilities shall not be viewed as a sign of democratization, but rather a tool of legitimation that helps preserving authoritarianism. Examining the Chinese example, Ma et al. (2005) argued that e-Government can simultaneously strengthen administrative control and promote economic development without empowering individual citizens in a democratic sense. In addition, concerns have been raised with regard to privacy and data protection practices that accompany digitalization of non-democratic states (Seifert and Chung 2009). Greitens (2013) suggested that “authoritarianism online” rests on three building blocks: control over the online content, citizen surveillance via online tracking, and the promotion of regime goals through various internet applications. E-Government features prominently in both surveillance and regime promotion, making it valuable in an authoritarian context (Stier 2015). Summing up, there is a potential complex of motivations to adopt e-Government, and those have been decoupled from the early “democratizing” perspectives.

### 3.2.2 *Stages of e-Government Development*

Layne and Lee (2001) put forward four stages of a growth model for e-Government: (1) cataloguing, (2) transaction, (3) vertical integration, and (4) horizontal integration. The first stage starts when a government opens simple websites that tell about the government, its structure, and functions. Next, the experimentation of public sector with digital tools proceeds to transactions, an interaction model where the user (citizen) interacts with the government via an electronic interface (service portal on a government website or mobile application) to receive public services ranging from a healthcare appointment to filing a tax declaration or registering a marriage. Government remains a service provider for citizens and businesses, but their interaction is “virtual” and online rather than in-person. The third stage is marked by deeper cooperation between various government departments. Different levels of government are also integrated, so that a citizen can contact one governmental body and complete any level of governmental transaction, often referred to as a “one-stop shopping”-public service provision.

The fourth stage of government digitalization is often connected to the idea of “government-as-a-platform” (GaaP). The concept of GaaP was coined by Tim O’Reilly (2011), a US (United States) based author, futurist, and entrepreneur, who envisioned significant benefits from shifting from “state as a provider” to “state as an enabler” of services. GaaP, which differs from previous e-Government initiatives in that the core digital infrastructure is shared between public and private sectors, is not a “platform for *government*,” but a platform for *governance*, where government is one of the participants, service producers, and innovators. A similar idea has been presented by Linders (2012) as “we-governance” and by Janssen and Estevez (2013) as “lean government”—government provides a platform on which stakeholders deliberate, while the public authorities retain their “orchestrating” functions. Another related concept, government 2.0 (analogous to web 2.0), was proposed by Taewoo Nam (2012)

who has been advocating for crowdsourcing, Application Programming Interfaces (APIs), and “citizen hacking” as means to improve the democratic quality and efficiency of government.

GaaP can be seen as a new package of ideas imported to public sector from business management. This time, the intellectual roots are in the “disruption theory” originating from the work of Christensen et al. (2015), which has become a mantra of Silicon Valley. Disruption stands for “a form of libertarianism deeply rooted in the technology sector, a sweeping ideology that goes well beyond the precept that technology can engage social problems to the belief that free market technology-entrepreneurialism should be left unhindered by the state” (Owen 2015, 7). The proponents of the concept emphasize its interactive character and the enabling potential (citizens as co-producers of public services) (O’Reilly 2011). Building new services from scratch means also that the old bureaucratic practices are not simply transferred into a digital form, but rather that procedures are renewed. The critics argue that the changing relationship between the state and society mediated by “big data,” software code and algorithms is a form of technocratic “solutionism” that effectively undermines democratic governance (Williamson 2016).

### 3.3 RUSSIAN GOVERNMENT’S DIGITALIZATION STORY

#### 3.3.1 *Towards an e-Government (2002–2009)*

In the early 2000s Russia’s backwardness in the field of digital technologies was obvious to the new Russian leadership with the public sector demonstrating almost no signs of progress in this sphere. While global leaders were gradually transitioning to the new digitization agenda, Russia only had to conduct a full-fledged public sector reform. This prompted the reformers to launch both reforms simultaneously, yet independent from each other. Under the Federal Target Program (hereafter FTP), “*Elektronnââ Rossiâ (2002–2010)*” (Electronic Russia) e-Government was first developed as a separate reform. In its initial stage, the concept embraced a large agenda of democracy promotion, a significant modernization of the general ICT infrastructure, including its public sector component. The approach seemed reasonable as both required substantial development before they could be merged. The “Electronic Russia” program included a full spectrum of measures, necessary to build the complex government Information Technologies (IT) infrastructure. Particularly, the measures included the development of the systems of identification and authentication as well as digital (paperless) workflow. In addition, the program prescribed the development of solutions to integrate various independently built state information systems to ensure a complex services delivery through the multifunctional centers. Yet in the first years of the program implementation the only visible result of the reform was the increased Internet presence of the federal government bodies through a network of interconnected departmental websites. The actual building of the e-Government infrastructure had not

begun almost until the end of the program. Throughout its implementation the program was plagued by multiple drawbacks, including critical underfunding, lack of coordination, inefficient use of budget funds as well as a comparatively low prioritization and insufficient political attention to the reform. Since its launch, the “Electronic Russia (2002–2010)” Program was revised at least five times, substantially narrowing down its scope and ambitious plans due to both, a very ambitious and loosely coordinated agenda as well as inefficiency of reform management and misappropriation of funds (Rudycheva 2011, Polenova 2011).

Only by 2006, reformers managed to complete the development of key nodal elements of the government IT infrastructure of the government—State Automated (Information) Systems “*Vybory*” (Elections, <http://www.cikrf.ru/gas/>), “*Pravosudie*” (Justice, <https://sudrf.ru/>), “*Zakonotvorčestvo*” (Lawmaking, <http://parlament.duma.gov.ru/>), and “*Upravlenie*” (Administration, <http://gasu.gov.ru/>)—and proceed to designing elements of e-Government, particularly the Single Portal of State and Municipal Services ([www.gosuslugi.ru](http://www.gosuslugi.ru)), launched in 2010. These systems automate certain significant political and administrative processes. Although being independent from one another and focused on specific tasks, these systems constitute the information backbone of any electronic government and their successful launch and further utilization demonstrate a significant step forward in regard with digitization of the government sphere. The overall inefficiency of the program was acknowledged by both the country leadership and key experts. In order to increase the effectiveness of the Program, in 2008 the Ministry of Communications of Russia conducted a review of the implementation of the Program. According to the report, many of the objectives of the Program have not been achieved. In particular, interdepartmental electronic interaction was not actually realized. In addition, standardization of IT solutions was not widely used, leading to the situation when the created hardware and software systems were not used to their full potential due to the lack of systems interoperability.

In this regard, in 2009, the Program was restarted and complemented by the independent “Conception of e-Government development until 2010,” emphasizing the strategic priority of e-Government. This was an important shift towards the recognition of the leading role of IT solutions in the future modernization of the national public sector. This restart coincided with the beginning of the presidency of Dmitry Medvedev that was marked by several modernization efforts. During 2008, a legal review had been conducted and new federal laws prepared. On February 9, 2009, the Federal Law 8-FZ “*Ob obespečenii dostupa k informacii o deâtel’nosti gosudarstvennyh organov i organov mestnogo samoupravleniâ*” (On the access to information on the activity of the state and local authorities) has been issued, together with an Order of the Government of Russia №478 from June 15, 2009, “*O edinoj sisteme informacionno-spravočnoj podderžki graždân i organizacij po voprosam vzaimodejstviâ s organami ispolnitel’noj vlasti i organami mestnogo*



*samoupravleniá s ispol'zovaniem informacionnotelekomunikacionnoj seti Internet*" (On the unified system of information and reference support of citizens and organizations on questions concerning their cooperation with the state and local authorities by means of the Internet), and the Presidential Decree N721 from September 9, 2009, has brought changes into the FTP "Electronic Russia 2002–2010" to enable a unified technical infrastructure for the Russian e-Government. The evaluation of the program's unsatisfactory outcomes coincided with the substantial revision of the results of the Public Administration Reform. By 2010 it was obvious that the outlined reform agenda was exhausted. Like the "Electronic Russia 2002–2010" Program, the public administration reform also failed to implement and consolidate new principles of public administration, based on the NPM approach. The initial strategy to build a triple-layer structure of functionally segregated government agencies and thus ensure organizational diversification of the Russian public sector did not come to fruition. It was planned to assign the policy creation and implementation function to ministries, the control and oversight function—to state services, and services provision function—to state agencies, which would be politically and administratively independent from each other. Instead, the reform resulted in the creation of a vertically integrated system of government with the dominant top-down vector of bureaucratic accountability. Further modernization in this direction had come to a logical standstill and required the revision of the strategy.

### 3.3.2 *Building e-Government (2011–2015)*

After six years, the Public Administration Reform had been demonstrating little evidence of improving the efficiency of the government and quality of public service. The reform failed to achieve most of the measurable targets that were laid in it. By the same token, the FTP "Electronic Russia 2002–2010" was openly regarded as a failure. In these circumstances, it has become evident that the approach to separately implement both modernization projects had proven its inefficiency. For the third phase of the Public administration reform it was decided to put the development of Information and Communication Technologies in the core of the government modernization project. Thus, Russia joined a plethora of countries in conversing its public administration into e-Government. To ensure that a bigger picture is not missed, the e-Government reform was harmonized with another overarching Federal Program, "*Informacionnoe obščestvo 2011–2020*" (Information Society, Government Decree N 1815-r from October 20, 2010), which set as its key objective the digitization of all spheres of the Russian society.

The focus of the reform was made on conversing public services, internal workflow and data government into a digital format. In the minds of reformers, e-Government would further extend the single-window access principle of public services delivery at the customer end through the united single portal of state and municipal services. The portal was aimed to provide information on



available services and government regulations, digital application forms, and payment services. To ensure access to multiple services from different federal, regional, and municipal government agencies, the portal should be integrated with the Unified Identification and Authentication System (Petrov et al. 2016, 26). Such ambitious goals determined a complete reformatting of the government IT back office.

The vector of further modernization was determined by the adoption in 2010 of the Federal Law No. 210-FZ “*Ob organizacii predostavleniiâ gosudarstvennyh i municipal’nyh uslug*” (On the organization of delivery of state and municipal services), which *de-jure* prohibited government agencies from requesting the previously collected and stored personal information of applicants. The clause made imperative interagency collaboration at least in the context of services delivery. In junction with policies to enforce the promotion of digital workflow, the main focus of the back-end modernization shifted towards the SMÈV (*Sistema mezhvedomstvennogo èlektronnogo vzaimodejstviâ*, System for Electronic Interagency Collaboration). Initially, it was perceived as an IT solution connecting the EPGU (*Edinyj portal gosudarstvennyh i municipal’nyh uslug*, Unified Portal of State and Municipal services) with similar regional portals and multi-function centers, on the one hand, with services providers—authorized government agencies, on the other. The functioning of the digital government infrastructure also prompted the development of the Unified System of Identification and Authentication in order to ensure proper user access. Finally, the approach included the synchronization of the system with the State Information Systems that were built in the previous period.

Thus, the next step in public administration reform was effectively converted into building e-Government in Russia. Yet despite such significant shift in the agenda, the overall approach seemed to remain intact. As with the earlier reform, it was decided to focus on the infrastructure development projects with the implicit expectation that they would foster policy and operational changes. In addition, the approach replicated the earlier and already proven faulty expectations that the infrastructural transformations will prompt the regions to catch up. The reformers assumed that regional government would take advantage of the developed infrastructure and utilize option of hosting its regional e-Government segments.

At the same time, reflecting on past experience, the decision was made to ensure a smooth transition to the predominantly online service delivery model. To ensure non-disruptive on-boarding, it was decided to enhance the functionalities of the already built territorial multifunction citizen service centers, which were tasked with promoting and facilitating citizen’s use of the online portal. However, since the centers were under the jurisdiction of the Ministry of the Economic Development, this decision did not eliminate the dual administrative control over the reform, which had plagued the reform process before. Under the new system, the division of authority over the reform was made as follows: the Ministry for Communication was predominantly tasked with the development of e-Government infrastructure and the Ministry for Economic

Development—with policy and oversight over the reform as well as the “off-line” on-boarding. This decision not only influenced the efficiency of coordination but also had a negative impact on the political capital necessary for the reform.

In designing the reform, key focus was made on developing normative standards, prescribing the reform’s end-points, and prioritizing infrastructure development over policy transformation. This allows defining the reformers’ approach as genuinely technocratic. Reformers refused to account to the existing capacity of the bureaucracy to influence implementation of the reform not only by slowing down its complicated and/or unfavorable aspects but also by resisting to certain policy proposals that undermine its control over certain policy domains. Following Pournelle’s famous *Iron Law of Bureaucracy* (Pournelle 2006), stating that in any organization some people work to further the organization’s goals, while others work for the organization itself, any evolutionary attempts to reduce the size of public administration or level of control over certain areas through any means of improvement and optimization, including digitization, would face with the administrative actions to curtail and diminish their effectiveness. Coupled with the lack of precise measurable indicators for the efficiency of the reform, the first reform was inadvertently set to demonstrate underperformance. Those implementation and performance indicators, proposed in the documents, did not justify the selected targets. For example, implementation has confirmed that the chosen reform methods would not lead to the conversion of 70 percent of all state and municipal services to the electronic format (Order of the Government No.2516-r, December 25, 2013).

The implementation of the reform in 2011–2013 revealed the deficiencies of the initial reform sign, as it struggled to achieve the designated goals. Despite the positive dynamics and ever-growing number of registered online citizens and users, coupled with advanced and well-designed United Portal of State and Municipal Services, the overall impact of digitization did not meet expectations. Most popular and frequently used online services were purely informational (i.e. required further offline actions to proceed) and the majority of registered online users opted for the option of simplified registration that excluded enhanced user verification and authentication. Subsequently, this permitted only limited access and functionality that, particularly, excluded the processing of payments and other operations that required the substantive utilization of personal and financial data (for more details refer to Zhrebtsov 2019).

From the operations perspective, the reformers failed to engage with regions, which in practice, resulted in the emergence of two parallel and often unsynchronized systems of e-Government portals—for the federal services, on the one hand, and for regional and municipal services, on the other. Speaking of the EPGU exclusively, less than fifteen percent of federal and less than ten percent of regional and municipal services were fully available electronically. The regular monitoring of regional e-Government development, conducted by the

Ministry for Economic Development revealed substantial discrepancy of the quality and quantity of services, available on regional portals. The reform implied the monopoly of the state-owned corporation, Rostelecom on providing hosting and infrastructure for e-Government. It was expected that regions would “rent” the provided infrastructure; yet the degree of compliance with this policy initiatives appeared to be low. Rich regions (such as Moscow, St. Petersburg) have already invested in the development of their own portals, and poor regions found the Rostelecom hosting prices too restrictive to use the infrastructure and realized that building local solutions is cheaper. Coupled with technical difficulties that affected the implementation of electronic workflow (for example, Internet bandwidth restricted access to regional databases and registries) that impacted the interagency collaboration, the first phase of e-Government reform in Russia was regarded inefficient.

As a result, substantial changes were made to the design of the reform. After conducting the inventory of existing services and analyzing users’ activities on the portal, the decision was made to focus on converting the most actively used services to a fully online format. The shift of focus from the extensive (quantity of services) to intensive (quality of services) development of the Portal was accompanied by the change from institution-oriented to user-oriented approach. Services, which were previously grouped by institutions, responsible for their delivery, started to be aggregated on the basis of user life situations, substantially improving the quality and user-friendliness of the portal.

Innovations, visible to the users, were supported by a considerable transformation of the government back-end functionality. In fact, the entire architecture of e-Government was reconsidered in order to put SMÈV—System for Electronic Interagency Collaboration—into the core of the infrastructure. In terms of the architecture design, the initial “hardware-based” approach, focused on the digitization and webification of the already existing infrastructure and processes, was replaced with the “solution based” principle that focused on supporting IT solutions fostering intra-governmental communication and data exchange. Reformers refocused on the creation of system of key IT gateways around key components of e-Government in an attempt to unite and synchronize previously developed objects of government IT-infrastructure.

The “bumpy” road to e-Government was noticed and reflected in Russia’s standing in international e-Government ratings. The e-Government development index, prepared by the United Nations on a biannual basis, marked a significant progress between 2010 and 2012, when Russia moved from 59th to the 27th place. Yet, between 2012 and 2016, Russia failed to improve its performance, falling to the 35th position with very limited positive dynamics in the index itself, allowing other countries to move forward. The situations started to improve in 2018, when the country moved to the 32nd place with substantial increase of its index score. This decade-long dynamic correlates with the ups and downs in Russia’s e-Government development process.

The period between 2011 and 2016 was marked by moderate actual growth and propagation of e-Government services. According to the official statistics,

the total number of registered users demonstrated exponential growth from just over 3 million in 2012 to 13 million in 2014, to 40 million in 2016. However, a more critical analysis reveals a quite different situation. When these data are compared with official demographics from Rosstat in the period between 2012 and 2014, the number of users registered on the EPGU appears to be less than 12 percent of the total population, older than 18 years and less than 18 percent of active internet-users from the same age group. Moreover, at least one-third of all registered users opted for the simplified registration, thus not having full access to the portal. All this reduces the number of Portal users with full and unrestricted access to only 8.3 percent of Russian citizens and 12.5 percent of internet-users.

As demonstrated by Hilov (2014), the reported data on activity dynamics was based on the number of submitted, not executed requests. According to the author, only 87% of the requests for federal services were executed and the numbers for regional and municipal were much lower—36% and 19% respectively. Services delivery also differed substantially between the top ten regions averaging 167 requests per 1000 people compared with bottom ten having only 13.8 requested per 1000 people. In addition, the quantity of recipients of fully electronic services remained relatively low during the same period of time. Only about 3.2% of Russian citizens opted for this option in 2015, while others still used the walk-in option (Dobrolyubova and Alexandrov 2016). In 2013, 63% of respondents did not interact with public authorities online because they “prefer a personal visit and personal contact” (Rosstat 2014).

In addition to the digitization of services, the e-Government reform proclaimed significant improvement of regulatory capacity of the public administration, positively affecting the business climate. It was expected that converting to the digital format would reduce the administrative and regulatory burden on business, thus enhancing the business climate and fostering economic growth. Yet existing evidence demonstrates that the business community remained disengaged with the government, despite all improvements in the IT-infrastructure. The 2015 annual report of the office of the business ombudsman to the President (Doklad 2015) stated that the government failed to impose any significant changes with respect to the existing regulatory burden. Despite the positive feedback on the EPGU, almost 52% of respondents outlined in 2015 that administrative burden has been increasing, accounting for 10 to 20 percent of the total company’s revenue. The business community indicated that the reform failed to streamline regulatory activities of the state agencies, as some still enforce regulations, the implementation of which would inevitably result in fines and other penalties.

It required a substantial review of the initial reform project in order for e-Government to catch up and become the leading form of public administration in Russia. The reform resulted in the creation of advanced and modern IT infrastructure of digital government with the most notable transformations occurring in the public services delivery aspect and particularly in the context of constant modernization of the EPGU. In this regard, late start (in

comparison with the leading countries) leveled the negative consequences of the technocratic approach. In reality, the very approach contributed to the rapid modernization of the IT infrastructure, as it did not account for how the developed infrastructure would be utilized by the bureaucratic apparatus. Nevertheless, the reform process revealed substantial flaws in the reform design and implementation, whose persistence at the following stages have the potential to become a very detrimental factor.

### 3.3.3 *Beyond the e-Government—Government as a Platform (2016–Now)*

The FTP “*Elektronnoe pravitel'stvo*” (Electronic Government) was concluded in 2016. The citizens gradually accepted the new form of interaction with regulators and bureaucrats, in particular young and middle-aged people found it convenient, and ever-growing Internet coverage (mobile first) made wider adoption possible (Shipov 2016). As electronic public services started to become normalized all over Russia, the most recent iteration of public sector digitalization—*Gosudarstvo kak platforma* (government as a platform)—had been presented as a concept in April 2018. The concept has been under development since 2016 at CSR (*Centr strategičeskikh razrabotok*, Center for Strategic Research), a think-tank curated by Alexei Kudrin, former Finance Minister and the current head of the Russian Audit Chamber, belonging to the political group of “reformers.” The document outlines how O'Reilly's concept could be transplanted into the Russian public administration. While it is not an official governmental program or strategy, it is worth noting that the leading political party “*Edinaâ Rossiâ*” (United Russia) has included GaaP into the program for the November 9, 2018, united election day in a few regions. While the idea is very new, it has already gained traction among the regional politicians and will most probably continue its way into the federal policy-making.

As discussed in Sect. 3.2.2, “government as a platform” is going a step further in comparison to e-Government, suggesting innovation in service delivery by allowing third parties to re-think public services without the direct intervention of authorities. The model for this is to provide application programming interfaces (APIs) to citizens and businesses who can innovate on the formats of service production and delivery. Hence, GaaP is “shifting services into new digital formats that will allow governments to continually gather huge reservoirs of data on citizens' everyday activities, interactions and transactions—data that can then be mined, analyzed and used as insights to shape services—whilst simultaneously encouraging citizens to become responsible participants in the coproduction and provision of those digital services” (Williamson 2014). This set of ideas can be found in the CSR “*Gosudarstvo kak platforma*” concept paper (2018). The concept links to the Digital Economy of the Russian Federation program 2018–2024 that focuses on enhanced adoption of digital technologies in economic and social spheres (for more, see Lowry 2020).

The justification of digital public administration is built around a number of explicit and implicit problem statements. First, it mentions lack of trust in state institutions. The lack of accountability and citizen control over public administration is regarded as a cause of inefficient bureaucracy. Corruption, mistakes, and heavy administrative burden are expected to be alleviated by GaaP. Second, lack of trustworthy data and ineffective, slow processes of data acquisition are considered to make the state slow to respond to various challenges. Authorities are presented as intermediaries between the citizens and their data who stall the efficiency and speed of public service delivery. The lack of horizontal, interdepartmental integration is seen as a further challenge. The resistance of the incumbent public administration system leads to “digital feudalism,” meaning that each public body develops its own digital systems and processes that are not interoperable. The concept also criticizes the Multi-Function Centers and a Unified Service Portal, which were introduced as a part of the Electronic Government program, claiming that they were a tactical win that turned into a strategic loss, since they preserve the existing inefficient system and block further development and genuinely new ways of public administration.

The CSR document is interesting because it presents GaaP as a solution to a number of problems in the current system of public administration. The concept states that poor public service delivery is the reason for the lack of innovation in Russian economy, while lack of reliable data and data analytics tools leads to suboptimal decision-making. The basic assumption is that the global competitiveness of a state is a direct consequence of the way the public administration is run, hence, introduction of GaaP is a way of ensuring Russia’s competitiveness in the global arena.

However, even more revealing is the analysis of implicit problems through the analysis of expected benefits. The two key characteristics of GaaP are being human-oriented (*čelovekoorientirovannyj*), yet human-independent (*čelovekonezavisimyj*). These are suggesting that the current system is not oriented towards the citizen but rather towards the state and its offices, while all the decisions are dependent on concrete public servants. The idea of automated, algorithmic, and big data-driven decision-making as fair, neutral, and citizen-oriented, emerges throughout the document. “Intellectual agents” (*intellektual’nye agenty*)—artificial intelligence (AI) driven decision-making algorithms—are expected to be at the core of public service. Bureaucratic process and personal responsibility in decision-making—both seen as problems of the current system—would therefore be substituted by an algorithmic process that eliminates personal contact. As a consequence, most of the public servants will be IT professionals and machine-learning specialists.

What is different in the CSR concept compared to the models developed by O’Reilly and other “visionaries” is the state-centric and hierarchical nature of developing and governing the transition to GaaP. Unfolding of the architecture, systems, and services is not simply curated by the state, but rather supervised. The state is the main developer and could involve third parties to develop additional services if it considers this necessary. There is only a marginal role for



the citizens who are re-conceptualized as users benefiting from the new GaaP. Each citizen is expected to acquire a “digital twin”—a set of data—already at birth and the amount of data constituting the digital representation of every person will grow with the time. The citizens therefore will be “datafied” (Hintz et al. 2018). Yet, no systems for citizen participation in GaaP development and maintenance are proposed. The concept lacks any instruments of accountability or citizen audit (for more on government data, see Chap. 22).

As a result, the problems outlined in the concept are not being addressed through deliberation or other forms of democratic participation, but automation and AI are taking the place of digital democracy. The word “democracy” (or its derivatives) does not emerge in the concept a single time. Focus on technology rather than democratic process is emblematic: the *technocratic* narrative of information technology as a source of increased efficiency for the state has been a prevailing ideology of the ruling elite since 2012 when Medvedev’s techno-political modernization agenda was curtailed.

### 3.4 REGIONAL AND LOCAL DIMENSION OF E-GOVERNMENT

The federal government has been the main driver of e-Government reforms and the main changes have happened at the federal level. Yet, also at the regional and local level, there have been various digital initiatives. Kabanov and Sungurov (2016, 85) studied the uptake of e-Government in the Russian regions. They argue that “the diffusion of e-Government itself was to a large extent the result of a vertical influence of the federal government.” This is well illustrated through examining different facets of e-Government. In case of public procurement, the new procurement law (94-FZ) introduced at the federal level mandated the creation of transparent and available information access. As a result, all regional governments created portals to implement the law, even though almost a half have only done so to fulfill the formal requirements (McHenry and Pryamonosov 2010). In the case of e-Government payments, there has been no unified legal provisions on their installment, hence, significant regional variation can be observed (McHenry and Borisov 2005). While today all the regional governments have Internet presence, the functionality of the websites differs considerably. Kabanov and Sungurov (2016) suggest that a more mature e-Government in a given region is a combination of several factors, including bureaucracy effectiveness, technological advancement, investment in ICT, and relatively democratic political regime. Techno-optimistic orientation of the regional governing elite, especially the governor, also seems to be important, at least judging from the cases of Sakha Republic (Yakutia, Ajsen Nikolaev), Moscow (Sergey Sobyenin), Belgorod oblast (Konstantin Poležnev), and so on.

Similar dynamics can be observed at the local level. While we have not observed relevant empirical studies in Russia, Johnson and Kolko (2010) compared the nation-level and the city-level e-Government initiatives in Central

Asia, concluding that local-level initiatives are more citizen-oriented and transparent. This probably is related to the fact that at the local level, governments are not mandated to develop electronic services or participation tools. A useful illustration is provided by the analysis of civic technology platforms, meaning digital platforms for citizen participation and engagement with the government, conducted by one of the authors. Civic technology is usually realized as an online or mobile application that allows citizen participation in urban management, planning and design through consultations, opinion polling, ratings, requesting repair, complaints, participatory budgeting, and other similar engagement forms. For the government, civic technology can perform several functions, from creating a new communication channel to get instant input on the bureaucratic performance and respond to the daily needs of the citizens with improved services, to a scalable method for collecting and analyzing popular needs, preferences, ideas, and values. According to our estimation, about half of the Russian regional capital deployed civic technology platforms over the past five years (2014–2019).

### 3.5 CONCLUDING REMARKS

This chapter traced the development of e-Government in Russia from 2002 to 2020 through the lenses of public administration reform. During the first period—2002–2009—an FTP “Electronic Russia” was launched in parallel with a major administrative reform. While there has been an overlap between the two, both reforms failed to implement the principles of New Public Management (NPM) to an extent that would yield them success. The second period—2010–2015—can be identified within the scope of the next FTP “Information Society 2011–2020,” and particularly, its key project “Electronic Government (2011–2015).” This project departed from an idea of e-Government as a complement or partial substitute to the “real” government and focused on the development of infrastructure for electronic public service delivery. Finally, the third period—2016–present—started the development of “government-as-a-platform” concept, that has so far not been implemented but raised much interest among various actors, as well as provoked debates regarding the future of data and digital infrastructures for its collection, processing, and storage.

These developments were aimed at serving several goals. The first aim was to improve the efficiency and decrease the cost of public administration, two central ideas of the NPM agenda. The projects cannot be regarded as pure “window-dressing,” as much of what has been achieved, in particular in the area of electronic service delivery, has had a positive effect on citizen–state interactions. In simple terms, for an average citizen in a non-conflictual situation, it has become more convenient, quick, and simple to communicate with government authorities. The e-Government project also had a pronounced political economy aspect as one of its goals has been to secure the country’s competitiveness internationally, appearing as a more attractive location to both



live and do business. Yet, the intentions did not match the reality and businesses noticed an increased administrative burden as a result of the innovations. Eventually, while driven by “good intentions,” the discrepancy between the plans and their implementation appeared large.

The review of the near two decades of digitalization of the public sector in Russia, performed through three consecutive federal programs/concepts, reveals an authentic style of conducting such reforms that can, at least partially, explain the observed discrepancy. First of all, there is a highly pronounced technocratism of planning and preparing the reform designs. Unlike in most democratic countries, e-Government reforms were designed with the state, rather than the citizen, at the center. Such unique style of the reforms can be regarded beneficial only for vast infrastructure-building projects, when it is important to enhance control over multifaceted implementation tasks in order to ensure a more or less balanced development of all components of the digital government infrastructure. Yet it seems that adhering to the same strategy at the following reform stages may result in multiple drawbacks and would require multiple corrections of the entire reform design.

Secondly, a significant level of centrality and directive management of the reforms is the characteristic of Russian e-Government implementation. The top-down approach was even embedded in the design of the reform. The ideas emanated from the federal center and were further adopted by the regions. There has been only limited opportunity for the subnational units to influence the progression of e-Government reforms. The initial inflexible approach did not propose cooptation strategies. Regions were given two options: to comply with the proposed solutions or to develop their own. This resulted in the emergence of two separate e-Government platforms—federal and regional. Moreover, the municipal level of self-government has been completely disregarded in the initial plans.

Finally, we identify the resistance of the incumbent public administration system (what is called “digital feudalism” in the CSR GaaP Strategy) and clash of ideas within the ruling elites with regard to the ways in which e-Government should be implemented and what is its ultimate purpose. The former is determined by the natural lack of the initiative of existing bureaucracy to adhere to the notion that digitization improved administration by reducing its size and streamlining key policies. The idea of seamless government, coupled with the reduced control over exclusive policy domains, does not sit well in the self-determination of current public administration leaders. The latter can be crudely reduced to the ideological disagreement between Medvedev, who started planning for the Electronic Government, and Putin, under whose government it has mainly be implemented.

The transition to the GaaP model has further exposed the flaws of the technocratic approach, as the emphasis is made on functional and policy changes and lesser on the transformation of infrastructure. The latter becomes necessarily distributed and uncontrollable from the single center. This undermines the entire top-down ideology of governance in Russia that critically modified the

course of the 2003–2013 public administration reform and significantly impacted the e-Government implementation at each development stage. The prolonged inability to adapt to the new principle of distributed and delegated governance over policy domains with blurred administrative boundaries will destine the new reform to follow the footsteps of its precursors.

## REFERENCES

- Christensen, C.M., M.E. Raynor, and R. McDonald. 2015. What Is Disruptive Innovation. *Harvard Business Review* 93 (12): 44–53.
- Dobrolyubova, E., and O. Alexandrov. 2016. E-government in Russia: Meeting Growing Demand in the Era of Budget Constraints. In *International Conference on Digital Transformation and Global Society*, 247–257. Cham: Springer.
- Doklad Prezidentu RF-2015. 2015. Commissioner for the Protection of the Rights of Entrepreneurs in Russia. Accessed November 11, 2019. [http://doklad.ombudsmanbiz.ru/doklad\\_2015.html](http://doklad.ombudsmanbiz.ru/doklad_2015.html).
- Dunleavy, P., H. Margetts, S. Bastow, and J. Tinkler. 2006. New Public Management Is Dead—Long Live Digital-Era Governance. *Journal of Public Administration Research and Theory* 16 (3): 467–494.
- Garson, G.D. 2006. *Public Information Technology and e-Governance: Managing the Virtual State*. Jones & Bartlett Learning.
- Greitens, S.C. 2013. Authoritarianism Online: What Can We Learn from Internet Data in Non-democracies? *PS: Political Science & Politics* 46 (2): 262–270.
- Heeks, R., and S. Bailur. 2007. Analyzing e-Government Research: Perspectives, Philosophies, Theories, Methods, and Practice. *Government Information Quarterly* 24 (2): 243–265.
- Hilov, P. 2014. Kak ulučšit' itogi raboty portala gosuslug, ili Nehitrye manipulácii s čislami [How to Improve the Results of the Public Services Portal, or Simple Manipulations with Numbers]. *D-Russia*, July 31. Accessed November 7, 2019. <http://d-russia.ru/kak-uluchshit-itogi-raboty-portala-gosuslug-ili-nexitrye-manipulyacii-s-chislami.html>.
- Hintz, A., L. Dencik, and K. Wahl-Jorgensen. 2018. *Digital Citizenship in a Datafied Society*. John Wiley & Sons.
- Homburg, V. 2004. E-Government and NPM: A Perfect Marriage? In *Proceedings of the 6th International Conference on Electronic Commerce*, 547–555. ACM.
- Janssen, M., and E. Estevez. 2013. Lean Government and Platform-Based Governance—Doing More with Less. *Government Information Quarterly* 30: S1–S8.
- Johnson, E., and B. Kolko. 2010. E-Government and Transparency in Authoritarian Regimes: Comparison of National-and City-Level e-Government Web Sites in Central Asia. *Digital Icons: Studies in Russian, Eurasian and Central European New Media* 3: 15–48.
- Kabanov, Y., and A. Sungurov. 2016. E-Government Development Factors: Evidence from the Russian Regions. In *International Conference on Digital Transformation and Global Society*, 85–95. Cham: Springer.
- Layne, K., and J. Lee. 2001. Developing Fully Functional e-Government: A Four Stage Model. *Government Information Quarterly* 18 (2): 122–136.

- Linders, D. 2012. From e-Government to We-Government: Defining a Typology for Citizen Coproduction in the Age of Social Media. *Government Information Quarterly* 29 (4): 446–454.
- Lowry, A. (2020). “Russia’s Digital Economy Program: An Effective Strategy for Digital Transformation?” In Gritsenko D. et al. (Eds): *The Palgrave Handbook of Digital Russia Studies*, Palgrave Macmillan.
- Ma, L., J. Chung, and S. Thorson. 2005. E-Government in China: Bringing Economic Development Through Administrative Reform. *Government Information Quarterly* 22 (1): 20–37.
- Maerz, S.F. 2016. The Electronic Face of Authoritarianism: E-Government as a Tool for Gaining Legitimacy in Competitive and Non-competitive Regimes. *Government Information Quarterly* 33 (4): 727–735.
- McHenry, W., and A. Borisov. 2005. E-Government Public-Private Partnerships: The ‘Gorod’ Payments System in Siberia. *Journal of Enterprise Information Management* 18 (5): 625–640.
- McHenry, W.K., and D. Pryamonosov. 2010. Emerging Electronic Procurement in Russia’s Regional Governments. *Journal of Public Procurement* 10 (2): 211–246.
- Nam, T. 2012. Suggesting Frameworks of Citizen-Sourcing Via Government 2.0. *Government Information Quarterly* 29 (1): 12–20.
- O’Reilly, T. 2011. Government as a Platform. *Innovations: Technology, Governance, Globalization* 6 (1): 13–40.
- Owen, T. 2015. *Disruptive Power: The Crisis of the State in the Digital Age*. Oxford Studies in Digital Politics.
- Petrov, O., M. Bunchuk, A. Stott, and Y. Hohlov. 2016. *Digital Government 2020: Prospects for Russia (English)*. Washington, DC: World Bank Group. <http://documents.worldbank.org/curated/en/562371467117654718/Digital-government-2020-prospects-for-Russia>.
- Polenova, I. 2011. Innovacii pod podozreniem [Innovations Under Suspicion]. *Rossijskaâ gazeta—Federal’nyj vypusk*, no. 196 (5572).
- Pournelle, J.E. 2006. Pournelle’s Iron Law of Bureaucracy [Online]. Accessed November 19, 2019. <http://jerrypournelle.com/archives2/archives2mail/mail408.html#Iron>.
- Rosstat. 2014. “Pričiny otkaza ot ispol’zovaniâ seti Internet pri polučenii gosudarstvennyh i munitsipalnyh uslug, po tipam poseleniâ i polu, po Rossijskoj Federatsii, Tablica 5.19”. Accessed July 2, 2020. [https://www.gks.ru/free\\_doc/new\\_site/business/it/fed\\_nablroc/PublishData%5CReports%5CReports\\_2014.html](https://www.gks.ru/free_doc/new_site/business/it/fed_nablroc/PublishData%5CReports%5CReports_2014.html).
- Rudycheva, N. 2011. ‘Èlektronnaâ Rossiâ: čem zakončilos’ i kto vinovat? [‘Electronic Russia’: How Did It End and Who Is to Blame?] *cnews.ru*. Accessed December 6, 2018. <http://www.cnews.ru/reviews/free/gov2011/articles/article2.shtml>.
- Seifert, J.W., and J. Chung. 2009. Using e-Government to Reinforce Government—Citizen Relationships: Comparing Government Reform in the United States and China. *Social Science Computer Review* 27 (1): 3–23.
- Shipov S. 2016. O kačestve perevoda gosudarstvennyh uslug v elektronnuû formu [On the Quality of the Transfer of Public Services in Electronic Form], December 12. Accessed November 22, 2019. <http://old.economy.gov.ru/wps/wcm/connect/economylib4/mer/resources/9ef8582f-fa8e-4f72-9476-d6ff5cea2e02/2.pdf>.

- Stier, S. 2015. Political Determinants of e-Government Performance Revisited: Comparing Democracies and Autocracies. *Government Information Quarterly* 32 (3): 270–278.
- Williamson, B. 2014. Knowing Public Services: Cross-Sector Intermediaries and Algorithmic Governance in Public Sector Reform. *Public Policy and Administration* 29 (4): 292–312.
- . 2016. Political Computational Thinking: Policy Networks, Digital Governance and ‘Learning to Code’. *Critical Policy Studies* 10 (1): 39–58.
- Zherebtsov, M. 2019. Taking Stock of Russian e-Government. *Europe-Asia Studies* 71: 1–29.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Russia's Digital Economy Program: An Effective Strategy for Digital Transformation?

*Anna Lowry*

## 4.1 INTRODUCTION

The impact of new technological innovations is all-pervasive today from altering consumer preferences in the direction of highly customized on-demand products to changing the way companies create, market, and deliver goods and services, in particular through increasing reliance on technology-enabled platforms. Currently, digital technologies are changing the business model of companies, especially in the banking and telecommunications sectors, while increasing efficiency and revealing new market opportunities. Even traditional industries increasingly employ methods for analyzing large volumes of data to make effective management decisions. The Internet of Things improves the quality of equipment operation, increases productivity of oil and gas fields, and makes urban infrastructure more energy efficient. In the next decade, the further development of such innovations as unmanned aerial vehicles (drones), augmented reality, block chain, robotics, and artificial intelligence will open up a wide range of opportunities for consumers, business, and governments (Aptekman et al. 2017).

In Russia, the digital transformation of the economy is becoming one of the main strategic directions of its development (Jakutin 2017). In his address to the Federal Assembly in December 2016, President Putin set the task of preparing a digital economy program. The President has repeatedly called attention to the challenges of Russia's digital transformation, most notably in his speech at the St. Petersburg Economic Forum in June 2017. This provided an

---

A. Lowry (✉)  
University of Helsinki, Helsinki, Finland

impetus for the subsequent discussion of the digitalization strategy at various discussion platforms in Russia. Within a month, almost all major Russian business associations and scientific communities held meetings, seminars, and conferences on digital issues. The public discussions became the basis of the organizational work on the formation of a digital transformation strategy for the Russian economy in the government's program (for more on digital government, see Chap. 3). Approved by the Presidential Council for Strategic Development and Priority Projects, the Digital Economy Program acquired the status of an official government document already in July 2017. On July 28, 2017, Prime Minister Medvedev signed a governmental order approving the program “*Cifrovaâ ekonomika Rossijskoj Federacii*” (Digital Economy of the Russian Federation).<sup>1</sup> Subsequently, national projects in 12 areas of strategic development were established.<sup>2</sup> One of these is the national program “Digital Economy of the Russian Federation,” approved by the Presidential Council for Strategic Development and National Projects<sup>3</sup> on December 24, 2018 (Pasport 2018), and created on the basis of the Digital Economy Program (2017).

#### 4.2 PUTTING “DIGITAL” IN PERSPECTIVE: THEORIES OF TECHNOLOGICAL CHANGE

Despite the widespread use of the term “digital economy,” it remains a fuzzy and contradictory concept. It is usually understood as all types of economic activity based on digital technologies, including e-commerce, Internet services, electronic banking, entertainment, and others. However, it is not clear where the precise boundary between digital and “analog” economies is now (Grammatchikov 2017). Additionally, economists note the contradiction in the term itself, suggesting that in economics, all processes have long been described, diagnosed, and projected using digits/numbers (Jakutin 2017, 32; Ivanov and Malineckij 2017, 4).

Digital transformation of the economy occurs under the influence of innovation waves (Aptekman et al. 2017, 21). The first wave of digital innovations, starting from the 1960s, involved automation of existing technologies and business processes. Starting from the mid-1990s, the rapid development of Internet technologies, mobile communications, social networks, and the emergence of smartphones have led to the widespread use of technology by end consumers. In the broader scientific context, these innovation waves, or interrelated radical breakthroughs, form a constellation of interdependent technologies defined as a *technological revolution*. Carlota Perez (2002, 2010) identifies five such revolutions since the initial Industrial Revolution in England. Each technological revolution is accompanied by a set of “best-practice” principles—a *techno-economic paradigm*—which guides a vast reorganization of economic and social institutions.

In Russian literature, digital transformation is often associated with the transition to the sixth technological order, or *tehnologičeskij układ* (Glaz'ev 1993, 2010). A technological order is defined as a complex of technologies characteristic of a certain level of development of production. Each technological order encompasses a closed cycle from the extraction of primary resources to all stages of their processing to the production of products that meet the relevant level of public consumption (Rodionov et al. 2017, 80). In this framework, digital economy is understood as a form of economic organization of society, resulting from scientific and technological progress, aimed at creating greater value with the use of technology of the sixth technological order and enabling its long-term sustainable development (Rodionov et al. 2017, 79). Digital transformation is conceptualized as the material embodiment of nano- and biotechnologies, artificial intelligence, the Internet of Things, robotics, and other modern technologies based on electronic devices (Jakutin 2017, 28). With regard to the Russian economy, its digital transformation is seen as part of a broader task of economic modernization, moving away from its raw-materials orientation.

### 4.3 RUSSIA ON THE GLOBAL DIGITAL MARKET

There are a number of studies that seek to identify the leaders of the digital economy and calculate its share in the gross domestic product (GDP) of different countries. According to the latest McKinsey study (Aptekman et al. 2017), Russia's digital economy accounts for 3.9 percent of its GDP, compared to 10.9% in the United States (US), 10% in China, and 8.2% in the European Union (EU, in 2015 prices). At the same time, digital transformation is one of the main factors of economic growth in Russia as well as globally. From 2011 to 2015, the total volume of Russia's digital economy increased by 59%, which means that it is currently growing at a rate that is 9 times faster than the country's GDP. Based on this considerable growth potential, the study suggests that it is possible to triple the size of Russia's digital economy by 2025 from the current 3.2 to 9.6 trillion rubles, which would bring Russia to the level of developed economies in terms of the relative share of digital economy in GDP (8–10%).

To assess Russia's relative position on the global digital market, it is possible to use relevant international indices. The Networked Readiness Index, developed by the World Economic Forum, measures countries' preparedness to reap the benefits of emerging technologies and to capitalize on the opportunities presented by the digital revolution (Baller et al. 2016). It is made up of four main categories—environment (political/regulatory and business/innovation), readiness (measured by information and communication technologies (ICT) affordability, skills, and infrastructure), usage (individual, business, and government), and impact (economic and social). Russia ranks 41st in the Networked Readiness Index 2016, far behind the leading countries such as Singapore, Finland, Sweden, Norway, the United States, the Netherlands,



Switzerland, the United Kingdom, Luxembourg, and Japan. Russia's relatively weak position in the ranking can be attributed to the gaps in the regulatory framework for the digital economy and the insufficiently favorable environment for innovation and doing business, and consequently, low ICT business usage (Programma 2017, 8).

Another relevant index is the International Digital Economy and Society Index (I-DESI) developed by the European Commission to measure the digital economy performance of EU28 Member States and the EU as a whole compared to 17 other countries (Wiseman et al. 2018). It is a composite index that comprises 5 dimensions: connectivity, digital skills, citizen use of Internet, business technology integration, and digital public services. Based on this index, Russia lags behind the EU average but is still ahead of China, Chile, Mexico, Turkey, and Brazil (Wiseman et al. 2018, 14). Russia ranked above the EU average in terms of human capital (digital skills) but fell behind in the other 4 dimensions. It received the lowest rating among the 45 countries in the study in terms of overall connectivity and was ranked below the EU bottom 4 in terms of business technology integration (for more, see Chap. 13).

#### 4.4 ANALYSIS OF THE DIGITAL ECONOMY PROGRAM: DEFINITIONS, GOALS, AND INDICATORS

This section provides an analysis of the program's content in terms of its definitions, goals, and indicators. It focuses on the 2017 state program as a conceptual document laying the framework for the subsequent national program (2018), which is more target oriented. The analysis also shows how the broadly formulated goals of the original program have been redefined and fine-tuned in the 2018 national program with more concrete tasks, indicators, and mechanisms of implementation.

##### 4.4.1 *Definition of the Digital Economy*

The state program defines digital economy as "an economic activity, in which the key factor of production is data in the digital form" (Programma 2017, 4–5). In classic economic theory, labor, capital, and raw materials are considered the main factors of production. In the context of innovative economy, technology and knowledge also play a key role in production. However, it is not clear why data in digital form should be considered the main factor of production (Ivanov and Malineckij 2017, 6). The authors of the program provide the following explanation: "Currently data become a new asset, mainly due to their alternative value, that is, as data are used for new purposes and realization of new ideas" (Programma 2017, 5). At the same time, the program does not specify these new purposes. A related criticism is that "data in the digital form" do not define the essence of today's digital economy since data have always



been used to describe and evaluate economic activity (Jakutin 2017, 32). A simpler and more straightforward definition of the digital economy would have been as an economy based on digital technologies. Consequently, strategic management of the digitalization processes of the Russian economy would entail, first, the management of the development of digital technologies and, second, the management of the processes of their deployment in the economic sphere (Jakutin 2017, 36).

#### 4.4.2 *Goals of the Programs*

The 2017 program outlines its three main goals as follows. The first goal is “creation of the ecosystem of the digital economy of the Russian Federation,” which ensures effective interaction between business, scientific and educational community, the state, and Russian citizens. This goal is weakly formulated and can hardly claim the status of a long-term target of government activities on digitalization. The “Strategy for the Development of the Information Society in the Russian Federation for 2017–2030” defines the “ecosystem of the digital economy” as “a partnership of organizations ensuring the continuous interaction of their technological platforms, applied Internet services, analytical systems, information systems of public authorities of the Russian Federation, organizations and citizens” (Strategiâ 2017, 5). Thus, the creation of the ecosystem of the digital economy entails the creation of “a partnership of organizations.” However, a partnership is not the main element of the digital economy (Jakutin 2017, 41). Regardless of whether enterprises-owners of digital technologies, Internet portals, and servers form or do not form a partnership, the economy does not cease to be digital.

The second goal is defined as “the creation of necessary and sufficient institutional and infrastructural conditions, the removal of existing obstacles and restrictions for the creation and (or) development of high-tech businesses and the prevention of the emergence of new obstacles and restrictions both in traditional industries and in new industries and high-tech markets” (Programma 2017, 2). This goal is too big and too compressed in its content. It can be subdivided into two separate strategic objectives: the formation of the institutional environment of Russia’s digital economy and the creation of its infrastructure.

The third goal is increasing competitiveness of Russian industries and the economy as a whole on the global market. However, this goal cannot be considered one of the directions of digitalization. Competitiveness is itself a result of the development of the digital economy. While improving competitiveness is a necessary task, it requires an active and diverse economic policy. The program lacks such a policy (Jakutin 2017, 45).

The national program “Digital Economy of the Russian Federation” (2018), developed on the basis of the 2017 program, redefines the goals as follows. The first goal is a three-fold increase in domestic spending on the development of the digital economy from all sources (by share in GDP) compared to 2017.

The second goal is “creating a sustainable and secure information and telecommunications infrastructure for high-speed transmission, processing and storage of large amounts of data that is accessible to all organizations and households.” The third goal is the use of predominantly domestic software by government agencies, local governments, and organizations. Thus, compared to the earlier program, the national digital economy program has more concrete goals. Consequently, the indicators have also been redefined accordingly. They are shown in Table 4.1.

The redefined and more concrete goals, with corresponding indicators, of the subsequent national program (2018) are a significant improvement on the original version of the program. In this regard, the shift from a very broadly formulated goal of creating the ecosystem of the digital economy to the more concrete objective of increasing domestic expenditures on the development of the digital economy, with fine-tuning of the necessary methodology, should be noted. Compared to the earlier version, the use of domestic software by government agencies is elevated to one of the main goals of the program. In the 2017 program, these measures were addressed under the rubric of information security with corresponding indicators for decreasing the share of foreign ICT equipment and software in the purchases of federal and regional government authorities and state-owned enterprises (SOEs). The new program uses different indicators for government bodies and SOEs but focuses exclusively on software, omitting ICT equipment. In sum, the program has been revised so that

**Table 4.1** Main indicators of the national program “Digital Economy of the Russian Federation” (2018)

<i>No.</i>	<i>Program indicators</i>	<i>2018</i>	<i>2019</i>	<i>2020</i>	<i>2021</i>	<i>2022</i>	<i>2023</i>	<i>2024</i>
1.1	Domestic spending on the development of the digital economy by share in GDP (%)	1.9	2.2	2.5	3.0	3.6	4.3	5.1
2.1	Share of households with broadband access to the Internet (%)	75	79	84	89	92	95	97
2.2	Share of socially significant infrastructure objects with broadband access to the Internet (%)	34.1	45.2	56.3	67.5	83.7	91.9	100
2.3	Availability of data processing centers in federal districts (quantity)	2	3	4	5	6	7	8
2.4	Russia’s share in the world volume of storage and data processing services (%)	–	–	1.5	2	3	4	5
2.5	Average downtime of public information systems as a result of cyberattacks (hours)	65	48	24	18	12	6	1
3.1	Cost share of domestic software purchased and (or) rented by federal, regional, and other government authorities, %	>50	>60	>70	>75	>80	>85	>90
3.2	Cost share of domestic software purchased and (or) rented by state corporations and companies with state participation, %	>40	>45	>50	>55	>60	>65	>70

there is a better fit between the goals, specific measures to be implemented, and target indicators. However, much of the original criticism regarding the lack of measures for streamlining the production of domestic ICT equipment remains valid. Similarly, there are no indications in the program that it is aimed at addressing import dependence in the component base of hardware or creating mechanisms to overcome the rigid sanctions regime applied to Russian high-tech companies (Jakutin 2017, 37).

#### 4.4.3 *Levels of the Digital Economy*

According to the program, the digital economy comprises three levels: *markets and industries*, where the interaction of specific subjects (suppliers and consumers of goods and services) takes place; *platforms and technologies*, where competencies for the development of markets and industries are formed; and *environment* that creates the conditions for the development of platforms and technologies and effective interaction of market actors and covers regulations, information infrastructure, personnel, and information security. The program focuses on “the two lower levels of the digital economy,” and specifically, the development of key institutions that create the conditions for the development of the digital economy (regulations, personnel and education, the formation of research and technological competencies) and basic infrastructural elements of the digital economy (information infrastructure and information security) (Programma 2017, 2–3).

The levels of the digital economy identified in the program do not correspond to the traditional micro-, meso-, and macro-levels established in economic theory (Jakutin 2017, 45). The first, “upper” level, according to the program, “markets and industries,” entails the interaction of specific subjects (suppliers and consumers of goods and services). In other words, it is the level of an enterprise or the micro-level. Referring to the micro-level as the “upper” level of the digital economy, the program puts established economic theory on its head. The two “lower” levels, according to the program, are platforms and technologies, and “the environment.”

The program states that it “focuses on the two lower levels of the digital economy” but in practice restricts itself to just one level, “the environment,” broken into two components—institutions and infrastructure (Programma 2017, 2–3). The program thus sees the basic directions of creating the digital economy as the development of various institutions and infrastructure. Omitted in this statement of objectives is the digital economy itself, or to use the program’s terminology, the entire second level—digital platforms and technologies. This omission is remarkable considering that the digital platform is generally recognized as the building block of the digital economy. It is defined as the system of algorithmic relationships of a significant number of market participants, united by a single information environment, which reduces transaction costs due to the use of a package of digital technologies and changes in the division of labor (Jakutin 2017, 47). The digital platform, thus, can

rightfully claim the status of the main “level” of the digital economy, without any reservations about the second, third or lower levels.

#### 4.4.4 *Cross-Cutting Technologies*

The program provides support for the development of “cross-cutting” technologies but does not offer a definition of this term. Nine technologies fall within the scope of the program, specifically, big data, neurotechnology and artificial intelligence, distributed registry systems, quantum technologies, new production technologies, industrial Internet, components of robotics and sensorics, wireless technology, and virtual and augmented reality technology (Programma 2017, 3). The list of technologies will be updated as new technologies emerge and develop. The program will also be supplemented with relevant sections and road maps in the process of the implementation of specific measures in the field of health, creation of “smart cities,” and public administration.

In the words of former Minister of Telecom and Mass Communications, Nikolaj Nikiforov, who presented the program at a meeting of the Council on Strategic Development and Priority Projects, cross-cutting technologies is “when a digital technology is developed once and can be used many times in various industries” (Zasedanie 2017). However, the program does not specify an economic mechanism that makes these technologies “cross-cutting.” If the technology was “once” developed by someone, what is the mechanism that will allow this technology to “get away” from its owner and find its “cross-cutting” application “in various industries”? Jakutin (2017, 50) raises a number of valid questions in this regard: Who will pay for it? Who will ensure its distribution? What about copyright and intellectual property rights? The state program does not provide any answers to these questions. The choice of the nine “cross-cutting” technologies listed in the program is likewise arbitrary. According to Sneps-Sneppe et al. (2018, 38), the nine cross-cutting technologies identified in the program represent a random collection of modern technologies, and hardly the most important ones. Furthermore, it is difficult to notice the manifestation of these technologies in the program.

Compared to the original version of the program, the revised national program (2018) represents an improvement in terms of introducing a number of concrete measures for the development of “cross-cutting” technologies, which are incorporated into the new federal project “Digital technologies.” These measures are aimed at achieving the goal of the national program to increase domestic expenditures on the digital economy and include (1) the creation of “cross-cutting” digital technologies predominantly on the basis of domestic research and development (R&D) and (2) the creation of an integrated system of financing projects for the development and implementation of digital technologies and platform solutions, including venture financing and other development institutions. The first objective encompasses a range of policies such as designing road maps for the development of promising cross-cutting digital

technologies, creation of digital platforms for conducting R&D in these technologies, support of Russian high-tech companies, which develop products, services and platform solutions on the basis of cross-cutting technologies for the digital transformation of priority industries, and forming demand for Russian digital technologies, products and platform solutions, in part by launching digital transformation of state corporations and companies with state participation.

#### 4.5 RUSSIA'S DIGITAL ECONOMY PROGRAM: MANAGEMENT SYSTEM

The program's management system can be characterized as flexible, with multiple centers of decision-making (Sneps-Sneppe et al. 2018; Ivanov and Malineckij 2017). In governance studies, a system with multiple semi-autonomous decision centers operating under an overarching set of rules is defined as polycentricity (Aligica and Tarko 2012; Carlisle and Gruby 2017). Despite the number of advantages ascribed to polycentric governance systems, including suitability for managing complex areas such as science, the concept of polycentricity has not been systematically applied in the study of innovation systems or science governance. This is somewhat surprising considering that the literature on science governance in Russia has framed the issue in terms of decentralization. At the same time, this literature acknowledges that the virtues of a decentralized science system are far from obvious in Russia or elsewhere since “[t]he best science is unapologetically elitist” (Graham and Dezhina 2008, vii). This section will briefly review these debates on the organization and support of science in Russia in the context of the Digital Economy Program. The objective is to assess the extent to which its management system resembles or differs from a polycentric structure by exploring its main attributes. These are: (1) the multiplicity of decision centers; (2) an overarching system of rules; and (3) a spontaneous order created by evolutionary competition between the various decision centers' ideas (Aligica and Tarko 2012, 254).

##### 4.5.1 *Multiple Decision Centers*

The most striking aspect of the program's management system is the multiplicity of decision centers and the range of participants involved in the program's development and implementation. The governmental commission for the use of information technologies to improve the quality of life and the conditions of doing business is responsible for the overall control over the implementation of the Digital Economy Program (Postanovlenie 2017). Its Sub-Commission for digital economy is in charge of reviewing action plans and monitoring their implementation, approving methodological recommendations and regulations as well as resolving disagreements between participants and reviewing contradictions in draft laws. Relevant ministries oversee their own areas.<sup>4</sup> The Ministry

of Digital Development, Communications and Mass Media of the Russian Federation oversees the formation of research and technological competencies,<sup>5</sup> information infrastructure, and security while the Ministry of Economic Development administers regulatory, personnel, and educational policy. 1.8 trillion rubles will be spent in 2019–2024 on the implementation of the national program for the development of the digital economy. More than 1 trillion of these funds will be allocated from the federal budget (Pasport 2018, 75).<sup>6</sup>

The Analytical Center for the Government of the Russian Federation acts as the project management office for the implementation of the Digital Economy Program. It provides organizational and methodological support for the implementation of the program, including the preparation of guidelines for the development of action plans and reports on their implementation. The Center also provides information and analytical support for the activities of the Sub-Commission and ensures the operation of a system of electronic interaction of the program's participants.

An autonomous non-profit organization (ANO) Digital Economy coordinates the participation of expert and business community in the implementation, development, and evaluation of the program's effectiveness. Created by Russian high-tech companies (Yandex, Mail.Ru Group, Rambler & Co, Rostec, Rosatom, Sberbank, Rostelecom, the Skolkovo Foundation, the Agency for Strategic Initiatives, and others), the organization functions as a platform for state-business dialogue. It forms and coordinates the activities of working groups and competence centers for the program's areas and evaluates the overall implementation of the program. In addition to ensuring the interaction with business and scientific community, its functions include support of digital technology start-ups and small/medium-sized enterprises (SME) as well as foresight and digital development forecasts.

Working groups prepare proposals for action plans and participate in evaluating the effectiveness of their implementation. Competence centers are responsible for the preparation and implementation of action plans. The ANO Digital Economy initially comprised working groups and competence centers in the following five areas: information infrastructure; formation of research and technological competencies; personnel and education; regulation; and information security. State corporations Rosatom and Rostec served as competence centers for the formation of research and technological competencies while Russian Venture Company headed the working group in this area. Russia's state nuclear corporation, Rosatom, oversaw the development of new production technologies, big data, virtual and augmented reality technologies, and quantum technologies. State corporation Rostec, which promotes the development, production and export of high-technology industrial products for civil and defense sectors, was responsible for the development of neurotechnology and artificial intelligence, industrial Internet, robotics and sensor components, wireless technology, and distributed registry systems (Sistema 2017). The competence centers and leaders of working groups for the other four areas

were the Skolkovo Foundation/MTS (regulation), the Agency for Strategic Initiatives/IC Company (personnel), Rostelecom/MegaFon (infrastructure), and Sberbank/InfoWatch (security).

#### 4.5.2 *A Single System of Rules*

The Russian government has made consistent efforts to develop an overarching set of rules governing the dissemination and use of information technologies in different spheres and to coordinate the various digitalization programs and initiatives within a comprehensive system of strategic planning. Thus, the Digital Economy Program is closely linked to the documents already in force on the strategic development of the Russian economy (Programma 2017, 4). It complements the goals and objectives of the National Technology Initiative and the adopted strategic planning documents, specifically the Forecast of Scientific and Technological Development of the Russian Federation for the Period until 2030, the Strategy for the Scientific and Technological Development of the Russian Federation (2016), the Strategy for the Development of the Information Society in the Russian Federation for 2017–2030, the priority project “Improving the organization of medical care through the introduction of information technologies” (2016), and other documents, including those of the Eurasian Economic Union. The adopted strategic planning documents provide for measures aimed at stimulating the development of digital technologies and their use in various sectors of the economy. For example, the adopted socio-economic development forecast of the Russian Federation envisions the active dissemination and widespread use of information technologies in the socio-economic sphere, public administration, and business (for more, see Chap. 3).

The Strategy for the Development of the Information Society in the Russian Federation for 2017–2030 is the closest strategic document to the Digital Economy Program in terms of content, with the goals of the Strategy being closely related to the program (Programma 2017, 4). Based on the Strategy, the program also takes into account its founding acts and legislative framework. These include the Federal Law No. 172-FZ “*O strategičeskom planirovanii v Rossijskoj Federacii*” (On Strategic Planning in the Russian Federation, 2014), “*Strategiâ nacional'noj bezopasnosti Rossijskoj Federacii*” (National Security Strategy of the Russian Federation, 2015), “*Doktrina informacionnoj bezopasnosti Rossijskoj Federacii*” (Information Security Doctrine of the Russian Federation, 2016) as well as related legal acts that determine the direction of the application of ICTs in Russia (Jakutin 2017, 30–31).

#### 4.5.3 *A Spontaneous Order?*

Despite the existence of multiple decision-making centers and an evolving overarching system of rules governing digitalization—key attributes of polycentric governance—the nature of the order generated by this system is



ambiguous and remains a subject of controversy. At the heart of this controversy is the question of whether the program's management system represents a move toward a more effective decentralized system of science governance or a step toward further bureaucratization of science. Theoretically, this question revolves around the nature of entry into the system—free, meritocratic, or spontaneous (Aligica and Tarko 2012, 254). Practically, the respective debate in Russia has centered on the role of the Russian Academy of Sciences (RAS) in overseeing digitalization.

The critics of the Digital Economy Program have been quick to note the absence of scientific organizations in its management system. They emphasize that the RAS, the main scientific organization responsible for determining research areas, including in the field of ICT, is not included in the management and implementation of the program. The absence of scientific organizations in the program's management system is seen as evidence of an established post-Soviet trend of technological development without the involvement of domestic scientific community (Ivanov and Malineckij 2017). The criticism goes further by suggesting that the program's flexible management system with multiple centers of decision making is ill suited for governing science in Russia. According to Ivanov and Malineckij (2017, 11), such an approach has been tried before and proven ineffective in managing Russia's scientific and technological complex. It leads to the growth of the bureaucratic apparatus and increases its costs while reducing the quality of policy.

An alternative view suggests that the absence of the RAS in the government's digital economy programs and initiatives is not coincidental, and that the Academy has traditionally been dismissive of Information Technologies (IT) professionals. As a result, information technologies were "pushed out" from the RAS. Currently, only a few IT sectors are represented in the RAS such as supercomputer computing and onboard software. According to Gorbunov-Posadov (2018), the academy cannot keep up with the pace of development of the IT industry, which puts its capacity to function as a universal body of national scientific expertise into question.

These opposing views were reflected in the controversial RAS reform and its public perception. The reform, launched in 2013, originally envisaged the dissolution of the RAS, which caused a negative reaction in scientific circles and led to a wave of protests across Russia. Without going into the details of the reform process, it suffices to note that significant changes in the management system of Russian science were made in 2018. The Ministry of Science and Higher Education of the Russian Federation was established in May 2018, with all institutes of the RAS subsequently falling under its jurisdiction. Amendments to the Law on Science and the Law on the RAS redefined and strengthened the role of the academy in the management system of Russian science. Specifically, the changes reaffirmed a key role of the RAS in the design and implementation of Russia's scientific and technological development strategy (Mehanik 2019).

Pursuant to the Decree of the Government of the Russian Federation No. 16 of January 17, 2018, the Ministry of Science and Higher Education formed



Councils in seven priority areas of scientific and technological development of the Russian Federation (IMEMO 2019). The first priority area and the name of the corresponding Council is “transition to digital, intelligent production technologies, robotic systems, new materials and methods of design, creation of big data processing systems, machine learning and artificial intelligence.” Its functions include formulating and monitoring of scientific and technological programs and projects in this area as well as providing expert and analytical support for the implementation of Russia’s scientific and technological development priorities. Among the members of the Council are academicians, representatives of leading research centers and universities, big business, federal executive bodies, and state corporations (RAS 2018).

Thus, the Council oversees digitalization within the framework of the Strategy for the Scientific and Technological Development of the Russian Federation but is far from the only institution responsible for the formation of Russia’s digital economy. Other programs and initiatives in this area include the Digital Economy Program and the National Technology Initiative, with their own teams and management systems. Additionally, most ministries have their own digitalization programs. Whereas critics insist that the duplication of functions and incontinency between various programs within this framework is a result of a poorly coordinated system of management (Chujkov 2019), it could also be argued that it is a result of a delicate compromise between the government, the RAS, and other stakeholders. Even though the role of the Academy has been strengthened, the existence of multiple decision-making centers prevents the monopolization of scientific expertise and allows competition between different ideas to take place. Thus, the polycentric structure of the Digital Economy Program’s management system is amplified on a broader scale of Russia’s digital economy governance where this program coexists with other digitalization initiatives.

## 4.6 CRITICISM OF THE PROGRAM AND WEAKNESSES OF THE GOVERNMENT’S DIGITALIZATION STRATEGY

### 4.6.1 *Imitation and Copying of Western Models*

In the post-Soviet economy, the practice of borrowing ideas and approaches from foreign programs has become widespread. According to Ivanov and Malineckij (2017, 4), the Digital Economy Program, which is based on the recommendations of the World Economic Forum, was no exception. This copying of Western models inevitably affects the content and quality of the program. The emphasis is not on essential, critical matters but on external issues such as places in the ratings and keeping up with technological trends. Furthermore, the program does not proceed from the ability to produce new types of products but from the interests of a “qualified consumer.” In the broader sense, the common criticism of the program is that it does not deal

with the economy as such or, more precisely, changing the technological base, which would lead to socio-economic transformations. The program focuses predominantly on the development of key institutions and infrastructure of the digital economy while “practically nothing is said about production, distribution or consumption” (Ivanov and Malineckij 2017, 4). As Loginov (2017) notes, “a lot and even too much is said about the ‘digital’ and practically nothing about the ‘economy.’” The program does not provide a clear answer as to how the “digital” would fit into the economy.

The fallacy of the catch-up logic of the program is highlighted by the government’s expert council in their conclusion on the program’s first draft. The goal of the program, according to the expert council, was not to advance Russia’s development but rather to raise the digitalization level of its economy to the current level of developed countries by 2025. This means that by that time Russia will need a new program for the development of the digital economy, since one of the fundamental characteristics of the ICT sphere is the rapid introduction of new technologies, the emergence of which cannot be foreseen today (Demidov 2017).

#### 4.6.2 *Emphasis on Services to the Detriment of Production*

Since the program is implicitly aimed at raising the digitalization level of the Russian economy to that of developed countries, it makes sense to briefly examine the industries and services that comprise the high-tech sector in developed economies. The US statistics, for example, distinguishes five high-tech manufacturing industries—pharmaceutical industry, semiconductor manufacturing, production of scientific and measuring equipment, production of communication equipment, and aerospace industry. The foundation of all these industries is electronics (Ivanov and Malineckij 2017, 8). There are also five service industries that comprise the high-tech sector of the US economy—business, financial, and communication services, education, and healthcare. Looking at the Digital Economy Program from this perspective, it is possible to conclude that it is focused on service industries while neglecting the high-tech manufacturing sector, the development of which is blocked in Russia.

One of the main criticisms of the program is that it does not provide measures for the development of Russian electronic components and systems (*èlementnaâ komponentnaâ baza*). At the same time, many of the program’s objectives require the development of electronic components (Loginov 2017). Specifically, the digital transformation of industry, or Industry 4.0, cannot occur without a national technological base, including the industry of domestic micromechanics and nanoelectronics (Sitnikov 2017). Micro-Electro Mechanical Systems (MEMS) top the list of technologies necessary for the development of Industry 4.0. In Russia, these technologies are developed within the framework of Rusnano’s programs.<sup>7</sup> Critics consider them ineffective, lamenting that Russia still has “ancient” technological competencies at the level of classical mechanics and limited laser processing capabilities. That is, it

is capable of producing parts with an accuracy of 0.1 mm on its equipment whereas the standard for global leaders in this field is 0.0001 mm.

One possible initiative in this regard could be the creation of a national 5G network based on Russian equipment (Loginov 2017). However, the program's activities in this field are limited to "assessing the capabilities" of the domestic industry to produce telecommunications equipment. As Loginov (2017) accurately points out, the domestic capabilities of building 4G networks were already assessed in 2011, but as a result, the networks were modernized using Chinese equipment. The program includes a number of target indicators for the development of domestic telecommunications industry, specifically increasing the share of domestic products in the purchases of software by federal and regional executive bodies and state-owned companies. However, in the absence of concrete measures for the revival of Russian telecommunications industry, it is unlikely that the program will meet these targets (Sneps-Sneppe et al. 2018, 39).

#### 4.6.3 *Preservation of Technological Dependence*

Most of the communications equipment and software in Russia is of foreign origin. Russia is critically dependent on the import of IT equipment (from 80% to 100% for various categories) and software (about 75%) (Aptekman et al. 2017, 43). In 2016, the volume of sales of smartphones in Russia amounted to about 30 million units; the sales of personal computers—about 5 million units. The share of products of Russian manufacturers, which are built almost completely on the basis of foreign components, is miniscule in these volumes, just a few percent (Betelin 2017, 24). As another example, the networks of Rostelecom, Russia's largest provider of digital services, have until recently been the arena of struggle between two American companies—Cisco Systems and Juniper Networks (Sneps-Sneppe et al. 2018, 37). Rostelecom's main project is a high-speed internet protocol (IP) network built entirely with the products developed by Juniper Networks.

The preservation of technological dependence runs counter to the Strategy of National Security and the Strategy for the Scientific and Technological Development of the Russian Federation (Ivanov and Malineckij 2017, 7). The critical dependence on imported components carries serious risks for the national security. It also blocks the development of many sectors of the domestic industry. The existing experience of using borrowed solutions in microelectronics indicates that Russian enterprises have access to technology and technical solutions with a lag of two or more generations, and the amount of payments for their use ranges from 30% to 80% of development costs and up to 50% in mass production (Betelin 2017, 23). This is one of the main reasons why the semiconductor industry in Russia is not significant in economic or social terms. There is a risk that the implementation of the Digital Economy Program and the related National Technology Initiative will not lead to Russia gaining any significant share of the new global high-tech markets. Without

developing domestic electronics industry, the transition to the digital economy can be considered only in the context of purchases of electronic equipment abroad, including for defense and security. This would require addressing an additional problem of “non-declared capabilities” or the detection of hidden functions of the supplied equipment, permitting unauthorized control (Ivanov and Malineckij 2017, 8).

#### 4.6.4 *Lack of Scientific Support*

One of the criticisms of the program’s management system is the absence of scientific organizations (Ivanov and Malineckij 2017, 11). With regard specifically to the ICT infrastructure, Sneps-Sneppé et al. (2018, 41) note that Russian scientific research institutes, industrial science, and professional scientists are not involved in addressing systemic issues of infrastructure development and the preparation of relevant conceptual documents. The lack of scientific support adversely affects the quality of the program, which does not provide sufficient justification for the key role of the digital economy in ensuring Russia’s economic leadership.

Available studies suggest that the products of the leaders of the global markets of semiconductors, electronic products, and software, such as INTEL, AMD, IBM, and Microsoft currently form the basis for the development of the digital economy (Betelin 2017, 24). In these conditions, the main risks and challenges for the formation of Russia’s digital economy stem from the lack of similar companies in Russia that carry proportionate economic and social weight. While the program envisions the creation of ten large high-tech companies by 2024, it lacks actual measures for stimulating domestic electronics industry and relies on modernization of the communications network based on imported equipment. Such modernization efforts are likely to result in the reduction of the size of the digital economy in Russia rather than its growth (Loginov 2017).

Even though the Strategy for the Scientific and Technological Development of the Russian Federation (2016) defines the key role of Russian fundamental science in ensuring the country’s readiness for grand challenges and timely assessment of the risks associated with scientific and technological development, in practice the program relies on the use of foreign scientific results and technologies (Strategija 2016; Ivanov and Malineckij 2017, 12). One of the stated objectives of the program is the creation of a support system for exploratory and applied research on the digital economy, which is supposed to ensure technological independence of each of the globally competitive cross-cutting technologies (Programma 2017, 11). However, relevant activities do not include basic (fundamental) research. Thus, the criticism of such an approach is that it cannot in principle ensure technological independence in ICT because new technologies can be created only on the basis of systematic results of exploratory and fundamental research (Ivanov and Malineckij 2017, 12).

#### 4.6.5 *Lack of Reliable ICT Infrastructure*

A number of studies note that the ICT infrastructure is relatively well developed in Russia, with digital services available for the majority of the country's population (Aptekman et al. 2017, 36). On this basis, some analysts even point out that it is “completely unnecessary” for the government “to try to control or stimulate this process” (Loginov 2017). This view suggests that Russian telecom companies are able to deal with the infrastructural issues on their own, at the level of their commercial needs.

Sneps-Sneppe et al. (2018) offer an alternative point of view from the perspective of telecom professionals. The basis of information and communication infrastructure, the information space of any country, is the next-generation network (NGN), which provides a user with universal broadband access to an unlimited range of ICT services. Has such an infrastructure been developed in Russia, and who is building it? The construction of next-generation networks in Russia has been carried out by private capital to make a profit from providing access to the Internet and related services. This is done without taking into account the task of creating the foundation of the country's digital infrastructure—a single telecommunications network of the Russian Federation, as required by the current law “*O svyazi*” (On Communications) and the interests of the state and society. The result, according to the authors, is the uncertainty of the architecture, location, and connectivity of the traffic exchange nodes of the composite network and the inability to manage it even in emergency situations. This “conglomerate of private fragments of the global Internet” cannot be used as an infrastructure for the networks that require high reliability and security of information exchange, which relates to the objectives of the Digital Economy Program (Sneps-Sneppe et al. 2018, 40–41). The ICT infrastructure cannot be developed solely on the commercial basis. It has to meet the needs of the state, governance, and national security, in addition to being an increasingly important factor in improving the quality of life of the citizens.

Examining the Digital Economy Program from this perspective, it is possible to make the following observations. First, despite the emphasis on the infrastructure development in the program and the key role of Rostelecom in this area, the main efforts are aimed at the provision of new ICT services. The program's activities do not include the development of technical means (Sneps-Sneppe et al. 2018, 40). The program is oriented toward the spread of the Internet and higher-level tasks such as satellite communications and 5G network without addressing the prior issue of the lack of a unified telecommunication network. Second, the risks associated with the ongoing modernization of private networks on the basis of next-generation technologies such as Software-Defined Networking (SDN), Network Function Virtualization (NFV), and 5G are not adequately addressed in the program. Third, the program focuses on the Internet, or regulation of IP packets, whereas the existing law “On Communications” is still oriented toward traditional networks and communication services. The actual meaning of such basic terms of the law as “federal

communications,” “a single (*edinaá*) telecommunication network,” and “a public telecommunications network” has changed dramatically. To date, this has not been reflected in the legal framework and mechanisms for regulating the development of the domestic telecommunications sector (Sneps-Sneppe et al. 2018, 40). Despite the long list of measures in the program aimed at improving legal regulation of the digital economy, these specific problems of the current legal framework are not addressed.

#### 4.7 CONCLUSION

The state program “Digital Economy of the Russian Federation” can be seen as the government’s latest attempt to approach the task of Russia’s modernization in new technological conditions. For Russia to fully harness the economic and social benefits of the digital revolution, digital technologies have to become the key factor in the modernization of Russian industries as well as the creation of completely new industries and markets, which requires a targeted and systemic state support based on a clear and coherent strategy. In this regard, the Digital Economy Program is an important milestone representing the Russian government’s concerted effort to envision the medium-term future of the digital economy in Russia and draft a comprehensive strategy in this area, even as it falls short in terms of its potential transformative effect on Russian industry.

Given the current state of development of domestic ICT equipment and software, the digitalization of Russian economy deserves the status of a strategic task. Such a strategic orientation, especially in the broader context of a shift from the management of hydrocarbon exports to technology governance, is extremely important. At the same time, the experience of post-Soviet development shows that the main problem lies not in ideas but in their implementation. One of the main reasons past economic initiatives were not successful is that they were made without sufficient scientific assessment based on very general considerations (Ivanov and Malineckij 2017, 3). As the analysis shows, some of the same mistakes are repeated in the case of the Digital Economy Program.

Even though the program’s management system with its multiple decision centers and an evolving overarching system of rules governing digitalization resembles a polycentric structure, which in theory is suitable for managing complex areas such as science, the advantages of this system in Russia’s case seem questionable. Alternatively, more attention should be paid to the nature of entry into this system. At present, the multiplicity of decision centers in the program’s governance structure masks the insufficient involvement of scientific organizations, which is reflected in the program’s content. The lack of scientific support adversely affects the quality of the program, which does not justify the role of the digital economy in ensuring Russia’s economic leadership or provide measures for stimulating domestic electronics industry.

Although the Strategy for the Scientific and Technological Development defines the key role of Russian fundamental science in the assessment of

challenges associated with scientific and technological development, in practice the program relies on foreign scientific results and technologies. Thus, the government attempts to address an important technological problem without using domestic scientific potential. This affects the content and quality of the program, which proceeds from the interests of a “qualified consumer” and focuses on the spread of the Internet and provision of new ICT services while neglecting the critical state of Russian electronic components and systemic issues of ICT infrastructure development.

The program is too concise and general, and consequently, does not provide sufficient justification for the key role of the digital economy in ensuring Russia's economic leadership or allow an adequate assessment of possible risks and challenges. The program defines multiple target indicators but does not provide evidence that the achievement of these indicators will reduce Russia's technological gap with leading countries. Furthermore, it lacks actual measures for stimulating domestic electronics industry and relies on the modernization of the communications network based on imported equipment. The critical dependence on imported components blocks the development of many sectors of the domestic industry and runs counter to the Strategy of National Security and the Strategy for the Scientific and Technological Development of the Russian Federation. Without developing domestic electronics industry, the transition to the digital economy can be considered only in the context of purchases of electronic equipment abroad, which is likely to result in the reduction of the size of the digital economy in Russia rather than its growth.

## NOTES

1. Unless otherwise noted, all translations are author's own.
2. Presidential Decree No. 204 of May 7, 2018 “*O nacional'nyh celáh i strategičeskikh zadáčah razvitiá Rossijskoj Federacii na period do 2024 goda*” (On the national goals and strategic objectives of development of the Russian Federation for the period up to 2024).
3. On July 19, 2018, the Council for Strategic Development and Priority Projects was reorganized into the Council for Strategic Development and National Projects (“*Ob uporiádčenií*” 2018).
4. The Digital Economy Program (2017) had five areas. In the process of its transformation into the national program (2018), the areas became federal projects and their number increased to six.
5. This area was changed to “Digital Technologies” in the national program. The federal project “Digital Public Administration” was also added to the areas overseen by the Ministry of Digital Development, Communications and Mass Media (Pasport 2018).
6. ICT analysts see this amount of funding as insufficient (Ustinova 2019). The largest amount of funds is allocated to information infrastructure whereas the funding of regulation and information security is quite modest. The final budget of the national program is also smaller compared to earlier estimates of 3.5 trillion rubles in total funding (Posypkina and Balenko 2018).



7. Rusnano was the largest investor in SiTime, “an industry leader in development of MEMS-based high-performance oscillators and silicon timing solutions” that was acquired by Megachips in October 2014 (Rusnano 2011; Yoshida 2014).

## REFERENCES

- Aligica, Paul D., and Vlad Tarko. 2012. Polycentricity: From Polanyi to Ostrom, and Beyond. *Governance: An International Journal of Policy, Administration, and Institutions* 25(2):237–262. <https://doi.org/10.1111/j.1468-0491.2011.01550.x>.
- Aptekman, Aleksandr, Vadim Kalabin, Vitaliy Klintsov, Elena Kuznetsova, Vladimir Kulagin, and Igor Yasenovets. 2017. Cifrovaâ Rossiâ: novaâ real’nost’ [Digital Russia: A New Reality]. Digital McKinsey report. [https://www.mckinsey.com/ru/~/\\_/media/McKinsey/Locations/Europe%20and%20Middle%20East/Russia/Our%20Insights/Digital%20Russia/Digital-Russia-report.ashx](https://www.mckinsey.com/ru/~/_/media/McKinsey/Locations/Europe%20and%20Middle%20East/Russia/Our%20Insights/Digital%20Russia/Digital-Russia-report.ashx).
- Baller, Silja, Soumitra Dutta, and Bruno Lanvin, eds. 2016. The Global Information Technology Report 2016: Innovating in the Digital Economy. *World Economic Forum*. <https://www.weforum.org/reports/the-global-information-technology-report-2016>.
- Betelin, Vladimir. 2017. Cifrovaâ èkonomika: navâzannye priority i real’nye vyzovy [Digital Economy: Imposed Priorities and Real Challenges]. *Gosudarstvennyj audit. Pravo. Èkonomika* [State Audit. Law. Economy] 3–4: 22–25.
- Carlisle, Keith, and Rebecca L. Gruby. 2017. Polycentric Systems of Governance: A Theoretical Model for the Commons. *Policy Studies Journal*. <https://doi.org/10.1111/psj.12212>.
- Chujkov, Aleksandr. 2019. Haos cifrovogo pravitel’sтва [Digital Government Chaos]. *Argumenty nedeli* [Arguments of the Week], February 7. <http://www.ras.ru/news/shownews.aspx?id=9fbb7ef8-7731-4734-8391-63670e4e39b9>.
- Demidov, Oleg. 2017. Ĉem programma ‘Cifrovaâ èkonomika’ pohoža na plan GOËLRO [How the Digital Economy Program Resembles the GOELRO Plan]. *RBC*, August 17. <https://www.rbc.ru/newspaper/2017/08/21/59957d2c9a794746fdafe7b4>.
- Glaz’ev, Sergej. 1993. *Teoriâ dolgosroĉnogo tehniko-èkonomičeskogo razvitiâ* [Theory of Long-term Technological and Economic Development]. Moscow: VlaDar.
- . 2010. Novyj tehnologiĉeskij układ v sovremennoj mirovoj èkonomike [New Technological Order in the Modern World Economy]. *Meždunarodnaâ èkonomika* [World Economy] 5: 5–27.
- Gorbunov-Posadov, Mihail. 2018. Cifrovaâ nauka v RAN [Digital Science in the RAS]. *Troickij variant—Nauka* [Trinity Option—Science] 249: 14, March 13. <http://trv-science.ru/2018/03/13/cifrovaya-nauka-v-ran/>.
- Graham, Loren, and Irina Dezhina. 2008. *Science in the New Russia: Crisis, Aid, Reform*. Bloomington and Indianapolis: Indiana University Press.
- Grammatchikov, Aleksej. 2017. Cifrovaâ real’nost’ [Digital Reality]. *Èkspert* [Expert] 29.
- IMEMO. 2019. Sovety po priority NTR [Scientific and Technological Development Priorities Councils]. [https://www.imemo.ru/Councils\\_priority\\_directions](https://www.imemo.ru/Councils_priority_directions).
- Ivanov, V. V., and G. G. Malineckij. 2017. Cifrovaâ èkonomika: ot teorii k praktike [Digital Economy: From Theory to Practice]. *Innovacii* [Innovations] 12: 3–12.
- Jakutin, Ju. V. 2017. Rossijskaâ èkonomika: strategiiã cifrovoy transformacii (k konstruktivnoj kritike pravitel’svennoj programmy ‘Cifrovaâ èkonomika Rossijskoj Federacii’)



- [The Russian Economy: A Digital Transformation Strategy (Toward Constructive Criticism of the Government Program ‘Digital Economy of the Russian Federation’)]. *Menedžment i biznesadministrirovanie* [Management and Business Administration] 4: 27–52.
- Loginov, Egor. 2017. Cifrovaâ ekonomika: celi i sredstva [Digital Economy: Goals and Means]. *LOYP*, November 16. <https://loyp.ru/vision/it-and-automation/digital-economy-goals-and-tools.html>.
- Mehanik, Aleksandr. 2019. RAN stanovitsâ glavnym naučnym ekspertom [RAS Becomes the Main Scientific Expert]. *Èkspert* [Expert] 5.
- “O nacional’nyh celâh i strategičeskikh zadačah razvitiâ Rossijskoj Federacii na period do 2024 goda [On the National Goals and Strategic Objectives of the Development of the Russian Federation for the Period until 2024].” 2018. Ukaz Prezidenta Rossijskoj Federacii [Decree of the President of the Russian Federation] No. 204, May 7, 2018. <http://kremlin.ru/acts/bank/43027>.
- “Ob uporoâdočeenii deat’nosti sovešatel’nyh i konsul’tativnyh organov pri Prezidente Rossijskoj Federacii [On Streamlining the Activities of Consultative and Advisory Bodies Under the President of the Russian Federation].” 2018. Ukaz [Decree] No. 444, July 19, 2018. <http://kremlin.ru/events/councils/by-council/1029/58071>.
- Pasport nacional’noj programmy “Cifrovaâ ekonomika Rossijskoj Federacii” [Passport of the National Program “Digital Economy of the Russian Federation”]. 2018. <http://government.ru/info/35568/>.
- Perez, Carlota. 2002. *Technological Revolutions and Financial Capital: the Dynamics of Bubbles and Golden Ages*. Cheltenham: Elgar.
- . 2010. Technological Revolutions and Techno-Economic Paradigms. *Cambridge Journal of Economics* 34 (1): 185–202. <https://doi.org/10.1093/cje/bep051>.
- Postanovlenie Pravitel’stva Rossijskoj Federacii “O sisteme upravleniâ realizaciej programmy ‘Cifrovaâ ekonomika Rossijskoj Federacii’” [Decree of the Government of the Russian Federation “On the System for Managing the Implementation of the Program ‘Digital Economy of the Russian Federation’”]. 2017. No. 1030, August 28, 2017. <http://government.ru/docs/29003/>.
- Posypkina, Aleksandra, and Evgeniya Balenko. 2018. Ocifrovannye trilliony: kak ocenili zatraty na cifrovuû ekonomiku [Digitized Trillions: Estimates of the Cost of the Digital Economy]. *RBC*, August 7. [https://www.rbc.ru/technology\\_and\\_media/07/08/2018/5b6863309a794771cab9e897](https://www.rbc.ru/technology_and_media/07/08/2018/5b6863309a794771cab9e897).
- Programma “Cifrovaâ ekonomika Rossijskoj Federacii” [Program “Digital Economy of the Russian Federation”]. 2017. Rasporâženie Pravitel’stva Rossijskoj Federacii [Order of the Government of the Russian Federation] No. 1632-r, July 28. <http://static.government.ru/media/files/9gFM4FHj4PsB79I5v7yLVuPgu4bvR7M0.pdf>.
- RAS. 2018. Zasedanie soвета po prioritetnomu napravleniû Strategii naučno-tehnologičeskogo razvitiâ [Council Meeting on the Priority Area of the Strategy for Scientific and Technological Development]. July 20. <http://www.ras.ru/news/shownews.aspx?id=ec653a0b-6112-4321-87d2-8d215897cd6f>.
- Rodionov, Dmitrij, Angi Shvediani, and Arsenij Bondarev. 2017. Cifrovaâ ekonomika: analiz razvitiâ v Rossijskoj Federacii [Digital Economy: Analysis of Development in the Russian Federation]. In Aleksandr Babkin, ed. *Tendencii razvitiâ ekonomiki i promyšlennosti v usloviâh cifrovizacii* [Economic and Industrial Trends in the Context of Digitalization]. St. Petersburg: Izdatel’stvo Politehničeskogo universiteta [Polytechnic University Publishing House], 68–93.

- Rusnano. 2011. RUSNANO Invests in MEMS. May 31. <https://en.rusnano.com/press-centre/news/88550>.
- “Sistema upravljeniâ realizacii programmy ‘Cifrovaâ èkonomika Rossijskoj Federacii’ [The Management System of the Program ‘Digital Economy of the Russian Federation’].” 2017. Analitičeskij centr pri Pravitel’stve Rossijskoj Federacii [Analytical Center under the Government of the Russian Federation]. <http://ac.gov.ru/files/content/14091/ce-sistema-upravleniya-pdf.pdf>.
- Sitnikov, Aleksandr. 2017. Promyšlennââ Rossiâ 4.0: Pered licom kraha [Industrial Russia 4.0: In the Face of collapse]. *Svobodnaâ pressa* [Free Press], December 4. <http://svpressa.ru/economy/article/187584/>.
- Sneps-Sneppe, Manfred, Vladimir Suhomlin, and Dmitrij Namiot. 2018. O Programme ‘Cifrovaâ èkonomika Rossijskoj Federacii’: kak sozdat’ informacionnuû infrastrukturu. *International Journal of Open Information Technologies* 6 (3): 37–48.
- “Strategiâ naučno-tehnologičeskogo razvitiâ Rossijskoj Federacii [Strategy for the Scientific and Technological Development of the Russian Federation].” 2016. Ukaz Prezidenta Rossijskoj Federacii [Decree of the President of the Russian Federation] No. 642, December 1. <http://kremlin.ru/acts/bank/41449>.
- “Strategiâ razvitiâ informacionnogo obšestva v Rossijskoj Federacii na 2017–2030 gody [The Development Strategy of the Information Society in the Russian Federation for 2017–2030].” 2017. Ukaz Prezidenta Rossijskoj Federacii [Decree of the President of the Russian Federation] No. 203, May 9. <http://kremlin.ru/acts/bank/41919>.
- Ustinova, Anna. 2019. V karman ‘Infrastruktury’ upadët bolee treti bûdžeta ‘Cifrovoy èkonomiki’ [More Than a Third of the Budget of the ‘Digital Economy’ will Drop in the Pocket of ‘Infrastructure’]. *ComNews*, February 12. <https://www.comnews.ru/content/117675/2019-02-12/v-karman-infrastruktury-upadet-bolee-treti-byudzeta-cifrovoy-ekonomiki>.
- Wiseman, Ian, Paul Foley, David Sutton, Lawrence Green, and Jake Moore. 2018. International Digital Economy and Society Index 2018. European Commission. <https://doi.org/10.2759/745483>.
- Yoshida, Junko. 2014. Japan’s MegaChips to Buy MEMS Timing Leader SiTime. *EE Times*, October 29. [https://www.eetimes.com/document.asp?doc\\_id=1324434#](https://www.eetimes.com/document.asp?doc_id=1324434#).
- “Zasedanie Soveta po strategičeskomu razvitiû i prioritnym proektam [Meeting of the Council for Strategic Development and Priority Projects].” 2017. July 5. <http://kremlin.ru/events/councils/by-council/1029/54983/work>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Law and Digitization in Russia

*Marianna Muravyeva and Alexander Gurkov*

## 5.1 INTRODUCTION

“The law is a seamless web,” states an old metaphor, meaning that law could be logically explained and that every new decision affects every legal proposition to a certain degree (Katsh 1993, 403). This metaphor, which originated in the common law context, has recently began to mean something else, that is, how we communicate and how we work with information. The shift from print to electronic information technologies provides the law with a new environment, one that is less fixed, less structured, less stable, and, consequently, more versatile and volatile. Law is a process that is oriented around working with information. As new modes of working with information emerge, the law cannot be expected to function or to be viewed in the same manner as it was in an era in which print was the primary communication medium. Going digital or online has profoundly affected the ways we practice law, as well as lawmaking and law functioning.

Russian state has been intensively digitalizing in the past decades. In 2009, Russian agencies, local governments, courts, and the Department of Justice were obliged to provide all information about their activities online, thus finalizing the process of going digital (Federal Law N 8-FZ 2009; Federal Law N 262-FZ 2008; Strategy of the Development of Information Society 2008). The first steps toward legal provisions for using digital information came in 1984, when the Union of Soviet Socialist Republics (USSR) issued its standard for unified systems of documentation—*GOST*—that outlined requirements for documents stored or created using computer technologies (USSR State Committee on Standards 1984). The 1984 standard responded to increasing

---

M. Muravyeva (✉) • A. Gurkov  
University of Helsinki, Helsinki, Finland  
e-mail: [marianna.muravyeva@helsinki.fi](mailto:marianna.muravyeva@helsinki.fi); [alexander.gurkov@helsinki.fi](mailto:alexander.gurkov@helsinki.fi)

demand on behalf of the Soviet legal system to handle electronic documents following the State Commercial Arbitration Court's guidance on the usage of e-documents as evidence and the Supreme Court's ruling allowing the use of e-documents in litigation and pleading (The State Arbitrage of the USSR 1979; The Plenum of the Supreme Court of the USSR 1982). The country entered the 1990s equipped with relevant legislation, which continued to be in force even during profound political and legal reforms. Taking a course toward democracy, access and openness of information became primary principles of legislation, at least on paper. At the same time, pressures from transitions to a market economy pushed legislation to accommodate models of electronic commerce, facsimile and electronic signatures, and other digital means of transactions (Art. 160.2 and 434.2 of the Civil Code of the Russian Federation; Federal Law N 1-FZ 2002). By the time the concept of open government, that is that citizens have the right to access the documents and proceedings of the government to allow for effective public oversight (Evans and Campos 2013), gained the attention of the Russian government in the late 1990s, Russian society and state agencies had sufficient experience in working with electronic documents and a good level of computer literacy (Vinogradova and Moiseeva 2015; Fedorov 2009).

Scholars call the increasing of electronic document processing “technicalization” or “electronification” (Gilles 2014). Legal scholarship, both in the sub-fields of law and technology (i.e., cyberlaw) and law and society (i.e., sociolegal studies), has struggled with theorization and analysis of technological change. Though largely ignored in sociolegal studies, the law's relationship to technology is central to the field of cyberlaw, where it is portrayed as linear: a new technology is presented to society and the law must move quickly to respond to the disorder technology creates (Jones 2018). The debate on “technological exceptionalism” in cyberlaw was started by Ryan Calo, who explained that technological exceptionalism occurs

when [a technology's] introduction into the mainstream requires a systematic change to the law or legal institutions in order to reproduce, or if necessary, displace, an existing balance of values. (Calo 2015, 552)

For any national legal system, this means that law needs to adapt to new technologies, which poses the question of to which degree this adaptation influences legal contents and legal values (Keen 2010). This question is specifically important for the Russian legal context in connection with contemporary problematic approaches to governance and democracy.

In this chapter, we will focus on legal transformations as a result of two important developments in Russia: Russia's adaptation of the concept of open government and Russia's joining digital economy. Both processes led to the development of e-justice, that included not only digitalization of legal documents, but development of new legal digital platforms, provision of safe legal environment for economic transactions online (such as blockchain) and

necessity to establish new means of internet control in relation to cybercrime and data protection.

## 5.2 OPEN GOVERNMENT PROJECT AND DIGITALIZATION OF LAW

In Russia, the concept of open government was introduced in 2002 by the federal target program “*Èlektronnaâ Rossiâ*” (Electronic Russia). The document stated that it aimed at

improving the quality of mutual communication between the state and society by expanding the access to information about activities of the state agencies, improving efficiency of providing state and municipal services, introducing unified standards of population services. (Federal Target Program “Electronic Russia” 2002)

The program followed the notion of open government as closely related to information status, where more information is published and, at some stage, the quality of information is an indicator of such openness. The program first provided legal foundations for extensive utilization of information and communication technologies (ICT) in regard to open government and available data, as well as increased communication among all stakeholders. At the time, the idea was closely linked with four major dimensions in open government: service provision to citizens and businesses, government performance improvement, social inclusion and development, and e-democracy and participation (Evans and Campos 2013). Russians quickly learned to be digital citizens (Rasskazova and Soldatova 2014). Digital citizens are generally identified as “those who use the Internet regularly and effectively” (Mossberger et al. 2008). Not only this, but digital citizenship means the ability to use technology competently; to interpret and understand digital content and to assess its credibility; to create, research, and communicate with appropriate tools; to think critically about the ethical opportunities and challenges of the digital world; and to make safe, responsible, respectful choices online (Ribble 2015). This became evident when digital platforms started working in Russia by 2010. The “Electronic Russia” program experienced a number of problems, including funding and absence of efficient cooperation between relevant agencies (Irkhin 2007). However, it provided a framework for development of e-platforms that facilitated access to state services and, as part of it, legal services. One of the first platforms—*Edinyj portal gosudarstvennyh uslug i funkcij* (Public Services Portal, <https://www.gosuslugi.ru/>), or *Gosuslugi* (StateService) for short—which started running in 2010, provided initial access to legal services such as facilitating the issuance of a variety of ID papers (international and domestic passports, driving license, and so on), or access to any court’s decision in relation to them. Russians were initiated into e-law by allowing them to review and pay traffic and other penalties via *Gosuslugi* online without dealing with the authorities in person. These days, the majority of state and law-related

actions could be initiated or done online with a *Gosuslugi* account, including launching a criminal or civil complaint, or submitting evidence to the commercial courts (see the next section). Since its launch and initial 335,000 users, *Gosuslugi* has developed into an e-service of everyday use with 86 million users and 582 million logins every day in 2018 (Tadviser 2019).

Together with *Gosuslugi*, main legal actors, as per the 2009 law, opened their webpages for interactional use. In 2006, *GAS* (*Gosudarstvennaâ avtomatizirovannaâ sistema*, State Automated System) “*Pravosudie*” (Justice, <https://sudrf.ru/>) was launched: it includes digital copies of decisions and judgments of all level courts in the Russian Federation. In 2016, commercial arbitration courts in the Russian Federation launched *Moi Arbitr* (My Arbitr, <https://my.arbitr.ru/>) portal, which allows for the submission of all paperwork related to a pending case online. In 2017, the Supreme Court of the Russian Federation also opened an online possibility (<http://www.supcourt.ru/appeals/>) to launch a complaint via its website using a *Gosuslugi* account. Once the possibility to use e-services became available, Russians started increasingly using them: in 2019, almost 70 percent of all complaints, addresses, and requests to state agencies are communicated online (Upravlenie Prezidenta po rabote s obrašeniâmi graždân i organizacij [Administration of the President for Work with Citizens and Organizations] 2019; for more, also see Chap. 22).

Political scientists point to a lack of democracy and classify the Russian regime as authoritarian (Ambrosio 2016). Linde and Karlsson suggest that authoritarian regimes set up e-government as a response to pressures of globalization, as well as to demonstrate modernity and legitimacy to the international community (Linde and Karlsson 2013). At the same time, others argue that this hypothesis does not account for variations of e-government across different types of authoritarian regimes. Maerz (2016), in her qualitative assessment of four post-Soviet authoritarian regimes, points to crucial differences of how e-government is used to legitimate authoritarianism. While the noncompetitive regimes of Turkmenistan and Uzbekistan create their web presences primarily for an international audience, she finds a surprising citizen-responsiveness on websites of the competitive regimes of Kazakhstan and Russia. Russians exercise their rights by extensive use of digital services and online participation in state, electoral, and judicial institutions, thus proving their interest in active citizenship (for more, see Chap. 3).

### 5.3 E-JUSTICE: DIGITALIZATION AND LEGAL PROCEDURE

The concept of e-justice can be interpreted in multiple ways. A broad definition of e-justice can cover ICT usage in the areas of crime prevention, administration of justice, and law enforcement (Xanthoulis 2010). Furthermore, e-justice for the administration of justice contains multiple subareas. These include usage of information technologies (IT) in general, electronic methods for communication (e.g., e-mail, videoconferencing), electronic case management

systems, and court room technology. E-justice can even offer citizens electronic services such as online access to case files. The Russian e-justice system developed via incorporating these subareas and trying to deal with difficulties in managing open access and data protection policies at the same time.

The development and implementation of an e-justice system entails, by its own nature, the reshaping of “institutions,” norms, and conventions that provide an implicit context for the performance of practices. In a process that Giovan Francesco Lanzara (2009) tries to capture with the concept of *assemblage*, e-justice systems are built linking and reshaping heterogeneous components and building blocks of technological, which are organizational and normative in nature. The new system comes from reusing, copying, adapting, and hooking together existing components, more than developing from scratch. In this process, different uses of technical, organizational, and normative components generate more or less visible shifts in their features and meanings of law and legal values, features and meanings (such as, for example, the very notion of justice) that are often invisible and taken for granted by the community of practitioners dealing with them. New actors, such as technological partners and network providers, make their appearance. Power and organizational borders alter, as “who-does-what” changes in the translation of procedures from paper to digital and from one form of digital to another (Velicogna 2011).

In Russia, the *assemblage* in terms of e-justice works quite efficiently. Russian e-justice system includes two key units. The first is a secured videoconference net, connecting all courts of the Russian Federation with direct access to the Internet through overt streaming video broadcasting channels, such as popular video hosting. The second is a group of portals of GAS “*Pravosudie*” on the Internet providing access for any person anywhere in the world with up-to-date information of the work of federal courts. The key principle of this portal’s functioning is to ensure transparency of justice, both in respect to procedures and access to the judicial acts in controversial cases. The system of commercial arbitration courts also has its own videoconference net and portal—*Moi Arbitr* (My Arbitrator). Both change ways and practices of administering justice and access to justice.

In terms of administering justice, the e-justice system in Russia allows for effective and cost-efficient notification of the date, time, and place of court hearings to all parties of a particular proceeding. There is a mailing system through e-mail on the portals of the GAS “*Pravosudie*,” *Moi Arbitr*, and *Gosuslugi*. One can download mobile applications supporting push notifications for new events and documents. Experts note that wide-scale adoption of these information technologies into work practices of the justice system has another advantage: it offers wide opportunities for court statistics to be automated and hence, early detection of court red tape and other procedural violations. When every judge in Russia is under restrictions to provide procedural documents in due time and up-to-date information of cases available on servers of the system, the court procedure and administration become more responsible and performance discipline sustainable on the proper level (Soloviev and Filippov 2013; Bykodorova 2015; Bonner 2018).



With electronic access to courtrooms both in civil and criminal justice that opened on January 1, 2017, Russian citizens could easily launch an e-complaint via already-existing systems *Gosuslugi* and *GAS "Pravosudie."* Since 2017, the number of complaints using *GAS "Pravosudie"* has doubled and now comprise more than 10 percent of all complaints to Russian courts (Epifanova 2019). The majority of complaints come from businesses. However, using digital platforms increased the demand for attorneys who now become intermediaries between citizens and courts: it is often them who file an electronic complaint, so their skill set has changed to include digital literacy and technical ability to navigate digital services. The possibility of launching a complaint online also generated a debate on the future of Russian justice system: if the country was heading toward "digital judges" and "digital attorneys." In January 2017, Vadim Kulik, the deputy head of the executive board of Sberbank, announced that legal robot, which Sberbank had launched in 2016, would result in 3000 positions being vacated.<sup>1</sup> German Gref, the chief executive officer (CEO) of Sberbank, also confirmed that they would stop hire lawyers without digital skills (Savkin 2017).

The most crucial improvement with introduction of e-justice as legal professionals see it is an automated process of assigning cases which should increase judicial independence and transparency (Nagornaja 2019). However, the consensus is that while artificial intelligence (AI) -based technologies are a positive improvement, they cannot substitute a human legal professional (Kurash 2017). At the same time, digital economies and legal provisions for online transactions have demonstrated that in the processes that could be automated via using algorithms, the usage of AI-based legal technologies is warranted. Russian government has been quite apt to push for legislation that supports commercial and business digital environments by introducing such notions as "digital rights" into its civil legislation and allowing "smart-contracts," which is essentially automated service for execution of legal contract. These changes have been happening at the background of Russian e-justice debate and are discussed in the next section in more detail.

#### 5.4 LAW AND DIGITAL ECONOMY: BLOCKCHAIN AND CROWDFUNDING

The original digitalization of economic transactions required fundamental changes in laws protecting data and ensuring the safety of emerging digital economies. Moving to cryptocurrency and online transactions using blockchain involved serious changes in civil, business, and commercial law that regulated market economy not only in Russia but also globally. Economic relationships involving cryptocurrency and blockchain tokens have become more organized and less volatile. Several countries are attempting to create a comfortable business and regulatory climate for prospective actors in this sphere (On the development of the digital economy 2017; Cryptocurrency

Offerings 2017). In October 2017, Vladimir Putin instructed the government and Central Bank of Russia to draft provisions regulating blockchain, cryptocurrency, smart-contracts, and tokens (Presidential Instruction On Digital Economy) by July 1, 2018. In March 2018, State Duma received draft laws “*Ob al'ternativnykh sposobakh privlëcheniâ investirovaniâ*” (On alternative means for attracting investments) and “*O cifrovyykh finansovykh aktivakh*” (On digital financial assets).

Discussions on the legal nature of blockchain tokens intensified in Russia as it became the subject matter in a bankruptcy proceeding in the case of *Car'kov v. Financial manager Leonov*. Car'kov, an insolvent individual, possessed a certain amount of bitcoins. A bankruptcy proceedings manager discovered the bitcoins and asked the commercial court of the city of Moscow to include them in bankruptcy assets. The court denied the request because Russian legislation does not regulate cryptocurrencies. The Ninth Commercial Appellate Court rectified this mistake. The court considered that Car'kov could exercise similar rights regarding bitcoins on his account as a property owner would exercise toward one's property. The court noted that the Russian civil procedure legislation establishes a list of property that cannot be levied. Cryptocurrency does not fall under such exceptions. Therefore, the court decided to include cryptocurrency in bankruptcy assets. Despite the issue being resolved by the courts, bankruptcy proceedings were only one sphere, alongside taxation and inheritance, affected by the lack of regulation of blockchain-based relations (Sannikova and Haritonova 2018, 88; Bessonova and Kasianov 2018, 69; Kuznecov and Chumachenko 2018, 100).

State Duma hesitated to pass the laws on the digital economy until in February 2019 Putin issued another Instruction setting the deadline for such laws for July 2019 (Presidential Instruction On implementing the Presidential Message to the Federal Assembly). In March 2019, State Duma amended the general part of the Russian Civil Code, the foundational source of civil law, with provisions aimed at regulating the digital economy. The legislator introduced Art. 141.1 “*Cifrovyye prava*” (Digital rights) to the Civil Code. Digital rights are a new object of civil rights in Russia. State Duma did not follow the draft law “On digital financial assets” or Russian legal commentaries suggesting to regulate cryptocurrencies as digital money, securities, or property (Sazhenov 2018, 108; Kuznecov 2018, 99; Fedorov 2018, 54). The amendment defines digital rights by using a model that is rather close to the definition of securities in article 142 of the Civil Code. That decision follows the line outlined by Putin and Russian Central Bank representatives, that ruble will remain the only legal tender currency in Russia.

These amendments to the Civil Code introduced regulations for the smart-contracts—computer protocols that facilitate the execution of a contract. Formally, Russian legislator implemented the Presidential Instruction—the Civil Code regulates smart-contracts. At the same time, this amendment does not change Russian contract law. It introduces smart-contracts as a contractual

provision and not as a separate type of contract. A provision that parties could have agreed for prior to the amendments.

Establishing the category of digital rights and regulating smart-contracts in the Civil Code laid a foundation for the development of further regulations on the digital economy. The Government of Russia announced the aim to develop this sphere in the 2016 strategy for the development of small and mid-size businesses. In Clause IV(4), the strategy declares a goal of developing new solutions for alternative sources of financing, including crowdfunding, for high-tech companies. The 2017 Presidential Instruction on Digital Economy required to draft the laws regulating Initial Coin Offering (ICO) by July 2018. ICO is a fundraising method used by companies primarily offering blockchain-connected products or services. The draft law stated the goal of following the approaches that successfully implement developed countries (Explanatory Note to Draft Law on Crowdfunding). By October 2019, the law “On attracting investments using the investment platforms (crowdfunding)” passed the third reading in the State Duma and enters into force from 2020. Before enactment of this law, there were already companies acting as crowdfunding platforms in Russia (Nekrasova and Shumejko 2017, 115). They needed to comply with the law by July 1, 2020.

A company can raise funds in an ICO by using different types of blockchain tokens. Most common types are utility tokens, investment tokens, and cryptocurrencies (Hacker and Thomale 2018, 108; Zetzsche et al. 2018, 11–12). The crowdfunding law only regulates utility tokens. Investment tokens and cryptocurrencies fall out of its scope and remain in the legal vacuum. The current law creates an ambiguity. On the one hand, it aims to regulate the relations in connection to investment—that is essential to attract investments. On the other hand, the law only defines utility tokens and avoids introducing investment tokens. The crucial component that distinguishes an investment token is the expectation of profits. In defining utility tokens, the Russian legislator excludes expectation of profits from what can be offered by utility tokens. The law thus creates a device whereby investors enter into investment relations without being able to receive an “investment” (in the true meaning of this term) in exchange for their contribution to a fundraising project. Such activity on behalf of the investors cannot be called investment. What they do is a purchase of goods or services paid upfront.

The law does not account for the technological realities of current blockchain crowdfunding platforms and excludes them from being recognized as an investment mechanism, denying legal protection to investors. Following Art. 13(8) of the Crowdfunding law, investments on the investment platform can only be done using noncash money. The Committee on Economic Policy, Industry, Innovational Development, and Entrepreneurship pointed out (Draft federal law N 419090-7 2018), that such limitation will exclude platforms offering Initial Coin Offering (ICO) services. Those platforms are technically not capable of handling regular money and can only operate with investors who exchange their money into cryptocurrency first. The circle

closes—following Art. 8(7) of the law, utility digital rights can only originate within the investment platform and investment platforms can only operate with noncash money. ICO platforms cannot operate with noncash money and thus cannot become investment platforms.

The initiatives for implementing the digital economy and creating the infrastructure for working with cryptocurrencies, smart-contracts, and ICO came from Vladimir Putin. In implementing these initiatives, State Duma failed to create a predictable regime that could compete with the leaders of digital economy like the United States, Switzerland, or Singapore. To reach the goal of securing alternative sources of financing for Russian small and mid-size businesses and reduce the capital flight from Russia, the legislation needed the introduction of investment digital rights. The law on crowdfunding could have done that. The Russian legislator took a cautious path by avoiding the regulation of investment tokens. Such partial regulation will likely alarm investors and start-ups from setting up their business in Russia.

## 5.5 CYBERLAW AND REGULATION OF RUNET

Moving online also requires new approaches to the regulation of cyberspace. Personal data protection becomes of primary concern (for more, see Chap. 6). The Russian government has tightened its control and supervision of cyberspace significantly in the last decade. The academic literature often sees this process in the context of containing opposition and political protest (Maréchal 2017; Ramesh et al. 2020). However, at the same time cyberspace faces challenges on its own and provides new opportunities for criminal or civil misbehavior, including the following: spreading of computer worms, viruses, bots, as well as other malware and spyware; illicitly accessing computers; exceeding authorized access; trafficking in information; enabling or facilitating unauthorized activities in cyberspace; and using information, communications systems, and networks to embezzle, commit fraud, stalk and harass, or invade the privacy of others (Ryan et al. 2011). Therefore, regulating cyberspace falls under a variety of control tools that the government uses for both censorship and crime prevention.

Several government authorities actively participate in regulating and supervising the telecommunications sector. The most important ones include: *Minkomsvâz*' (Ministry of Communications and Mass Media), *Roskomnadzor* (Federal Service for Supervision of Communications, Information Technology and Mass Media), *Rossvâz*' (Federal Communications Agency), and *Rospëchat*' (Federal Agency for Press and Mass Communications of the Russian Federation). As a result of administrative reform, conducted in 2004, ministries define state policy and perform regulatory activities, while state services and agencies perform executive and supervisory functions (Bogdanovskaya et al. 2016).

*Roskomnadzor* is the main watchdog over the Runet and manages the information controls regime in Russia. It is tasked with a wide range of

competences, including silencing of mass media and audiovisual platforms, as well as management of a list of operators. In December 2011, the Ministry of Communications issued a new administrative regulation “*O vedenii reestra operatorov, osušestvláúših obrabotku personal’nyh dannyh*” (On introducing the register of operators processing personal data), which significantly increased data protection control. In 2012–2016, Federal Law N 149-FZ “*Ob informacii, informacionnyh tehnologiáh i o zúšite informacii*” (On Information, Information technologies and Protection of Information) was significantly amended to accommodate changes in relation to (1) a package of protectionist legislation prohibiting promotion of nontraditional sexual relations among minors and dissemination of information harmful to health and development of minors and (2) a package of security legislation known as “Yarovaya laws” (for more, see Chap. 6). The latest 2019 controversial amendment added Art. 15.1-1 limiting access to “indecent” information that insult human dignity and offend public decency, express blatant disrespect for the public, the state, official state symbols of the Russian Federation, Constitution of the Russian Federation (RF), or state agencies. These amendments produced an increasing amount of complaints to *Roskomandzor* and Russian courts.<sup>2</sup>

Nathalie Maréchal (2017) argues that Russia does not view internet governance, cybersecurity, and media policy as separate domains, which enable strong information controls. Other scholars identify Russian policies as “decentralized control” due to the lack of direct ownership of Internet Service Providers (ISP) by government authorities. This lowers their ability to unilaterally roll out technical censorship measures, instead pushing the state to enact controls via law and policy, compelling their network owners to comply, which subsequently significantly increases censorship (Ramesh et al. 2020).

## 5.6 CONCLUSIONS

Digitalization of law and legal services has positive and negative effects on human rights and everyday lives of citizens. Following global going online, Russia has achieved impressive results in providing e-services, as well as access to state and private digital information and resources. Access to e-courts removes certain barriers in accessing justice for vulnerable groups and makes litigation more transparent and effective. Digital citizens have a wide range of strategies to navigate cyberspace to improve their quality of life. However, these achievements have come at significant cost for law, the legal system, as well as public and private individuals, especially in an authoritarian political framework.

The ongoing legalization of judicial or procedural phenomena by the creation of e-justice or e-procedural norms also represents a strong move toward what is here called “formalization” or even hyperformalization, to an extent never before seen in history (Gilles 2014). This hyperformalization is needed for smoothing the work of ICTs and for efficiency of administering justice

online, but it often lacks flexibility and has a profound impact on quality and content of law. In the Russian case, law had been formalistic before the digital turn; it has become even more so since. This hyperformalization is positive for business and market economy, especially in a global dimension, but might be harmful for private citizens.

Digitalization of law has brought a new level of surveillance, censorship, and information control that has not been available before. The law once again serves as an instrument of political manipulation, which leads to even further formalization of procedures and uses of e-justice to curtail freedoms of speech and other human rights. High levels of securitization will demand a further increase in censorship and surveillance as Russia heads toward creating an internet “kill switch.” This would allow the Russian state to disconnect the Runet from the global network “in case of crisis,” without specifying what such a crisis might entail beyond vague allusions to the internet being shut off from the outside (Duffy 2015; Nocetti 2015). This uncertainty and mistrust of due process and the government’s intentions create further anxiety in civil society (for more, see Chap. 8), which consolidates its activism online, but feels a tightening surveillance and prosecution of its activities due to instrumental use of digital law. In this respect, Russia is an example of successful usage of e-government and e-law by authoritarian regimes as it leverages globalization for its own political ends.

## NOTES

1. <https://www.interfax.ru/business/545109>.
2. According to the *Roskomnadzor*’s reports, the amount of complaints increased significantly between 2012 and 2013 from 26,287 to 86,274; by 2019, 154,914 complaints had been filed. See Federal Service for Supervision of Communications, Information Technology and Mass Media of the Russian Federation. Report on the processing of communications from the citizens of the RF for 2018, available here: <https://rkn.gov.ru/treatments/p436/>.

## REFERENCES

- Ambrosio, Thomas. 2016. *Authoritarian Backlash: Russian Resistance to Democratization in the Former Soviet Union*. London: Routledge.
- Bessonova, T.V., and M.R. Kasianov. 2018. Problema pravovoj prirody kriptovalûty [Problem of the Legal Nature of Cryptocurrency]. *Vestnik ekonomiki, prava i sociologii* [Bulliten of Economics, Law, and Sociology] 2: 68–70.
- Bogdanovskaya, Irina, Mikhail Bashirov, Alexander Vishnevsky, Sergei Danilov, Vitaly Kalyatin, and Alexander Savelyev. 2016. *Cyber Law in Russia*. Netherlands: Wolters Kluwer.
- Bonner, A.T. 2018. Èlektronnoe pravosudie: real’nost’ ili novomodnyj termin? [E-Justice: Reality or a Newfangled Term?]. *Vestnik graždanskogo processa* [Bulliten of Civil Proceedings] 8 (1): 22–38.



- Bykodorova, L.V. 2015. Proekt 'Elektronnoe pravosudie': sostoanie i problemy razvitiia [The E-Justice Project: Status and Development Challenges]. *Uspehi sovremennoj nauki i obrazovaniia* 5: 86–88.
- Calo, Ryan. 2015. Robotics and the Lessons of Cyberlaw. *California Law Review* 103: 513–563.
- Duffy, Natalie. 2015. *Internet Freedom in Vladimir Putin's Russia: The Noose Tightens*. Washington, DC: American Enterprise Institute.
- Epifanova, A. 2019. Internet-interv'ù s L.A. Úhnevichem, rukovoditelem FGBU 'Informacionno-analitičeskij centr podderžki GAS Pravosudie': 'Elektronnoe pravosudie – budušee rossijskoj sudebnoj sistemy' [Online Interview with L.A. Yuhnevich, Head of the FSBI 'Information and Analytical Support Center GAS Justice': 'E-Justice – the Future of the Russian Judicial System']. *Consultant Plus*. <http://www.consultant.ru/law/interview/yuhnevich/>.
- Evans, Angela M., and Adriana Campos. 2013. Open Government Initiatives: Challenges of Citizen Participation. *Journal of Policy Analysis and Management* 32 (1): 172–185.
- Fedorov, Alexander. 2009. Media Education in Russia: A Brief History. In *Issues in Information and Media Literacy: Criticism, History and Policy*, ed. Marcus Leaning, 167–188. Santa Rosa, CA: Informing Science Institute.
- Fedorov, D.V. 2018. Tokeny, kriptovalúta i smart-kontrakty v otečestvennyh zakonoproektah s pozicii inostrannogo opyta [Russian Draft Laws on Tokens, Cryptocurrency and Smart-Contracts Regulation from the Perspective of Foreign Experience]. *Vestnik graždanskogo prava [Bulletin of Civil Law]* 18 (2): 30–74. <https://doi.org/10.24031/1992-2043-2018-18-2-30-74>.
- Gilles, Peter. 2014. Civil Justice Systems and Civil Procedures in a Changing World: Main Problems, Fundamental Reforms and Perspectives-A European View. *Russian Law Journal* 2 (1): 41–56.
- Hacker, Philipp, and Chris Thomale. 2018. Crypto-Securities Regulation: ICOs, Token Sales and Cryptocurrencies Under EU Financial Law. *European Company and Financial Law Review* 15: 645–696. <https://doi.org/10.1515/ecf-2018-0021>.
- Irkhin, Iu V. 2007. 'Electronic Government' and Society: World Realities and Russia (A Comparative Analysis). *Sociological Research* 46 (2): 77–92.
- Jones, Meg Leta. 2018. Does Technology Drive Law: The Dilemma of Technological Exceptionalism in Cyberlaw. *Journal of Law, Technology & Policy* 46 (2): 249–284.
- Katsh, Ethan. 1993. Law in a Digital World: Computer Networks and Cyberspace. *Villanova Law Review* 38 (2): 403–486.
- Keen, Andrew. 2010. Why We Must Resist the Temptation of Web 2.0. In *The Next Digital Decade. Essays on the Future of the Internet*, ed. Berin Szoka and Adam Marcus, 51–57. Washington, DC: TechFreedom. <https://nissenbaum.tech.cornell.edu/papers/The-Next-Digital-Decade-Essays-on-the-Future-of-the-Internet.pdf>.
- Kurash, Anton. 2017. Kakoe budušee ždet turistov v èpohu cifrovoj èkonomiki [What Is the Future for Lawyers in the Epoch of Digital Economy]. *CNEWS.ru*, October 20. [https://www.cnews.ru/articles/2017-10-20\\_kakoe\\_budushchee\\_zhdet\\_yuristov\\_v\\_epohu\\_tsifrovoj\\_ekonomiki](https://www.cnews.ru/articles/2017-10-20_kakoe_budushchee_zhdet_yuristov_v_epohu_tsifrovoj_ekonomiki).
- Kuznecov, Yuri Vladimirovich. 2018. Kriptoaktivy Kak Dokumentarnye Cennye Bumagi [Cryptoassets as Definitive Securities]. *Zakon [Law]* 9: 96–105.
- Kuznecov, M.N., and I.N. Chumachenko. 2018. Cifrovye prava – novelty v rossijskom zakonodatel'stve [Digital Rights – Novels in Russian Legislation]. *Problemy èkonomiki i úridičeskoj praktiki [Problems of Economic and Legal Practice]* 4: 97–100.

- Lanzara, Giovan. 2009. Building Digital Institutions: ICT and the Rise of Assemblages in Government. In *ICT and Innovation in the Public Sector: European Studies in the Making of E-Government*, ed. Francesco Contini and Giovan Lanzara, 9–48. London: Palgrave Macmillan.
- Linde, Jonas, and Martin Karlsson. 2013. The Dictator's New Clothes: The Relationship Between E-Participation and Quality of Government in Non-democratic Regimes. *International Journal of Public Administration* 36 (4): 269–281.
- Maerz, Serphine. 2016. The Electronic Face of Authoritarianism: E-Government as a Tool for Gaining Legitimacy in Competitive and Non-competitive Regimes. *Government Information Quarterly* 33 (4): 727–735.
- Maréchal, Nathalie. 2017. Networked Authoritarianism and the Geopolitics of Information: Understanding Russian Internet Policy. *Media and Communication* 5 (1): 29–41.
- Mossberger, Karen, Caroline J. Tolbert, and Ramona S. McNeal. 2008. *Digital Citizenship: The Internet, Society, and Participation*. Cambridge, MA: MIT Press.
- Nagornaja, Marina. 2019. Advokaty i ūristy ob iskusstvennom intellekte v sudoproizvodstve [Attorneys and Lawyers Commented the Usage of AI in Litigation]. *Advokatskaâ gazeta* [Attorney's Gazette], April 9. <https://www.advgazeta.ru/novosti/advokaty-i-yuristy-ob-iskusstvennom-intellekte-v-sudoproizvodstve/>.
- Nekrasova, Tatiana Petrovna, and Ekaterina Vladimirovna Shumejko. 2017. Èkonomičeskaâ ocenka kraudfandinga kak metoda privlečeniâ investicij [Economic Estimation of Crowdfunding as a Type of Fundraising]. *Naučno-tehničeskije vedomosti SPbGPU* [Scientific-Technical Gazette of SPbGPU] 10 (5): 114–124.
- Nocetti, Julien. 2015. Contest and Conquest: Russia and Global Internet Governance. *International Affairs* 91 (1): 111–130.
- Ramesh, Reethika, Ram Sundara Raman, Matthew Bernhard, Victor Ongkowijaya, Leonid Evdokimov, Anne Edmundson, Steven Sprecher, Muhammad Ikram, and Roya Ensafi. 2020. *Decentralized Control: A Case Study of Russia*. Paper submitted for presentation at the Network and Distributed Systems Security (NDSS) Symposium 2020, San Diego, CA, USA, February 23–26, 2020. <https://doi.org/10.14722/ndss.2020.23098>. <https://mbernhard.com/papers/russia.pdf>.
- Rasskazova, Elena I., and Galina V. Soldatova. 2014. Assessment of the Digital Competence in Russian Adolescents and Parents: Digital Competence Index. *Psychology in Russia: State of the Art* 7 (4): 65–74.
- Ribble, Mike. 2015. *Digital Citizenship in Schools: Nine Elements All Students Should Know*. Eugene, OR: International Society for Technology in Education.
- Ryan, Daniel J., Maeve Dion, Eneken Tikk, and Julie J.C.H. Ryan. 2011. International Cyberlaw: A Normative Approach. *Georgetown Journal of International Law* 42 (4): 1161–1198.
- Sannikova, Larisa Vladimirovna, and Julija Sergeevna Haritonova. 2018. Pravovaâ Sušnost' Novyh Cifrovyh Aktivov [Legal Nature of New Digital Assets]. *Zakon* [Law] 9: 86–95.
- Savkin, Aleksej. 2017. Otvet Grefu. Počemu èlektronnoe pravosudie nevozmožno [Reply to Gref. Why e-Justice Is Impossible [to Us]]. *Forbes.ru*, December 5. <https://www.forbes.ru/tehnologii/353819-otvet-grefu-pochemu-elektronnoe-pravosudie-nevozmozhno>.
- Sazhenov, A.V. 2018. Kriptoalŭty: dematerializaciâ kategorii vešej v graždanskom prave [Cryptocurrencies: Dematerialisation of Legal Category of Things in Civil Law]. *Zakon* [Law] 9: 106–121.



- Soloviev, Andrey, and Yury Filippov. 2013. Course of Justice of Arbitration Courts in the Russian Federation. *Law and Modern States* 3: 65–72.
- Tadviser. 2019. Auditoriâ i statistika portala gosuslug [Audience and Statistics of the Public Services Portal]. May 4. <http://www.tadviser.ru/a/452220>.
- Upravlenie Prezidenta po rabote s obrašeniâmi graždan i organizacij [Administration of the President for Work with Citizens and Organizations]. 2019. Informacionno-statističeskij obzor rassmotrennyh v oktâbre 2019 goda obrašenij graždan, organizacij i obšestvennyh ob'edinenij, adresovannyh Prezidentu Rossijskoj Federacii [Information and Statistical Review of Appeals of Citizens, Organizations and Public Associations Examined in October 2019 Addressed to the President of the Russian Federation]. November 11. <http://letters.kremlin.ru/digests/225>.
- Velicogna, Marco. 2011. Electronic Access to Justice: From Theory to Practice and Back. *Droit ET Cultures*. *Revue internationale interdisciplinaire*, 61. <https://journals.openedition.org/droitcultures/2447>.
- Vinogradova, N., and O.A. Moiseeva. 2015. Open Government and ‘E-Government’ in Russia. *Sociology Study* 5 (1): 29–38.
- Xanthoulis, N. 2010. Introducing the Concept of E-Justice in Europe: How Adding an ‘E’ Becomes a Modern Challenge for Greece and the EU. *Effectius Communication* 1 (1): 1–10.
- Zetzsche, Dirk A., Ross P. Buckley, Douglas W. Arner, and Linus Föhr. 2018. The ICO Gold Rush: It’s a Scam, It’s a Bubble, It’s a Super Challenge for Regulators. *SSRN Electronic Journal* 63 (2): 1–30. <https://doi.org/10.2139/ssrn.3072298>.

#### LEGAL SOURCES

- Canada. Ontario Securities Commission. 2017. CSA Staff Notice 46-307 “Cryptocurrency Offerings.” August 24. [https://www.osc.gov.on.ca/en/SecuritiesLaw\\_csa\\_20170824\\_cryptocurrency-offerings.htm](https://www.osc.gov.on.ca/en/SecuritiesLaw_csa_20170824_cryptocurrency-offerings.htm).
- Car'kov v. Leonov*. 2018. Commercial Court of the City of Moscow. <http://kad.arbitr.ru>.
- . 2018. Ninth Commercial Appellate Court. <http://kad.arbitr.ru>.
- Civil Code of the Russian Federation dated November 30, 1994. [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_5142/](http://www.consultant.ru/document/cons_doc_LAW_5142/).
- Draft Federal Law 419059-7 of March 28, 2018. 2018. O cifrovych finansovyh aktivah [On Digital Financial Assets]. Registered to the State Duma, March 28. <https://sozd.duma.gov.ru/bill/419059-7>.
- Draft Federal Law 419090-7 of March 28, 2018. 2018. Ob al'ternativnyh sposobah privlečeniâ investirovaniâ (kraudfandinge) [On Alternative Means for Attracting Investments (Crowdfunding)]. Registered in the State Duma, March 28. <https://sozd.duma.gov.ru/bill/419090-7>.
- Federal Law of January 10, 2002, N 1-FZ. 2002. Ob èlektronnoj cifrovoj podpisi [On electronic digital signature]. Dated January 10. <http://base.garant.ru/184059/>.
- Federal Law of July 27, 2006, N 149-FZ. 2006. Ob informacii, informacionnyh tehnologiâh i o zaštite informacii [On Information, Information Technologies and Protection of Information]. Rev. 1 May 2019. [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_61798/](http://www.consultant.ru/document/cons_doc_LAW_61798/).
- Federal Law of December 22, 2008, N 262-FZ. 2008. Ob obespečeniî dostupa k informacii o deâtel'nosti sudov v Rossijskoj Federacii [On Access to Information about the Courts in the Russian Federation]. <https://rg.ru/2008/12/26/sud-internet-dok.html>.

- Federal Law of February 9, 2009, N 8-FZ. 2009. Ob obespečenii dostupa k informacii o deatel'nosti gosudarstvennyh organov i organov mestnogo samoupravleniâ [On Access to Information on the Activities of State and Local Governments]. Rev. 28 December 2017. [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_84602/](http://www.consultant.ru/document/cons_doc_LAW_84602/).
- Federal Law of October 14, 2014, N 139-FZ. 2014. O vnesenii izmenenij v Federal'nyj zakon 'o zašite detej ot informacii, pričînâušej vred ih zdorov'û i razvitiû' i otdel'nye zakonodatel'nye akty Rossijskoj Federacii [On Amending the Federal Law 'On Protection Children from Information Harmful to Their Health and Development' and Other Laws of the Russian Federation]. [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_133282/](http://www.consultant.ru/document/cons_doc_LAW_133282/).
- Federal Law of July 6, 2016, N 374-FZ. 2016. O vnesenii izmenenij v Federal'nyj zakon 'O protivodejstvii terrorizmu' i otdel'nye zakonodatel'nye akty Rossijskoj Federacii v časti ustanovleniâ dopolnitel'nyh mer protivodejstviâ terrorizmu i obespečenâ obšestvennoj bezopasnosti [On Amending the Federal law 'On Countering Terrorism' and Other Laws of the Russian Federation to Provide Additional Measures to Counter Terrorism and Ensuring Public security]. <https://rg.ru/2016/07/08/antiterror-dok.html>.
- Federal Law of March 18, 2019, N 30-FZ. 2019. O vnesenii izmenenâ v Federal'nyj zakon 'ob informacii, informacionnyh tehnologiâh i o zašite informacii' [On Amending the Federal law 'On Information, Information Technologies and Protection of Information']. [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_320400/#dst100008](http://www.consultant.ru/document/cons_doc_LAW_320400/#dst100008).
- Federal Law of August 2, 2019, N 259-FZ. 2019. O privlečenii investicij s ispol'zovaniem investicionnyh platform i o vnesenii izmenenij v otdel'nye zakonodatel'nye akty Rossijskoj Federacii [On Alternative Forms of Attracting Investments via Using Investment Platforms and Amending Other Laws of the Russian Federation]. <http://publication.pravo.gov.ru/Document/View/0001201908020032>.
- Ministry of Digital Development, Communication, and Mass Media of the Russian Federation. 2002. Federal Target Program "Elektronnaâ Rossiâ [Electronic Russia]," adopted by the Decision of the Government of the Russian Federation N 65 from January 28. <https://digital.gov.ru/ru/activity/programs/6/>.
- Ministry of Digital Development, Communication and Mass Media of the Russian Federation. 2011. The Instruction N 346 "Ob utverždenii Administrativnogo reglamenta Federal'noj služby po nadzoru v sfere svâzi, informacionnyh tehnologij i massovyh kommunikacij po predostavleniû gosudarstvennoj uslugi 'Vedenie reestra operatorov, osušestvlâuših obrabotku personal'nyh dannyh' [On the Endorsement of the Administrative Regulation of the Federal Service for Supervision of Communications, Information Technology and Mass Media to Provide the State Service 'On Introducing the Register of Operators Processing Personal Data']". Dated December 21. <https://rg.ru/2012/03/30/operatory-site-dok.html>.
- President of the Russian Federation. 2017. Instruction on the Results of the Meeting Regarding Using Digital Technologies in the Financial Sphere. Dated October 10 (Presidential Instruction on Digital Economy). <http://kremlin.ru/acts/assignments/orders/558991>.
- . 2019. List of Instructions for the Implementation of the Presidential Message to the Federal Assembly, dated February 27 (Presidential Instruction on Implementing the Presidential Message to the Federal Assembly). <http://kremlin.ru/acts/assignments/orders/59898>.

- Republic of Belarus. Presidential Decree No. 8 “On the Development of the Digital Economy,” dated December 21, 2017. <http://pravo.by/document/?guid=12551&p0=Pd1700008&p1=1>.
- State Duma of the Russian Federation. On the Draft Law N 419090-7 “O privilečienii investicij s ispol’zovaniem investicionnyh platform i o vnesenii izmenenij v otdeľnye zakonodatel’nye akty Rossijskoj Federacii [On Alternative Means for Attracting Investments (Crowdfunding)].” Opinion of the Committee of State Duma on Economic Politics, Industry and Innovational Development. <https://sozd.duma.gov.ru/bill/419090-7>.
- . 2018. Committee on Financial Markets. Explanatory Note to the Draft Federal Law “O privilečienii investicij s ispol’zovaniem investicionnyh platform i o vnesenii izmenenij v otdeľnye zakonodatel’nye akty Rossijskoj Federacii [On Alternative Means for Attracting Investments (Crowdfunding)].” (Explanatory Note to Draft Law on Crowdfunding), registered in the State Duma, March 28. <https://sozd.duma.gov.ru/bill/419090-7>.
- Strategy of the Development of Information Society in the Russian Federation (approved by the President of the Russian Federation February 7, 2008) N Pr-212. <https://rg.ru/2008/02/16/informacia-strategia-dok.html>.
- Strategy of the Development of Small and Medium-Sized Businesses in the Russian Federation till 2030 (approved by the Government of the Russian Federation June 2, 2016) N 1083-p. [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_199462/f3fa9da4fab9fba49fc9e0d938761ccffdd288bd/](http://www.consultant.ru/document/cons_doc_LAW_199462/f3fa9da4fab9fba49fc9e0d938761ccffdd288bd/).
- The Plenum of the Supreme Court of the USSR. 1982. Resolution N 7 “O sudebnom rešenii [On Court Decision]”. Dated July 9. <http://base.garant.ru/10105790/>.
- The State Arbitrage of the USSR. 1979. Instructions N I-1-4 “Ob ispol’zovanii v kačestve dokazatel’stv po arbitražnym delam dokumentov, podgotovlennyh s pomošč’ü elektronno-vyčislitel’noj tehniki [On the Use as Evidence in the Arbitration Proceedings Papers Prepared Using Computer Technology]”. Dated June 29. <https://bazanpa.ru/gosarbitrazh-sssr-instruktivnye-ukazaniia-ni-1-1-4-ot29061979-h208683/>.
- USSR State Committee on Standards. 1984. Regulation N 3549GOST 6.10.4-84 “Unificirovannye sistemy dokumentacii. Pridanie űridičeskoj sily dokumentam na mašinnom nositele i mašinogramme, sozdavaemym sredstvami vyčislitel’noj tehniki. Osnovnye položenia [Unified Systems of Documentation. Conferment of Legal Force to Documents on Software and Machinogramme Created by Computers. General Guidelines]”. Dated October 9. <http://docs.cntd.ru/document/9010879>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International Licence (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Personal Data Protection in Russia

*Alexander Gurkov*

## 6.1 INTRODUCTION

Data protection is a recent area of law in Russia. The Russian State Duma enacted data protection laws only in 2006. Before that, the Russian Constitution's (1993) articles 23 and 24 laid the foundations for data protection. Starting in 2014, the Russian legislator introduced major amendments to data protection regulations, allowing for more control by governmental agencies over data flow.

The ideas of the Russian legislator are not unique in the global arena and were in some form implemented in other jurisdictions. This chapter uses EU conceptions of personal data protection as a point of reference. In 2018, the EU 2016 General Data Protection Regulation (GDPR) took effect and influenced the development of the data protection sphere around the globe. As one of the most comprehensive data protection legislations implemented in the world, the GDPR is a good point of comparison.

After the introduction (Sect. 6.1), the chapter provides an overview of the legal framework of data protection in Russia (Sect. 6.2). This lays the foundation for the next sections, which explain three important changes in Russian data protection legislation. These changes provided governmental agencies in Russia with more control over transferring information: introduction of a data localization requirement (Sect. 6.3), the Yarovaya law (Sect. 6.4), and regulations aimed at creating a sovereign internet (Sect. 6.5). The chapter ends with a section analyzing the influence of a political case on the understanding of personal data by the Federal Service for Supervision of Communications, Information Technology and Mass Media (Roskommnadzor) and showing the

---

A. Gurkov (✉)  
University of Helsinki, Helsinki, Finland  
e-mail: [alexander.gurkov@helsinki.fi](mailto:alexander.gurkov@helsinki.fi)

vague nature of legislative definitions that gives public authorities vast freedoms in the application of regulations (Sect. 6.6).

Many Russian data protection legislative initiatives fall outside of world trends. Yet, some initiatives align Russian legislation with global trends, with the caveat that changes can be implemented when the government needs them to win a political case. This chapter shows the growing role and authority of Roskomnadzor, which will soon receive the potential to control the entirety of internet traffic in Russia and the ability to isolate the Russian internet. Some requirements of Russian data protection legislation are unprecedented in the world and are very costly for companies. Overall, the Russian legislator and various enforcement agencies act not with the aim of protecting individual rights in the sphere of personal data protection but with the aim of providing Russian authorities with more power to monitor and control the flow of data in Russia. This can be a legitimate aim given the fast development of personal data threats, but such an aim should be stated clearly and openly.

## 6.2 GROUND RULES

### 6.2.1 *Legal Framework*

Articles 23 and 24 of the Russian Constitution (1993) already show that the main subjects to which data protection legislation is directed are data subjects and data operators. These same ideas were reflected in the legislation.

Article 23 provides that “Everyone is entitled to privacy of personal life, personal and family secrets, protection of one’s honor and good name.” Privacy is the right to control information about oneself. The right to privacy is a universal human right and is recognized as such by the Universal Declaration of Human Rights and the European Convention of Human Rights. It is the foundation for the right to data protection. The right to data protection originates from privacy but is not a universal human right. It is aimed toward operators of personal data to ensure its fair processing. Correspondingly, article 24 of the Russian Constitution addresses operators of personal data. It requires that the “collection, storage, usage, and distribution of information on private life are not permitted without the approval of a person.” Before the enactment of specialized legislation, in December 2005 Russia ratified the 1981 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Council of Europe Convention). The Council of Europe Convention is a foundation on which several countries have built their data protection legislation.

In July 2007, the State Duma passed two laws dedicated to data protection: Federal Law No. 149-FZ “*Ob informacii, informacionnyh tehnologiáh i o zâsite informacii*” (On information, information technologies and data protection, Data Protection Act) and Federal Law No. 152-FZ “*O personal’nyh dannnyh*” (Personal Data Law). The provisions of these acts were conventional and similar to those of the 1995 European Data Protection directive (Garrie and

Byhovskiy 2017, 239). The Personal Data Law is the principal law regulating this sphere in Russia. It sets the purpose of personal data protection—securing the rights and freedoms of a person and a citizen in processing one’s data (article 2).

Up until 2014, Russian data protection regulations did not stand out from the Council of Europe Convention. Following the terrorist acts in the city of Volgograd in 2013, the State Duma passed an anti-terrorist packet of legislation. A part of that package was the Federal Law of July 21, 2014, No. 242-FZ (Localization law), which introduced the localization requirement (more on that in Sect. 6.3). Apart from the Russian legislator, several authorities are competent to create data protection regulations. The Russian President, the Russian government, and Federal Services take active roles in this sphere (for more, see Chap. 3).

### 6.2.2 *Enforcing Authorities*

Among public authorities, Roskomnadzor plays the most active role. Dmitry Medvedev established Roskomnadzor in 2008 (Decree of the President No. 1715). Roskomnadzor reports to the Ministry of Digital Development, Communications and Mass Media (Ministry of Communications). It has many important competencies such as monitoring mass media and keeping the registries of data operators and prohibited websites (Resolution of the Government on Roskomnadzor). When it comes to specific powers of Roskomnadzor, the vector of activity of this Federal Service derogates from the direction in which personal data protection is aimed—securing individual rights and freedoms. Following article 23 of the Data Protection Act, Roskomnadzor can investigate and initiate control and supervision of data operators, without regard to violation of personal rights of individuals. It acts without regard to whether those individuals whose data is processed have any claims to data operators. As a result, the activity of Roskomnadzor is directed toward the protection of data as such and not toward the protection of individual rights affected by data processing (Tereshhenko 2018, 146).

Apart from Roskomnadzor, a few other authorities exercise their power in enforcing data protection policy in Russia. The Office of the Prosecutor is responsible for prosecuting criminal actions related to infringement of data protection. The Federal Service on Technical and Export Control is responsible for supervising the safety of personal data within the informational infrastructure of Russia.

### 6.2.3 *Main Categories of Data Protection Legislation*

The main categories that define data protection legislation in Russia are data, personal data, data operators, data processing, and transfer of personal data. Article 2(1) of the Data Protection Act defines information as any data irrespective of its form of representation. Following article 3(1) of the Personal

Data Law, personal data is any information directly or indirectly related to a certain or identifiable individual (data subject). The law will not protect data that does not relate to an identifiable individual (anonymized data). Following this definition, it could be hard to differentiate between technical data and personal data, as almost any transaction made on the internet will constitute personal data (Bauer et al. 2015, 2).

When it comes to establishing the criteria of what counts as an identifiable individual, Roskomnadzor's practices may create some ambiguity. For example, in 2017 the Pension Fund of Russia leaked information containing full names and surnames of its clients, their taxpayer numbers, and information about their pension savings. As per the response of the Pension Fund, these do not constitute a data breach, as such data does not allow to identify a person (Tereshhenko 2018, 152). Roskomnadzor did not respond to this breach with any action. As much as the Pension Fund wanted to keep the breach harmless, information that contains names, surnames, and identity numbers is without a doubt personal data. Senior officials of Roskomnadzor stated in a 2015 commentary to the Personal Data Law that an individual taxpayer number allows to clearly identify a natural person (Gafurova et al. 2015, 16).

The Personal Data Law differentiates between categories of personal data. According to article 10 of the law, a special regulation applies to data relating to racial and national identity, political views, religious or philosophical beliefs, health conditions, and intimate life. The processing of such data can only be done in cases prescribed by the law, for example, if a data subject gives written consent to processing the data.

Following article 3(2) of the Personal Data Law, an operator is an authority, a company, or an individual that organizes and (or) performs processing of personal data. An operator also defines the purpose of personal data processing and composition of personal data to be processed, as well as actions toward personal data. Data protection legislation applies to all operators of data and third parties authorized by the operators. A general rule is that data operators need to notify Roskomnadzor of their intent to process data before engaging in data processing (article 22). There are certain cases where such notification is not necessary, for example, where data processing is done under labor legislation, if the data only includes a surname, name, and paternal name of the data subject, or if the data subject revealed the data in open access.

When collecting personal data, operators need to inform subjects about certain required aspects of data processing. For example, following article 18.1 (1) (2), operators need to publish a data processing policy. The law takes a reasonable approach by imposing this obligation on operators that are legal entities. In practice, this means that natural persons, as well as individual entrepreneurs, do not need to publish their processing policy.

Data operators need to set up security measures. According to article 18.1 of the law, data operators are free to choose measures that they need to take to comply with the law. The recommended measures under the law are



appointing a data protection officer, implementing certain organizational and technical measures aimed at securing the data, and performing internal control and audit.

What is interesting is that the list of such measures does not include an obligation to notify of a data breach, to either Roskomnadzor or data subjects. There was an attempt to amend the legislation and introduce the obligation to notify Roskomnadzor, Ministry of Internal Affairs, and even relevant data subjects of data breaches, but the draft law has not been passed by the State Duma since 2017 (Draft law No. 416052-6).

Data processing is any action or combination of actions associated with personal data (with or without the means of automation), including collection, recording, systematization, storing, extracting, and transferring. Processing should be adequate, relevant, and not excessive to the purpose for which the data is processed. Following article 5(7) of the Personal Data Law, one of the principles of data processing is that once the goal for which the information was processed is reached, the operator needs to anonymize or destroy the data unless there was any agreement to the contrary. At the moment, there are no detailed rules on how data should be destroyed. However, corresponding amendments authorizing Roskomnadzor to establish such detailed rules are being considered by the State Duma (Draft law “On termination of personal data”).

The consent of data subjects is an essential part of processing personal data. Following article 9 of the Data Protection Law, an individual should give one’s written consent for data processing. The consent should be specific, informed, and deliberate. It can be acquired in any form that can confirm that it was given, including filling online forms. The data subject can later change one’s mind and revoke consent for data processing. Data operators bear the burden of providing proof that a data subject provided her consent.

Following article 9(4)(4) of the law, in certain cases, including when processing data related to political views, religious beliefs, health conditions, and intimate life, consent should be given in writing. The written form of consent should include the purpose of data processing. The law does not specifically require that the data processor ask a data subject to provide separate consent for each purpose of data processing. Data processors often construe this provision in such a way as to list different purposes of data processing in one form. Yet, since construing the law in the other direction is possible, there is a material risk that Roskomnadzor will require written consent from a data subject for every purpose of data processing. This was the case in a dispute between a limited liability company (LLC) Skartel and Roskomnadzor (*LLC Skartel v. Roskomnadzor Administration of the Central Federal Circuit*). The commercial court of the city of Moscow and then the appellate court confirmed the position of Roskomnadzor. The clients of Skartel signed terms and conditions that listed certain purposes of data processing. After doing that, some of the clients made additional agreements online. Such agreements included more purposes of data processing. The courts agreed with Roskomnadzor that consent for

such additional purposes of data processing, following verbatim reading of the law, should have also been given in paper-based writing form. To address this situation, the Ministry of Communications drafted amendments to the law on data protection that, among other measures, would allow receiving single consent of a person for multiple purposes of data processing (Draft law “On single consent form”). This is one of the examples where the aim of amendments is to ease the burden for data operators, as opposed to creating numerous new regulations introducing limitations and obligations in the sphere of data protection, as will be shown in further sections of this chapter.

Personal data can be processed without the data subject’s consent in certain cases (article 6). For example, consent is not needed when data processing is necessary for a professional journalistic activity or when it is necessary for the enforcement of a court or a public authority decision.

#### 6.2.4 *Transfer Outside of Russia*

Data operators can transfer personal data outside of Russia. Before making such transfer, the operator has to make sure that the rights of the personal data subject will receive adequate protection in the receiving country of the transfer. Article 12(1) of the Personal Data Law provides that all signatories to the Council of Europe Convention provide adequate protection to personal data. Apart from this, Roskomnadzor keeps a regularly updated list of countries that provide such protection (Order of Roskomnadzor on the list of countries with adequate personal data protection).

#### 6.2.5 *Territorial Scope of Application*

The internet spreads across national borders. Russian citizens can access websites of operators located all around the world (except for those blocked by Roskomnadzor). This does not mean that all of those operators need to comply with Russian localization requirements. The Data Protection Law does not specifically establish the territorial scope of its application. At the same time, when defining operators of personal data, the law does not limit operators to only companies registered in Russia. In view of Roskomnadzor, the Personal Data Law is binding upon foreign companies that process personal data in Russia (Roskomnadzor 2019a). The territorial scope is defined by data processing that (1) either takes place or is aimed at Russia or (2) concerns the data of Russian citizens. What is important is not where a company/person is based but the territory at which the actions of such a company or a person are directed. Companies incorporated outside of Russia may nevertheless be subject to Russian data protection regulations. In a similar fashion, article 3 of the GDPR establishes that its data protection requirements are binding not only for companies established in EU member states but also for companies located anywhere in the world if they process the data of EU citizens. The importance

of the territorial aspect of Russian data protection regulations is amplified with the adoption of the localization requirement for data operators.

### 6.3 LOCALIZATION REQUIREMENT

The personal data localization requirement was a part of the 2014 anti-terrorist legislation package (Localization Law). Before the enactment of these amendments, there were no limitations on localization—processing and storing information of Russian citizens could be done on servers located anywhere in the world (Garrie and Byhovskiy 2017, 242). The purpose of the localization requirements, according to the head of Roskomnadzor, is to “provide an extra protection for Russian citizens both from misuse of their personal data by foreign companies and from surveillance of foreign governments” (Savelyev 2016, 138; Zharov 2014).

From an economic standpoint, the introduction of the localization requirement is a self-imposed sanction that seriously weakens Russia’s ability to attract investments (Bauer et al. 2015, 3). The localization rules affect many companies, including giants like Apple, Microsoft, Google, Facebook, and Twitter as well as big companies such as eBay, PayPal, [Booking.com](#), and Reddit (Zhuravlev and Brazhnik 2014, 26). When enacted, these regulations disincentivized some companies from entering the Russian market. Such was the case with Spotify, which canceled its plans to launch services in Russia in 2015 due to the localization requirement (Garrie and Byhovskiy 2017, 244).

The law imposes obligations for data operators and provides new competences to Roskomnadzor. When collecting and processing online data regarding Russian citizens, an operator must use databases (servers) that are located in Russia. Roskomnadzor received expanded competences while the entities that it supervises lost some guarantees. Following article 3 of the Localization Law, Roskomnadzor in its control and supervision over personal data protection no longer follows the guarantees provided to legal entities and sole entrepreneurs by the Federal Law “*O zashite prav ūridičeskikh lic i individual’nykh predprinimatelej*” (On the protection of businesses). In practice, this means more freedom to Roskomnadzor and less control over its actions from other public authorities. For example, the Public Prosecution Office controls public authorities by approving their plans for inspections of businesses. Following Section II of the Roskomnadzor Inspection Rules, Roskomnadzor now plans its inspections without coordination with the Prosecution Office and has more freedom in making changes to inspection plans.

Roskomnadzor has defined priority spheres of interest where it most diligently monitors compliance with localization requirements. These spheres include, but are not limited to, recruiting agencies, credit companies, hotel businesses, and insurance companies (Roskomnadzor 2017). In these niches, by the very nature of business (recruiting agencies) or due to legislative requirements (insurance and credit companies), companies have to collect customers’ personal data.

### 6.3.1 *Subjects of the Obligation*

Following article 18(5) of the Personal Data Law, when collecting personal data of Russian citizens, data operators should provide for recording, systematization, accumulation, and storage of data by using databases (servers) located in Russia. It is important to note that the localization requirement is limited to only some of the actions that constitute data processing—collecting the personal data of Russian citizens. Correspondingly, other actions of data processors, including usage, anonymization, erasure, and destruction, are not subject to this requirement.

Roskomnadzor has issued a clarification on when a data operator needs to comply with regulations. Such instances include using a domain name that is connected to Russia, like ru, рф, or su; having a Russian-language version of a website; and/or performance in Russia of a contract made on a website. In practice, this means that if an online store offers delivery to Russia, it needs to use a Russian server to process the data of Russian citizens.

### 6.3.2 *Registry of Infringers*

Roskomnadzor keeps a constantly updated Registry of Infringers of the Rights of Personal Data Subjects. In August 2016, it filed a claim to include the social network LinkedIn in the Registry of Infringers for failures to comply with the localization requirement and other data protection laws (*Roskomnadzor v. LinkedIn Corporation*). After winning the case in the court of first instance and the court of appeal, Roskomnadzor blocked LinkedIn. LinkedIn is not the only major internet service that received the attention of Roskomnadzor. According to the commentaries of Roskomnadzor representatives, Facebook and Twitter also did not comply with the regulations. However, a differentiated treatment was given to LinkedIn due to “repeated reports of data leaks from LinkedIn” (Bondarev et al. 2016). Perhaps the Russian government expected LinkedIn to comply given that LinkedIn located its servers in China to avoid the ban (Mozur and Goel 2014). Twitter and Facebook failed to comply with localization requirements in China and were banned there.

### 6.3.3 *Amplification of Fines for Infringement*

Article 13.11 of the Russian Code of Administrative Offences (CAO) establishes penalties for the infringement of Russian data protection regulations. Currently, it does not contain penalties for failing to comply with the localization requirement. Because of this, when Roskomnadzor was trying to pressure Twitter and Facebook into localizing their databases, the federal service had to fine the companies only for failing to provide information about the localization of their databases—an infringement provided in article 19.7 of the CAO. The maximum fine in this article is 5000 rubles (approximately 70

euros). Correspondingly, Twitter and Facebook were fined 3000 and 5000 rubles, respectively.

To influence this situation, the State Duma is considering the Draft Federal law “On amending the Code of Administrative Offences of Russia.” The draft introduces special provisions for the violation of the data localization requirement and substantially increases fines—up to 18 million rubles (approximately 252,000 euros). Roskomnadzor will likely not attempt to block Twitter and Facebook for several reasons. First, Twitter and Facebook already demonstrated in the Chinese market that they are not willing to compromise under the risk of a ban. Second, blocking them will cause a bigger international response than that of LinkedIn. Third, Roskomnadzor does not have the technical means to properly implement a ban against such giants, as the futile attempt to block Telegram messenger demonstrated (discussed in Sect. 6.4).

## 6.4 YAROVAYA LAW

In 2016, the State Duma enacted two laws that are commonly referred to by the name of one of their authors—Irina Yarovaya—Federal Law 374-FZ and Federal Law 375-FZ (Yarovaya law). As per the Yarovaya law, organizers of data distribution are bound to store transferred information and provide Russian enforcement authorities with encryption keys (for more, see Chap. 5).

### 6.4.1 *Storing Requirement*

According to the newly introduced article 10.1 of the Data Protection Act, from July 2018, organizers of data distribution on the internet should, first, store text messages, voice communications, images, audio, video, and other messages of users in Russia for six months and, second, store all these messages’ and users’ metadata for one year. To top this off, in April 2018 the government of Russia issued a Resolution binding telecommunications providers to store all internet traffic data for 30 days (Resolution on Internet Traffic). As per the report of the Analytical Credit Rating Agency of October 2018, the aggregated cost for implementing these measures just for Russian mobile networks will exceed 250 billion rubles (approximately 3.5 billion euros) (Tishina 2018). The volume of stored data for 2019 is estimated at 60 exabytes (60 billion gigabytes), which is challenging to implement (Kolomychenko 2016).

A more controversial part of these amendments is the duty of the organizer of data distribution to provide state intelligence and surveillance authorities with access to the above-listed information. Data organizers will have to provide Russian enforcement authorities access to sensitive information without a court order. The aforementioned April 2018 Resolution of the Government, in clause 4, officially includes technical means of data accumulation into communications equipment of intelligence and surveillance operations. By this inclusion, the Resolution provides unmonitored access for enforcement

authorities to stored data of telecommunication providers. Communications equipment of enforcement authorities is constantly connected to data accumulation centers. Authorities do not need to ask for access to this information or even notify service providers. The GDPR does not provide for any comparable duty. Such obligation is clearly aimed at easing state control and not toward the protection of individual rights for personal data.

Similar regulation for internet organizers of data distribution was issued on October 29, 2018, by the Decree of the Ministry of Communications. Clause III(4) of the Decree sets up a upfront requirement for data distribution organizers—technical means should provide search, processing, and transfer of stored data to FSB (*Federal'náá služba bezopasnosti*, Federal Security Service). Roskomnadzor keeps a Registry of Organizers of Data Distribution. As of October 2019, the registry contains 182 entries. Among the companies that are listed as organizers (and, correspondingly, bound to comply with the technological requirement of providing access to the Federal Security Service) are services like social network VKontakte, public email services Mail.ru and Mail.Yandex, cloud storage service Disk.Yandex, dating service Tinder, and classified advertisements website Avito. Being on that list and refusing to provide access to data can lead to blocking of the corresponding company's website.

Even before enactment of the Data Protection Act and Personal Data Law, regulations required Russian mobile operators to install devices providing access to Russian enforcement authorities to messages transmitted over mobile networks. These provisions were the subject of a dispute resolved by the European Court of Human Rights (ECtHR) in the case of *Roman Zakharov v. Russia*. Roman Zakharov (applicant), the editor-in-chief of a publishing company, filed a claim against Russian mobile telecom companies for violating his right to privacy of telephone communications. The mobile companies provided access for the FSB to install equipment intercepting all telephone communications. After losing this case in Russian courts, on October 20, 2006, Zakharov applied to the ECtHR.

In its judgment of December 4, 2015, the ECtHR noted that the legislation in question requires mobile operators to install equipment allowing the FSB to intercept communications of all users. The FSB does not need to notify users or telecom companies of such intrusion. The ECtHR indicated that the interception of telephone conversations can be justified by the aims of protection of national security, public safety, and prevention of crime. Such was the case in Russia. At the same time, legislation should provide adequate safeguards against abuses and guarantees that such a system will only be used when these measures are necessary. In view of the ECtHR, Russian legislation allowed such secret measures “in respect of a very wide range of offenses.” Telephone conversation interceptions can be applied not only in regard to suspects but also toward persons that might possess information about an offense. The secrecy of interceptions was subject to court control. As a general rule, any interception needed a prior court order. Yet, some information, for example, about undercover agents or about the organization and tactics of conducting

operational-search measures, could not be submitted to a court. As a result, courts were not able to assess how reasonable the measures were. Courts could also order measures that were very wide in scope—like authorizing the interception of all conversations in the area where a crime was committed, without limiting it to specific persons. Enforcement authorities are not bound to notify telecom users that their conversations are intercepted. In light of the above-mentioned argument, the ECtHR found that Russian legislation “did not provide adequate and effective guarantees against arbitrariness and the risk of abuse.”

The very same day that the ECtHR made this ruling, the State Duma approved the draft law amending the Federal Constitutional Law “On the Constitutional Court of Russia.” The amendments allow the Constitutional Court to consider whether enforcement of an ECtHR decision will be contrary to the Russian Constitution and allow refusal of performing such a decision.

#### 6.4.2 *Encryption Keys*

Having access to stored information does not necessarily allow enforcement authorities to reach their goals. The majority of transferred data is encrypted. To get access, for example, to the messages of the users, the enforcement authorities will need to possess encryption keys. Following article 4.1 of the Data Protection Act, when organizers of data distribution use encoding, they have to provide the FSB with keys for decoding electronic messages. The most notorious case based on the implementation of this rule was the conflict between FSB and Telegram messenger. In July 2017 Roskomnadzor included Telegram into the registry of organizers of data distribution. FSB requested Telegram to provide it with encryption keys. Telegram refused and Roskomnadzor applied to the Taganskij district court of Moscow to fine and block Telegram (*Roskomnadzor v. Telegram Messenger Limited Liability Partnership*). The court ruled in favor of Roskomnadzor. The Supreme Court of Russia upheld the decision. For technical reasons, Roskomnadzor was not able to block Telegram. In its crusade against the messenger, Roskomnadzor blocked over 50 virtual private network (VPN) services and anonymizers (Tereshhenko 2018, 148). The services of Yandex, Viber, Google, and VKontakte had interruptions or were blocked for some time in the implementation of these measures (Suharevskaja 2018). Yet, these measures turned futile.

The FSB requested Yandex, Russia’s largest technology company and fifth largest search engine worldwide, to provide it with encryption keys (Kolomychenko 2019). Yandex offers over 70 services in Russia that include public email, cloud storage, and online map services. At first, Yandex made a public refusal to provide FSB with the keys. Later, Yandex and the head of Roskomnadzor reported that Yandex and the FSB were able to find a solution to comply with the Yarovaya law but did not disclose the details of such solution (Kuznecova and Vyrodova 2019).



## 6.5 SOVEREIGN RUNET

### 6.5.1 *Russian Informational Security*

Since 2016, the protection of personal data is no longer a priority direction of Russian informational security doctrine. Personal data protection lost its place to countering the threats of informational security from foreign countries and actors. This conclusion can be made by analyzing the 2016 Presidential Decree of Vladimir Putin, which set up a new Doctrine of informational security in Russia (Doctrine). Following Clause III of the Doctrine, the President sees the main threats to Russian informational security coming from hostile geopolitical, military-political, terrorist, extremist, and criminal aims of unnamed foreign countries and actors. The Doctrine is predominantly focused on establishing protection and responses in the military sphere. The Doctrine replaced the 2000 Doctrine of Informational Security, which was also introduced by Putin. What is interesting is that the 2000 Doctrine set the protection of interests of a person as the first goal.

The 2016 Doctrine aims to protect the “critical informational infrastructure” of Russia. In 2019, in the implementation of the Doctrine, the State Duma has introduced amendments to the Data Protection Act and the Federal Law on Communications (Sovereign Runet law). The amendments introduce a set of measures aimed at ensuring the stable operation of the Russian internet (Runet). According to article 56.1(1) of the Law on Communications, the obligation to ensure safe, steady, and integral functioning of the Runet falls on the communications operators and owners of communications networks. Roskomnadzor will be carrying out primary state policies in this area (for more on Runet, see Chap. 16).

### 6.5.2 *Runet Law*

Internet providers need to install in their network the technical means (black boxes) for countering threats to stability, security, and integrity of the internet in Russia (article 46(5.1)). Roskomnadzor will provide the black boxes. The same article directly relieves internet providers from the obligation to limit access to prohibited websites. This is now the function of the black boxes.

Roskomnadzor receives centralized control over the entire Runet in cases of discovering a threat to the functioning of the networks (article 65.1). The government of Russia is yet to define what types of threats qualify for empowering Roskomnadzor with centralized control (article 65.1(5)). According to the head of Roskomnadzor, even a mere ban of a website already constitutes such a threat (Suharevskaja 2019). Thus, for now, it is not clear what should be the scale of the threat to transfer centralized control to Roskomnadzor. The legislator, by changing the heading of the encompassing chapter of the law, signals that a non-exceptional threat could be sufficient. The name changed from “Managing communication networks in cases of emergency and the state of emergency” to “Managing communication networks in certain cases.” Thus,



the state of emergency was downgraded to “certain cases.” Roskomnadzor will have centralized control over Rунet beyond emergency cases.

Once the black boxes are installed, the Russian government will be able to control domestic traffic and, if needed, turn off incoming foreign traffic.

## 6.6 A NEW INTERPRETATION OF PERSONAL DATA

In December 2018, Roskomnadzor presented its new vision of personal data by including cookies in the scope of the term. Cookies collect certain data about the users to, for example, tailor advertisements to the user’s location and browsing activity. The use of cookies on a website in terms of data protection is a controversial issue at the moment. The Russian legislation does not define the term “cookies.” The legal analysis of cookies stems from the definition of personal data in Russian law. As discussed earlier, to be considered personal data, user data needs to allow a natural person to be identified. Roskomnadzor representatives themselves, in a 2015 commentary to the Personal Data Law, stated that the data should not be considered personal data if it does not allow identifying a natural person without the use of additional information (Gafurova et al. 2015, 15). It took a political case for Roskomnadzor to change the opinion.

The use of cookies was one of the subject matters in the dispute involving the “Smart voting system” of Russian political activist Alexei Navalny (*Roskomnadzor v. Gandi SAS*). Navalny’s goal was to prevent the domination of United Russia party candidates in the regional and municipal elections of 2019. The system was built with the idea of uniting pro-opposition votes in each voting district for a single candidate that has the highest chance of winning the election against United Russia’s representative. Voters could register on the website and on the day of elections would receive a text message with the name of an opposition candidate that has the highest winning chance. The website, <https://2019.vote/>, was registered to Gandi SAS. Roskomnadzor claimed a violation of data protection legislation by the website and applied to a court. Among the violations, Roskomnadzor stated that by using the services of Google Analytics and Yandex Metrica the website collected and processed personal data of its users.

Google Analytics and Yandex Metrica collect the data of users. Such data can include the location of the user, the device used to access a website, browser, and internet protocol (IP) address. This data itself, without the use of additional information, does not allow identifying a natural person. Nevertheless, in the eyes of Roskomnadzor and the court, using cookies through the services of Google Analytics and Yandex Metrica constituted data collection and processing. Navalny appealed the decision but with no success. In this understanding, Roskomnadzor goes against its commentaries on the scope of personal data. At the same time, if compared to the way other countries apply data protection legislation with regard to cookies, the measure is appropriate. For example, the GDPR specifically states that cookies may allow identifying a natural person (Recital 30).

## 6.7 CONCLUSION

The law gives a vague definition of personal data. It allows Roskomnadzor to include in personal data new types of data without ever needing to amend legislation. The inclusion of cookies into the scope of personal data is one such example. This step, although following the understanding of personal data in the GDPR, differs from Roskomnadzor's former understanding of the term. The duty of data distributors to store users' data substantially eases monitoring of data for Russian enforcement authorities. This duty is not an invention of the Russian legislation. The 2006 EU Data Retention Directive introduced similar measures. The major difference from Russia was that the EU Directive required data operators to store the metadata (e.g., telephone numbers and IP addresses), not the data itself. Russian legislation, apart from that, creates convenient conditions for the enforcement authorities to obtain access to data. In 2014, the Court of Justice of the European Union invalidated the Directive for violating fundamental rights (*Digital Rights Ireland v. Minister for Communications*).

Fines for breaches of data protection legislation will be increased to provide Roskomnadzor with an additional instrument of pressure. Blocking websites can be an effective measure, but blocking giants like Google will be noticeably harmful to the Russian economy itself, as many Russian companies use Google cloud services. Trying to ban Twitter and Facebook might prove futile since Roskomnadzor was not able to block a much smaller messaging application, Telegram. Once the black boxes are fully implemented, Roskomnadzor will have much more capabilities in blocking services and websites with great precision. At the same time, the law "On the Sovereign Runet," despite being enacted, still needs substantial time before it can be properly implemented.

Among the expected novelties of Russian legislation is the introduction of the Big Data concept. Big Data allows re-identifying a person from a data set that seems to have no direct link, as well as extracting personal data that an individual did not provide, through the analysis of vast amounts of information (Gruschka et al. 2019, 5027). An example of such re-identification is the 2006 release by Netflix of a data set including a user ID and movie ratings connected to such ID. By itself, this data does not allow identifying a person. When combined with other information however, such as user movie ratings of the Internet Movie Database, the data allowed identifying a Netflix customer (Narayanan and Shmatikov 2008, 121–124). Currently, Big Data does not fall within the scope of personal data in Russia. At the same time, the definition of personal data in article 4 of the GDPR allows including Big Data in the scope of personal data (Bonatti and Kirrane 2019, 7).

The Russian legislator is very active in the sphere of data protection. Almost all novelties grant new powers to controlling authorities and increase the burden of compliance even for companies located outside of Russia if their activity is aimed at Russia. Roskomnadzor plays a central role in this sphere. Roskomnadzor does not need to comply with the rules and limitations for

conducting inspections that are obligatory for other public authorities. Instead, the Federal Service follows a set of rules especially established for its activities. In the nearest future, Roskomnadzor will strengthen its position by receiving the power to exercise centralized control over Rунet. Legislative grounds for the state monitoring over the data flows grow alongside the technical capabilities of Russian government to exercise such control.

## REFERENCES

- Bauer, Matthias, Lee-Makiyama Hosuk, Erik van der Marel, and Bert Verschelde. 2015. *Data Localisation in Russia: a Self-Imposed Sanction*. Brussels: European Centre for International Political Economy (ECIPE).
- Bonatti, Piero A, and Sabrina Kirrane. 2019. Big Data and Analytics in the Age of the GDPR. In *2019 IEEE International Congress on Big Data (BigDataCongress)*, 7–16. <https://doi.org/10.1109/BigDataCongress.2019.00015>.
- Bondarev, Denis, Dmitrij Nosonov, and Anna Balashova. 2016. Roskomnadzor načal blokirovku LinkedIn [Roskomnadzor Started Blocking LinkedIn]. *RBC*, November 17. [https://www.rbc.ru/technology\\_and\\_media/17/11/2016/5829cb809a7947c578b9cfd](https://www.rbc.ru/technology_and_media/17/11/2016/5829cb809a7947c578b9cfd).
- Gafurova, Alfija Nasibullovna, Elena Vladimirovna Dorotenko, and Jurij Evgenevich Kantemirov. 2015. *Federal'nyj zakon "O personal'nyh dannyh."* Antonina Arkadevna Priezževa, ed. Moscow: Redakciã "Rossijskoj gazety."
- Garrie, D., and Irene Byhovskiy. 2017. Privacy and Data Protection in Russia. *Journal of Law Cyber Warfare* 5: 235–255. <https://doi.org/10.1136/bmjopen-2019-EMS.28>.
- Gruschka, Nils, Vasileios Mavrocidis, Kamer Vishi, and Meiko Jensen. 2019. Privacy Issues and Data Protection in Big Data: a Case Study Analysis Under GDPR. In *2018 IEEE International Conference on Big Data (Big Data)*, 5027–5033. <https://doi.org/10.1109/BigData.2018.8622621>.
- Kolomychenko, Marija. 2016. Operatoram vystavili sčët [Operators Were Issued an Invoice]. *Kommersant*, December 26.
- . 2019. FSB potrebovala ključì šifrovaniã perezpiski pol'zovatelej u 'Yandex' [FSB Requested Encryption Keys for Users' Messages from Yandex]. *RBC*, June 4. [https://www.rbc.ru/technology\\_and\\_media/04/06/2019/5cf50e139a79474f8ab5494b](https://www.rbc.ru/technology_and_media/04/06/2019/5cf50e139a79474f8ab5494b).
- Kuznecova, Evgenija, and Julija Vyrodova. 2019. Yandex podtverdil naličie rešeniã po ključam šifrovaniã dlã FSB [Yandex Confirmed the Existence of a Decision on the Encryptions Keys for FSB]. *RBC*, June 7. [https://www.rbc.ru/society/07/06/2019/5cfa2a169a7947affada5b1a?from=from\\_main](https://www.rbc.ru/society/07/06/2019/5cfa2a169a7947affada5b1a?from=from_main).
- Mozur, Paul and Vindu Goel. 2014. To Reach China, LinkedIn Plays by Local Rules. *The New York Times*, October 5. <https://www.nytimes.com/2014/10/06/technology/to-reach-china-linkedin-plays-by-local-rules.html>.
- Narayanan, Arvind, and Vitaly Shmatikov. 2008. Robust De-Anonymization of Large Sparse Datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 111–125. IEEE. <https://doi.org/10.1109/SP.2008.33>.
- Roskomnadzor. 2017. Podvedeny itogi realizacii federal'nogo zakona o lokalizacii baz personal'nyh dannyh rossijskih graždan na territorii Rossii [Summed Up the Results of Implementation of the Federal Law on Localization of Databases of Personal Data

- of Russian Citizens on the Territory of Russia]. <https://rkn.gov.ru/news/rsoc/news49466.htm>.
- . 2018. Analitičeskij obzor meždunarodnogo opyta po lokalizacii baz dannyh, soderžaših personal'nye dannye graždan [Analytical Review of International Practices of Localising Databases Containing Personal Data of Citizens]. [https://pd.rkn.gov.ru/docs/Obzor\\_po\\_lokalizacii.docx](https://pd.rkn.gov.ru/docs/Obzor_po_lokalizacii.docx).
- . 2019a. Otvety na voprosy v sfere zašity prav sub'ektov personal'nyh dannyh [Answers for Questions in the Sphere of Protection of Personal Data Subjects' Rights]. <https://rkn.gov.ru/treatments/p459/p468/>.
- . 2019b. Reestr narušitelej prav sub'ektov personal'nyh dannyh [Registry of Infringers of the Rights of Data Subjects]. <https://pd.rkn.gov.ru/registerOffenders/>.
- . 2019c. Reestr organizatorov rasprostraneniâ informacii [Registry of Organizers of Data Distribution]. [https://reestr.rublacklist.net/distributors\\_main/](https://reestr.rublacklist.net/distributors_main/).
- Savelyev, Alexander. 2016. Russia's New Personal Data Localization Regulations: A Step Forward or a Self-Imposed Sanction? *Computer Law & Security Review: the International Journal of Technology Law and Practice* 32 (1): 128–145. <https://doi.org/10.1016/j.clsr.2015.12.003>.
- Suharevskaja, Alena. 2018. Ot blokirovki Telegram v Rossii postradali 400 internet-resursov [400 Internet Resources Suffered from Blocking the Telegram]. *Vedomosti*, May 7. <https://www.vedomosti.ru/technology/articles/2018/05/07/768819-ot-popitok-zablokirovat-telegram-postradali-400>.
- . 2019. Gosduma prinâla zakon o suverennom runete [State Duma Passed the Law on Sovereign Runet]. *Vedomosti*, April 16. <https://www.vedomosti.ru/technology/articles/2019/04/16/799258-gosduma-zakon>.
- Tereshhenko, Ljudmila Konstantinovna. 2018. Gosudarstvennyj kontrol' v sfere zašity personal'nyh dannyh [State Control in the Sphere of Personal Data Protection]. *Pravo* 4: 141–161. <https://doi.org/10.17323/2072-8166.2018.4.142.161>.
- Tishina, Julija. 2018. Zakon Yarovoï soberet dividendy [Yarovaya Law will collect dividends]. *Kommersant*, October 12: 2018.
- Zharov, Aleksandr. 2014. Kontrol' za informacionnoj sredoj—ëto absolûtno normal'no [Control Over the Infosphere is Normal]. <https://82.rkn.gov.ru/news/news70654.htm>.
- Zhuravlev, Mikhail, and Tatiana Brazhnik. 2014. Problems for Internet Business and Users Caused by New Russian Legislation. *Information Law Journal* 5 (4): 25–32.

#### LEGAL SOURCES

- Decree of the President of Russia N 1715. 2008. O nekotoryh voprosah gosudarstvennogo upravleniâ v sfere svâzi, informacionnyh tehnologij i massovyh kommunikacij [On Certain Questions of State Governance in the Sphere of Communication, Informational Technologies and Mass Communications], 3 December. <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=125069213307289911155398059&cacheid=05820B83A55F59EE0E94998C14648BF4&mode=splus&base=LAW&n=129958&rnd=0.21288845240581#1p3rvlx5azg>.
- Decree of the President of Russia N 646. 2016. Ob utverždenii Doktriny informacionnoj bezopasnosti Rossijskoj Federacii [On Establishing the Doctrine of

- Informational Security of Russian Federation], 5 December. <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=50238875209243857070365725&cacheid=3011D334331B32E8AC7532B1E6A41400&mode=splus&base=LAW&n=208191&rnd=C93498B21CA171595106BE62A5A5A7AF#2cicpk6ophu>.
- Digital Rights Ireland Ltd v. Minister for Communications, Marine and Natural Resources and Others and Kigital Landesregierung and Others*. 2014. Court of Justice of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62012CJ0293>.
- Draft Federal law N 01/05/08-19/00093621. O vnesenii izmeneniâ v Federal'nyj zakon 'O personal'nyh dannyh' v časti utočneniâ trebovanij k uničtoženiû personal'nyh dannyh [On Amending the Federal Law 'On Personal Data' in Clarifying the Duty to Confirm the Termination of Personal Data] (Draft law On Termination of Personal Data). <https://regulation.gov.ru/projects#npa=93621>.
- Draft Federal law N 729516-7. O vnesenii izmenenij v Kodeks Rossijskoj Federacii ob administrativnyh pravonarušeniâh [On Amending the Code of Administrative Offences of Russia]. <https://sozd.duma.gov.ru/bill/729516-7>.
- Draft law N416052-6. O vnesenii izmenenij v Federal'nyj zakon 'O personal'nyh dannyh' i stat'û 28.3 Kodeksa Rossijskoj Federacii ob administrativnyh pravonarušeniâh [On Amending the Federal Law 'On Personal Data' and Article February 28 of the Code of Administrative Offences of Russia]. [http://asozd2.duma.gov.ru/main.nsf/\(Spravka\)?OpenAgent&RN=416052-6](http://asozd2.duma.gov.ru/main.nsf/(Spravka)?OpenAgent&RN=416052-6).
- EU Regulation N 2016/679. 2016. On the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, 27 April (GDPR).
- European Parliament and Council Directive N 2006/24/EC. 2006. On the Retention of Data Generated or Processed in Connection with the Provision of Publicly Available Electronic Communications Services or of Public Communications Networks and Amending Directive 2002/58/EC, 15 March (EU Data Retention Directive).
- Federal Constitutional Law N 1-FKZ. 1994. O Konstitucionnom Sude Rossijskoj Federacii [On the Constitutional Court of Russian Federation], 21 July. <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=59026883909335713102225273&cacheid=EBB9123F2F2DE437A09152E8759719BB&mode=splus&base=LAW&n=303524&rnd=75A319AFD029A32757FF30715DAF729C#1rcvv770we5>.
- Federal Law N 126-FZ. 2003. "O svâzi [On Communications], July 7. <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=50238875209243857070365725&cacheid=8145F3B2B11EC4A6313C8BB2FA35483B&mode=splus&base=LAW&n=323999&rnd=C93498B21CA171595106BE62A5A5A7AF#2joroiejcyy>.
- Federal Law N 242-FZ. 2014. O vnesenii izmenenij v otдел'nye zakonodatel'nye akty Rossijskoj Federacii v časti utočneniâ porâdka obrabotki personal'nyh dannyh v informacionno-telekommunikacionnyh setâh [On Amending Certain Legislative Acts of the Russian Federation in Clarifying the Order for Processing Personal Data in the Informational-Telecommunications Networks], 21 July (Localization law). [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_165838/](http://www.consultant.ru/document/cons_doc_LAW_165838/)
- Federal law N 294-FZ. 2008. O zašite prav ūridičeskikh lic i individual'nyh predprinimatelej pri osušestvlenii gosudarstvennogo kontrolâ (nadzora) i municipal'nogo

- kontrolá [On Protection of Rights of Legal Entities and Individual Entrepreneurs Subject to State Control (Supervision) and Municipal Control], 26 December (Federal Law on the Protection of Businesses). <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=50238875209243857070365725&cacheid=CA7B5A15556F739CC816E000C5CB51A6&mode=splus&base=LAW&n=330806&rnd=C93498B21CA171595106BE62A5A5A7AF#2gxvycneh54>.
- Federal Law N 90-FZ. 2019. O vnesenii izmenenij v Federal'nyj zakon 'O svázi' i Federal'nyj zakon 'Ob informacii, informacionnyh tehnologiáh i o zašite informacii' [On Amending the Federal Law 'On Communications' and Federal Law 'On Information, Information Technologies and Protection of Information'], May 1 (Sovereign Runet law). <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=50238875209243857070365725&cacheid=F80E7209AB399E0609FB56D046A80279&mode=splus&base=LAW&n=323815&rnd=C93498B21CA171595106BE62A5A5A7AF#1xj47cvtt2e>.
- Draft Federal law N 04/13/09-19/00095069. O vnesenii izmenenij v Federal'nyj zakon 'O personal'nyh dannyh' [On Amending the Federal Law 'on Personal Data'] (Draft Law on Single Consent Form). <http://www.consultant.ru/cons/cgi/online.cgi?req=doc;base=PRJ;n=187549#043229329098463776>.
- LLC Skartel v. Roskomnadzor Administration of the Central Federal Circuit*. 2016. Commercial court of the city of Moscow. <http://kad.arbitr.ru>.
- Order of the Federal Service for Supervision of Communications, Information Technologies and Mass Media N274. 2013. Ob utverždenii perečná inostrannyh gosudarstv, ne ávláúšihšá storonami Konvencii Soveta Evropy o zašite fizičeskikh lic pri avtomatizirovannoj obrabotke personal'nyh dannyh i obespečivaúših adekvatnuú zašitu prav sub'ektov personal'nyh dannyh [On Approving the List of Foreign States Non-Members of the Council of Europe Convention for the Protection of Individuals With Regard to Automatic Processing of Personal Data and Providing Adequate Protection of Rights of Personal Data Subjects], 15 March (Order of Roskomnadzor on the List of Countries with Adequate Personal Data Protection). <http://base.garant.ru/70368490/53f89421bbdaf741eb2d1ecc4ddb4c33/#ixzz61mGs68jR>.
- Resolution of the Government of Russia N 146. 2019. Ob utverždenii Pravil organizacii i osušestvleniá gosudarstvennogo kontrolá i nadzora za obrabotkoj personal'nyh dannyh [On Adoption of the Rules of Organisation and Execution of State Control and Supervision Over Processing of Personal Data], 13 February (Roskomnadzor Inspection Rules). <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=50238875209243857070365725&cacheid=8567EB662DD7CB0DB503A0C9C925737B&mode=splus&base=LAW&n=318256&rnd=C93498B21CA171595106BE62A5A5A7AF#2axuel56aqm>.
- Resolution of the Government of Russia N 228. 2009. O Federal'noj službe po nadzoru v sfere svázi, informacionnyh tehnologii i massovyh kommunikacij [On a Federal Service for Supervision of Communications, Information Technologies and Mass Media], 16 March (Resolution of the Government On Roskomnadzor). <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=463319870561769819592886&cacheid=F5E2970579ED033F234F23220E4E794F&mode=splus&base=LAW&n=334166&rnd=C93498B21CA171595106BE62A5A5A7AF#94nspil12vc>.
- Resolution of the Government of Russia N 445. 2018. Ob utverždenii Pravil hraneniá operatorami svázi tekstovyh soobšenij pol'zovatelej uslugami svázi, golosovoj informacii, izobraženij, zvukov, video- i inyh soobšenij pol'zovatelej uslugami svázi [On

Adoption of the Rules of Storing by Communications Providers of Text Messages of Communications' Users, Voice Information, Images, Sounds, Video and Other Messages of the Users of Communication], 12 April (Resolution on Internet Traffic). <http://www.consultant.ru/cons/cgi/online.cgi?req=doc&ts=50238875209243857070365725&cacheid=EFC0F91BA6AC6AE25FAB42585CF0556F&mode=splus&base=LAW&n=325767&rnd=C93498B21CA171595106BE62A5A5A7AF#rjff75k49>.

- Roman Zakharov v. Russia*. 2015. European Court of Human Rights. [https://hudoc.echr.coe.int/eng#{\"itemid\":\[\"001-159324\"\]}](https://hudoc.echr.coe.int/eng#{\).
- Roskomnadzor v. Gandi SAS*. 2018. Taganskij district court of Moscow. <https://www.mos-gorsud.ru/rs/taganskij/cases/docs/content/3a5e78f7-cb95-414d-985a-8ab2347bec9a>.
- . 2019. City of Moscow court. <https://www.mos-gorsud.ru/mgs/cases/docs/content/368d1a9a-0d98-43da-8b49-a1001f08c126>.
- Roskomnadzor v. LinkedIn Corporation*. 2016. City of Moscow court. <https://www.mos-gorsud.ru/mgs/cases/docs/content/c364d1d9-e30c-4ffa-aabb-327c8977adab>.
- Roskomnadzor v. Telegram Messenger Limited Liability Partnership*. 2018. Taganskij district court of Moscow. <https://mos-gorsud.ru/rs/taganskij/services/cases/civil/details/2cc72aea-39e7-4f8e-adc9-37d170966efa?caseFinalRangeDateFrom=13.04.2018&caseFinalRangeDateTo=13.04.2018&formType=fullForm>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.







# Cybercrime and Punishment: Security, Information War, and the Future of Runet

*Elizaveta Gaufman*

## 7.1 INTRODUCTION

Cybersecurity is notoriously hard to define (Salminen 2018), especially given the possible macro and micro incarnations from cyber war to individual password theft (Nissenbaum 2005). This chapter operates from the assumption that digital security is a complex process of intertwining of practices and technology that ensure the undisturbed functioning of information and communication technologies for the everyday needs of individuals, protected from breaches of confidentiality and anonymity. While most countries' daily functions depend on their invulnerability to cyber disruptions, a state's potential to discipline and punish its citizens through digital surveillance has hardly been underestimated as well (Dupont 2008; Teboho Ansorge 2011; Morozov 2011). This chapter explores the influence of digitalization on security in Russia, touching upon the issues of governmental control, surveillance, information war, as well as the issue of (Russian) internet sovereignty. This chapter aims to show the discrepancies in Russian cyber politics at home and abroad, highlighting its struggle for more internet regulation that is seen by the Russian government as a panacea against perceived external attempts at regime change. At the same time, this chapter shows that despite seemingly formidable "cyber army" capabilities for external use, domestic surveillance and attempts to build a Great Russian Firewall are still lacking even though the law on the isolation of the Runet has been passed in April 2019 (for more, see Chap. 2).

---

E. Gaufman (✉)  
University of Groningen, Groningen, Netherlands  
e-mail: [e.gaufman@rug.nl](mailto:e.gaufman@rug.nl)



An image of a hacker with a thick Russian accent hastily typing something into the computer has become a staple trope in popular culture from James Bond movies to Late Night Comedy shows. Despite the fact that cybernetics was considered for a long time a “reactionary pseudoscience that appeared in the United States of America (USA)” (Peters 2016), Soviet Union and later Russia discovered “possible military applications for computers.” Since then, Russia has been consistently ranked as one of the dominant cyber powers around the world (Clarke and Knake 2014) ostensibly capable to disrupt democratic elections, organize a protest movement, or shut down a government. As in many cases in Soviet Union, military needs propelled the development of an industry that soon spilled over into civilian use with unexpected consequences. Russian Internet, or Rунet, has become a threat to “regime stability” in Russia to such a degree that Soviet practices of dissident citizen surveillance and banning of anti-regime statements have found their way back to Russia today.

Externally, Russia enjoys a status of a cyber superpower (Musgrave 2016), but domestically it is still struggling to create a fully national, digitally sovereign Rунet (Ristolainen 2017) reinforcing the digitized reason of state regime (Bauman et al. 2014). So far, governmental attempts at policing and censoring the digital space have made its infrastructure more vulnerable to traffic disruption for private users and have slowed down the development of Russia’s internet industry. Moreover, increased government control that aims to rid the Russian cyberspace of “servers hosted in California,” has negative consequences for civil society, freedom of speech and privacy that would potentially restrict the services and flow of information to regular Russian internet users. This is, however, so far of little concern to the Russian government that aims at creating a fully Russian Net by the end of 2021 that would be completely independent from international internet infrastructure (Vedomosti 2016)

Even though destruction of critical infrastructure has always been a staple part of military strategy, it has been gradually conceptualized as a matter of national security during the Cold War and especially after 9/11 (Collier and Lakoff 2008; Aradau 2010). In the digital age, the notion of critical infrastructure has been expanded to mean not just power plants and bridges, but also banking systems and e-government with both “old” and “new” infrastructure susceptible to outside threats that, in turn, take the form of not only missiles and troops, but also computer malware and the proverbial hackers. If in the times of Lenin, one had to physically take over the telegraph station in order to organize a successfully revolution (Lenin 1969), nowadays a person with a messenger app like Telegram or WhatsApp installed in a smartphone far away from the social movement can seemingly do as much damage (Kow et al. 2016).

Third World War, according to numerous Russian scholars, is supposed to be informational-psychological (Samokhvalova 2011; Vladimirov 2013; Markov and Nevolina 2018; Kiselev 2015), which stands in sharp contrast with the Western fears of “Digital Pearl Harbors” and “Weapons of Mass Disruptions” (Hansen and Nissenbaum 2009). Russian Military Doctrine has

emphasized the need to develop armed forces and means for an “information confrontation” since 2010 (Kremlin 2010). This belief is partially rooted in various conspiracy theories that argue that “the West” is on a quest to destabilize Russia through corrupting Russian core values and its society (Gaufman 2017; Yablokov 2018). Russian conspiracies have their counterpart conspiracy in the West—the already-retracted Gerasimov Doctrine that is supposed to explain Russian information warfare as part of long-standing Soviet origin “active measures” that are supposed to undermine Western countries (Sanovich 2017; Galeotti 2018). The US Defense Intelligence Agency’s report on Russian Military Power even talks of how “Information Confrontation” is “strategically decisive and critically important to control its domestic populace and influence adversary states,” encompassing “Informational-Technical” (defense, attack, and exploitation) and “Informational-Psychological” (changing people’s behavior or beliefs in line with Russian’s government agenda) strategies (Defense Intelligence Agency 2017).

Moreover, continuous securitization of terrorist attacks has paved the way for governmental overreach in surveillance and the expansion of cyber capabilities—and not only in Russia (Eriksson and Giacomello 2007). Encryption technology and the Internet in general have been framed as a cesspit of, for example, terrorists and pedophiles—a rhetoric remarkably similar among surveillance proponents in the United States and in Russia (Gaufman 2017; Monsees 2019). Discursively linking the Internet with crime has worked remarkably well across the world, justifying the surveillance and policing of “trembling creatures” who use it. No wonder that the notion of security has become inextricably linked with the cyberspace. This chapter outlines the digital turn in the Russian government’s understanding of security that is primarily concerned with regime stability and curbing outside influence. Hence, its focus is on governmental control of the Internet, surveillance, cyber war, and the struggle for global internet governance à la Russe.

## 7.2 FREEDOM OF SPEECH VS. THE GOVERNMENTAL CONTROL OF THE RUNET

Most analysts contend that the Arab Spring was a driving force behind the Russian government’s attempts to put the Russian Internet under more stringent control emulating a Chinese model (Soldatov and Borogan 2015). Images of toppled dictators whose grip on their countries seemed unwavering must have sent shockwaves through the Kremlin, where it was now obvious that the Internet could be more than just kitchen talk 2.0. Evidence of massive resources that have been invested in regulating and penetrating the Russian blogosphere since 2010–2011 shows that the Russian leadership was also keenly aware of the influential role played by new media in shaping public opinion and wasted no time trying to “manage” them. The government’s attitude toward new media would appear to be encapsulated by the famous phrase used in early

2012 by Stanislav Govoruhin, then head of Putin's reelection campaign staff, who described the Internet as "a rubbish-dump controlled by *GosDep* (the US State Department)." And yet, the Russian government wastes no time making sure that the "dump" is under control.

Even before the protest wave of 2011–2012, there has been evidence of the government's involvement in Runet albeit not on a grand scale. During the 2011–2012 elections, Distributed Denial-of-Service (DDoS) attacks on oppositional websites, seemingly with state involvement, were registered by numerous independent organizations (Mikhaylova 2012). The 2012 "Kremlingate" scandal also showed that the Russian authorities had in fact gone much further than merely obstructing oppositional media, and that millions of rubles had been spent by the government with the aim of channeling online discussions in the desired direction (Karimova 2012; RFE/RL 2015; for more on history of this period of Runet, see Chap. 16).

The hacked correspondence between then head of the Agency for Youth Affairs Vasily Yakemenko and his deputy Kristina Potupchik demonstrated as early as 2011 and 2012 that a significant amount of budgetary funds was being spent on paying an "army of bots"—people paid to write online comments and posts on themes of interest to the government. These online warriors reportedly took their cue at least in part from the current discourse on *Russia Today* (RT) and *Pervyj kanal* (Delovoi Peterbrug 2014). An ad hoc 50-ruble commentary has transformed into a large company—the now-infamous Internet Research Agency—that employs people on a regular basis and supplies pro-Kremlin content at home and abroad. Pro-Kremlin-paid internet commentators are the frequent butt of jokes. For example, a cartoon by an oppositional caricaturist Ėlkin shows an internet user measuring his online speed based on the number of "Kremlin-bot" comments appearing on a particular post (Radio Svoboda 2014). The amount of financing that went and is still going into paying for pro-Kremlin commentators and bloggers shows that the Kremlin considers online public sphere an important battlefield. Only the battlefield has also moved outside of Russia as well, with the so-called Kremlin trolls "invading" other countries (see below Cyber War).

The Russian government has been steadily trying to tighten the grip on the Internet for the fear of violent regime change reminiscent of the Arab Spring. The Federal Service for Supervision of Communications, Information Technology and Mass Media or Roskomnadzor that is tasked with implementation and enforcement of laws on mass media is the Russian federal executive body that in practice carries out censorship in media and telecommunications, encompassing electronic media, information technology (IT), and telecommunications. Technically, Roskomnadzor is also supposed to be overseeing compliance with the law protecting the confidentiality of personal data, but, in reality, it cooperates with state law enforcement agencies such as FSB (*Federal'naâ služba bezopasnosti*, Federal Security Service) in order to carry out surveillance tasks (Soldatov and Borogan 2013; Ermoshina and Musiani 2017), often under the guise of anti-terrorism measures.

However, a more radical to date attempt to enforce the “anti-terrorist” concerns on the Runet was far from successful. “Telegram” is a cloud-based instant messaging app developed by the creator of Russia’s most popular social network VKontakte Pavel Durov. Telegram is one of the most popular messaging services in Russia and especially in Moscow, with many media personalities and celebrities having their own “channel.” On April 13, 2018, Telegram was banned in Russia by a Moscow court, due to its refusal to grant the Federal Security Service access to encryption keys needed to view user communications as required by federal anti-terrorism law. On the morning of April 16, Roskomnadzor began sending out requests to providers to block Telegram.

At first, the department demanded to stop access only to the Internet Protocol (IP) addresses of the servers of the messenger itself. In the first days, millions of Amazon and Google cloud services IP addresses were added to the registry of banned websites, which Telegram ostensibly used to bypass blocking. Gradually, these addresses were unlocked, but Roskomnadzor’s attempts created a lot of disruptions in the everyday life of ordinary Russian businesses from flower delivery to online education and had a “Barbra Streisand effect” on the popularity of Telegram itself: Telegram’s traffic increased by a third in the first month, while the number of app downloads for Android jumped twice (Meduza 2019a). The Telegram ban debacle has shown that Russia is deeply integrated into the global digital ecosystem, and because international finance and commerce rely heavily on automated solutions that generate cross-border traffic, it would be difficult and costly to create an Internet kill switch. Moreover, with a wide availability of Virtual Private Network (VPN) technology, it is easy to circumvent the ban by getting access to transnational traffic, thus making Runet far from being behind a great wall.

From a governmental point of view, however, this represents a major security risk. Most of the legislation aimed at tightening the regulation of the Runet is presented as a means to protect the public from dangerous information or counter terrorist activity; for instance, the 2012 laws “*O zâsite detej ot informacii, pričînâûsej vred ih zdorov’û i razvitiû*” (On Protecting Children from Information Harmful to Their Health and Development and Other Individual Legislative Acts of the Russian Federation on the Issue of Limiting Access to Unlawful Information on the Internet). The same packaging was applied to the so-called Yarovaya law of 2016, a set of Internet regulations that took effect in July 2018. According to the new regulations, Internet and telecom companies are required to disclose communications and metadata, as well as “all other information necessary,” to authorities, on request and without a court order. However, amid the attempts of Roskomnadzor to ban Telegram, even some governmental officials and the head of *Russia Today* Margarita Simonyan kept using the messenger. Ordinary citizens have quickly realized that noncompliance with Internet laws has led to selective “like” and “share” persecution (Verkhovsky 2018, for more on digital law, see Chap. 5).

Several court cases highlight a selective application of punishment for digital “crimes.” Aleksandr Gozenko was convicted for “inciting hate speech” after

posting four comments in VKontakte, where he allegedly wanted to “organize a *vata*<sup>1</sup> Holocaust.” Other convictions, which information and analysis center SOVA for the monitoring of xenophobia deemed as “Inappropriate enforcement of anti-extremist legislation,” included posts or memes that contained political messages opposing the annexation of Crimea or proclaiming independence of some Russian subjects of federation (Verkhovsky 2018). One of the examples of “Like, Share, Repost, Prison” became particularly prominent when famous Russian rapper Oxxxymiron publicized a case against a 23-year-old Maria Motuznaya, who posted several anti-religious memes (Novaya gazeta 2018). It is also notable that the overwhelming majority of cases of “digital extremism” that are being prosecuted were committed in VKontakte, which is obligated by law to share private information with the law enforcement agencies. At the same time, social networks (Facebook and Twitter) presumably do not cooperate with Russian security forces and are being prosecuted for non-compliance with the Federal Law No. 242-FZ, which obligates foreign companies to store data of Russian users on servers within Russia (Burgess 2019).

Persecution and punishment of digital “crimes” are in part made possible due to cooperation between the Russian government and the internet infrastructure owners (Sivetc 2018) such as hosting providers or other large IT companies. A similar infrastructure connection was observed during the crisis in Ukraine, when no sophisticated information warfare tools could be necessary given that Russia could gain physical control over the internet infrastructure in Crimea (Giles and Geers 2015). Most importantly, Russian government’s hold on the Runet is administered through control on three infrastructural levels. Firstly, Rostelecom, one of the main providers of broadband internet, already cooperates with Roskomnadzor, which means that Rostelecom would filter and block the content that violates Yarovaya law for instance. Secondly, given that local companies and platforms such as Yandex or VKontakte are much more popular than Google and Facebook, Roskomnadzor can regulate access to information through the Netoscope project, a database of unlawful websites on the Runet, in which both Yandex and VKontakte and a host of other Russian IT giants participate. Netoscope markets itself as a database of “malicious” websites and offers the service of “checking” a domain name to establish whether it can harm a user or not. Lastly, Roskomnadzor also has secured the cooperation of Technical Center “Internet,” which operates Main Registry of Runet’s Domain Name System (DNS) (Sivetc 2018), which would be morphed into a national one under the new “sovereign internet law” (Stadnik 2019). But even without the sovereign internet, Roskomnadzor can already have a say in the information flows as DNS translates Uniform Resource Locators (URL) into IP address and acts as a type of “phonebook” that Roskomnadzor is allowed to edit. Thus, governmental control of the Runet is much more tangible, and “surgical strikes” of digital speech prosecution create a lot of press coverage, but at the same time the Russian government lacks the resources so far for comprehensive digital surveillance.

### 7.3 SURVEILLANCE

During the Soviet era, a seemingly omnipotent ability of the state to spy on the population became a source of numerous precautions and jokes. Soldatov and Borogan (2015) even preface their book *The Red Web* with the Russian saying “It’s not a telephone conversation”; this reflected a Soviet-era fear of surveilled device communication and general distrust of technology that was operated by the state. Jokes about microphones hidden in ashtrays and electric plugs where “Comrade Major” is listening to the conversations people have in the privacy of their kitchens have survived to this day. It turned out, however, that “Comrade Major” has been listening not only in the Soviet Union.

In 2013, Edward Snowden leaked a cache of documents to the media that documented the existence of classified surveillance programs run by the US National Security Agency (NSA) in cooperation with secret services of other major Western powers. Snowden revelations became a turning point in the discourse on digital security and surveillance (Bauman et al. 2014). Proponents of the “net delusion” argument (Morozov 2011; Paltmaa and Vuori 2009; Dupont 2008; Golkar 2011), who argued that the Internet represents an easy tool for the government to surveil and control the population, were vindicated given the scope and reach of the surveillance programs and their potential for governmental overreach, including opposition repression as opposed to “Twitter Revolution” techno-optimists who believed in the purely emancipatory power of the Internet.

Soldatov and Borogan note that the Russian state does not have the capabilities to carry out mass surveillance that would be comparable to the scope of the NSA’s reach (Soldatov and Borogan 2015). Russian “SORM” (*Sistema tehničeskikh sredstv dlâ obespečeniâ funkcij operativno-razysknyh meropriâtij*, System for Operative Investigative Activities) is the system designated for lawful interception interfaces of telecommunications and telephone networks that enables the targeted—but not mass—surveillance of both telephone and Internet communications in Russia. SORM was first implemented in 1995 to allow access to surveillance data for the FSB, but during Vladimir Putin’s first week in office on January 5, 2000, the law was amended to allow seven other federal security agencies (apart from the FSB) to access to data gathered via SORM, including Ministry of Internal Affairs, Border patrol and customs, Police and Russia’s tax police. The legality of the SORM legislation has always been questioned, and in December 2015 the European Court of Human Rights ruled unanimously that that Russian legal provisions “do not provide for adequate and effective guarantees against arbitrariness and the risk of abuse which is inherent in any system of secret surveillance,” especially given that this risk “is particularly high in a system where the secret services and the police have direct access, by technical means, to all mobile telephone communications,” thus violating Art. 8 of the European Convention on Human Rights that stipulates a right to respect for one’s “private and family life, his home and his correspondence” (European Court of Human Rights 2015).



Nevertheless, even the aforementioned “Yarovaya Law” and the expansion of the grounds for addition to the list for blocked Internet sites in Russia were supplemented with further legislation and measures. The national project “*Cifrovaá òkonomika*” (Digital Economy) is yet again framed as an initiative that would ensure that the Runet does not contain dangerous sites for children, and at the same time identify customers of communication services and protect the users of the Internet of Things devices. An innovation in the trillion ruble national project is an obligatory installation of domestic antiviruses on personal computers manufactured and imported to Russia that already raises concerns that this measure would further commercial interests of select few governmentally approved companies that would most likely cooperate with security services (Zhukova et al. 2018). In the course of the project, a national traffic filtering system with a “white list of Internet-friendly resources for children” is supposed to be created by December 31, 2021, executed by Roskomnadzor and security agencies such as the Ministry of Internal Affairs and the FSB.

Also, according to the state program “*Informacionnoe obšestvo*” (Information Society), the goal is to have 90% of Internet traffic transmitted domestically (as opposed to 70% in 2014). Russian Minister of Culture Medinsky even stated that in the future, he guarantees that people will have to show their passport to “enter Internet,” not just in Russia, but around the world as well (Kommersant 2019). Medinsky offered the one option for the Internet policing that is being developed in the Digital Economy national project that would potentially work like parental control on devices—only the Russian government and Roskomnadzor being the “parents.” Another option, however, remains, which is the Chinese “Great Firewall”—completely encapsulating and monitoring the Runet, creating a Russian digital panopticon with endless possibilities for governmental overreach (Veeraraghavan 2013; Teboho Ansoorge 2011). Experts agree that both options seem to be on the table in the Kremlin (Soldatov and Borogan 2013; Zhukova et al. 2018). In either case, these measures will increase the possibilities and capabilities for mass surveillance.

#### 7.4 CYBER WARFARE VS. INFORMATION WARFARE

An apocalyptic vision of cyber war that involves financial market crashes, power plant meltdowns and financial collapse came to be as a warning tale from military experts amid the overwhelming enthusiasm about the Internet (Clarke and Knake 2014). This vision, however, has been amended in the policy world by the notions of “hybrid war” and “active measures” (Charap 2015; Johnson 2018; Seely 2017; Biscop 2015). The latter one references a Soviet-era term for the actions conducted by Soviet security services such as KGB (*Komitet gosudarstvennoj bezopasnosti*, Committee for State Security,) by means of media manipulation and various degrees of violence (Johnson 2018). Most scholars contend that “hybrid war” is not new or unique to Russia (Galeotti 2016; Renz 2016; Polese et al. 2016) and should not lead to panic over

YouTube cartoons about girls and bears<sup>2</sup> that can somehow indoctrinate its audience to love Putin (Galeotti 2017). While cyber warfare has been viewed in Clausewitzian terms of critical security infrastructure strikes (Rid 2012), information warfare or its more Russia-specific designations such as “hybrid war” and “active measures,” common among Kremlin critical journalists, had had much less coverage until 2007.

The Estonian Bronze Soldier controversy has become patient zero in the modern cyber war discussion (Hansen and Nissenbaum 2009). In 2007, Estonian authorities decided to remove *Alõisa*, a bronze statue in the center of Tallinn that commemorated the Soviet soldiers who fought against Nazi troops in World War II. The statue was widely seen as a symbol for Soviet occupation by many Estonians (Brüggemann and Kasekamp 2008). After the statue was relocated to a military cemetery, there were several waves of protest, both in Estonia and in Russia (Herzog 2011). Demonstrations in front of the Estonian embassy organized by the pro-Kremlin youth movement *Nashi* (Our people), an attack on the Estonian ambassador in Moscow, and finally a cyberattack on the Estonian government showed a certain degree of popular outrage, in which the role of the Russian government was seen as encouraging, if not sponsoring (Ottis 2008; Herzog 2011).

Ottis identified a clear political angle of the malicious traffic that was the core of cyberattack (Ottis 2008). For instance, malformed queries directed at the Estonian government included Russian-language swearwords calling the Estonian prime minister a “faggot” and a “fascist.” At the same time, instructions of how to attack Estonian governmental websites were scattered across Russian-language websites and even relatively primitive denial of service attack (the so-called ping flood) could have caused considerable trouble. While Ottis did not identify Russian government’s direct involvement into the attacks, he also notes that it did nothing to mitigate them, that is, to dial down the public outrage about Estonia and condemn the “patriotic” cyberattacks. Herzog also emphasizes that the Estonian cyberattack milestone showed that virtually untraceable “hacktivists” may now possess the ability to disrupt or destroy government operations. Alternatively, Rid (2012) has pushed against the designation of Estonian cyberattacks as “cyber war” as it fails to meet Clausewitz’s war criteria of being violent, instrumental, and politically attributed. Balzacq and Cavely (2016) also offer the conceptualization of “cyber incidents” instead of “cyber war” with the latter notion being the result of a securitization process of deliberate disruptions of normalized cybersecurity practices.

While other cyberwarfare weapons, such as Stuxnet,<sup>3</sup> have significantly changed the way policy and academia discuss cyber war (Langner 2011; Collins and McCombie 2012), it is Russian “disinformation campaigns” that have made headlines all over the world, spurring the development of different working groups and projects that assess the impact of pro-Russian narratives in Western countries, such as North Atlantic Treaty Organization’s (NATO), NATO Strategic Communications Centre of Excellence (StratCom COE), Center for European Policy Analysis, Australian Renewable Energy Agency



(ARENA), and numerous computational propaganda projects. The most widely publicized ones had at best murky methodology (such as Hamilton 68) or were quickly discredited (such as PropOrNot). However, as Sanovich notes,

tools like bots and trolls were developed ... to jam unfriendly and amplify friendly content and the inconspicuousness of trolls posing as real people and providing elaborate proof of even their most patently false and outlandish claims. The government also utilized existing, independent online tracking and measurement tools to make sure that the content it pays for reaches and engages the target audiences. Last but not least, it invested in the hacking capabilities that allow for the quick production of compromising material against the targets of its smear campaigns. (Sanovich 2017)

According to several journalistic investigations (Delovoi Peterburg 2014; Seddon 2014; RFERL 2015), there is a special “troll army,” that is, a team of fake internet bloggers who are hired to promote pro-Kremlin discourse. After the leak of the “bot manuals,” even a regular internet user is able to track identical comments that pollute social networks (Gunitsky 2015). Kremlin trolls are not “classic” trolls identified in the literature: even though Kremlin trolls may end up emotionally provoking the audience, their main purpose is an ideological one, while regular trolls are usually devoid of ideology (Hardaker 2010). Kremlin trolls are real people who are paid to promote Kremlin-friendly discourse.

NATO StratCom COE identifies six criteria for Kremlin trolls (2016), including being consistently pro-Russian and posting repetitive messages and not purpose-made context. Some of these criteria are helpful, but again require close-reading, which is usually impossible in large-scale data sets and still employ a rather high level of bias (NATO StratCom COE 2016). Moreover, these criteria are not applicable to targeted advertising on Facebook that was allegedly part and parcel of the Russian interference in the American presidential elections in 2016. Apart from IP data that StratCom COE used, it is hard to determine whether the person is a paid troll or not definitively. This was one of the caveats that was problematic for PropOrNot and partially for Hamilton 68. Moreover, the cited study analyzed specifically comments to popular Latvian news agencies, such as DELFI, and this methodology is not always applicable to the analysis of social networks due to design of these platforms.

Attribution in cyberspace remains one of the main challenges. It normally follows the “cui bono” (“to whom is it a benefit?”) logic, but one would always be uncertain about the identity of the perpetrator(s) (Baezner and Robin 2017), especially given that some cyber activists do not act on governmental orders—which has been the official line of the Russian government when accused of acts of cyber war. “Patriotic Hacking” is not unique to Russia as evidenced by #OPvijaya cyberattack on Pakistani governmental websites and governmental initiatives in several Asian countries to promote cyber privateering (Hare 2017). However, it was an issue related to Russian foreign policy that put cyber war on international security and NATO’s radar.

The “hacktivist” defense was used also in the case of alleged Russian interference in the US elections in 2016. US intelligence agencies and several private cybersecurity firms identified two groups with alleged ties to FSB and GRU (*Glavnoe razvedyvatel'noe upravlenie*, Main Intelligence Directorate), that were involved in the hacking of the Democratic National Committee (DNC). Two groups, Advanced Persistent Threat (APT) 29, aka Cozy Bear, and APT 28, aka Fancy Bear, penetrated the servers of the DNC and leaked private communications that had a damaging effect on Donald Trump’s rival in the presidential campaign—Hillary Clinton (Polyakova and Boyer 2018). During the 2018 Helsinki summit between President Trump and President Putin, Russian president insisted that the Russian state has never interfered and does not interfere in elections in other countries. Further, he admitted that some Russians could have sympathized with Trump, since during his election campaign he was in favor of improving ties with Moscow, and those people acted on their sympathies (Khimshiashvili 2018).

The Report by the US Office of the Director of National Intelligence “Assessing Russian Activities and Intentions in Recent US Elections” claimed that it was President Putin who directed the hacking activities, with the Central Intelligence Agency (CIA) and Federal Bureau of Investigation (FBI) confirming this with “high confidence” and the NSA with “moderate confidence.” Moreover, the Report went even further and documented not only hacking activities and trolling by the Internet Research Agency, but also some of the articles and reporting by *Russia Today* and *Sputnik* (Defense Intelligence Agency 2017). The report’s assessment of RT’s “Occupy Wall Street” documentary and critique of the US political system as “anti-American rhetoric” played into the hands of Russian interference skeptics: if a country prides itself in its freedom of speech, what kind of damage is a documentary on an anti-establishment movement supposed to do to a democracy? As Sanovich notes, maintaining the reputation of mainstream media and ensuring their objectivity, fairness, integrity and professionalism would be a much more effective defense against any kind of “active measures” (Sanovich 2017).

Sanovich’s argument rings especially true because the Russian government also relies on flooding technique in its information war effort, that is, governmentally affiliated mass media as well as Internet Research Agency trolls provide such an overwhelming amount of contradictory information that is difficult to parse (Roberts 2014). As Farrel and Schneier note, the goal is “to seed public debate with nonsense, disinformation, distractions, vexatious opinions and counter-arguments” (Farrell and Schneier 2018, 2019), and create, in Russian terms, “info-noise” that would fragment social reality and undermine public deliberation with a flurry of “alternative facts.” The problem here is that flooding is not a uniquely Russian or Chinese technique; it has become common place even in established democracies, with the United States being the birthplace of the “alternative facts” catchphrase in the first place. Can flooding on foreign ground be considered a cybercrime that warrants punishment? This is probably the assumption that led the Director of National Intelligence (DNI)

report to include *Russia Today* and *Sputnik* into the report on Russian interference into the American elections. Sanctioning content, however, is an authoritarian technique that most democratic countries cannot afford, and IT giants only recently began filtering outright false information such as anti-vaccine conspiracies (Matsakis 2019).

At the same time, the long arm of the Kremlin is often overestimated. The botched assassination attempt on the former Russian secret agent Skripal, in Salisbury, is a case in point, as the alleged killers' identities were established in a matter of days because of inadequate data protection (Bellingcat 2018). The "information noise" on the assassination attempt was performed in a much more efficient way with some journalists describing no less than 19 theories of the assassination and the head of *Russia Today* conducting an interview with the disclosed alleged assassins who pretended to be fitness coaches and dutifully described the beauty of Cathedral's spear in Salisbury.

While some departments of Russian Secret Services seem to possess a remarkable expertise (or a possibility to outsource it) in cyber warfare techniques including hacking, others have much less impressive record in information security. At the same time, the prosecution of hacker Hell (Sergej Maksimov) in Germany, who hacked and released private emails of the Russian oppositional politician Alexei Navalny, shows that outside of Russia cybercrime often ends in punishment. Similarly, in the United States, former US attorney for the Southern District of New York Preet Bharara prosecuted the infiltration and cyber theft of personal information from the database of J.P. Morgan Chase (Reuters 2015), as well as drug trafficking cases on the so-called dark web. These and several other cases show that despite the need to trace intricate trails of digital evidence across multiple international jurisdictions attribution and punishment for cybercrimes is possible.

## 7.5 INTERNET SOVEREIGNTY

After Edward Snowden blew the whistle on the scope of American mass surveillance, the US government charged him with theft of government property and two counts of violating the Espionage Act of 1917 through "unauthorized communication of national defense information" (Espionage Act, section 793(d)) and "willful communication of classified communications intelligence information to an unauthorized person" (Espionage Act, section 798(a)(3)). Eventually, he found temporary political asylum in Russia, a country hardly famous for its liberal internet governance, but probably one of the very few countries that does not have an extradition agreement with the United States and granted Snowden asylum. By harboring Snowden, Moscow seemingly had a good hand to renegotiate global Internet governance as he provided first-hand data of governmental surveillance overreach that extended well beyond US borders and threatened other countries' digital security.

The Russian government has a realist state-centric understanding of the cyberspace that is supposed to have its borders, sovereignty, and

nonintervention (Nocetti 2015), a type of digital Westphalia (Zinovieva 2013) with separate “national” Internets and principles of nonintervention. This view is not uniquely held by Russia; the Snowden revelations did indeed push many countries around the world to engage in digitized geopolitics, where cyberspace is a battlefield and each country needs to build up their cyber defenses (Bauman et al. 2014). Hence, it was rather frustrating for the Russian authorities that the United States had a joint responsibility with Internet Corporation for Assigned Names and Numbers (ICANN) to carry out the Internet Assigned Numbers Authority (IANA) overseeing global IP address allocation until 2016. This perception of the Internet was vocalized by Putin, who called it a “CIA project” and by numerous Russian officials who had blamed Russian waves of protest on social networks “whose servers are hosted in California.” With ICANN’s headquarters being indeed in California, its independent nongovernmental status is questioned by other countries as well (Becker 2019).

Runeset can, however, potentially do without the overseas servers. Most researchers note that Runet is a self-contained linguistic and cultural environment with its own well-developed search engines, social networks, and messenger services and software products often imported to other countries (Price 2017; Asmolov and Kolozaridi 2017). Introduction of Cyrillic domain addresses was a notable breakthrough after years of Latin script domination. The preparation and testing of the .рф domain started in 2007 by registrar RU center and proceeded as an application to ICANN. In January 2010, ICANN announced that the domain was one of the first four new non-Latin country code top-level domains to have passed the Fast Track String Evaluation and the domain became operational on May 13, 2010, with two websites: for the president of Russia and for the government of Russia. As of the moment of writing, the .ru domain is still approximately five times more popular than its Cyrillic counterpart. While the introduction of Cyrillic domain names could be seen (and framed) as an emancipatory move to decolonize the Internet (Leslie 2012; Farivar 2011), so far it just implemented a higher level of Russian state influence that does not necessarily have an emancipatory agenda.

Moreover, Russian government tried to forge alliances with China, Saudi Arabia, Egypt, and United Arab Emirates in order to promote a more centralized and controlled vision for the global Internet, specifically trying to introduce global Internet governance at an International Telecommunications Union (ITU) conference in 2012. This attempt was unsuccessful given the opposition from most Western countries including the United States, whose representative insisted that the conference was not supposed to deal with the issue of internet governance or the fact that it was supposed to be carried out through ITU, because that would potentially open the door to content censorship (Fitzpatrick 2012). Even further, when Russian Minister of Communications Nikiforov provided remarks in Brazil’s NETmundial conference calling to hand over the power from ICANN to ITU in light of the Snowden revelations about American mass surveillance, his speech was not even included into the conference documents. Even though several authoritarian countries are eager to

support Russian proposals for a more regulated cyberspace, with the United States so far standing behind an “Internet Freedom” agenda (Price 2017), it is unlikely that Russian suggestions will be implemented. Moreover, Deputy Head of the Ministry of Digital Development, Communications and Mass Communications of the Russian Federation Aleksej Volin remarked during a MediaForum in Shanghai, that Russia and China are looking at creating “alternative” social networks and messengers that would rival its Western analogues (RIA Novosti 2018) if Twitter, YouTube, and Facebook keep on filtering Russian and Chinese media out.

## 7.6 CONCLUSION

Cybersecurity à la Russe is marked by the authoritarian nature of the state that is primarily concerned by the question of regime survival. This logic motivates both external and internal double-pronged strategy of digital security, or, as Yatsyk succinctly puts it, “to hack abroad and ban at home” (Yatsyk 2018). While externally Russia enjoys an image of a cyber superpower, seemingly capable of unseating heads of state, Russian government’s attempts at controlling the cyberspace have not been quite as successful yet; that is why the government relies not only on filtering content, but also on flooding the information space both at home and abroad. Governmental surveillance capabilities are much less formidable, and filtering content is not particularly effective as the recent struggle to ban Telegram showed. Current attempts by the Russian government at making Runet independent from foreign traffic are especially worrisome because without the reliance on the “servers in California,” the ones in Moscow can be switched off albeit with significant political and economic costs. In the end, using the Internet is a matter of national security, according to Putin:

They [the Western intelligence agencies] are sitting there, it’s [the Internet] their invention. And everyone listens, sees and reads what you say, and accumulates defense information. And [once we have sovereign internet] they won’t. (Putin 2019)

This chapter provided the overview of the digital security strategy in Putin’s Russia. The law that is supposed to isolate Runet is officially on “ensuring safe and sustainable functioning of the Internet” on the Russian territory, echoing digital security’s stated main concerns (Meduza 2019b). At the moment of writing, it seems that the main purpose of the Russian government is to emulate the Chinese model and create the self-sustaining sovereign Runet that is independent of foreign infrastructure. This is in part motivated by the considerations of regime stability and the overwhelming perception in the Russian government that the Internet and social media could be used as a “crowbar” for regime change facilitating social mobilization of opposition groups. Viewing the Internet as a tool of criminals is not unique to Russian authorities but, at

the same time, this perception leads to the continuous securitization of cyberspace in Russia and legitimizes acts of information war and cyberattacks. Moreover, existing difficulty in prosecuting digital security offences will likely leave alleged cybercrimes of Russian secret services without punishment.

## NOTES

1. Somewhat derogatory term that is supposed to mean the supporters of the Kremlin's Ukraine policy, including Crimea annexation. The court decision "translated" this term as "Russian patriots," thus interpreting "vata" as a social group against said hate speech was used.
2. "Maša i medved" (Maša and the Bear) is a Russian fan fiction version of the Goldilocks fairytale, and is extremely popular on YouTube with over 26 million channel subscribers.
3. Stuxnet is a malicious computer worm, allegedly developed by American and Israeli intelligence service and first uncovered in 2010. Stuxnet is believed to be responsible for causing substantial damage to Iran's nuclear program.

## REFERENCES

- Aradau, Claudia. 2010. Security That Matters: Critical Infrastructure and Objects of Protection. *Security Dialogue* 41 (5): 491–514.
- Asmolov, Gregory, and Polina Kolozaridi. 2017. The Imaginaries of RuNet. *Russian Politics* 2 (1): 54–79.
- Baezner, Marie, and Patrice Robin. 2017. *Cyber and Information Warfare in the Ukrainian Conflict*. ETH Zurich.
- Balzacq, Thierry, and Myriam Dunn Cavelti. 2016. A Theory of Actor-network for Cyber-Security. *European Journal of International Security* 1 (2): 176–198.
- Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon, and Rob B.J. Walker. 2014. After Snowden: Rethinking the Impact of Surveillance. *International Political Sociology* 8 (2): 121–144.
- Becker, Manuel. 2019. When Public Principals Give up Control over Private Agents: The New Independence of ICANN in Internet Governance. *Regulation & Governance* 13: 561–576.
- Bellingcat. 2018. Full Report: Skripal Poisoning Suspect Dr. Alexander Mishkin, Hero of Russia. Accessed September 21, 2019. <https://www.bellingcat.com/news/uk-and-europe/2018/10/09/full-report-skripal-poisoning-suspect-dr-alexander-mishkin-hero-russia/>.
- Biscop, Sven. 2015. Hybrid Hysteria. *Egmont Institute Security Policy Brief* (64).
- Brüggemann, Karsten, and Andres Kasekamp. 2008. The Politics of History and the "War of Monuments" in Estonia. *Nationalities Papers* 36 (3): 425–448.
- Burgess, Matt. 2019. Russia Takes Aim at Facebook and Twitter in a Bid for Online Control. *WIRED*. <https://www.wired.co.uk/article/russia-suing-facebook-twitter>.
- Charap, Samuel. 2015. The Ghost of Hybrid War. *Survival* 57 (6): 51–58.
- Clarke, Richard Alan, and Robert K Knake. 2014. *Cyber War*: Tantor Media, Incorporated.
- Collier, Stephen, and Andrew Lakoff. 2008. The Vulnerability of Vital Systems: How 'Critical Infrastructure' Became a Security Problem. In *The Politics of Securing the Homeland: Critical Infrastructure, Risk and Securitisation*, 40–62. Routledge



- Collins, Sean, and Stephen McCombie. 2012. Stuxnet: The Emergence of a New Cyber Weapon and its Implications. *Journal of Policing, Intelligence and Counter Terrorism* 7 (1): 80–91.
- Defense Intelligence Agency. 2017. Russia Military Power. Building a Military to Support Great Power Aspirations.
- Delovoi Peterburg. 2014. Trolli iz Ol'gino perechali v novyj četyrëhëtažnyj ofis na Savuškina [Olginio Trolls Moved to a New 4-story Office Building on Savushkina]. *dp.ru*. October 28. Accessed March 22, 2015. [http://www.dp.ru/a/2014/10/27/Borotsja\\_s\\_omerzaniem\\_mo/](http://www.dp.ru/a/2014/10/27/Borotsja_s_omerzaniem_mo/).
- Dupont, Benoit. 2008. Hacking the Panopticon: Distributed Online Surveillance and Resistance. In *Surveillance and Governance: Crime Control and Beyond*, 257–278. Emerald Group Publishing Limited.
- Eriksson, Johan, and Giampiero Giacomello. 2007. *International Relations and Security in the Digital Age*. Routledge.
- Ermoshina, Ksenia, and Francesca Musiani. 2017. Migrating Servers, Elusive Users: Reconfigurations of the Russian Internet in the Post-Snowden Era. *Media and Communication* 5 (1): 42–53.
- European Court of Human Rights. 2015. *Roman Zakharov v. Russia*. Strasbourg: European Court of Human Rights.
- Farivar, Cyrus. 2011. *The Internet of Elsewhere: The Emergent Effects of a Wired World*. Rutgers University Press.
- Farrell, Henry, and Bruce Schneier. 2018. Common-Knowledge Attacks on Democracy. *Berkman Klein Center Research Publication* (2018-7).
- . 2019. Democracy's Dilemma. *Boston Review*. <https://bostonreview.net/forum-henry-farrell-bruce-schneier-democracys-dilemma>.
- Fitzpatrick, Alex. 2012. U.S. Refuses to Sign UN Internet Treaty. *CNN Business*. <https://edition.cnn.com/2012/12/14/tech/web/un-internet-treaty/index.html>.
- Galeotti, Mark. 2016. Hybrid, Ambiguous, and Non-linear? How New is Russia's 'new way of war'? *Small Wars & Insurgencies* 27 (2): 282–301.
- . 2017. Masha and the Bear are Not Coming to Invade your Homeland. In *Moscow's Shadows*. <https://inmoscowshadows.wordpress.com/2018/11/17/masha-and-the-bear-are-not-coming-to-invade-your-homeland/>.
- . 2018. I'm Sorry for Creating the 'Gerasimov Doctrine'. *Foreign Policy*, 5.
- Gaufman, Elizaveta. 2017. *Security Threats and Public Perception: Digital Russia and the Ukraine Crisis*. Basingstoke: Palgrave Macmillan.
- Giles, Keir, and Kenneth Geers. 2015. Russia and its Neighbours: Old Attitudes, New Capabilities. In *Cyber War in Perspective: Russian Aggression against Ukraine*, 19–28. CCDCOE, NATO Cooperative Cyber Defence Centre of Excellence.
- Golkar, Saeid. 2011. Liberation or Suppression Technologies? The Internet, the Green Movement and the Regime in Iran. *International Journal of Emerging Technologies and Society* 9 (1): 50.
- Gunitsky, Seva. 2015. Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability. *Perspectives on Politics* 13 (01): 42–54.
- Hansen, Lene, and Helen Nissenbaum. 2009. Digital Disaster, Cyber Security, and the Copenhagen School. *International studies quarterly* 53 (4): 1155–1175.
- Hardaker, Claire. 2010. Trolling in Asynchronous Computer-mediated Communication: From User Discussions to Academic Definitions. *Journal of Politeness Research* 6 (2010): 215–242.



- Hare, Forrest B. 2017. Privateering in Cyberspace: Should Patriotic Hacking Be Promoted as National Policy? *Asian Security* 15: 1–10.
- Herzog, Stephen. 2011. Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses. *Journal of Strategic Security* 4: 1–13.
- Johnson, Robert. 2018. Hybrid War and Its Countermeasures: A Critique of the Literature. *Small Wars & Insurgencies* 29 (1): 141–163.
- Karimova, Anastasia. 2012. Kremlëvskaâ blogodel'nâ [Kremlin Blog Maker]. *Kommersant*. Accessed February 3, 2019. <http://www.kommersant.ru/doc/1868022>.
- Khimshiashvili, Polina. 2018. Razminočnyj sammit: kak prošla vstreča Putina i Trampa v Hel'sinki [Training Summit: How the Meeting between Putin and Trump in Helsinki Went]. *RBC, Politika*, July 16. [https://www.rbc.ru/politics/16/07/2018/5b4c98d99a794739ea4c2e0a?from=materials\\_on\\_subject](https://www.rbc.ru/politics/16/07/2018/5b4c98d99a794739ea4c2e0a?from=materials_on_subject).
- Kiselev, V. 2015. Gibridnaâ vojna kak novyj tip vojny budušego [Hybrid War as a New Type of a Future War]. *Armejskij sbornik* 12.
- Kommersant. 2019. Medinsky predskazal vseмирnyj perehod na 'Internet po pasportu'. *Kommersant*. [https://www.kommersant.ru/doc/3859694?utm\\_source=facebook.com&utm\\_medium=social&utm\\_campaign=amplifir\\_social](https://www.kommersant.ru/doc/3859694?utm_source=facebook.com&utm_medium=social&utm_campaign=amplifir_social).
- Kow, Yong Ming, Yubo Kou, Bryan Semaan, and Waikuen Cheng. 2016. Mediating the Undercurrents: Using Social Media to Sustain a Social Movement. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*.
- Kremlin. 2010. Ukaz Prezidenta Rossijskoj Federacii ot 05.02.2010 № 146 O Voennoj doktrine Rossijskoj Federacii [Decree of the President of the Russian Federation from 05.02.2010 №146 On Military Doctrine Of Russian Federation]. <http://kremlin.ru/acts/bank/30593>
- Langner, Ralph. 2011. Stuxnet: Dissecting a Cyberwarfare Weapon. *IEEE Security & Privacy* 9 (3): 49–51.
- Lenin, Vladimir. 1969. Sovety postoronnego [Advice of a Stranger]. *Complete Works*, 34.
- Leslie, Christopher. 2012. *Decolonizing the internet*. London, England: SAGE Publications Sage UK.
- Markov, Yevgeniy, and Anna Nevolina. 2018. Rossiâ kak glavnyj ob'ekt sovremennyh informacionnyh vojn [Russia as the Main Target of Modern Information Wars]. *Historia provinciae—žurnal regional'noj istorii* 2 (3): 12–48.
- Matsakis, Louise. 2019. Facebook Will Crack Down on Anti-Vaccine Content. *Wired*. <https://www.wired.com/story/facebook-anti-vaccine-crack-down/>.
- Meduza. 2019a. Telegram v Rossii: kak na nëm otrazilas' popytka blokirovki? Skol'ko čelovek im pol'zuetsâ—i rastët li auditoriâ [Telegram in Russia: How Did the Blocking Affect it? How Many People Use it and Does it Audience Grow]. Accessed December 15, 2019. <https://meduza.io/feature/2018/08/07/telegram-v-rossii-kak-na-nem-otrazilas-popytka-blokirovki-skolko-chelovek-im-polzuetsya-i-rastet-li-auditoriya>.
- . 2019b. Zakonoproekt ob izolácii rossijskogo segmenta interneta predstavili v okončatel'noj redakcii. Vot kak èto budet rabotat' [The Bill on Isolating the Russian Segment of the Internet was Presented in its Final Version. This is How it Will Work]. Accessed December 15, 2019. <https://meduza.io/feature/2019/04/11/zakonoproekt-ob-izolyatsii-rossijskogo-segmenta-interneta-predstavili-v-okončatel'noj-redaktsii-vot-kak-eto-budet-rabotat>.
- Mikhaylova, Lidya. 2012. Internet as a Battlefield in Russian Elections. *The Examiner*. Accessed August 6, 2019. <http://www.examiner.com/article/internet-as-a-battlefield-russian-elections>.

- Monsees, Linda. 2019. *Crypto-Politics: Encryption and Democratic Practices in the Digital Era (Routledge New Security Studies)*, Routledge New Security Studies. Abingdon: Routledge.
- Morozov, Evgeny. 2011. *The Net Delusion: How Not to Liberate the World*. Penguin UK.
- Musgrave, Paul. 2016. If You're Even Asking if Russia Hacked the Election, Russia Got What it Wanted. *The Washington Post*, November 28.
- NATO StratCom COE. 2016. Internet Trolling as a Tool of Hybrid Warfare: The Case of Latvia. Full Report. <http://www.stratcomcoe.org/internet-trolling-hybrid-warfare-tool-case-latvia-0>.
- Nissenbaum, Helen. 2005. Where Computer Security Meets National Security. *Ethics and Information Technology* 7 (2): 61–73.
- Nocetti, Julien. 2015. Contest and Conquest: Russia and Global Internet Governance. *International Affairs* 91 (1): 111–130.
- Novaya gazeta. 2018. Oksimiron prizval vyjti na miting v podderzku figurantov del iz-za memov vo 'VKontakte' [Oksimiron Urged to Go to a Rally in Support of Defendants in Cases on Memes in VKontakte]. *Novaya gazeta*, August 13. <https://www.novayagazeta.ru/news/2018/08/13/144149-oksimiron-prizval-vyyti-na-miting-v-barnaule-v-podderzhku-figurantov-del-ob-ekstremizme-iz-za-memov-vo-vkontakte>.
- Ottis, Rain. 2008. Analysis of the 2007 Cyber Attacks Against Estonia from the Information Warfare Perspective. In *Proceedings of the 7th European Conference on Information Warfare*.
- Paltemaa, Lauri, and Juha A. Vuori. 2009. Regime Transition and the Chinese Politics of Technology: From Mass Science to the Controlled Internet. *Asian Journal of Political Science* 17 (1): 1–23.
- Peters, Benjamin. 2016. *How Not to Network a Nation: The Uneasy History of the Soviet Internet*. MIT Press.
- Polese, Abel, Rob Kevlihan, and Donnacha Ó. Beacháin. 2016. Introduction: Hybrid Warfare in post-Soviet Spaces, is There a Logic Behind? *Small Wars & Insurgencies* 27 (3): 361–366.
- Polyakova, Alina, and Spencer P. Boyer. 2018. The Future of Political Warfare: Russia, the West, and the Coming Age of Global Digital Competition. Brookings – Robert Bosch Foundation Transatlantic Initiative, The New Geopolitics: Europe.
- Price, Monroe. 2017. The Global Politics of Internet Governance. In *Technology and World Politics: An Introduction*. Routledge
- Putin, Vladimir. 2019. Vstreča s predstavitelami rossijskih informagentstv i pečatnyh SMI [Meeting with representatives of Russian news agencies and print media]. *Kremlin.ru*. Accessed September 21, 2019. <http://kremlin.ru/events/president/news/59865>.
- Renz, Bettina. 2016. Russia and 'Hybrid Warfare'. *Contemporary Politics* 22 (3): 283–300.
- Reuters. 2015. U.S. Charges three in Huge Cyberfraud Targeting JPMorgan, Others. <https://www.reuters.com/article/us-hacking-indictment/u-s-charges-three-in-huge-cyberfraud-targeting-jpmorgan-others-idUSKCN0S21VM20151110>
- RFERL. 2015. Russian Trolls' Vast Library Of Insulting Images. Accessed July 20, 2015. <http://www.rferl.org/content/russian-trolls-vast-insult-cartoon-database/26938435.html>.

- RIA Novosti. 2018. Volin rasskazal o napravleniâh sotrudničestva Rossii i Kitaâ v medi- asfere [Volin Spoke about Russian-Chinese Areas of Cooperations in Media]. <https://ria.ru/20181104/1532107675.html>.
- Rid, Thomas. 2012. Cyber War Will Not Take Place. *Journal of Strategic Studies* 35 (1): 5–32.
- Ristolainen, Mari. 2017. Should ‘RuNet 2020’ Be Taken Seriously? Contradictory Views about Cyber Security Between Russia and the West. *Journal of Information Warfare* 16 (4): 113–131.
- Roberts, M. E. 2014. *Fear, Friction, and Flooding: Methods of Online Information Control*. Doctoral dissertation, Harvard University.
- Salminen, Mirva. 2018. 2.10 Digital Security in the Barents Region. In *Society, Environment and Human Security in the Arctic Barents Region*, 187–204. Routledge.
- Samokhvalova, Vera. 2011. Specifika sovremennoj informacionnoj vojny: sredstva i celi poraženîâ [The Specifics of Modern Information Warfare: The Means and Targets of Destruction]. *Filosofiâ i obšestvo* 3: 54–73.
- Sanovich, Sergey. 2017. Computational Propaganda in Russia: The Origins of Digital Misinformation. Working Paper.
- Seddon, Max 2014. Documents Show How Russia’s Troll Army Hit America. *BuzzFeed News*. Accessed September 21, 2019. <http://www.buzzfeed.com/maxseddon/documents-show-how-russias-troll-army-hit-america#.nxEvk1YVWZ>.
- Seely, Robert. 2017. Defining Contemporary Russian Warfare: Beyond the Hybrid Headline. *The RUSI Journal* 162 (1): 50–59.
- Sivetc, Liudmila. 2018. State Regulation of Online Speech in Russia: The Role of Internet Infrastructure Owners. *International Journal of Law and Information Technology* 27 (1): 28–49.
- Soldatov, Andrei, and Irina Borogan. 2013. Russia’s Surveillance State. *World Policy Journal* 30 (3): 23–30.
- . 2015. *The Red Web: The Struggle Between Russia’s Digital Dictators and the New Online Revolutionaries*. Hachette UK.
- Stadnik, Ilona. 2019. A Closer Look at the ‘sovereign Runet’ law. *Georgia Tech*. <https://www.internetgovernance.org/2019/05/16/a-closer-look-at-the-sovereign-runet-law/>.
- Radio Svoboda. 2014. Karikatura dnâ na Radio Svoboda [Caricature of the Day on Radio Svoboda]. Accessed December 15, 2019. <http://www.svoboda.org/content/article/26648017.html>.
- Teboho Ansorge, Josef. 2011. Digital Power in World Politics: Databases, Panopticons and Erwin Cuntz. *Millennium* 40 (1): 65–83.
- Vedomosti. 2016. Runet budet polnost’û obosoblen k 2020 godu [Runet will be Completely Independent by the Year 2020]. *Vedomosti*. <https://www.vedomosti.ru/technology/articles/2016/05/13/640856-runet-obosoblen>.
- Veeraraghavan, Rajesh. 2013. Dealing with the Digital Panopticon: The Use And Subversion of ICT in an Indian Bureaucracy. In *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development*, Full Papers—Volume 1.
- Verkhovsky, Alexander. 2018. Nepravomernoie primenenie antièkstremskogo zakonodatel’stva v Rossii v 2017 godu [Improper Use of Anti-Extremist Legislation in Russia in 2017]. SOVA. Center for Information and Analysis.

- Vladimirov, Aleksandr. 2013. Osnovy obšej teorii vojny. Čast' I: Osnovy teorii vojny [Foundations of Common War Theory. Part I, Foundations of War Theory]. *Kadet.ru*.
- Yablokov, Ilya. 2018. *Fortress Russia: Conspiracy Theories in the Post-Soviet World*. John Wiley & Sons.
- Yatsyk, Alexandra. 2018. To Hack Abroad and Ban at Home: The Kremlin's Cyber Activism. In *PONARS Policy Memo*.
- Zhukova, Kristina, Vladislav Novyi, and Yulia Tishina. 2018. RuNet za stenoi: Kak nacproekt 'Cifrovaâ èkonomika' povliâet na pol'zovatelej interneta [RuNet Behind the Wall: How National Project 'Digital Economics' will Influence Internet users]. *Kommersant*. <https://www.kommersant.ru/doc/3773489>.
- Zinovieva, Yelena. 2013. Cifrovoj Vestfal'? [Digital Westphalia?]. In *MGIMO opinion*.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Digital Activism in Russia: The Evolution and Forms of Online Participation in an Authoritarian State

*Markku Lonkila, Larisa Shpakovskaya, and Philip Torchinsky*

## 8.1 INTRODUCTION: EVOLUTION OF ONLINE ACTIVISM IN RUSSIA

The development of digital technology, particularly Internet, social media applications, and mobile communications has in many ways changed the nature of activism: citizens' ways of addressing and resolving social, cultural, and political issues. For an individual citizen it is today cheaper and faster to seek, debate, and distribute news, facts, and falsehoods worldwide concerning a wide variety of issues.

Digitalization has also enabled new, “connective” and horizontal modes of mobilizing citizens, which has changed the role of social movement organizations (Bennett and Segerberg 2012). Numerous examples from the Zapatistas, Occupy Wall Street, Arab Spring, and the #metoo-movement to color revolutions of Eastern Europe and the Russian opposition protests of 2011–2013 have demonstrated the importance of online actions in informing and mobilizing citizens. These actions may be carried out by one person or by twenty

---

M. Lonkila (✉)  
University of Jyväskylä, Jyväskylä, Finland  
e-mail: [markku.lonkila@jyu.fi](mailto:markku.lonkila@jyu.fi)

L. Shpakovskaya  
Higher School of Economics (HSE University), Saint Petersburg, Russia

P. Torchinsky  
Independent scholar, Helsinki, Finland

million people; they may—depending on the context—be legal or heavily sanctioned, result in praise or imprisonment, start revolts, and overthrow governments (cf. Gibson and Cantijoch 2013; Theocharis 2015; Earl 2016; Kaun and Uldam 2018).

On the darker side, digital technology may also be used to obstruct and annihilate human and political rights as the persecution of Rohingyas in Myanmar or Russia's meddling in the 2016 United States (US) elections have illustrated. Moreover, digital technology also enables completely new ways of monitoring citizens both by profit-seeking enterprises and governments. Video surveillance, automatic face recognition, and accumulating databases on users' health, consumption habits, and movements enable new modes of control: data given out voluntarily or unknowingly on social media platforms make it possible to predict users' sexual orientation, political affiliation, ethnicity, and many other things with a high degree of accuracy (Kosinski et al. 2013).

In democratic countries the misuse of digital technology can be exposed and countered by independent professional media and democratic political institutions. In authoritarian countries lacking such counterforces, new digital media have provided governments with unprecedented tools for regulating and controlling citizens' on- and offline behavior.

Russia is a specific example of an authoritarian country with a well-educated population, widely available broadband access and a social media ecosystem dominated by domestic applications. Russia is, for example, one of the few countries worldwide, where Facebook is not the leading social network site, losing clearly in popularity to its Russian counterpart VKontakte ("In contact," more commonly known as VK). In political terms, Russia is an example of "electoral authoritarianism": a system of political governance where unfair elections are organized to furnish the ruling elite with a veneer of democratic legitimacy (cf. Gel'man 2017).

During his first term in office, President Vladimir Putin subjected Russian traditional media to state control while the Russian-language sector of the Internet (often dubbed Runet by the Russians—for more, see Chap. 16) remained practically free. Before the opposition protest wave in 2011–2013, lively discussions on social, cultural, and political issues took place on the Runet; well-known opposition activists from across the political and cultural spectrum deliberated on the LiveJournal blogging and social networking site, which, prior to the protest wave, was considered the hub of political debate in Russia (Etling et al. 2010).

The magnitude of the opposition mass protests in fall 2011, that erupted in response to the falsification of the results of the parliamentary elections and swapping of chairs by Putin and Medvedev,<sup>1</sup> came as a surprise to protesters and the Kremlin alike, which for the first time felt the political force of social media. The years 2011–2018 were marked by an intensive, state-led campaign to regulate Runet and curtail freedom of expression, which we have dubbed the "occupation of Runet" (Lonkila et al. 2020; for more, also see Chaps. 5 and 2).

The occupation marked a transition from lively online political debate and activism to a mode of oppressed activism in which expressing openly anti-Kremlin views in Russia has become risky. This has resulted in a “nymphosis” of activism: many former anti-government protesters have left politics (e.g., Pussy Riot member Maria Alyokhina) or turned inwards to family life in a Soviet manner; others have emigrated (e.g., Yevgeniya Chirikova, Boris Akunin and Ilya Ponomarev) or turned to less dangerous topics. However, as suggested by Svetlana Erpyleva (2019), a new generation of Russian activists may be emerging which merges politics and solving concrete, daily life problems.

Compared to the situation ten years ago, in 2020 Russia has only a handful of anti-Kremlin activists openly expressing their views on Rунet. LiveJournal, which banned political agitation in 2017, has lost its position as the hub of activist debate to Facebook, YouTube, Telegram, and Instagram.

In the next section we will present our notion of online activism, define the focus of this chapter, describe the variety of forms of online activism and discuss these with reference to the theory of connective action (Bennett and Segerberg 2012). In Sect. 8.3 we will first present survey results concerning Russians’ participation in various forms of activism and then investigate in detail two of the most noteworthy recent cases of contentious online activism in Russia. These two cases address first, the campaign conducted by Alexei Navalny and his FBK (*Fond bor’by s korrupciej*, Anti-Corruption Foundation), and second, the battle by the Telegram messenger service to provide online communication services that are protected against state monitoring. Telegram is a messenger application which works on many platforms, among them mobile applications (Apple’s iOS, Google’s Android) and desktop applications (Windows, Linux, and MacOS). It offers communication via text messages or voice calls and claims to be the most secure messenger on the market because of its custom encryption protocol and end-to-end encryption in secret chats. This means that the content of a secret chat can only be decrypted by the recipient of the message but not by a third party, including Telegram personnel. This feature makes Telegram a pivotal application for activists challenging the powers of the Russian state.

## 8.2 THEORIZING ONLINE ACTIVISM

### 8.2.1 *Defining Online Activism*

We define online activism, modifying the term “digitally networked participation” by Yannis Theocharis (2015, 6) to cover *citizens’ voluntary actions to raise awareness about or exert pressure in order to solve a political, cultural, or social problem.*<sup>2</sup> Our definition thus covers a wide variety of issues from social and environmental problems to human rights, local disputes, and more. In terms of organizational forms, it governs a continuum of actions and activists ranging from lone hackers and sporadic flashmobs organized by anonymous



individuals to established movements with their entrenched social movement organizations.

The definition excludes institutionalized party politics and politicians, as well as political actions by the state (e.g., state-organized trolling, individuals affiliated with or sponsored by the state, covering also indirect sponsorship and informal approvals),<sup>3</sup> but includes actions by citizens, such as opposition leader Alexei Navalny who have been excluded from institutionalized politics but who nevertheless try to influence the political process.

The attribute *online* refers to a mode of web-based activity that has become possible and ubiquitous thanks to digital technology, Internet, social media, and mobile communications.<sup>4</sup> Although our focus is on activities conducted completely or partly on the Internet, we do not consider online to be an ontologically separate sphere since the boundaries between on- and offline are becoming increasingly blurred.

Some forms of online activism resemble and overlap with their offline counterparts. A politician may, for example, be contacted either through social media, via email, or personally, and a petition can be signed both on a website and on paper. Notably, our definition includes posting, commenting, sharing, and “liking” various items in social media, but not merely reading a post or watching a video.<sup>5</sup>

Other forms of online activism are, however, qualitatively different from the traditional means of protest and are only feasible online, for example, creating, reworking, and distributing Internet memes or hacking into a computer database. Similarly, some forms of *offline* activism have characteristics which cannot be transferred online—for example, the feeling of a riot policeman’s stick hitting a citizen’s jaw.

In this chapter we first present in detail two cases of contentious action which explicitly challenge the Kremlin. We have selected these cases because they are among the most prominent and well-known forms of Russian online activism, and have also managed to incite related street protests. In addition, we will present examples of visible and significant, but non-political forms of activism.<sup>6</sup>

Although the focus of this chapter is on online activism, one should remember that an important part of the political activism in Russia is still conducted entirely offline. During the campaigns of opposition leader Alexei Navalny, for example, volunteers distribute printed leaflets in the staircases of apartment blocks in Russian cities to inform people about forthcoming street protests.

### 8.2.2 *Types of Online Activism*

There are multiple types of online activism. The list of new forms is continuously growing with the development of technology, and various forms have been actively employed by both international and Russian activists. Among the most prevalent forms are the *posting, debating, and sharing of relevant information* online in various social media applications such as social networking

sites. Another important form of online activism is *mobilizing and coordinating* actions, for example, setting up an event or group site on Facebook. Through *witnessing* activists transmit information about events ignored by the state-controlled media in Russia, for example, by streaming videos of opposition street protests in real time. The video *On vam ne Dimon* (He is not Dimon to you) published by Alexei Navalny's team and accusing prime minister Medvedev of corruption also utilized *social media doxxing*: finding and publishing private information about an individual—this time the prime minister of Russia—on the Internet.<sup>7</sup>

*Crowdfunding and crowdsourcing* have been used, for example, to collect money to fund Boris Nemtsov's pamphlets about Putin, to support the independent channel TV Rain (*Dožd'*), to raise money for Navalny's anti-corruption project *RosPil*, to pay the fines imposed by the court on the Russian liberal magazine *New Times* and to investigate the downing of Malaysian Airline flight MH17 (cf. Sokolov 2015).

Still other forms of online activism include, among others, *leaktivism* (e.g., wikileaks), *hashtag activism* (raising awareness of an issue across various social media platforms; e.g., the #metoo movement, #Navalny2018) and *hacking* and *distributed denial of service (DDoS) attacks*.

To manage this growing multitude of types of online activism, we propose, modifying Sandor Vegh's (2013) classification, to divide online activism into *communicative activism* and *technoactivism*. *Communicative activism* refers primarily to human-to-human interactions: exchanging information and raising awareness of societal problems and issues among people. The second form of communicative activism includes mobilizing and organizing people to act either on- or offline—for example, to sign an e-petition or to participate in a street protest. Communicative activism usually takes place on widely available platforms, such as popular social networking or video sharing sites. Since it requires no sophisticated technical skills, it is the most common type of online activism.

By *technoactivism* we refer to the actions by humans to manipulate technological systems. These may include hacking into a central bank database, programming bots, or mounting digital resistance as in the case of the instant messaging service Telegram's efforts to avoid blocking by the Russian state (see Sect. 8.3). A second form of technoactivism is data activism, by which we mean the use of either publicly available or open, but not widely known, datasets to bring about a change in society. In comparison to communicative activism, technoactivism typically presupposes technological know-how and competences, which exceed those of an average Internet user.

Russian examples of data activism include exposing corrupt state-sponsored purchases, such as buying luxury cars for the Ministry of Emergency Situations instead of fire trucks<sup>8</sup> or publishing data on expensive property belonging to modestly salaried Russian state employees. Still another example concerns using data available in a specific industry (e.g., a list of blocked Internet Protocol (IP) addresses and websites) to publish unfair or erroneous actions by

state agencies such as the Internet watchdog Roskomnadzor (see <http://rkn.gov.ru/>).

In empirical cases, different forms of online activism may blend into a combination of these types. In their anti-corruption campaigns Navalny's staff, for example, combines forms of communicative activism (YouTube videos and blog posts) with forms of data activism and social doxxing (using public databases to identify and disclose assets and properties of Russian politicians or oligarchs at home and abroad).

### 8.2.3 *Online Activism as Connective Action*

We relate online activism to Lance Bennett's and Alexandra Segerberg's theory of connective action (Bennett and Segerberg 2012). The authors contrast traditional collective activism to the "connective" variety, the latter being only possible via new digital media.

In traditional collective action the advocates of a cause share the same collective action frame and the actions are coordinated by a social movement organization in a top-down manner. To put it bluntly, the members of the traditional communist movement shared the Marxist ideology and the movement's problem consisted of selling this common ideology and action frame to followers.

In connective action, by contrast, the participants may find their own, easily personalized action frame and entry point to activism with no obligation to adhere to a clear-cut ideology. The volunteers and supporters may only share a vague and inclusive action frame (e.g., "we are the 99%," "for fair elections")<sup>9</sup> and their grass-root actions are not dictated from above but there is room for creativity and improvisation.

Bennett and Segerberg (2012, 756) distinguish between three forms of connective action. In the first ("self-organizing networks") the action is completely grass-roots based and mobilized horizontally by the users via Internet without a central coordinating organization. In the second form ("organizationally enabled networks"), there is an organization coordinating action in the background but giving leeway for users to find their own, personal ways to participate. In the third form ("organizationally brokered networks"), there is strong organizational coordination of action.

In our empirical cases of online human and technoactivism presented in the next section, the three-fold classification above can be thought of as a variable of increasing organizational coordination. In the crowdfunding instances of Russian activism (e.g., saving the magazine *New Times*, initiatives conducted through [change.org](http://change.org)), there is usually little or no organizational coordination since the action consists of donating money through a ready-made online platform. In other instances, such as in the campaigns by Navalny's team described in the next section, hierarchical organization coordination is combined with a horizontally networked group of volunteers.

### 8.3 ONLINE ACTIVISM IN TODAY'S RUSSIA

In this section we first present empirical data on the on- and offline forms of activism in Russia based on the 2016 European Social Survey data in comparison to four European countries. Second, we illustrate communicative and technoactivism based on two case studies. The two cases are selected because we consider them to be among the most prominent and successful campaigns so far in a struggle against the Russian state's "occupation" of Runet (Lonkila et al. 2020). The first case is an example of communicative activism conducted by Alexei Navalny and his Anti-Corruption Foundation and the second an example of technoactivism conducted by the Telegram messenger service.

#### 8.3.1 *Empirical Data on Russian Activism*

Table 8.1 summarizes Russians' participation in various forms of activism based on the results of the eighth round of the European Social Survey in 2016—the first year when a question explicitly measuring online participation ("have you posted or shared anything about politics online") was added to the survey.

According to the table, the Russians were lagging behind in most of the traditional forms of activism compared to Germany, France, the United Kingdom (UK), and Finland, with the exception of working in a political party or action group, where the Russians were as passive as the citizens of the four European countries. In addition, the Russians were only slightly less keen to wear a campaign badge or sticker than the Germans. They also took part in lawful public demonstrations less frequently than the French and Germans, as often as the British but more frequently than the Finns.

**Table 8.1** European Social Survey questions on various forms of activism (European Social Survey Round 8 Data 2016)

	<i>Germany</i>	<i>Finland</i>	<i>France</i>	<i>UK</i>	<i>Russia</i>
Voted in the last national election	72.1	75.3	54.9	70.2	46.8
Contacted politician or government official last 12 months	15.7	18.8	12.9	17.3	4.9
Worked in a political party or action group last 12 months	4.1	3.3	2.8	3.1	2.9
Worked in another organisation or association last 12 months	29.0	38.1	13.7	7.4	3.8
Worn or displayed campaign badge/sticker last 12 months	5.2	19.8	10.1	9.7	4.3
Signed a petition last 12 months	35.4	35.6	30.5	44.1	7.4
Took part in a lawful public demonstration last 12 months	10.3	3.8	14.9	5.3	5.2
Boycotted certain products last 12 months	33.3	36.9	30.8	21.1	2.3
Posted or shared anything about politics online last 12 months	21.5	20.9	20.7	29.9	4.7

Most interestingly from the viewpoint of this chapter, only 4.7 per cent of the Russians—three to four times fewer than the Germans, French, Finns, and the British—had posted or shared anything about politics online during the 12 months preceding the survey.

However, the mean percentages presented in Table 8.1 hide the polarization of Internet use: heavy Internet users are typically young urban Russians, while Internet use is less prevalent in the rural areas and among the elderly. Moreover, the European Social Survey (ESS) questions do not cover the wide variety of non-political forms of civic activism. According to Sobolev and Zakharov (Sobolev and Zakharov 2018), for example, increasing numbers of Russians have been participating in recent years in charity, volunteering, and also in actions to improve their immediate surroundings.

### 8.3.2 *Communicative Online Activism: Alexei Navalny and the Anti-Corruption Foundation*

Alexei Navalny is a Russian lawyer, anti-corruption fighter, and political activist born in 1976, who rose to fame on the Russian political scene during the opposition mass protests in 2011. In 2019 he remains the only credible challenge to Vladimir Putin from outside the political establishment and the only opposition leader who can mobilize nation-wide demonstrations in major Russian cities.

Navalny's online activism is conducted and coordinated by his professional social media team at the Anti-Corruption Foundation on several platforms such as his blog (<https://navalny.com/>) Facebook, VKontakte, Twitter, Odnoklassniki, Instagram, Telegram, and YouTube (for more, see Chap. 16). In his campaigning, Navalny has utilized several variants of online activism ranging from data activism and crowdsourcing (the anti-corruption project *RosPil*, <https://fbk.info/projects/>), witnessing via YouTube videos, to hashtag activism (#Navalny 2018), social media doxxing, and educating users on information security issues (NavalnyLIVE/cloud YouTube channels).

According to Dollbaum et al. (2018), Navalny's campaign for the 2018 presidential elections, from which he was banned, combined a strictly hierarchical coordination of action by the Anti-Corruption Foundation and its regional offices with the work of a large network of volunteers all over the country. The core of the campaign consisted of a broad anti-corruption stance, which allowed various political actors with a common interest in opposing the ruling regime to participate. In terms of "organizationally enabled connective action" (Bennett and Segerberg 2012), the campaign offered a low threshold for participating:

*It required little prior knowledge, and participation was framed as fun, hip, and sociable. Each of the 80 regional offices recruited several dozens of active volunteers, most in their teens and early twenties, who distributed flyers, gathered signatures,*

*and registered supporters. Furthermore, the offices evolved into hubs for civic activity, connecting to other oppositional activists on the ground, hosting lectures, film screenings, and discussions. Besides nurturing a collective identity and strengthening social ties, this activity was explicitly aimed at involving young people in political discourse, combating apathy and depoliticization.* (Dollbaum et al. 2018, 5)

One indication of Navalny's success in reaching out to young Russians is the new law signed by Putin on December 28, 2018, which clearly connected to the fact that the street protests of 2018 saw the participation of many teenagers: The law punishes the organizers of unsanctioned public gatherings with participants under 18 years of age with 15 days' imprisonment or fines (Radio Free Europe/Radio Liberty 2018).

However, although Navalny's campaign utilized a wide variety of social media platforms and its broad anti-corruption message gave supporters much leeway for personalized connective action (e.g., in the form of constructing and sharing Internet memes), its hierarchical organization led to an inbuilt tension in the campaign. At the heart of this tension was the clash between the logics of goal-oriented political action and a movement of volunteers and activist recruited through street protests. (Dollbaum et al. 2018, 6).

A unique feature of Navalny's online presence is a series of exchanges of YouTube videos with the Russian political elite. The Russian oligarch Alisher Usmanov as well as the head of the Russian National Guard, Viktor Zolotov, have responded to Navalny's provocative YouTube videos exposing their alleged corruption by publishing their own YouTube video replies—to which Navalny has retaliated with further videos. This exchange of public videos stands in stark contrast to Putin's and Medvedev's total ignoring of Navalny in their public appearances.<sup>10</sup>

Navalny's 2019 campaign "*umnoe golosovanie*" ("smart voting", <https://vote2019.appspot.com/>) targeted the 2019 Moscow city council elections and some regional elections, which happened at the same date, September 8, 2019. In the related instructional YouTube video, he urged people to vote for the candidate of the party—with the exception of the ruling party United Russia—which polled the most votes during the last election in their voting district. Exact candidates to vote for were suggested by Navalny's team, which followed the results of their own polls. The suggestions were sent to voters by email, made available via Telegram bot, and at the campaign website.<sup>11</sup>

In all, the particularity of communicative online activism in Russia consists of a cat-and-mouse game between activists and the Kremlin. In this game, the Kremlin has succeeded in recreating an atmosphere of fear where all anti-government expression online in Russia has become risky.

Alexei Navalny is one of the few who, thanks to his popularity, can afford to run this risk, and continues to speak directly to the people through social media, thereby circumventing his ban on state-controlled media. Navalny's political campaigning strategy seems to more or less consciously implement a strategy dubbed "the cute cat theory" of online activism by Ethan Zuckerman

(2007). According to Zuckerman, under authoritarian conditions opposition activists should rely on popular platforms (on which non-political pictures of cute cats are posted). Due to the popularity of these platforms, their shutting down by the government is risky since it may annoy a large part of population—also those previously not interested or involved in politics.

### 8.3.3 *Technoactivism: The Example of Telegram*

In addition to Navalny's campaigns, the battle waged by the Telegram messenger service against the Russian state has been among the most noteworthy events of Russian online activism in recent years. In this conflict the Russian state tried to block the messenger service, whose global image and marketing campaigns focus on encryption and privacy. In particular, Telegram assures its users that, unlike other messengers, it is able to protect the users' chats from strangers' eyes and denies any cooperation with secret services. In line with this, the company refused to collaborate with the Russian security service. It therefore allowed activists to continue publishing and distributing their anti-governmental views anonymously.

The case of Telegram constitutes the most significant example of a successful struggle against Internet control by the increasingly authoritarian Russian state. Telegram used its knowledge and understanding of Internet protocols, as well as mechanisms for updating smartphones from mobile application stores to circumvent blocks. Telegram combined this with a major crowd-sourcing initiative to fight for the free exchange of information protected against state monitoring.<sup>12</sup>

This section sheds light on the legal, technological, and societal aspects of the struggle which also had an offline form of mobilization: On April 30, 2018, thousands of protesters marched in Moscow and threw paper planes—the symbol of Telegram—to protest against the state's decision to block the service. Because of its visual nature, the action succeeded in gaining media attention and in showing support for Telegram. However, unlike the technological online resistance described later in this article, this offline public support action had no sequels and was ignored by the Kremlin.

#### 8.3.3.1 *Telegram's Legal Battle Against the Russian Security Service*

Although state pressure on free expression and on Telegram's founder, Pavel Durov, have a longer history, the actual start of the conflict between the company and Russian state can be traced back to July 2017. In July, the Russian federal security service FSB (*Federal'náá služba bezopasnosti*) required Telegram to create a way for the FSB to intercept communications on Telegram. To be more precise, the FSB asked Telegram to hand over the encryption keys, that is, digital passwords, without which it is impossible to read communication content. The security service justified its requirement by the need to decrypt terrorist messages sent via Telegram in connection with the terror attack on a St. Petersburg metro train on April 3, 2017. Telegram responded



by stating that the company did not have the keys because the application keeps them only on users' devices. In addition, the founder and Chief Executive Officer (CEO), Pavel Durov, noted that the FSB's request was contradictory to the protection of privacy of communication guaranteed by the Constitution of Russia.

In October the FSB filed a formal complaint with the court, which fined Telegram for non-compliance with the FSB's request (Bryzgalova 2017). The FSB defended its position claiming that providing the FSB with a technical capability to decode messages still required the FSB to seek a court order to read correspondence from specific individuals (Pis'mennye vozraženíá FSB 2017). On March 20, 2018, Russia's Supreme Court rejected Telegram's appeal, after which the Russian Internet watchdog Roskomnadzor announced that the messaging service had 15 days to provide the required information to the security agencies—otherwise access to Telegram in Russia would be blocked.

### 8.3.3.2 *Technological Resistance by Telegram*

On April 13, 2018, the Taganskij court in Moscow ruled that access to Telegram in Russia should be blocked due to the failure of Telegram to provide the FSB with the encryption keys. In technical terms, the FSB required Telegram to rewrite their messaging application from scratch to enable the FSB to read all messages sent via Telegram. The requirement was based on the federal law “on information, information technologies, and the protection of information.” Refusing to comply, Telegram deemed the law and its implementation unconstitutional.

How did Telegram resist the state's attempts to block the use of the service?

When Roskomnadzor told the Internet service providers (ISPs) the addresses of the Telegram servers, the ISPs disabled the connections to these servers. As a response, Telegram assigned them different addresses, making it challenging to discover the new addresses and to communicate their location to the ISPs fast enough. (ISPs in Russia are obliged to download a register of addresses to block daily, and Telegram can change addresses several times per hour).

However, Telegram cannot assign random addresses to its servers because they must be in a range owned by the company at which Telegram keeps the servers, such as Google or Amazon. Thus, Roskomnadzor's attempts to block large ranges of addresses belonging to these companies led to a temporary block not only of Telegram but also of many other websites. Google and Amazon, for example, provide hosting for many companies worldwide, including companies operating in Russia. Internet services not related to Telegram were merely affected because they had servers in the same range of addresses as Telegram.

As a wealthy company Telegram could afford to rent many large ranges of addresses from giant hosting providers. Blocking all of them would have meant collateral damage to Internet services, which are essential for many people in

Russia. Thus, Roskomnadzor was able to block only some of them, and Telegram used the remaining part.

In addition to the actions described above, Telegram took several steps to avoid blocking imposed by Roskomnadzor.

First, the company encouraged users worldwide to run so-called proxy servers, that is, intermediary services with ample capacity to forward Telegram traffic to actual Telegram servers. Pavel Durov, the CEO of the company, even announced a grant program promising financial support to individuals who develop and run proxies for Telegram users on their own or rented servers.

Second, Telegram encouraged people to use virtual private networks (VPN). VPN allows establishing an encrypted connection from a laptop or smartphone to a location outside their country. A VPN server there serves as an intermediary allowing connection to Telegram from that location.

Third, Telegram uses so-called push updates (similar to message notifications in messengers such as WhatsApp) to notify the Telegram application of any server address changes. If Roskomnadzor had blocked the push notifications, it effectively would have blocked all notifications from Apple and Google servers to all applications on all Android and iOS smartphones in Russia. It would have disrupted many services, including popular online banking applications, which Roskomnadzor did not dare.

In sum, Telegram's technoactivism is a form of activism intended to resist attacks on civil rights, such as freedom of speech and freedom of communication. Technoactivism often requires extensive technical expertise and money to build a technical solution and a relatively large community ready to support, popularize, crowdfund, and help technically with its implementation. Its success depends on technical abilities, expertise, and the limitations of its opponents.

### 8.3.4 *Non-contentious Forms of Online Activism*

Our definition of online activism includes forms of action which are not contentious or political in nature. They do not directly challenge state power but are rather targeted at resolving social, cultural, or local problems. Such activities are relatively common in Russia; they address a wide variety of issues and usually do not require an organization to coordinate operations. These activities may, however, become politicized and transformed into protests when, for example, the discussions approach the fields of healthcare, education reforms, taxation, or parental interests; when residents start opposing the planning of new garbage dumps nearby, or when apartment owners begin to mobilize against the replacement of a neighborhood park with an apartment block. Nevertheless, some topics such as lesbian, gay, bisexual, and transgender (LGBT) rights, gender identities or sexual and domestic violence have already been politicized in official discourse in Russia regardless of the initial nature of the public debate or intentions regarding contentious mobilization.

The range of non-political issues and social problems addressed by online activists covers a wide variety of everyday problems from animal rights to parental movements, car owner rights, and so on. Below we will illustrate some of the most noteworthy examples of non-political online activism related, first, to environmental and housing issues and, second, to women's and LGBT rights.

The issues related to *environmental topics and problems related to real estate ownership rights* (e.g., five-storey building renovations in Moscow) were not originally politicized in public discourse. Activism around these topics usually begins as an attempt to solve local problems and becomes politicized in the course of events (cf. Erpyleva 2019). Numerous small local environmental initiatives in the middle of the 2010s, mainly aimed at cleaning green zones in urban areas, shared the ideology of “small steps,” which implied the idea of making life better by improving the immediate surroundings. One of the first big ecological movements was the defense of the Khimki forest (Moscow region) 2007–2011. It became politicized relatively quickly, but involved negotiations with the authorities, communication with them, and even their sporadic support for the movement. The garbage protests (2018–2019) in Moscow region and Shies (Arkhangelsk region) had clear anti-government significance right from the outset, and with this agenda and the use of social media (Facebook and VK) and thematic sites (Шлес.рф, Bellona.rf) they easily reached a nationwide audience.

Examples of the movement defending real estate ownership rights include joint action by apartment owners of the same block of flats, who create groups on the social networking site VKontakte to solve various housing management problems, such as maintenance and repair of the building's infrastructure (water pipes, heating, elevators, etc.) or construction of a playground in the yard. This type of activism has been common in campaigns organized by local residents against urban construction projects and for the protection of parks and green urban zones in Russian cities (see Gladarev and Lonkila 2012 and 2013 for an example in St. Petersburg). In Moscow, protests against the plan initiated by the city government to demolish and rebuild whole neighborhoods of Soviet-era tenements were coordinated through thematic Internet sites (for example, <http://renovation.tbcc.ru>) and Facebook groups in 2017–2018 (Rosenblat 2018).

The disputes concerning *women's and LGBT rights* present, by contrast, an example of online activism on a topic that has already become highly politicized as part of conservative and nationalist political rhetoric, also at the state level. Domestic violence and LGBT rights have been discussed not only by liberal activists, but also by conservatives, who reported websites to the Russian Internet watchdog Roskomnadzor for allegedly containing prohibited “gay propaganda.” In particular, the group *Deti-404. LGBT-podrostki* (Children 404. LGBT teens) on the popular Russian social network site VKontakte was blocked by a court order in 2015 after being found guilty of propagating “non-traditional sexual relationships.” Elena Klimova, the founder of the group and a project

bearing the same name, was sentenced to pay fines and she and other participants of the project became targets of online hate speech (Children-404 n.d.).

Another case of activism in defense of women's and LGBT rights was the *#yaNeBoyusSkazat* (I'm not afraid to speak) movement—the Russian equivalent of #metoo—in 2017, which was a hot topic among Russian users of Facebook. Victims shared their accounts of sexual harassment in an attempt to create visibility for the sexual and domestic violence agenda (Zhigulina 2016; *Dviženie #MeToo god spustâ* 2018). These actions were repeatedly commented on by high-ranking state officials and Duma (the lower house of the Federal Assembly of Russia) deputies, who denied the relevance of the issue, referring to traditional Russian family values, such as patriarchal family relations.

The examples presented above of online activism demonstrate its significance in protecting human rights, solving everyday problems, and making the authorities aware of them. They also highlight the thin and easily permeable line between non-political and political activism in Russia. (cf. Erpyleva 2019)

## 8.4 CONCLUSIONS

In this chapter we have illustrated through selected cases the ways digitalization has affected activism in Russia. The two cases of contentious activism presented above describe variants of “organizationally enabled” connective action, where central coordination is combined with grass-root activism in digital media. In the case of the communicative activism of Alexei Navalny, the coordination was implemented by his team at the Anti-Corruption Foundation. Although Navalny's team also engages in data activism—for example, when investigating the property of Russian politicians abroad—the ultimate aim of its digital activism is to gain support and raise awareness in order to exert pressure on the government and ultimately to gain political power.

Telegram and Pavel Durov lack similar political ambitions. The technoactivism of Telegram showed that with sufficient technical expertise and financial resources it is possible to develop relatively sophisticated and distributed protection against the blocking of web resources by the state. Before the battle between Telegram and the Kremlin, all efforts of the Russian state to block Internet content had been successful: the torrent tracker [rutracker.org](http://rutracker.org), for example, was blocked due to multiple copyright violations, and the service remains inaccessible from Russia unless its user connects to it via VPN. The success of Telegram showed technoactivists that digital technology can be used not only for state monitoring and control, but also to protect freedom of expression and users' right to private communications.

Both of these two cases have been rare examples of visible and contentious online activism enabled by digital technology in Russia. In both cases hierarchical coordination was combined with grass-root actions by citizens who could develop their own ways of participating under fairly general slogans against corruption (Navalny) or for freedom of expression (Telegram). Both

campaigns have also managed to recruit young Russians into contentious online activism.

In addition, our examples of the non-contentious forms of online activism illustrate the flexible and contested line between non-political and political forms of activism. Some topics, such as those related to sexuality, marriage, and religion have already become politicized in official discourse and through legislation while other, at first sight non-political problems, such as those related to parenting or housing, may become politicized when people start to view them as examples of bad governance.

In a country as large as Russia, nationwide contentious action is not realistic without the Internet and modern digital technology. The acid test for online activism is, however, how to influence the societal and political affairs *offline*. Jennifer Earl (2016) suggests that online activism has added to the traditional repertoire of social movements an alternative, “flash-based” power—rapid, temporally limited, and massive, but not necessarily continuous mobilization—which may also die out quickly. According to Earl, online mobilization may draw a greater number of people to flash activism, which reduces the cost of participating in otherwise high-risk offline demonstrations. This kind of flash-based power was manifested at the beginning of the Russian opposition mass protests in 2011 and it has been shown to be able to overthrow governments, for example, during the Arab Spring—even though many of the uprisings were subsequently repressed.

In the traditional model, the power of protest emanates from continuous mobilization and pressure exerted upon the state. This requires transforming grievances into stable political programs, institutions, and structures and thus a transition from connective activism to more traditional forms of collective action. Such a transformation was attempted in Russia, for example, during the protest wave in 2012, when over 80,000 people participated in the online elections of the opposition coordinating council. However, both as a result of internal tensions within the council between the nationalists, leftists, and liberals and the tightening repression by the state, the resistance faded at the end of the one-year term and the council was dissolved (Toepfl 2018). Another and partly successful attempt to transform online actions into offline political capital and structures was Navalny’s initiative of “smart voting”, which very likely contributed to the poor performance of United Russia in the Moscow city council elections on September 8, 2019.

In 2019, with the Russian state continuously introducing new constraints on freedom of expression, online participation in Russia has become risky (Lonkila et al. 2020). As a consequence, many activists have ceased to participate in online discussions, many have moved to social media platforms based outside Russia, such as Twitter or Facebook, and others have opted for emigration. Still others have directed their energy and attention towards the non-political problems of everyday life.

However, Russians’ struggles to solve local daily life problems are often the results of policy failures and the online connections made through social media

between similar local struggles elsewhere may result in the generalization and politicization of individual and local grievances (cf. Gladarev and Lonkila 2012, 1386–7; Erpyleva 2019). Digital technology offers both new means to mobilize people and share these grievances, as well as new tools to monitor and repress them. The outcome of this tension between emancipatory and repressive aspects of digitalization is uncertain and merits further research.

## NOTES

1. This expression refers to the statement in the fall 2011 announcing that Prime Minister Putin would run for president in 2012 and, if successful, would appoint the then president Medvedev to prime minister.
2. Cf. Theocharis' original formulation: "digitally networked participation can be understood as a networked media-based personalized action that is carried out by individual citizens with the intent to display their own mobilization and activate their social networks in order to raise awareness about, or exert social and political pressures for the solution of, a social or political problem" (Theocharis 2015, 6; see also Van Deth 2014; Ohme et al. 2018).
3. There are individuals and groups of citizens involved in "online vigilantism" with diverse ideological convictions and ties with state organs in Russia. An account of such organized groups as the *Molodežnaâ služba bezopasnosti* (Youth Security Service) sponsoring an emergent "cyber Cossack movement" and the *Liga bezopasnogo Interneta* (Safe Internet League) can be found in Daucé et al. (2019). The authors also discuss the hearings at the Russian Civic Chamber on a bill on "*kiberdružiny*" (cyber patrols). They find a tension between politically involved organizations and дума members supporting the bill and the experts criticizing the bill for its inefficiency.
4. A non-exhaustive list of terms in literature trying to cover the phenomenon of online activism includes digital activism, cyberactivism, Internet activism, web activism, digital campaigning, online organizing, electronic advocacy, e-campaigning, social media activism, and e-activism.
5. On the debate on slacktivism and "liking" in social media see Earl 2016, 374–5; Theocharis 2015, 8–9.
6. Our focus does not imply that we consider non-political forms of activism less important in Russia. First, many forms of social and cultural activism are indispensable as such—e.g., in taking care of social or health care services not provided by the state. In addition, the non-political forms may function as substitutes for political action; as ways to create alternative cultural framings which mirror, ridicule, and contest dominant cultural codes (Flikke 2017); and as platforms to form horizontal ties in civil society (Gladarev and Lonkila 2013) which may later on serve as precondition for explicitly political resistance. Finally, as explained later in this chapter, the boundary between political and non-political activism is fluid and contested: from the viewpoint of the Kremlin any independent action organized by civil society may potentially threaten the current status quo and turn into contentious action.
7. In September 9, 2020, the video had 36.2 million views, see [https://www.youtube.com/watch?v=qrwlk7\\_GF9g](https://www.youtube.com/watch?v=qrwlk7_GF9g).

8. <https://autoassa.ru/novosti/mchs-zakupaet-slishkom-roskoshnye-avtomobili/>.
9. “We are the 99%” was the slogan of the Occupy social movement, referring to the income inequality in the United States. “The movement for fair elections” refers to the mobilization of citizens against the rigging of Duma elections in the fall 2011.
10. The strange and unique exchange between Navalny and one of the Russia’s richest oligarchs Alisher Usmanov started from a video *On vam ne Dimon* ([https://www.youtube.com/watch?v=qrwlk7\\_GF9g](https://www.youtube.com/watch?v=qrwlk7_GF9g)) published in March 2017. In this video, which by September 9, 2019, had 31.8 million views, implied that Usmanov had bribed Dmitry Medvedev—something that Usmanov denied. In response to this denial Navalny published a follow-up video with almost 5 million views (<https://www.youtube.com/watch?v=xn0Ah0J5p5Y>). Usmanov, in a move unheard of a Russian billionaire, replied to Navalny in his own YouTube video, which, however, got only 12,735 views (<https://www.youtube.com/watch?v=XfWB1cKtFws>), whereas Navalny’s further reply to Usmanov’s reply had collected 3.8 million views (<https://www.youtube.com/watch?v=YwlrKfLeRfs>).
11. Though it is difficult to measure accurately the results of smart voting, it seem to have worked in the *Mosgorduma* (Moscow City Duma) elections in 2019: United Russia lost 13 seats ending up with 25 seats in the 45-seat council, whereas the Communist Party was the greatest beneficiary with 13 seats—8 seats up from previous elections (cf. Pertsev 2019).
12. For a critical look on the political history of Telegram see Maréchal (2018); for a comparison between political uses of Telegram in Russia and Iran see Akbari and Gabdulhakov (2019).

## REFERENCES

- Akbari, Azadeh, and Rashid Gabdulhakov. 2019. Platform Surveillance and Resistance in Iran and Russia: The Case of Telegram. *Surveillance & Society* 17 (1–2): 223–231.
- Bennett, W. Lance, and Alexandra Segerberg. 2012. The Logic of Connective Action. Digital Media and the Personalization of Contentious Politics. *Information, Communication & Society* 15 (5): 739–768.
- Bryzgalova, Ekaterina. 2017. Sud oštrafoval Telegram za otkaz sotrudničat’ s FSB. *Vedomosti*, October 16: 2017.
- Children-404. n.d. Wikipedia. Accessed 27 April 2018. <https://en.wikipedia.org/wiki/Children-404>.
- Daucé, Françoise, Benjamin Loveluck, Bella Ostromoukhova, and Anna Zaytseva. 2019. From Citizen Investigators to Cyber Patrols: Volunteer Internet Regulation in Russia. *Laboratorium: Russian Review of Social Research* 11 (3): 46–70.
- Dollbaum, Jan Matti, Andrey Semenov, and Elena Sirotkina. 2018. A Top-down Movement with Grass-Roots Effects? Alexei Navalny’s Electoral Campaign. *Social Movement Studies* 17 (5): 618–625. <https://doi.org/10.1080/14742837.2018.1483228>.
- Dviženie #MeToo god spustâ [Movement #MeToo a Year Later]. 2018. BBC News, Russkaâ služba [BBC News, Russian service], August 23. <https://www.bbc.com/russian/features-45247219>.



- Earl, Jennifer. 2016. Protest Online: Theorizing the Consequences of Online Engagement. In *The Consequences of Social Movements*, ed. Lorenzo Bozi, Marco Giugni, and Katrine Uba, 363–400. New York: Cambridge University Press.
- Erypyleva, Svetlana. 2019. *The New Local Activism in Russia: Biography, Event, and Culture*. Doctoral dissertation, Faculty of Social Sciences, University of Helsinki.
- Etling, Bruce, Karina Alexanyan, John Kelly, Robert Faris, John G. Palfrey, and Urs Gasser. 2010. *Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization*. SSRN Scholarly Paper ID 1698344. Rochester, NY: Social Science Research Network.
- European Social Survey Round 8 Data. 2016. Data File Edition 2.1. NSD - Norwegian Centre for Research Data, Norway – Data Archive and Distributor of ESS Data for ESS ERIC. <https://doi.org/10.21338/NSD-ESS8-2016>.
- Flikke, Geir. 2017. ‘Monstrations for Mocracy’: Framing Absurdity and Irony in Russia’s Youth Mobilization. *Demokratizatsiya: The Journal of Post-Soviet Democratization* 25 (3): 305–334.
- Gel’man, Vladimir. 2017. Why Not Authoritarian Modernization in Russia. In *Authoritarian Modernization in Russia. Ideas, Institutions and Policies*, ed. Vladimir Gel’man, 1–21. New York: Routledge.
- Gibson, Rachel, and Marta Cantijoch. 2013. Conceptualizing and Measuring Participation in the Age of the Internet: Is Online Political Engagement Really Different to Offline? *The Journal of Politics* 75 (3): 701–716. <https://doi.org/10.1017/s0022381613000431>.
- Gladarev, Boris, and Markku Lonkila. 2012. The Role of Social Networking Sites in Civic Activism in Russia and Finland. *Europe-Asia Studies* 64 (8): 1375–1394. <https://doi.org/10.1080/09668136.2012.712272>.
- . 2013. Justifying Civic Activism in Russia and Finland. *Journal of Civil Society* 9 (4): 375–390. <https://doi.org/10.1080/17448689.2013.844450>.
- Kaun, Anne, and Julie Uldam. 2018. Digital Activism: After the Hype. *New Media & Society* 20 (6): 2099–2106. <https://doi.org/10.1177/1461444817731924>.
- Kosinski, Michal, David Stillwell, and Thore Graepel. 2013. Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences* 110 (15): 5802–5805. <https://doi.org/10.1073/pnas.1218772110>.
- Lonkila, Markku, Larisa Shpakovskaya, and Philip Torchinsky. 2020. The Occupation of Rунet? The Tightening State Regulation of the Russian-Language Section of the Internet. In *Freedom of Expression in Russia’s New Mediasphere*, ed. Marielle Wijermars and Katja Lehtisaari, 17–38. Routledge.
- Maréchal, Natalie. 2018. *From Russia With Crypto: A Political History of Telegram*. 8th USENIX Workshop on Free and Open Communications on the Internet (FOCI 18). Accessed 30 March 2020. <https://www.usenix.org/conference/foci18/presentation/marechal>.
- Ohme, Jakob, Claes H. de Vreese, and Erik Albaek. 2018. From Theory to Practice: How to Apply van Deth’s Conceptual Map in Empirical Political Participation Research. *Acta Politica* 53 (3): 367–390. <https://doi.org/10.1057/s41269-017-0056-y>.
- Pertsev A. 2019. Strategiä ‘umnogo gosovaniä’ okazalas’ pobednoj ili ne očen’? *Meduza*, September 9. Accessed 6 April 2020. <https://meduza.io/feature/2019/09/09/strategiya-umnogo-golosovaniya-okazalas-pobednoj-ili-ne-ochen>.

- Pis'mennye vozraženîâ FSB [FSB written objections]. 2017. October 16. [http://agora.legal/fs/a\\_delo2doc/63\\_file\\_Telegram\\_FSB.pdf](http://agora.legal/fs/a_delo2doc/63_file_Telegram_FSB.pdf).
- Radio Free Europe/Radio Liberty. 2018. Putin Endorses Law Establishing Jail Time For Luring Minors Into Protests (blog). December 28. <https://www.rferl.org/a/putin-endorses-law-establishing-jail-time-for-luring-minors-into-protests/29681317.html>.
- Rosenblat, Steve. 2018. Katkom po Moskve: polnaâ istoriâ stolichnoj renovacii v lichâ i faktah [On a Roller across Moscow: Full History of the Capital's Renovation in Persons and Facts]. *Novye Izvestiâ*, December 30. <https://newizv.ru/article/general/30-12-2018/katkom-po-moskve-polnaya-istoriya-stolichnoy-renovatsii-v-litsah-i-faktah>.
- Sobolev, Anton, and Alexey Zakharov. 2018. Civic and Political Activism in Russia. In *The New Autocracy: Information, Politics, and Policy in Putin's Russia*, ed. Daniel Treisman, 247–276. Washington, DC: Brookings Institution Press.
- Sokolov, Alexander. 2015. Russian Political Crowdfunding. *Demokratizatsiya: The Journal of Post-Soviet Democratization* 23 (2): 117–149.
- Theocharis, Yannis. 2015. The Conceptualization of Digitally Networked Participation. *Social Media + Society* 1 (2). <https://doi.org/10.1177/2056305115610140>.
- Toepfl, Florian. 2018. From Connective to Collective Action: Internet Elections as a Digital Tool to Centralize and Formalize Protest in Russia. *Information, Communication & Society* 21 (4): 531–547. <https://doi.org/10.1080/01369118X.2017.1290127>.
- Van Deth, Jan W. 2014. A Conceptual Map of Political Participation. *Acta Politica; London* 49 (3): 349–367. <http://dx.doi.org.ezproxy.jyu.fi/10.1057/ap.2014.6>.
- Vegh, Sandor. 2013. Classifying Forms of Online Activism: The Case of Cyberprotests against the World Bank. *Cyberactivism*. August 21. <https://doi.org/10.4324/9780203954317-9>.
- Zhigulina, Olga. 2016. Fleshmob #ÂNeBoûs'Skazat' [#Flashmob I'mNotAfraidtoSay]. *Tjournal*, June 7. <https://tjournal.ru/flood/30989-fleshmob-yaneboyusskazat>.
- Zuckerman, Ethan. 2007. The Connection between Cute Cats and Web Censorship. July 16. <http://www.ethanzuckerman.com/blog/2007/07/16/the-connection-between-cute-cats-and-web-censorship/>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Digital Journalism: Toward a Theory of Journalistic Practice in the Twenty-First Century

*Vlad Strukov*

## 9.1 INTRODUCTION

The digital turn has made a profound impact on journalism, ranging from the ways in which journalists collect and display information to how journalistic items are perceived by the publics in regional, national and transnational contexts. Among other things, the proliferation of digital technologies has allowed for a number of transformations, including new genres of journalistic output (for example, reports organized and presented as questionnaires), new forms of collaboration among journalists (for example, files sharing and remote uploads of content which makes communication and reporting instantaneous) and new methods of carrying out journalistic investigation (for example, the use of databases and information available on digital networks in the public domain). Moreover, new models for journalistic entrepreneurship emerged (for example, setting up media outlets in “non-geographic” areas such as offshore areas and tax-free zones and outsourcing content production to individuals in other countries). At the same time, new regimes of exploitation imposed by owners of the media outlets and resistance by journalists became apparent (for example, zero-hour contracts and situations when journalists are exposed online making them objects of public shaming and threats).

In addition to the changes in terms of how journalists work, there have been changes in terms of journalistic agency, institutions and drivers of innovation.

---

V. Strukov (✉)  
University of Leeds, Leeds, UK  
e-mail: [v.strukov@leeds.ac.uk](mailto:v.strukov@leeds.ac.uk)

For example, the increased speed with which reports are released is a hallmark of digital journalism, and it has led to an even greater competition among different media outlets, each aiming to be the first to report an event. Contrary to this trend, some media outlets have chosen to focus on “slow news,” that is, analytical reports which are aimed at reflective consumption by the users.<sup>1</sup>

Innovative organization of the news flow and innovative use of new technologies have helped re-define the relationship between content producers and content consumers. For example, on one level (micro-)blogging is just a new form of journalistic output. On another, it refers to a new relationship among producers and users of news items. As a result, the traditional notion of “audiences” has been re-considered to include networked, de-centralized and geographically unbound agency. These audiences are not simply more “active,” rather they are more dynamic and diverse in terms of how they relate to news items and reports.

Similarly, as a result of digitalization, there are entirely new players on the field such as media institutions and tech companies. The former include organizations that focus on other sectors but utilize sophisticated tools that affect other media. This is evident in the proliferation of Russian media interests in other countries.<sup>2</sup>

In terms of technical companies, Microsoft and Google have been influential in the Russian Federation (the RF), especially after the introduction of localized versions of their software. Their Russian competitors, Mail.ru and Yandex, have been backers of journalistic innovation such as live streaming. For example, Yandex, which builds products and services powered by machine learning, has a video stream for live and on-demand video on the company’s streaming content platform, Yandex.Live. By circuiting live-streaming in digital realms controlled by Yandex, the company has increased demand for new content, including journalistic outputs and entertainment pieces (for more on social media, see Chap. 19).

Not all Russian services have been built as “alternatives” to western technologies, that is, Yandex versus Google. There are many examples of transnational convergences and collaborations, too. In terms of live streaming and video content sharing, Rutube, which belongs to Gazprom, is a competitor to YouTube; however, in terms of built-in videos, its strategic partner is Facebook. At the same time, Rutube is used by Russia Today (RT), the government-backed television and online platform, which has been accused of disinformation and propaganda. RT uses Rutube as one of its main channels of content dissemination. The analysis of these digital ventures—in this case Facebook-Rutube-RT—reveals a somewhat unexpected mix of national and transnational corporate and government interests.<sup>3</sup>

Mail.ru has benefited from the mutability of digital media, for example, when they make use of convergent flows of news reporting and banking. This is when the news agenda is organized in ways that advance public interest in financial instruments, and vice versa. This reveals not only a convergence across

platforms but also across perceptions of media genres and information per se, thus pushing boundaries between different kinds of journalism.<sup>4</sup>

To account for all the changes in journalism that had occurred thanks to the proliferation of digital technologies would be an impossible task. Hence, in this chapter, I reflect on the processes of digitalization of journalism, on the one hand, and, on the other, on digital forms of investigative journalism. The latter means journalism which is native to digital realms and which utilizes digital-only means to conduct research and publish reports. So, my account supplies not a survey of technical innovations and cultural forms, but a conceptualization of transition from legacy to digital journalism in the RF and Russophone world. To confirm, I pay special attention to how in journalistic practice, the use of digital technologies had emerged from being an auxiliary tool to being the main—and only—method of producing, delivering and consuming news. My approach allows the following definition of “digital journalism.” The term designates the transition from one technological base to another and the transformations in the profession and practice of journalism which had occurred during the process. The term does not designate the broad field of contemporary journalism which is extremely diverse in terms of technologies, forms, “audiences” and other factors.

Western scholarship has focused on the economic and technological implications of the shift (see, for example, Jones and Salter 2011), often citing challenges in terms of identity politics, power structures and professional networks (see, for example, Anderson 2013; Bradshaw and Rohumaa 2013). A critique of Western neoliberal order from the perspective of the changing dimensions of journalistic profession is available in a number of publications, too (see, for example, Franklin 2017). Most recent debates have been about the automation of news (Diakopoulos 2019) in the context of populist political campaigns in the USA (Bucher 2018; Wahl-Jorgensen 2019). In their most recent publication, Bob Franklin and Lily Canter (2019) offered a classification of possible fields of application of digital technologies in journalism, thus broadening the notion of journalism per se. This corpus of literature complements numerous critical anthologies assessing skillsets of journalists in the digital era (for example, Hill and Lashmar 2013; Zion and Craig 2014). These publications reveal the complexity of digital journalism as a phenomenon; they also signpost the developments exclusively in the Western context. Hence, my discussion contributes to the existing debate by deliberately internationalizing the phenomenon of digital journalism and offering alternative modes of conceptualization. These modes stem from the analysis of the context, producing an original paradigm (Sects. 2 and 3). Moreover, the emphasis is on the transnational characteristics of Russian digital journalism, thus avoiding the redundancy of the “West-versus-the rest” approach (Sect. 4). Finally, the proposed typology (Sect. 5) helps categorize digital journalism and also social, political and cultural phenomena in the Russian context, thus offering a more universal model for consideration.

The chosen understanding of digital journalism has informed the selection of the cases and the organization of the discussion. To confirm, the first subsection provides a theorization of Russian digital journalism from the perspective of its evolution and types of activity. In subsequent subsections I analyze cases that shed light on pivotal moments in the development of Russian digital journalism. In the conclusion I summate the discussion, arguing that the digital turn has provided Russian journalists with new opportunities such as setting a transnational media company and building and engaging with translocal communities in the RF and abroad, as well as new challenges such as increased surveillance by the state and security services and new regimes of exploitation such as unregulated job markets.

The discussion is based on my research of Russian digital media and journalism<sup>5</sup> and on interviews with journalists and editors which I collected during a major study of contemporary Russian media in 2014–2018.<sup>6</sup> The discussion is additionally informed by my survey of literature on new media, digital media and contemporary journalism available in specialized publications.<sup>7</sup> I am grateful to all the journalists, editors and media practitioners who had agreed to talk to me about their transition to digital journalism.

## 9.2 “ALTERNATIVE” JOURNALISM

Initial studies of digital journalism (e.g., Thorsen and Jackson 2017) focused on the ways in which journalistic materials were produced and presented to the public. Journalists had to make a choice about which platform to use to publish their story. This practice was multimodal insofar as it included multiple platforms to deliver content and also multimedia to present it. For example, writing in *Novaya Gazeta* about local elections,<sup>8</sup> Lilit Sarkisian uses text, photographs, scans and videos to provide a report about the role of political parties in the RF. The piece is written in the documentary style whereby the analysis of the situation is mixed with documentation and evidence. All citations are carefully attributed and all pictures are geo-tagged thus making the user feel like they are part of the investigation. The piece includes multiple hyperlinks enabling the user to check some other facts to view related content. The piece is easily sharable on multiple platforms. All of these elements of digital journalism are incorporated in the story, thus making it not only about the use of technologies but also about the ways in which to narrate about an event or a social concern.

Journalists would also invite comments and feedback from the users and would customize their outputs to meet expectations of specific groups of users. Between 2005 and 2015, user commenting was a common feature in online media outlets; it has been gradually phased out as the media outlets shifted discussions and user interactivity onto social media, making them responsible for the user-generated content, on the one hand, and on the other, making them part of the story-telling. So, when posting texts online journalists would use hyperlinks to connect their story to others and to build news archives. For

example, Sarkisian folds her story about local elections in *Novaya Gazeta's* publications about United Russia, the dominant party in the RF which has been accused of corruption on all levels. She links her argument to other stories and requires that the user should carry out the work of putting the evidence together by following this and other stories. Thus, the political stance of *Novaya Gazeta* emerges not from a single publication but from a database of publications on a specific topic.

Thus, in the period of early digital journalism, multimediality, interactivity and hypertextuality were key methods with the help of which to produce content, including engagement with users (for more on hypertext, see Chap. 15). Eventually, digital journalism emerged to encompass a wide variety of ways in which digitalization has influenced news production. Nowadays, digital journalism also incorporates related areas and forms of activity, arrangement and engagement, including communication among journalists, their work environment, and so on. This means that digital journalism should be considered as an entirely new practice and institution of journalism, not just a particular practice of writing and publishing. In many ways, digital journalism has supplanted “analogue” journalism of the twentieth century.

Some commentators have described these changes as “the death of journalism,” meaning that journalism as it was known in the twentieth century had ceased to exist. For others, just like with the previously announced death of the novel and death of cinema,<sup>9</sup> the digital turns mean a re-interpretation and reinvigoration of journalism. To go on with the analogy, just like celluloid cinema is perhaps dead, but post-celluloid, digital cinema is thriving, supplying new genres, stories and visual regimes, and using new platforms for content distribution, digital journalism is an emerging and expanding field of activity aimed at informing the public about current events and providing political, social and cultural commentary along with organizing and maintaining new spaces for information sharing and collaboration among the publics, in the national and transnational, and local and global settings.

One of the principal outcomes of the death and re-birth of journalism in its digital phase is the emergence of “alternative journalism.” I define alternative journalism in the following way. The difference between professional and alternative journalists is in how people understand their objectives and acceptable levels of responsibility. The former group—professional journalists—includes any kind of journalists whereby individuals, associations of individuals and officially accredited companies engage in journalism as their primary activity. For example, it can be an individual with a university degree in journalism, or someone without formal education in journalism,<sup>10</sup> for whom still journalism is a professional occupation. They can be members of a professional society such as the Russian Association of Journalists, or, they can belong to an informal network of individuals and companies involved in similar activities. They can be on a permanent contract with one company or work part-time or as freelancers for a number of media outlets.



The latter group—alternative journalists—encompasses individuals and companies that are responsible for news content but who do not consider themselves reporters per se. For example, it can be an arts organization—like London-based Calvert and its equivalents in Russia such as Afisha.ru and The Village—that informs the public about events concerning contemporary arts and culture in the national and international context. Or, it can be an individual who makes regular posts on current affairs in social media and attains a high level of visibility and credibility in their circles. For example, in the late 2010s, Dr. Ekaterina Schulmann emerged from an academic active on social media into an important, liberally minded political commentator appearing on federal channels.

Indeed, in the twentieth century there were individuals and organizations that attempted to create their own news flows,<sup>11</sup> yet it is with the arrival of the digital era that the opportunity to build their own news flow and provide media content to a niche or general audience became available. As the Schulmann example demonstrates, the boundaries between professional and alternative journalism are fluid and transitions from one to another are enhanced thanks to the digital media. Some organizations like universities encourage alternative journalism when it serves the needs of the organization. Others, for example, banks are nervous about the release of any data by their employees.<sup>12</sup>

To be absolutely clear, the difference between professional and alternative journalism is not that of quality, but that of the relationship of an individual or an organization to the broader journalistic field. In other words, alternative does not mean “amateur,” a term which implicitly designates poor quality of content. Instead “alternative” stands for the new ways of organizing production and circulation of content which is possible thanks to digital technologies.

In this framework, alternative is also different from grassroots journalism. In the early new media parlance and digital criticism, the term meant journalistic practice stemming from the activities of “ordinary users.” It was believed that these users were happy to “share” their (local) insights and independently produce content with professional media companies. Eventually, it became apparent that grassroots journalists would not only collaborate but also compete with professional journalists in terms of salaries, contracts, awards, visibility, authority, and especially symbolic capital. These were no longer grassroots reporters but media content producers of significant influence in their own right. The shift was noticeable in how major media companies such as the (British Broadcasting Corporation) Russian Service went from inviting user comments, that is, building news stories on the basis of “grassroots journalism,” to disabling user comments altogether, that is, aiming to maintain “a professional stance” as a marker of journalistic quality. This way they differentiated themselves from the range of new media outlets that had carved out their share of the media market in a direct threat to legacy media outlets such as the BBC.

Thus, alternative journalism signifies new arenas of journalistic activity, both in terms of production and consumption of materials, and new forms of content, user engagement and circulation patterns. In the beginning, alternative journalism carried hallmarks of mainstream digital culture, that is, it was markedly different from professional journalism. However, eventually, the boundaries between the two became increasingly blurred. This was one of the transformations that led to the decline of legacy journalism in the late 1990s–early 2000s. Some Russian media outlets easily adapted to the new realities of digital journalism; others were less successful and have disappeared from the Russian market or have developed into completely new projects. In the end, what has remained is digital journalism: nowadays virtually all existing Russian media outlets function according to the logic of digital journalism. This allows me to suggest that in the RF all journalism is digital journalism, if not in terms of technology used but in terms of structure and processes.

### 9.3 ALL JOURNALISM IS DIGITAL JOURNALISM

The transition to digital journalism means more than a greater use of digital tools. It encompasses major transformations of media flows, systems of authority and trust, business arrangements, everyday practices of journalistic work, for example, opportunities to work remotely, and so on. In the RF, the transition to digital journalism occurred at the same time as in developed economies in the Anglophone West, which means that the processes and practices of digital journalism are not dissimilar in these countries.

For example, because of the changing fabric of the journalistic profession including the spread of digital technologies, we see the rise of influential female journalists in the RF and the United Kingdom. For example, Èlina Tikhonova is a business and culture reporter on RBC (Russian Business Consulting), a principal Russian-language media outlet for business reporting, and Laura Kuenssberg is a political editor on the BBC, the United Kingdom's most important public broadcaster. The authority of these journalists had been established thanks to their activity on social media such as Twitter and Facebook.<sup>13</sup> To confirm, having built a reputation in social media, they gained greater visibility in their respective media outlets. In return, the media outlets have started to use the authority of these journalists to advance their agenda in social media, which signals a convergence of digital spaces and tools. Their case exemplifies a transfer of alternative and professional strands of journalism within their professional career. The fluidity of agendas, forms of reporting, modes of expressing an opinion, and relationship to and within their media outlets points to a new system of journalism.

This new system of journalism provides individuals with new opportunities. For example, both Tikhonova and Kuenssberg have used their professional reputation in order to advance emancipatory agenda. Kuenssberg has promoted the issue of gender equality and diversity, making it one of the most visible social concerns in the United Kingdom. Conversely, Tikhonova took

part in the Russian spin-off of the global #metoo campaign, urging RBC and other journalists to boycott reporting from the State *Duma* (lower house of the Federal Assembly of Russia) after allegations of sexual harassment against its deputy Leonid Sluckij became public. Kuenssberg and Tikhonova have operated in realms that are highly politicized in the United Kingdom and RF, thus straddling the traditional arenas of reporting and activism. We observe a convergent of national and transnational realms of journalism and activism, and a transfer of agendas from essentially the journalistic domain to that of broader societal concerns (for more on digital activism, see Chap. 8).

This case demonstrates that currently the processes and practices of digital journalism in the RF and other western countries are not dissimilar. Yet, there is a big difference in terms of the general evolution of journalism and what it means to the respective societies. The point I wish to emphasize here is that in the RF, the rise of digital journalism coincides with the rise of Russian journalism per se. To confirm, modern Russian journalistic practice is based on the neoliberal form of journalism that was imported from the West as part of Gorbachev's perestroika of the 1980s and Yeltsin's privatization campaigns of the 1990s. This journalistic practice had supplanted the system of media organization and journalism that had existed in the Union of Soviet Socialist Republics (USSR). The transfer was complete by the start of the twenty-first century when digital technologies were becoming mainstream. So, in the United Kingdom, the transfer to digital journalism was a gradual process of transformations of journalistic practice; in the RF it signified a radical break from the tradition of Soviet journalism.

To elaborate, these reforms introduced during the perestroika period and the 1990s included the abolishment of censorship, greater freedom of expression and more emphasis on the protection of journalists. This was a positive outcome of the reforms. The negative outcome was in that these reforms put journalists on a collision course with private business which, in order to grow, employed aggressive and sometimes brutal methods of control. These reforms also gave rise to unregulated lobbying and the use of illegal and semi-legal promotional campaigns, especially during political elections. Early digital journalism—in the spirit of digital utopianism—attempted to eradicate two problems at the same time: the old practices of Soviet journalism, on the one hand, and on the other, the new practices installed as a result of the neoliberalization of journalism in the 1990s. The attempt was partially successful: propagandistic features of Soviet media were carried over to Russian state-funded television channels and also to the international broadcaster RT,<sup>14</sup> and commodification of information in the 1990s gave rise to sensationalist and click-bait media. At the same time, Russian contemporary understanding of privacy is informed by the notions and practices formulated in the digital realm, which, to remind the reader, remained completely unregulated for a significant period of time, relying on self-regulation instead.

As a result, many problems of contemporary Russian journalism are accounted for by the gap between legacy and digital journalism, and between

professional and alternative journalism. For example, the safety of journalists in Russia is a recurring concern. Western media and scholarship have addressed this issue from the perspective of the oppression of journalists by the state (see, for example, Oates 2006). The case of Anna Politkovskaya, who was murdered in 2006, is indicative. However, researchers have overlooked other aspects of oppression, resistance and safety such as corporate controls over journalists, privacy, wellbeing and intellectual property. Indeed, my interviewees had complained about their experience of working in small and medium-size media outlets.<sup>15</sup> In terms of the digital realm, they noted that, due to the lack of training provided by the media outlet owners, they are exposed to threats such as harassing in social media, data breaches, illegal file sharing, and so on.

How did these concerns develop? What were the pivotal moments in the development of digital journalism? In the subsequent sections I answer these questions from the perspective of the evolution of digital journalism (Sec. 4), and from the perspective of its form and functionality (Sec. 5).

#### 9.4 HISTORICAL OVERVIEW OF RUSSIAN AND RUSSOPHONE DIGITAL JOURNALISM

I have established that in case of the RF, the emergence of digital journalism is a complex process that signifies a lot more than the transition to new technologies employed in the production, circulation and consumption of journalistic items. In this regard, what has the evolution of digital journalism been like? Is it possible to identify significant trends and phases that help us understand these transformations?

In previous publications, I have argued that the proliferation of digital technologies in the RF includes four distinct stages, each defined by the type and frequency of use.<sup>16</sup> In this section, I intend to use the historical periodization of the evolution of digital technologies to develop a periodization of digital journalism in the RF. I identify four stages that correspond to and underpin four stages in the development of digital journalism that I outline below.

The first phase—the early 1990s—was characterized by the experimental use of digital technologies. At that time, scientific labs, artistic collectives and creative individuals began to use advanced digital technologies. Soviet-era computers had become most obsolete, with users relying on technologies imported mostly from the West. Users would engage in the exchange of data, including news, across the Russophone space of the internet. Cross-border, politically unhindered exchange of information was particular to this period of the evolution of digital technologies. For example, the art collective known as net.art were responsible for building first international networks, sharing information, pieces of news and pieces of code.

During the second phase, the experimental users of the 1990s became established in their professional circles, including journalism, giving rise to what I have labeled (Strukov 2014) the elite user of the late 1990s–early 2000s. For

example, in the 1990s Anton Nosik was based in Israel working as a programmer and running a number of internet-based projects among Russian speakers. In the 2000s, thanks to his proven record of successful media projects, he relocated to Moscow in order to direct major web-based news agencies such as Lenta.ru. Together with other elite users, all of whom were journalists and programmers living in large urban centers and being in charge of the strategic development of media, culture, science and technology, Nosik was responsible for building what was to emerge as the Runet. During this phase, technological innovation provided elite users with significant symbolic capital (for more on history of Runet, see Chap. 16).<sup>17</sup>

The third phase relates to the late 2000s when digital technologies including mobile phones became commonplace, and different kinds of users started using digital technologies for work, socializing and networking. The mass user challenged the authority of the elite user, effectively diversifying Russian digital system. During this phase, the Russian government became more active on the internet, launching a series of “national projects” aimed at stimulating economic and cultural activity in certain sectors of the digital technologies. The government was responsible for the technological upgrade of the Russian media system. For example, it set deadlines for the digital switch-over, compelling Russian companies, media outlets and users to accept new technologies such as digital television (see Strukov 2011). During this period, individuals such as Nosik switched from building their authority online to monetizing their symbolic capital. For example, Nosik was the director of high-profile investment projects concerning digital media such as his company SUP which purchased *LiveJournal* and transferred it to the RF.

The most recent period—the late 2010s—is characterized by “total” digitalization. Around that time, digital technologies and media had been firmly established as the main means of communication among the majority of users, with “old” media and non-digital technologies increasingly playing an auxiliary role, especially in urban centers. During this phase, the government has been extremely active on the digital field launching a few initiatives that have effectively nationalized the Runet, for example, precluding “foreign” companies to own solely media in the RF. The purpose of this activity was to make the Runet less transparent to the Western observers and to protect the economic interests of the Russian political elite. During this period, the role of the elite users such as Nosik has diminished whilst the new trends have been set by Russian major tech corporations such as Yandex and social media influencers such as Yury Dud’, the editor of the principal sports outlet Sports.ru who had built notoriety due to posting his controversial interviews with celebrities on YouTube. In fact, this example reveals the merger between tech and media giants such as YouTube and individual content producers such as Dud’. It blurs the boundaries between individual and corporate agency, between news reporting and lifestyle media, between customized and universally available content, and so on.

These stages of technological and media development correspond to the stages in the development of Russian and Russophone digital journalism, including:

- (a) exchanges of essential information and news items through email and messaging on personal and professional networks, including transnational exchange; emerging digital networks remain completely unregulated until the intervention of the government and security services at the start of the twenty-first century;
- (b) the emergence of alternative media outlets taking advantage of the unregulated realms of the early internet; these encompass web sites and mail lists that circulate information and news items to a target group such as subscribers to a service; by the start of the new millennium these services emerge into big players on the Runet; state-backed and corporate media scramble to increase their presence on the internet in order to catch up with services such as Lenta.ru;
- (c) lifestyles media and media outlets based on personalized media flows such as *LiveJournal* proliferate, effectively diluting the impact of online investigative journalism; with the launch of (Russophone) social media at the end of the decade, the media landscape is entirely transformed with legacy media such as official television channels playing a catch-up game with digital media; and
- (d) the government begins to regulate aggressively the digital realm in order to protect the interests of large digital corporations<sup>18</sup>; it introduces legislation which effectively “nationalizes” the Runet, that is, makes it possible to separate domestic and transnational media flows. Alongside government regulation, digital media corporations such as RBC and Mail.ru bid to increase their share on the internet, including mergers and collaborations with global media companies. For example, Mail.ru becomes one of the flow amplifiers for the BBC World Service and Yandex taxi service merges with Uber. At the same time new formats of news and lifestyle media emerge on the internet, especially on services that enable streaming audio-visual content such as YouTube. They impact the processes of information gathering and presentation styles on legacy media; digital natives begin to dominate professional and alternative forms of journalism.

This cross-check of technological, social, political and cultural developments allows an historical, dynamic consideration of digital journalism. In this system, the ways in which digital journalism works become apparent. It reveals the realms and modes of digital journalism, the role of government and corporate regulators, and the role and expectation of digital audiences. It also signposts areas of innovation which can be used in both the progressive and regressive manner by individual, state-aligned and corporate agents. In this process, the question of practice of digital journalism becomes important. In the final section of the chapter, I attempt to conceptualize these practices from the perspective of their function, not form or outreach or frequency.

## 9.5 TYPOLOGICAL OVERVIEW OF RUSSIAN AND RUSSOPHONE DIGITAL JOURNALISM

In the previous section I apprehended the realms of digital journalism by considering the principal areas of impact in the historical context. In the concluding section, I consider digital journalism in its most contemporary form by analyzing a number of interrelated phenomena. I account for the nature and configuration of each of them by introducing a particular case. It is meant to reveal current debates and help me relate back to the discussion presented at the start of the chapter. Thus, I wish to argue that digital journalism defines new spaces of activity and problematizes existing social, political and cultural concerns such as the notion of privacy and geographical distribution of data.

Digital journalism has problematized the notion of media and media outlet through the use of new platforms. Platforms are digital realms identified by a particular distribution model, content organization and visual language that allow new modes of production and distribution of content. For example, Telegram is an instant messenger that was created by entrepreneurs Pavel Durov and his brother Nikolaj in the early 2010s. Since then it has emerged into one of the most powerful platforms for messaging, micro-blogging, storytelling and channeling of information including audio-visual materials. Created to assist communication, Telegram is nowadays used by many journalists to enhance their professional activities such as secure communication with other journalists. Telegram advocates complete privacy of communication, that is, information distributed on its platforms cannot be filtered by security sources.<sup>19</sup> For many journalists in the RF, Telegram symbolizes freedom of communication and freedom of speech. And so, Telegram is considered by many to be a means to protect human rights. As a result, Telegram is used to launch and sustain independent alternative media outlets such as Telegram groups, for example, LGBTQ+ (lesbian, gay, bisexual, transgender, and queer) groups. (The highly private nature of Telegram means that it is also used to deliver questionable content such as pornography.)

Groups on messengers are a form of news delivery which, when applied at a mass scale, can be employed as a powerful tool for distribution of content. They are related to news aggregators, meaning they provide information including news in a structured and/or customized way. These are algorithms and networks that allow the collection and distribution of items on a massive scale; news aggregators blur the boundaries between original and unoriginal/re-published content, thus posing the question of authorship and intellectual property in the digital age. The proliferation of aggregators in Russian and Russophone media is due to the weakness of the Russian law and its ability to protect intellectual property. At the same time, news aggregators advance the culture of sharing, collaboration and mobilization by creating a sense of commonality and belonging among users. In some cases, the use of aggregators has enabled media startups to grow so that eventually they are able to produce their own content. A good example is the Riga-based Meduza who, in the



beginning, re-posted reports and news items from established media, for example, *Kommersant*, and eventually developed into an independent information producer, sharing content across a range of platforms and outlets.

The emergence of news platforms such as *Meduza* is possible to increase datafication of all aspects of life. Datafication is the process through which, on one level, journalists make use of digital tools such as computer-assisted reporting, digital indexing and database researching, and on another, they present their findings in the form of data such as data visualizations. In other words, datafication defines the omnipresence of data—the data turn in journalism—whereby journalists use data to present information about the world and to conceive of the world as data. In terms of journalistic output, nowadays there is less emphasis on story-telling and more on organizing information as banks of data whereby the user is expected to do their own research and arrive at a conclusion. In this respect, there is a growing problem with verification of information, resulting in abuses of data and spread of conspiracy theories. There is also a problem with the assumed neutrality of data: in the early 2000s, in reporting, data was considered a means to achieve impartiality; in the late 2010s, data is seen to contain its own ideologies, impacting how data is gathered, processed and stored. Recently, the rise of affective journalism—the use of deeply personal experience such as sexual problems to account for the changes in the world—can be attributed, in many ways, to the backlash against datafication of journalism.

Datafication accounts not only for new technologies and new ways of structuring information and communication but also for new ways of thinking about ourselves and our world. Reading the world as data results in the new position of the subject in the physical world and the world of data whereby the boundaries between the two become increasingly blurred. In some cases, this new ambiguity reveals the complexity of the use of digital tools in journalism such as mapping and surveillance tools and recognition tools. Mapping and surveillance tools are gadgets and applications that help journalists gather information that is otherwise not available. For example, Alexei Navalny, who, in the West, is routinely described as the Russian opposition leader, uses leaked and hacked documents, and open source investigation to counteract corrupt elements in the Russian government. For example, he uses drones to survey properties of the members of Russian political establishment. He incorporates footage obtained by these means—which would be illegal in the West—into his investigative reports about the wealth and corruption of Russian nomenclature which he releases on his channel on YouTube. This kind of practice occupies a gray zone from the point of view ethics of journalism and legal framework (arguably, Navalny uses loopholes in existing legislation). And recognition tools are applications that allow journalists to identify subjects and maintain effective networks. For example, those working in big media organizations have reported using apps that help them catalog contacts including their own colleagues. For example, they use feature recognition tools to “recall” the names and positions of their peers and contacts. *Findface* was a media startup

that launched a free service in 2016 enabling users to identify passers-by by taking their pictures and linking the individuals to the profiles on social media. Very soon Findface closed its operations; however, online communities discussed how their services were acquired by security and commercial enterprises. For example, in June 2018, S7 Airlines, the chief competitor of Aeroflot, started using face recognition tools in its lounges allowing passengers to check in automatically. This event was reported neutrally in Russian progressive media,<sup>20</sup> meaning that digital innovation of this type has been securitized in popular imagination.

These developments signify that in digital journalism, various platforms, tools and databases are employed to carry out journalistic investigations and produce and deliver content across a wide range of networks. This creates and sustains a constantly evolving news world so that the user is continuously engaged in this world on all available platforms. This form of transmedia storytelling (Jenkins 2007) blurs the boundaries between “real” events, media events and mediated events, on the one hand, and on the other, advances new social interactions and cultural phenomena. All of them foreground digital journalism as a new system of complex social and political realities.

## NOTES

1. By occupying a particular section of the market, namely, the publication of analytical reports in the evening, in just three years Meduza has emerged from a niche media outlet to one of the most important Russophone content producers.
2. For example, Calvert is a London-based arts foundation directed by Nonna Materkova, a Russian entrepreneur and investor. In addition to exhibitions showcasing the arts of the New East, Calvert runs an online magazine about art and culture in the former communist countries. Thanks to an appealing design of the journal and innovative approach to news selection—they tend to write about new trends in fashion, music and architecture, as well as provide critical reflections on social and political transformations in the region—Calvert Journal has become an influential media outlet. Their impact is evident on how the Guardian and other British media re-publish items from Calvert Journal and/or respond to the available items. In other words, through its innovative focus on the creative industries of the New East which has been attained with the help of digital collaborations with artists, musicians and photographers from the region, the Calvert Journal has broadened the agenda of other media. Indeed, writers working for the Calvert Journal also work as freelancers for other media, thus building a specific circuit of exchange whereby journalists no longer have to be present in an office and instead work from multiple locations for multiple media outlets.
3. In fact, mail.ru is used as an amplifier by the BBC World Service.
4. For example, news items published on mail.ru tell a story about events in Russia and abroad, and this is a traditional concern of journalism; however, these items are linked to financial data, including Mail.ru online banking services, thus, impacting the range and applicability of traditional reporting and other activities such as investment and banking, and so the realm in which journalism exists in

the digital era is big and complex. From introducing “native advertising” to developing and incorporating elements of machine learning and artificial intelligence, Mail.ru, Yandex and digital startups have transformed the processes and practices of journalistic work in the Russian Federation, as well as in other countries through their subsidiaries.

5. See, for example, Strukov 2011 and 2014.
6. Fifty semi-structured interviews with media practitioners were collected. Their duration varies between sixty and ninety minutes.
7. See, for example, interviews with Russian journalists and media practitioners published in *Studies in Russian, Eurasian and Central European New Media* ([www.digitalicons.org](http://www.digitalicons.org)).
8. <https://www.novayagazeta.ru/articles/2019/07/05/81136-sobolinaya-ohota>.
9. See, for example, Boxall 2015.
10. I deliberately do not call these individuals “amateurs” insofar as their activities are professional albeit lacking recognized formal qualifications.
11. The notion of citizen journalism is a related concept.
12. Stories about how Sberbank controls the use of mobile phones by its employees are a common feature in Russian media.
13. This assertion is based on my interviews with the RBC and BBC journalists and editors (2014–2019).
14. See, for example, Hutchings 2018.
15. Citation.
16. The classification is based on my analysis of Russian transition from Soviet to Western digital technologies and computational system presented in Strukov 2014.
17. See my discussion of Nosik’s *LiveJournal* activities in Strukov 2010.
18. For example, the so-called Yarovaya Law which introduced “counter-terror and public safety measure” but in fact allowed Russian companies and security services to control economic aspects of media flows such as data mining and sale of personal data.
19. Telegram’s refusal to share the code with the government had led to attempts to shut down the service in 2018 which were unsuccessful.
20. <https://www.the-village.ru/village/city/news-city/355649-pass>.

## REFERENCES

- Anderson, Chris W. 2013. *Rebuilding the News: Metropolitan Journalism in the Digital Age*. Philadelphia: Temple University Press.
- Boxall, Peter. 2015. *The Value of the Novel*. Cambridge: Cambridge UP.
- Bradshaw, Paul, and Liisa Rohumaa. 2013. *The Online Journalism Handbook: Skills to Survive and Thrive in the Digital Age*. London: Routledge.
- Bucher, Taina. 2018. *If ... Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.
- Diakopoulos, Nicholas. 2019. *Automating the News: How Algorithms are Rewriting the Media*. Cambridge, MA: Harvard University Press.
- Franklin, Bob. 2017. *The Future of Journalism: In an Age of Digital Media and Economic Uncertainty*. London: Routledge.

- Franklin, Bob, and Lily Canter. 2019. *Digital Journalism Studies: The Key Concepts*. London: Routledge.
- Hill, Steve, and Paul Lashmar. 2013. *Online Journalism: The Essential Guide*. New York: SAGE.
- Hutchings, Stephen. 2018. International Broadcasting and 'Recursive Nationhood'. In *Russian Culture in the Age of Globalization*, ed. Vlad Strukov and Sarah Hudspith. London: Routledge.
- Jenkins, Henry. 2007. *Transmedia Storytelling 101*. blogpost, March 21. [http://henryjenkins.org/blog/2007/03/transmedia\\_storytelling\\_101.html](http://henryjenkins.org/blog/2007/03/transmedia_storytelling_101.html). Accessed 10 Oct 2019.
- Jones, Janet, and Lee Salter. 2011. *Digital Journalism*. New York: SAGE.
- Oates, Sarah. 2006. *Television, Democracy and Elections in Russia*. London: Routledge.
- Strukov, Vlad. 2010. Russian Internet Stars: Gizmos, Geeks, and Glory. In *Celebrity and Glamour in Contemporary Russia: Shocking Chic*, ed. Helena Gosילו and Vlad Strukov. London: Routledge.
- . 2011. Digital Switchover or Digital Grip: Transition to Digital Television in the Russian Federation. *The International Journal of Digital Television* 2 (1): 67–85.
- . 2014. The (Im)Personal Connection: Computational Systems and (post-) Soviet Cultural History. In *Digital Russia: The Language, Culture and Politics of New Media Communication*, ed. Michael Gorham, Ingunn Lunde, and Martin Paulsen. London: Routledge.
- Thorsen, Einar, and Daniel Jackson. 2017. Seven Characteristics Defining Online News Formats: Towards a Typology of Online News and Live Blogs. *Digital Journalism* 6 (7): 847–868.
- Wahl-Jorgensen, Karin. 2019. *Emotions, Media and Politics*. Hoboken, NJ: Wiley.
- Zion, Lawrie, and David Craig, eds. 2014. *Ethics for Digital Journalists: Emerging Best Practices*. London: Routledge.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Digitalization of Russian Education: Changing Actors and Spaces of Governance

*Nelli Piattoeva and Galina Gurova*

## 10.1 INTRODUCTION

Digital technology has become an integral part of public schooling across countries since the 1980s (Selwyn 2018), driven by the combination of technological innovation and political determination for efficient governance. Russia is no exception and the recently intensifying introduction of Information and Communication Technology (ICT) into different aspects of education manifests Russia's convergence with the rest of the world. Digitalization increasingly draws the attention of the government, non-state sector and philanthropic organizations (for more, see Chap. 3). They view technology as a solution to a wide range of problems including the lacking and often uneven resources of educational institutions, low or unequal learning outcomes, outdated and unmotivating pedagogies or lack of a consistent monitoring of student progress. In addition, ICT and related skills are perceived as a defining feature of future professional and societal life as well as a means to ensure efficient governance. As a novel and strengthening focus of education policy and pedagogical practice, as well as an increasing source of private revenue, the introduction and actual use of digital education technologies deserve critical academic scrutiny. Scholars of education policy call for studies on the implications of digitalization for education governance,

---

N. Piattoeva (✉)  
Tampere University, Tampere, Finland  
e-mail: [nelly.piattoeva@uta.fi](mailto:nelly.piattoeva@uta.fi)

G. Gurova  
Moscow School of Management SKOLKOVO, Moscow Oblast, Russia

seeing it in the context of the profound turn in the governance of education to “new modes of government and governing where power is not confined to the state or to the market but is exercised through a plethora of networks, partnerships and policy communities who ‘consensually’ work with stakeholders to produce more flexible, responsive forms of service delivery” (Wilkins and Olmedo 2018, 5).

We focus particularly on two interrelated changes in education governance that have been widely attributed to the rise and operation of digital education technologies. First is the reconfiguration of old and proliferation of new, particularly non-state actors (e.g. Williamson 2017; Hartong 2018) in performing regulation on behalf of or in collaboration with the national governments. Second, we are interested in how digital education technologies act as connecting devices between actors and how they constitute or reconstitute spaces of governance (e.g. Gulson and Sellar 2019; Hartong 2018).

We distinguish between four forms of education digitalization. First, ICT is a resource for teaching and learning in the format, for instance, of online courses that use digitalized textbooks and other virtual materials. Second, digitalization rises to prominence through the teaching of technology as a subject or an extracurricular activity in its own right or as a cross-curriculum theme. In addition, there is overall a growing emphasis on acquiring ICT-related knowledge and skills as a core learning competence (curriculum in coding or robotics). Third, we categorize datafication defined as the “transformation of different aspects of education (such as test scores, school inspection reports, or clickstream data from an online course) into digital data” (Williamson 2017, 5) as a distinct but entangled manifestation of digitalization. Fourth, digitalization entails resourcing educational institutions with hardware, software and other digital infrastructure, so it changes the material environment of education in unprecedented ways. What brings these four forms of digitalization together is the fact that the actors who make them possible “gain increasing control over the field of judgment in education” (Takayama and Lingard 2018, 2). Thus, scholars claim that the governance of education is displaced towards new digitalized sites of expertise (Williamson 2017): “while ICT have become translated into the field of education, they simultaneously act as a core medium through which new actors have become authorized as key players to shape output- and accountability-based policy and practice” (Hartong 2018, 135).

Lewis et al. (2016) have urged researchers to acknowledge the complexity of emerging power relations and governance structures and to examine them beyond topographical imaginaries. This is particularly so due to the proliferation of data collection and use enabled by digital technologies that have turned metrics, calculations and comparisons into new means of governance in education enabling new visibilities and proximities. We deploy Harvey’s (2012, 77) useful distinction between the topological and the topographical in the ensuing analysis: “[i]n topographical mapping, the boundaries of state power appear as commensurate with a clearly defined territorial boundary, and such categorical mappings are echoed in the spatially nested structures of administrative

division.” Topologization, by contrast, draws attention to multiple new spatial figures where borders “do not coincide with the edges of a demarcated territory, and where it is the mutable quality of relations that determines distance and proximity, rather than a singular and absolute measure” (Harvey 2012, 77–78). Datafication produces novel connections between the governing and the governed actors and thus “create(s) and sustain(s) dynamic political and moral spaces” (Harvey 2012, 88). In other words, new possibilities for action and the exercise of power (Lewis et al. 2016) are facilitated by the fact that “presence and proximity (are) no longer simply a question of physical distance” (Allen 2011, 295; Gulson and Sellar 2019).

Situated in relation to an international body of literature and the two arguments pertaining to the contribution of digitalization to the (changing) governance of education, we proceed as follows: we first map the general education policy context within which the earlier and current digitalization efforts have unfolded. The next two sections present analyses of the ongoing policies and practices of education digitalization. First, we show how digitalization changes the character of traditional actors and enables new actors and actor assemblages to enter the scene of education governance and provision. Second, we look specifically at datafication as extending spaces of governance in both a topographical and a topological manner. Topographically, some practices of datafication follow established administrative structures and enable tighter vertical control over regions and education institutions by the federal authorities. But datafication also generates spaces that overcome topographical distance through relationality and connectedness. These manifest, first, in the possibilities of “intimate” governance (see Gorur 2018) reaching into individual subjectivities and, second, intensifying proximities to the global level of education governance bypassing the national authority. Needless to say, we are only able to scratch the surface of the ongoing developments, partly due to the scarcity of existing research and partly because we are dealing with a rapidly moving target. The analysis builds on diverse sources, including our own and others’ studies on the digitalization and datafication of Russian education, as well as recent policy documents, media reports and websites of central actors.

## 10.2 POLICY CONTEXT

Information technologies first entered the Soviet schools as a focus of teaching. Already in 1959, some schools in Moscow and Novosibirsk, the cities best resourced with computers, started teaching the basics of programming and computational mathematics in the name of international competition and efficiency of national economic planning. Political and economic prerogatives, coupled with increased accessibility of computers, transferred education in technology from a subject taught in specialized and elite schools into a compulsory curriculum area. Soviet schools started to teach the course “Principles of Information Science and Technology” in 1985, and in 1990/1991, it was declared a compulsory subject for grades 10 and 11 throughout the Soviet



Union. The course and the overall introduction of computer technology to schools raised enormous interest among educators, and the proposed new curriculum included topics related to both the technical and mathematical sides of computing, as well as discussion on the role of computers in society more widely. The “computerization” of education was viewed as a necessary step to keep up with progress and to start a pedagogical change away from the “chalk and talk” teaching method (Muckle 1988). The compulsory ICT course retained its socio-political and economic relevance: “Progress in modern electronics, computer technology, and robot technology is not only a critical constituent of the scientific and technological revolution, but also an area in which two societal and economic systems come into direct confrontation” (Vinokurov and Zuev 1985 as cited in Monakhov 1986, 143). In other words, teaching in and with technology was a means of making a contribution to the Cold War arms and space race between the Union of Soviet Socialist Republics (USSR) and the “capitalist West.” It is also important to mention the anticipated impact of technology on students’ worldviews. The newly introduced school subject had to foster a “communist upbringing” and to enhance students’ understanding of the world through objective mathematical models (Monakhov 1986, 148). Prerequisites for international collaboration existed long before the USSR “opened up” to the West, as Soviet pedagogues and programmers continuously studied, for example, US-developed programmed instruction and other international experiments in computer-based learning (Davydov and Rubtsov 1991; Afinogenov 2013; Tatarchenko 2019). Moreover, despite the fact that the late-Soviet developments in education technology were short-lived due to the collapse of the Soviet regime, these experiences continue to shape current expertise in and imaginaries of digital technology and its role in society (Tatarchenko 2019).

In 1992 the new Law on Education permitted education institutions more freedom in choosing curriculum and pedagogy, and in making financial and operational decisions. They were also allowed and even encouraged to seek private sources of funding. At the same time, a severe economic crisis caused abrupt cuts in state subsidies and pushed schools and universities to raise money through tuition fees, tutoring services and even by renting out premises. Many administrative and fiscal responsibilities were transferred from central to regional and local authorities in order to enable regionally and locally tailored solutions and in some cases survival strategies. In practice, decentralization led to increasing inequalities between regions and within them—between rural and urban areas—and made the education sector less transparent to the federal center (Polyzoi and Dneprov 2010).

In the 2000s the Russian Ministry of Education and Science issued several strategic and legislative documents that stressed the role of education in ensuring national economic growth, global competitiveness and human capital development, promoted the introduction of market mechanisms into the education sector, and called for the efficiency, transparency and accountability of education institutions (Gounko and Smale 2007). Tackling economic

deficiencies in education, the government used loans from the World Bank for particular education reforms, including those enhancing “efficient use of digital learning resources and electronic tools” (World Bank 2004), to complement federal funding for education. The government then introduced measures of centralized control, such as state standards, accreditation, licensing and centralized examinations, and a scheme of funding tied to the attainment of nationally determined indicators and outcomes. Reforms were designed to increase governmental and organizational efficiency, stimulate cost optimization, reduce space for lobbying and corruption in public funds allocation and ensure the overall realization of state priorities through outcome monitoring (Yastrebova 2013; OECD 1999). Further prerequisites for the digitalization and datafication of Russian education were thus created on the one hand by the opening up and commercialization of the education sector that started in the early 1990s, and on the other hand by the state’s embracement of the New Public Management (NPM) paradigm since the 2000s.

A significant latest leap in digitalization policies was prompted by the re-election of Vladimir Putin and the publication of his “decrees” of May 2018. These include the task of “ensuring an accelerated implementation of digital technologies in the economic and social spheres” (*Prezident* 2018; see also Kolesnikova 2018).<sup>1</sup> Specifically for education, the task is to create a “modern and safe digital education environment which ensures high quality and access to education at all levels” (*ibid.*). The decree continues and expands the “Digital educational environment” project (*neorusedu.ru*) launched in 2016, but takes digital education to the next level. The aim of education modernization for 2018–2024 is to ensure international competitiveness of Russian education and Russia acquiring a position among top-10 leading countries with the best quality of education according to international education rankings (Government of Russia 2018). “Development of digital education environment” is outlined as one of ten priority sub-projects, while the other nine sub-projects also feature different aspects of digitalization (*ibid.*). What is worth highlighting is the aim to establish a federal center for digital education transformation, to create a centralized federal platform that would compile information on and services in education, the related call to increase the provision of online courses and digitalize education administration and federal support for an increasing number of in-school and extra-curricular activities related to teaching ICT.

### 10.3 THE RISE OF NEW ACTORS AND ACTOR ASSEMBLAGES

In this section we document some examples of the entry of for-profit actors and philanthropies into the field of education by means of education digitalization. Their contribution is vital for the federal government to realize its political prerogatives of digitalization. At the same time, for-profit organizations are becoming increasingly attracted to the education sphere due to the prospect of new revenues and the opportunity to reach out to young people as future

employees and consumers. Growing proximity to the decision-makers through new actor assemblages enables these actors to communicate their visions of the future, and their political and economic interests, to legislative and executive bodies. Both more traditional education actors, such as textbook publishers and new actors, such as Internet service providers, the banking sector and industry, promote digitalization. Co-operation with multinational technology providers and international intergovernmental organizations manifests the growing entanglement between national and international actors.

Publishing houses promote digital learning materials and develop online platforms for teachers and students. A major education publisher *Prosvěsienie* (Enlightenment, <https://prosv.ru/>)—exclusive supplier of standardized education literature in the Soviet Union—has lately regained its central position and holds a 40 per cent share of the country’s educational market (Prosvěsienie 2017). Some commentators claim that it has (re)monopolized the textbook market (ibid.) and that its substantial revenues come solely from state contracts (Bryzgalova 2017; Becker and Myers 2014). By now *Prosvěsienie* has digitalized the entire spectrum of its textbooks, though questions are raised about the actual availability of ICT infrastructure in schools and the danger of growing inequalities among schools and students as to their access to digitalized education products. *Prosvěsienie* contracted Microsoft to enable access to digital education on Microsoft tablet personal computers (PC), but the agreement was terminated due to international sanctions on the company’s former chair of board Arkady Rotenberg (*Microsoft zamorozil* 2014). However, the task was taken over by Samsung with the successful sale of tablets starting in the summer of 2017. In 2017, *Prosvěsienie* also signed a US\$ 1.1 million deal with the Russian Internet service provider Yandex to develop an online platform for schoolchildren, teachers and parents with self-proclaimed elements of machine learning and personalization. Yandex has been actively developing education services and products, including prep materials across compulsory school subject areas (Gerden 2017; *Analiz dannykh* n.d.; *Yandex učebnik* n.d.).

Several large Russian high-tech companies with mixed ownership (such as *AFK Sistema*, *Rosnano*, *Sberbank*, *Bazovyy Element*) have launched influential philanthropies that claim to improve school and higher education particularly via access to and provision of education technologies. *Bazovyy èlement’s* (Basic Element) charity *Vol’noe Delo* (Voluntary Work) runs a large-scale program for schools on new pedagogical methods (<http://volnoe-delo.ru/directions/education/inzhenery-novogo-pokoleniya/>), and *Rosnano* (Russian Corporation of Nanotechnologies) offers schools and higher education institutions an online platform with distance education courses in science, technology, engineering and mathematics (STEM) subjects, calling it “a large-scale online project that forms the professions of the XXI century.” The project includes support and recommendations to teachers (<https://edunano.ru/stemford/>).

For-profit players, government actors, academic institutions and intergovernmental organizations form novel assemblages that increase their influence

and opportunities for action in the arena of education digitalization. An example of such an assemblage can be found in the flagship project Competencies of the 21st Century of Sberbank's (a state-owned Russian banking and financial services company headquartered in Moscow) philanthropy *Vklad v budušee* (Investment in the Future, <https://vbudushee.ru/>). The project sponsored the preparation of a research report that makes "recommendations for the transformation of Russian school that would enable to close the gap between the education system and the demands of real life." The report was prepared in 2018 by a major Russian think tank in education policy, the Institute of Education at the Higher School of Economics, in co-operation with the Organization for Economic Co-operation and Development (OECD) Education 2030 group and the United Nations Educational, Scientific and Cultural Organization (UNESCO) experts (*Kompetencii 21 veka* n.d.). The report has been widely cited in the Russian media and presented on various public and government forums. The number one "new literacy" advocated in the report is digital literacy (*Kompetencii i gramotnost'* n.d.). Simultaneously, Sberbank's CEO announced that the corporation is developing a digital learning platform to be ready for use in 2019. The platform will be open to schools free-of-charge, and will enable personalized learning, pupil's choice of pedagogy, study outside of school, and continuous monitoring of educational achievement (*Sberbank rabotaet* 2018).

Other active players include EdTech startups, small and medium-sized education businesses, startup accelerators (e.g. Skolkovo Innovation Center or Russian Venture Company, RVC, <https://www.rvc.ru/en/>), and business forums that promote the EdTech agenda, such as the yearly EdCrunch exhibition (<https://2019.edcrunch.ru/>). In 2018 the head of RVC announced the creation of a new investment fund that will focus solely on education technologies (Futur'e 2018); and a government representative commented at the Open Innovations Forum in autumn 2018 that the government will stimulate EdTech initiatives and assist their access to schools, universities and state-owned companies in order to facilitate their development, since Russia has good potential for becoming a global-level player on the EdTech market.

#### 10.4 DATAFICATION EXTENDING SPACES OF GOVERNANCE

Education digitalization manifests particularly as a process of encoding ever more complex educational processes into software products (Williamson 2017), which has led to and is entangled with another major development, namely education datafication. Digitally produced or analyzed and visualized data can be inserted into databases, allowing different actors and their performances to be measured, evaluated and re-presented, and decisions to be made on the basis of data and their analysis. Education administration at different levels of governance and across educational institutions is increasingly data-driven, underpinned by the need to both produce and use indicators, data analytics and other forms of "objective evidence." This is the development to

which we next turn. We perceive datafication as both entangled with and distinct from digitalization. Datafication manifests as a process of data collection and deployment in its own right, for instance, through the exercise of quality evaluation and testing of learning outcomes and intensifies through the deployment of digital technologies, such as students' engagement with electronic teaching materials and games and teachers' reporting in electronic journals. Datafication is likely to intensify in the coming years, opening the door further to new actors and actor assemblages (as discussed in the previous section) and extending governance practices topologically, that is, cutting across established spatialities and composing new proximities and continuities, and thus spaces of governance, by means of datafication (Allen 2011).

In the environment of both the intended and the unintended diversification of education (see Sect. 10.2), the federal government has realized the potential of controlling education by means of data. The proclaimed demand for output data reproduces arguments about the need to increase efficiency and accountability of federal and sub-national executive authorities, to close the policy implementation gap and to fight against corruption and thus to pave the way for meritocracy and equality of opportunity (Piattoeva 2018). The development of centralized examinations and national surveys of education quality were strongly recommended by international actors such as the OECD and the World Bank. Russia participated in international large-scale assessments of learning outcomes to compare its educational achievements to international standards and to students' performance in other (Western) countries (PISA, TIMSS, PEARLS; PIAAC; TALIS).<sup>2</sup> On a smaller scale, managerial and market approaches prompted educational institutions to become more customer-oriented, to collect regular feedback from students and parents and to test students to monitor their progress "objectively." All these activities involve the gathering and analysis of data of rising quantity and breadth with the help of computers and software (for more on government data outside education, see Chap. 22).

The key driver and the first manifestation of the datafication of education, the Unified State Exam (USE), was introduced on an experimental basis in 2001 and was launched nation-wide in 2009. The examination combined the functions of the school graduation test and the national university entrance test, then gradually became a central source of information on educational achievement. The USE now serves as a means of external quality control of schools and universities, promotes national education standards and closer proximity between the official curriculum and actual classroom practices (Piattoeva 2015). Since 2009, USE as a measure of quality and a source of data about schools has been supplemented with the annual VPR (*Vserossijskie proveročnye raboty*, All-Russia Examinations) and a sample-based NIKO (*Vserossijskie Nacional'nye issledovaniâ kačestva obrazovaniâ*, National Study of Education Quality). These studies multiply federally driven education datafication and show how the federal center intensifies new topographically bordered proximities between the federal authorities and the regions through data,

rendering different actors not only amenable to control by making them more transparent, but also by attaching sanctions and rewards to quantitative outcomes, thus guiding the actors “softly” towards particular political ends (Piattoeva 2015; Hartong and Piattoeva 2019). In this manner, the government makes its presence felt at a distance, enabling “powers of reach” and “powers of connection” to create specific political spaces (Allen 2011).

The emergence of government-sponsored datafication has given rise to state-level organizations responsible for data-driven education quality control, such as Rosobrnadzor (Federal Service for Supervision in Education and Science), that gradually gained such powers that it rivaled the Ministry of Education and Science in the decisions about closing down education institutions deemed inefficient in terms of assessment results. In this sense, internal state structures, too, are being adjusted—and empowered or disempowered—by data-driven education governance. Simultaneously, experts in education measurement, psychometrics and software are gaining in power: for example, a small private association, the Moscow Center for Continuous Mathematical Education ([www.mccme.ru](http://www.mccme.ru)) has gained the status of a prominent expert after developing NIKO (see above) and publicizing the ranking tables of Russian schools on a contractual basis with the federal government. Simultaneously, as regions, schools and even individual teachers are increasingly controlled through the practices of data production, small-scale paid-for services emerge to offer, for example, commercial diagnostics of student achievement, paid-for student academic contests and a variety of local ranking exercises providing documentary proof of “high performance.” In this manner, government-sponsored datafication initiatives, carrying high stakes, feed the emergence of supplementary datafication services to help students, teachers and schools to manage the pressure, amplifying data collection exercises and the volumes of data produced (see Gurova et al. 2018).

While intensifying data collection through national tests intends to make educational affairs in regions and schools transparent and thus legible to federal-level governance and its attempts to standardize and unify, other developments speak of simultaneous differentiation in the system. In 2016, the city of Moscow initiated its participation in the “PISA for schools” international large-scale assessment (*Lučšie iz lučših* 2016). Following the prototype of the OECD’s Programme for International Student Achievement (PISA), PISA for schools enables school-to-school comparisons (Lewis et al. 2016). This benchmarked Moscow against schools in top-ranked countries and metropolises and enabled the Moscow authorities to proclaim that “Moscow school education is among the six best systems in the world, in terms of reading and mathematical literacy and among the top 20 in terms of scientific literacy.” The study not only highlighted that the quality of education in Moscow is much higher than in Russia on average, but also, and importantly for this paper, showed how local authorities can initiate alternative or complementary data collection exercises for their own political and administrative aims (Six facts 2017). Through PISA for schools, the Moscow administration bypassed topographically defined



administrative borders and initiated topological relations with one of the key actors in global education governance and established proximity between its (successful) system of education and world “best performers,” while distancing itself from the rest of the country. Moscow’s success in PISA for schools was partly attributed to its advances in education digitalization, and a recent initiative promotes partnering between Moscow schools and schools around Russia to disseminate Moscow’s experience on a school-to-school basis, setting up Moscow as an example to emulate. The “Moscow electronic school” project has been marketed as an outstanding innovation even capable of arousing international interest in Russian education among world education leaders (<https://www.mos.ru/en/news/item/48603073/>; <https://hundred.org/en/innovations/moscow-electronic-school>). In this example, we see how commensurative practices enable topological relations that produce new continuities between disparate education systems by locating them on a common metric—Moscow alongside international high-performing systems and cities, but also connecting schools across regions, bypassing the usual regional and municipal levels of education governance. But these new continuities also condition the production of discontinuities within the national space of governance—that is, marking out Moscow as a system in its own right and distinct from the rest of the country due to documented international success. These new dis/continuities facilitated by data create new spaces of and opportunities for educational governance—possibly contradicting the federal officials’ efforts to create a unified national education space.

Finally, in an attempt to envisage the future, we want to highlight the intensifying interrelationship between digitalization and datafication, enabling their mutual enhancement and complex governance arrangements that will increasingly work on and pervade individual subjectivities. The government-initiated organization Agency for Strategic Initiatives, ASI (Four Years of Agency for Strategic Initiatives 2017) plays the role of the government’s champion of the digitalization of all economic and social spheres and aspires to be the moderator for other private and public actors. It enjoys significant financial resources and symbolic support from Vladimir Putin and the Presidential Administration, and makes recommendations in the format of roadmaps to major actors such as federal and regional ministries and professional associations. ASI runs the project of the University for the National Technological Initiative (<https://asi.ru/news/85128/>) in which digital platforms and tools mediate all educational activities from the school level to adult education (*Koncepciã universiteta n.d.*). In 2018 ASI showcased digital education in a pilot education event for over one thousand participants. To gain admission, participants had to participate in several online tests, questionnaires and computer games, which assessed their performance and personal qualities by means of artificial intelligence (AI). AI simultaneously used these data for training the algorithm. The successful participants received personal online profiles, appraisal of their individual potential and recommendations for further learning in one of the six professional directions—presumably those which, in ASI’s estimation, would be most relevant in



the future: data analyst, technologist, entrepreneur, organizer, community leader and ecosystem architect. During the event, every participant's activities were continuously assessed and digitally documented. Their biometric data such as stress levels were also stored. On the basis of these data, participants were awarded points which they could use to gain access to education activities; they were also grouped according to their profiles and given recommendations for individual educational trajectories. The system analyzed which contacts would be most useful for each participant and connected participants with each other. After the event the organizers boasted about the vast amount of data collected (through audio and video records, participants' logs into the digital platform, bracelets that tracked biometric data, and so on). The data are to be used in the assessment of all participants' competencies and literacies and to make recommendations for their future jobs and education, as well as for the further development of artificial intelligence to guide future educational activities. The aim is that digital tools would enable direct co-ordination of personalized learning activities rather than continue to organize educational institutions that "teach everyone the same way" in an outdated "industrial époque" fashion. The tools and approaches piloted by ASI are expected to provide guidance for the development of other educational institutions, primarily universities, but also schools, professional colleges and extracurricular education organizations.

## 10.5 CONCLUSION

This chapter has documented the ongoing expansion of education digitalization and datafication that affects how Russian education is governed—who the important actors are in setting and implementing education priorities and how these priorities are put into practice. Digitalization creates space for an array of new actors to have a say in Russian education, though it must also be noted that the prerequisites for their involvement have been established by the federal state throughout the post-Soviet period and even earlier. Digitalization has led to an increased role of the philanthropic, business and voluntary sectors of society in the processes of education policy-making and delivery, while simultaneously changing the nature of and instruments available for more traditional players such as textbook publishers or executive-level authorities.

Whereas on the one hand, digital technologies help to re-center national authorities in the governance of education, the processes of digitalization, enjoying considerable support from current national policies across sectors and public discourse, do not entirely emanate from and are therefore not entirely controlled by the national authorities. The examples documented here show how they also unfold as a loose and spontaneous grassroots process, as a development promoted and steered by multiple public, private, mixed, individual and collective actors and their respective interests. Therefore, we propose that further digitalization of Russian education is likely to produce two co-existing realities: one in which certain aspects of education are

re-centralized by means of digital technology and in turn re-center the state in the activity of education governance, and the other, in which the proliferation of digital technologies leads to further diversification, raptures and inconsistencies in the education system. Both, however, manifest and generate novel governance relations between actors that elude description in a solely traditional topographical manner. As actors create complex arrangements between old and new state, private for-profit, academic and international intergovernmental organizations, they make the sphere of education governance in Russia more complex.

If the plans of the government materialize, Russian education system will soon produce increasing amounts of data far beyond what has been deliberately generated through systems of examinations and national assessments of education quality and learning outcomes. This also means that the governance of education will shift from human actors using data to govern to the governed actors engaging with the data that govern—what Williamson (2017) has called “digital education governance.” More data production will be possible by means of students’ continuous engagement with the digital environment through online learning resources and databases, as well as the more regulated and regular participation of schools and other educational institutions in quality monitoring exercises. The production, analysis and utilization of (numerical) data within the new regimes of education governance present a whole range of mechanisms that enable governance at a distance. In addition, new kinds of connectivities will emerge and effectively change the co-ordinates of governance as data increasingly reach individuals and groups that may have been beyond (topographical) reach (see Lewis et al. 2016). Local authorities and schools are now in a situation where they are rendered quantifiable and visible through intimate data (Gorur 2018). Transparency potentially renders them legible and amenable to control and intervention at any time. Moreover, by means of producing and publicizing data on themselves, institutions are guided towards aligning their work with a particular set of expectations (Piattoeva 2015). The plans to create individual digital portfolios for every teacher and student reflect a desire to reach these individuals by uploading their data and tracking their activities at every step of their educational lives. Penetrating the motivation structures and choices of teachers and students (e.g. by tracking and AI-analyzing their activities in the digital educational environment; by gamifying education and assigning scores and bonuses for certain activities) seek a similar intimate effect on subjectivity.

## NOTES

1. All translation from Russian are ours unless stated otherwise.
2. These abbreviations stand for the following international large-scale assessments: OECD’s Programme for International Student Assessment (PISA), OECD’s PIAAC (Programme for the International Assessment of Adult Competencies); IEA’s (International Association for the Evaluation of Educational Achievement)

TALIS (Teaching and Learning International Survey); IEA's Trends in International Mathematics and Science Study (TIMMS), and IEA's Progress in International Reading Literacy Study (PIRLS).

## REFERENCES

- Afinogenov, Gregory. 2013. Andrei Ershov and the Soviet Information Age. *Kritika: Explorations in Russian and Eurasian History* 14 (3): 561–584.
- Allen, John. 2011. Topological Twists: Power's Shifting Geographies. *Dialogues in Human Geography* 1 (3): 283–298.
- Analiz dannyh. n.d. [Data Analysis]. [https://academy.yandex.ru/events/data\\_analys/](https://academy.yandex.ru/events/data_analys/). Accessed 8 Jan 2019.
- Becker, Jo, and Steven Lee Myers. 2014. Putin's Friend Profits in Purge of Schoolbooks. *The New York Times*. <https://www.nytimes.com/2014/11/02/world/europe/putins-friend-profits-in-purge-of-schoolbooks.html>. Accessed 14 Jan 2019.
- Bryzgalova, Ekaterina. 2017. Samsung vypustil obrazovatel'nyj planšet na baze učebnikov Prosvěšeniâ [Samsung Launched an Educational Tablet on the Basis of Prosvěšeniâ's Textbooks]. *Vedomosti*. <https://www.vedomosti.ru/technology/articles/2017/08/29/731444-samsung-obrazovatel'nyj-planshet>. Accessed 14 Jan 2019.
- Davydov, V., and V. Rubtsov. 1991. Trends in the Informatization of Soviet Education. *Soviet Education* 33 (2): 58–70.
- Four Years of Agency for Strategic Initiatives' Work Were "Not in Vain" – Putin. 2017. [https://asi.ru/eng/news/detail.php?ELEMENT\\_ID=84440](https://asi.ru/eng/news/detail.php?ELEMENT_ID=84440). Accessed 15 Dec 2018.
- Futur'e, Èduard. 2018. RVK zajmetsâ vnedreniem virtual'noj real'nosti [RVC Will Work on the Implementation of Virtual Reality]. <https://www.ridus.ru/news/272584>. Accessed 15 Dec 2018.
- Gerden, Eugene. 2017. *Edtech: Russian Publisher Prosveshchenie Announces a Digital Educational Platform*. Publishing Perspectives. <https://publishingperspectives.com/2017/06/prosveshcheniye-yandex-digital-education-partnership-russia/>. Accessed 14 Dec 2018.
- Gorur, Radhika. 2018. Escaping Numbers? Intimate Accounting and the Challenge to Numbers in Australia's 'Education Revolution'. *Science & Technology Studies* 31 (4): 89–108.
- Gounko, Tatiana, and William Smale. 2007. Modernization of Russian Higher Education: Exploring Paths of Influence. *Compare* 37 (4): 533–548.
- Government of Russia. 2018. Pasport nacional'nogo proekta 'Obrazovanie' [Passport of the National Project 'Education']. <http://mo.mosreg.ru/download/document/1524233>. Accessed 24 Nov 2018.
- Gulson, Kalervo, and Sam Sellar. 2019. Emerging Data Infrastructures and the New Topologies of Education Policy. *Environment and Planning D: Society and Space* 37 (2): 350–366.
- Gurova, Galina, Helena Candido, and Xingguo Zhou. 2018. Effects of Quality Assurance and Evaluation on Schools' Room for Action. In *Politics of Quality in Education: A Comparative Study of Brazil, China, and Russia*, ed. Jaakko Kauko, Risto Rinne, and Tuomas Takala, 137–160. London: Routledge.
- Hartong, Sigrid. 2018. Towards a Topological Re-Assemblage of Education Policy? Observing the Implementation of Performance Data Infrastructures and 'Centers of Calculation' in Germany. *Globalisation, Societies and Education* 16 (1): 134–150.

- Hartong, Sigrid, and Nelli Piattoeva. 2019. Contextualizing the Datafication of Schooling – A Comparative Discussion of Germany and Russia. *Critical Studies in Education* (online first).
- Harvey, Penelope. 2012. The Topological Quality of Infrastructural Relation: An Ethnographic Approach. *Theory, Culture & Society* 29 (4-5): 76–92.
- Kolesnikova, Ksenija. 2018. “Klass BigData. Kak bol’sie dannye i ‘cifra’ izmenât rossijskie školy [The Class of Big Data. How Big Data and the Digital Will Change Russian Schools].” *Rossijskaâ gazeta* 7717 (254). <https://rg.ru/2018/11/12/kak-bolshie-dannye-i-cifra-izmeniat-rossijskie-shkoly.html>. Accessed 8 Jan 2019.
- Kompetencii 21 veka [Competencies of 21st Century]. n.d.. <https://vbudushee.ru/>. Accessed 15 Dec 2018.
- Kompetencii i gramotnost’ n.d. = Kompetencii i gramotnost’: operacional’naâ ramka [Key Competencies and Literacy: Conceptual Frame]. n.d.. <https://ioe.hse.ru/21skills/ks/list>. Accessed 15 Dec 2018.
- Koncepciâ universiteta n.d. = Koncepciâ universiteta nacional’noj tehnologiĉeskoj iniciativy 2035 [Concept of the University of the National Technology Initiative 2035]. n.d.. <https://2035.university/upload/iblock/0a9/0a9231d2eeca8e5165a2cf38a5e488f.pdf>. Accessed 8 Jan 2019.
- Lewis, Steven, Sellar Sam, and Bob Lingard. 2016. PISA for Schools: Topological Rationality and New Spaces of the OECD’s Global Educational Governance. *Comparative Education Review* 60 (1): 27–57.
- Luĉšie iz luĉših 2016 = Luĉšie iz luĉših: ĉego dobilis’ moskovskie školy [Best of the Best: What Moscow Schools Have Achieved]. 2016. <https://www.mos.ru/news/item/17830073/>. Accessed 20 Dec 2018.
- Microsoft zamorozil 2014 = Microsoft zamorozil programmu po sozdaniû planšeta dlâ rossijskih škol’nikov [Microsoft Stopped the Programme for Developing a Tablet for Russian Schoolchildren]. 2014. <https://www.interfax.ru/russia/405195>. Accessed 13 Jan 2019.
- Monakhov, V.M. 1986. The Content of the Basic Information Science and Computer Technology Course. *Soviet Education* 28 (10–11): 142–155.
- Muckle, James. 1988. *A Guide to the Soviet Curriculum: What the Russian Child Is Taught in School*. Beckenham: Croom Helm.
- OECD. 1999. *Country Background Report for the Russian Federation. Reviews of National Policies for Education*. Paris: OECD Publications.
- Piattoeva, Nelli. 2015. Elastic Numbers: National Examinations Data as a Technology of Government. *Journal of Education Policy* 30 (3): 316–334.
- . 2018. Doing Equality Through Greater Transparency? Troubling Surveillance Expansion in the Russian School System. In *Equity in and Through Education: Changing Contexts, Consequences, and Contestations*, ed. Stephen Carney and Michele Schweisfurth, 157–173. Rotterdam: Sense Publishers.
- Polyzoi, Eleoussa, and Eduard Dneprov. 2010. A Framework for Understanding Dramatic Change: Educational Transformation in Post-Soviet Russia. In *Post-Socialism Is Not Dead: (Re) Reading the Global in Comparative Education*, ed. Iveta Silova, 155–179. Emerald Group Publishing Limited.
- Prezident 2018 = Prezident podpisal Ukaz “O nacional’nyh celâh i strategiĉeskikh zadaĉah razvitiâ Rossijskoj Federacii na period do 2024 goda” [President Signed Decree “On the National Goals and Strategic Development Tasks of the Russian Federation for the Period Until Year 2024”]. 2018. <http://kremlin.ru/events/president/news/57425>. Accessed 24 Nov 2018.

- Prosvěšenie. 2017. <https://prosv.ru/eng>.
- Sberbank rabotaet 2018 = Sberbank rabotaet nad sozdaniem obrazovatel'noj platformy dlâ škol [Sberbank Works on the Development of Educational Online Platform for Schools]. 2018. <https://tass.ru/obschestvo/5274119>. Accessed 4 Oct 2019.
- Selwyn, Neil. 2018. Technology as a Focus of Education Policy. In *The Wiley Handbook of Education Policy*, ed. Rosemary Papa and Shadow W.J. Armfield, 459–477. Chichester: John Wiley & Sons.
- Six Facts About Moscow Schools: City Education in 2016. 2017. <https://www.mos.ru/en/news/item/19625073/>. Accessed 8 Jan 2019.
- Takayama, Keita, and Bob Lingard. 2018. Datafication of Schooling in Japan: An Epistemic Critique Through the 'Problem of Japanese Education'. *Journal of Education Policy* 34 (4): 449–469.
- Tatarchenko, Ksenia. 2019. Thinking Algorithmically: From Cold War Computer Science to the Socialist Information Culture. *Historical Studies in the Natural Sciences* 49 (2): 194–225.
- Wilkins, Andrew, and Antonio Olmedo. 2018. Introduction: Conceptualizing Education Governance: Framings, Perspectives and Theories. In *Education Governance and Social Theory*, ed. Andrew Wilkins and Antonio Olmedo, 1–17. London: Bloomsbury.
- Williamson, Ben. 2017. *Big Data in Education: The Digital Future of Learning, Policy and Practice*. London: Sage.
- World Bank. 2004. Project Appraisal Document on a Proposed Loan in the Amount of US\$ 100 million to the Russian Federation for an E-learning Support Project in Support of the First Phase of the Government's Education Modernization Program. Washington, DC: World Bank <http://documents.worldbank.org/curated/en/603841468760158792/text276500RU.txt>.
- Yandex učebnik [Yandex Textbook]. n.d.. <https://education.yandex.ru/>. Accessed 8 Jan 2019.
- Yastrebova, O.K. 2013. O principah formirovaniâ normativov poduševoogo finansirovaniâ obrazovatel'nyh programm vyššego obrazovaniâ [On the Principles of Forming Per Capita Normatives of Funding of Educational Programmes of Higher Education]. Informacionno-metodičeskij seminar 'Realizaciâ obrazovatel'nyh programm vyššego obrazovaniâ' [Information and Methods Seminar on 'The Implementation of Higher Education Programmes']. <https://docplayer.ru/30287522-O-k-yastrebova-konsultant-minobrnauki-rossii.html>. Accessed 28 Dec 2018.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Digitalization of Religion in Russia: Adjusting Preaching to New Formats, Channels and Platforms

*Victor Khroul*

## 11.1 INTRODUCTION

Facing religious life and religious practices that are traditionally conservative or even archaic, the “digital” has not yet transformed the field of religion in Russia as radically and visibly as some other areas, such as business, media, education, or culture. Nevertheless, the analysis of the digital in the religious sphere does not fit into simple statements, such as that religion is ancient, traditional and therefore—“natural,” while media are modern, upgrading and therefore—“artificial”; it is far more complex (Lundby 2014).

Helland (2000) has made an important and heuristically promising distinction between “online religion” and “religion online”: religion online means the adoption of digital formats for conveying traditional religious information (dogmatic texts, worships, preaching, institutional information of all kinds), whereas online religion engages users in spiritual activity via the Internet, and this activity may be not in line with traditional religious practices and sometimes is in open opposition to them. This distinction, when applied to Russian religious life, gives a picture that is overwhelmingly dominated—quantitatively and qualitatively—by religion online, i.e. traditional discourse “repacked” into digital form and distributed through digital channels; online religion is marginal and almost invisible. The Russian Orthodox Church (ROC) more and more effectively uses digital technologies, but still utilizes the Old Slavonic

---

V. Khroul (✉)

Lomonosov Moscow State University, Moscow, Russia



language during liturgies. Muslim and Jewish communities use smartphone apps to calculate the correct time for prayers but pray in Arabic or Hebrew as in ages before. The inner, sacral religious space remains untouched by the “digital.”

Normatively, digitalization as such does not contradict the dogmatic of any traditional religion. In Christianity, Judaism, Islam and Buddhism, it is theologically considered to be a neutral process with good or bad consequences depending on human will. Therefore, functionally digital technologies are seen by religious communities first of all as one more facility (channel, tool, space, network) for effective preaching, or *Propaganda Fidei* (the Propagation of the Faith) (Campbell 2005).

This chapter consists of three basic units. The first discusses religious organizations in Russia. The second analyzes religious digital practices, while the third section examines challenges for digitalization in religious sphere. Starting from a short description of the Russian religious landscape, we analyze normative and practical aspects of digitalization in the context of religion and then examine problematic areas of this process in Russia—the digital remapping of sacred and profane, the marginalization of religious minorities, forms of anti-digital resistance and extremism in the digital space.

## 11.2 RUSSIAN RELIGIOUS LANDSCAPE

The Constitution of the Russian Federation is considered by experts to be liberal and democratic. It provides equal rights: “The state shall guarantee the equality of rights and liberties regardless of sex, race, nationality, language, origin, property or employment status, residence, attitude to religion, convictions, membership of public associations or any other circumstance. Any restrictions of the rights of citizens on social, racial, national, linguistic or religious grounds shall be forbidden”; and also the freedom of religion “Everyone shall be guaranteed the right to freedom of conscience, to freedom of religious worship, including the right to profess, individually or jointly with others, any religion, or to profess no religion, to freely choose, possess and disseminate religious or other beliefs, and to act in conformity with them” (Constitution of the Russian Federation 1991).

The Government generally respects these rights in practice; however, in some cases authorities impose restrictions on certain (religious) groups.

The Russian law on religion (1997) recognized for all citizens the right to freedom of conscience and faith. It underlined the spiritual contribution of Orthodox Christianity to the history of Russia, and respect to Christianity, Islam, Buddhism and Judaism as so-called traditional religions.

When it comes to determining the numbers of followers of these religions, different approaches often give contradictory results. Moreover, the most natural approach, which is based on self-identification data, works well in most Western countries but fails in Russia. In practice, only a minority of citizens actively participate in any religion. Many who identify themselves as members

of a religious group participate in religious life rarely or not at all. There is no single set of reliable statistics about the religiosity of the Russian population.

According to the Pew Research Center, 71% of Russians are Orthodox Christians, 15% are not religious, 10% are Muslim, 2% are Christians of other denominations, and 1% belonged to other religions (Religious Belief 2017). But those who claim themselves to be Orthodox Christians, do not fit any traditional criteria of religiosity, such as church attendance and familiarity with basic dogmas of their faith. Radically different results are obtained by estimating the number of practicing adepts. For example, even though up to 70–80% of the Russian population identify themselves as Russian Orthodox, less than 10% of them attend church services more than once a month and only 2–4% are considered to be integrated into church life. Moreover, the coverage in mainstream media strengthens the ethnic background of the religious identity. According to the Levada-Center, a correlation between “I am Russian” and “I am an Orthodox believer” has become stronger over the last two decades (Obščestvennoe mnenie 2013, 118). Russian sociologist D. Furman suggested that the increase in ideological uncertainty and eclecticism, with beliefs in reincarnation and astrology, ufology, energy vampires, witches, shamans and so on, demonstrates that atheism still dominates in Russia (Furman and Kaarijnen 2006).

The Russian government evidently favors “traditional” religions, and most of all the ROC with budget financing of constructing and restoring church buildings and educational and social projects, which faces critique in the public sphere. For example, human rights activists quote the Russian Constitution and insist that the ROC and other religious organization should be separate from the state. Non-traditional religions, on the other hand, are marginalized, suppressed and even persecuted as sects (for example, Jehovah’s Witnesses).

According to the SOVA Center for Information and Analysis, the trend of increasingly restrictive policies toward Protestants and new religious movements, especially Jehovah’s Witnesses intensified in 2019:

Persecution of Jehovah’s Witnesses has become more large-scale and severe. Criminal prosecution for continuing the activities of an extremist organization, de facto for continuing the profession of religion, has already affected more than 300 people. 18 of them were sentenced, half of them to prison time, including three who received six years in penal colony. This is the first time since the Jehovah’s Witnesses organization was banned that its believers were tortured during criminal investigations. Numerous rough searches and arrests and confiscation of community property continued. (Sibireva 2020)

Experts do not expect any liberalization in government policy as the year 2020 started off with new imprisonment sentences and instances of Muslim communities that suffer as a result of the enforcement of so-called anti-extremism legislation. In addition, religious groups continue to face problems in the construction of new and continued use of existing buildings, risk

criminal prosecution based on the restrictions on missionary activities and are confronted with discrimination.

### 11.3 DIGITALIZATION AND RELIGION: NORMATIVE ASPECTS

The impact of digitalization on religious organizations and practices in Russia is best understood in the framework of mediatization. The notion of mediatization has been applied to religion by Danish scholar Stig Hjarvard (2008). He suggested that in the digital era religion can no longer be studied separately from the media, because (a) media are for most people the primary source of their religious knowledge and religious imagination; (b) some social functions of religion are now primarily the functions of media; and (c) religious institutions use media logic and media framing for their actions (Hjarvard 2008).

There are three main ways of mediatization of religions:

- Media allow, enable and assist the self-presentation of religions, observe their activities in the public interest by maintaining religious formats (broadcasting services, funerals, weddings, etc.)
- Media cover religious life (news reports, feature stories, etc.) and may have a critical approach towards some social activities or religious institutions.
- Media outlets may use religion for their own aims: selectively importing well-known religious symbols into entertainment, keeping out sacral meanings and secularizing the essence of religion. This process is out of the control of religious authorities and therefore causes many complaints and conflicts (Thomas 2015).

The first way of mediatization mentioned above is more or less self-evident and depends on the goodwill of media institutions and on audience demand. In most cases it keeps the religious format “untouched” and the media are used more as a channel of transmission rather than actively interacting with the subject. The second and the third ways presume a more active role of journalists covering religion. The process becomes more important and at the same time more problematic. Conflict and scandals are rooted in misunderstanding or in poor reporting on religious issues.

The historical analysis of religious media in Russia explicitly shows two stages: (a) a rapid development of all religious media (1990–1997) and (b) their stratification after the division of religions in 1997 into so-called traditional (Orthodox, Muslim, Jewish and Buddhist) and non-traditional (Catholic, Protestant, Hindu, new religious movements and others). Orthodox media are supported by the state, on national and regional levels. For example, Orthodox TV channel “Spas” is included into a number of federal channels transmitted all over Russia. Some of “non-traditional” religious media decided to choose the strategy of “self-silencing.”

The situation in Russian news media and public sphere regarding religious issues differs from the situation in traditional Western democracies. The differences are rooted in the understanding of press and religious freedoms. To illustrate: while up to a million French people gathered to express their solidarity with the Charlie Hebdo journalists who were killed in Paris in January 2015 by terrorists who claimed to be Muslims, a few days later 1 million Russian citizens—mostly Muslims and Orthodox Christians—came together on the streets of Grozny (the capital of Chechnya) to show their support for “Islamic values.”

In the Russian context, the mediatization of religion faces (1) ignorance towards ethics and social accountability of digital media practitioners, (2) a normatively disoriented audience with a low level of media literacy and religious practice, and (3) a predominantly secular public sphere with problems in social dialogue processing.

In ethical perspective, the Congress of Russia’s Journalists adopted a Code of Professional Ethics (1994). Journalistic standards listed in the Code are similar to those adopted by journalists worldwide. However, its norms are hardly applied or respected by the majority of journalists.

TV remains the most important medium, and it does not appear that it will lose its prominence in the near future. Russia has become a “watching nation” instead of a “reading nation,” therefore for any actor seeking to have an impact on the general audience TV remains a strategic resource. Yet, contrary to European “success stories,” the history of the attempts to create Public TV in Russia and implement it into the existing media system in the last two decades has been marked by a series of failures.

The lack of journalistic self-reflection, the low level of media’s comprehension of their social mission and the ignorance concerning possible consequences sometimes led external structures (political, economic, social) to raise their warning voices. For example, the State Duma (Russian Parliament) on January 23, 2015, called upon all journalists for more accurate and professional coverage of religious life in Russia and abroad. “The State Duma calls on all media and all journalists in Russia and foreign countries in covering events of a religious nature to be guided by the principles of ‘do no harm,’ to refer to the publication of materials that may affect and offend the religious feelings of citizens with special responsibility and sensitivity,” the Duma statement says (Gosduma 2015).

The main dysfunctions in the coverage of religious life in Russia have been confirmed by different researchers (Kashinskaja et al. 2002; Khroul 2012):

- a biased approach among journalists, tolerated by their colleagues;
- a lack of education on religious issues and therefore a lack of understanding of what is really going on;
- an urgent need of specialized media focused on religious life;
- secular media’s dependence on political and influential Russian Orthodox Church elites;
- the marginalization of religious minorities in the public sphere.

**Table 11.1** Religions and digital media normative expectations

	<i>Religions</i>	<i>Digital media</i>
Pluralism	<ul style="list-style-type: none"> <li>• Try to ensure religious values transparency, availability of texts representing them clearly;</li> <li>• Seek correct articulation of their faith, use adequate symbolic systems, language and cultural codes.</li> </ul>	<ul style="list-style-type: none"> <li>• Give platforms for complete spectrum of religions and normative models (with respect to minorities);</li> <li>• Optimize channels and information flows.</li> </ul>
Dialogue	<ul style="list-style-type: none"> <li>• Tolerate other approaches to religion with which they are not in agreement;</li> <li>• Use the framework of common cultural code;</li> <li>• Commit themselves to participate in the dialogue, send experts to be active in the public sphere.</li> </ul>	<ul style="list-style-type: none"> <li>• Organize and support the search for new subjects of the dialogue;</li> <li>• Mediate, moderate, create forums for discussions;</li> <li>• Expand—quantitatively and qualitatively—the space for dialogue in various forms of communication.</li> </ul>
Consensus	<ul style="list-style-type: none"> <li>• Are seeking the common good;</li> <li>• Are optimizing the “preaching,” the presentation of their vision from the perspective of consensus.</li> </ul>	<ul style="list-style-type: none"> <li>• Consider consensus to be one of the most important goals of media;</li> <li>• Are peacemakers during conflicts and tensions;</li> <li>• Develop openness and solidarity.</li> </ul>

From a religious perspective, the lack of knowledge about and experience of religious life among digital media practitioners gives much more space for myths and stereotypes in digital platforms. Moreover, not only the mass media but also religions themselves have to contribute to agenda setting and to elaboration of digital mediatization mechanisms in this very sensitive sphere. In addition to difficulties of translation from the archaic language of the religious ghetto into a modern one and problems with understanding the internal functionality of religious organizations, there are some social expectations religions do not meet.

At least two problematic areas in Russian society—“religious illiteracy” of journalists and “media illiteracy” among faith communities—could be optimized with the clarification of mutual expectations from the perspective of “pluralism—dialogue—consensus” logic (Habermas 1989) (see Table 11.1).

#### 11.4 RELIGIOUS RESPONSES TO THE CHALLENGE OF DIGITALIZATION

Digitalization of religion is even more complex in Russia because of its poly-confessional and poly-ethnic social structure. The set of values promoted by ROC is questioned by many Russians. Yet, the ROC remains one of the most highly trusted social institutions and some anti-ROC campaigns and scandals (“Pussy Riot” punk prayer in Moscow Cathedral and others) have not significantly decreased the trust in the ROC. Experts agree that, “a common trope

for self-positioning of the Church is that the ROC is a ‘state-shaping’ religion, and as such it weaves its own historical narrative with the narrative of the Russian state” (Suslov et al. 2015). Researchers emphasize the political and geopolitical components of Russian Orthodoxy and the importance of the concept of “symphony”—harmonious relations of mutual support and mutual non-interference—between Church and state (Engström 2014; Papkova 2011; Simons and Westerlund 2015).

In order to make ROC more active in the digital space, Patriarch Kirill after his election and enthronization in 2009 announced the establishment of a new *Sinodal’nyj informacionnyj otdel* (Synodal Department of Information). In 2010, an Orthodox video channel on YouTube (<http://www.youtube.com/user/russianchurch>) was launched, and the Department of religious journalism and public relations at Russian Orthodox University was established.

Not all of more than 1000 Orthodox media outlets (most of them have digital versions) are in line with the ROC position, and some of them have a different approach in commenting on everyday life. Some non-official outlets, like the magazine *Tat’ânin Den’* and journal *Foma*—both founded in 1995—are not official and enjoy a larger degree of freedom of discussions than what is allowed at the official resources. Web portal “*Pravoslavie i mir*” (Orthodox Christianity and the World, [www.pravmir.ru](http://www.pravmir.ru)), launched in 2004, is currently the leading Orthodox multimedia portal publishing news and analytical reviews, comments and interviews, audio, video, info graphics. The audience of the portal is around 2.5–3 million visitors per month, or 100–120 thousand per day.

According to Anna Danilova, the Editor-in-Chief of Pravmir.ru, there are several essential negative presuppositions in Orthodox religious identity that affect the missionary work within digital media. “Still for a religious community the process of exploring new media normally is connected with at least these potential obstacles: (1) tendency of any religious institution to be conservative in everything including the media; (2) unclear impact of the new media on the psychological state, society and interpersonal relationships; (3) tendency to interpret many innovation as ‘diabolic ones’ (one of the best cases of which was shown in the fear of many people in Russia to accept personal tax identification code, even though the Church has officially stated that it had nothing to do with the number of the Antichrist),” writes the Orthodox journalist (Danilova 2011, 20).

Chief editor of the portal “Bogoslov.ru”, archpriest and theologian Pavel Velikanov, mentioned three pros for digital activity of the Church: (1) the possibility of Christian witnessing, the ability to communicate with people looking for answers to their questions in social networks; (2) the possibility of Christian charity—according to the priest, “charitable organizations are active in networks and live through networks,” and (3) the rapid dissemination of information. Contras, according to the theologian, are the reverse side of the pros: (1) it is very difficult to verify information; it often comes from not-trustworthy and strange sources; (2) discussions are conducted in a manner that is not appropriate for Christians; (3) people spend a lot of time on the social networks and come into the real world “just to eat” (Khroul 2015; quotations below see

ibid.). Danilova considered as positive the fact that social networks make it possible to get out of the “ghetto” of just the Orthodox audience and to understand the agenda, to find out what people are now interested in. A negative point is the lack of information accuracy and difficulties with verification: “fakes” rapidly spread through social networks. On the negative side Danilova also mentioned the fact that social networking presumes too quick a reaction: “People react while they still do not really understand the situation, and relationships become strained,” Danilova said and called for general “Internet hygiene.”

Well-known Russian Orthodox journalist Sergej Hudiev suggested that it is difficult to divide the “plusses” and “minusses,” because most of the advantages are at the same time disadvantages. The advantage of anonymity is that many people are able to overcome the exclusion zone between them and the clergy, but the disadvantage is that the question of anonymity removes inhibitions of the people in the network: they cease to control what they say.

Russian TV commentator Elena Žosul, speaking about the advantages, noted that social networks are main sources of news; they allow to establish useful contacts and professional relationships and allow quick collective reflection about what is happening. On the negative side, she mentioned “the overflow of information and inability to concentrate on some issue, therefore long texts are so unpopular in the network.”

In order to prevent cybercrimes and the use of the digital space for pedophilia, pro-Orthodox organization “*Liga bezopasnogo interneta*” (League for a Safe Internet) was established in 2011 with support from the Ministry of Communication of the Russian Federation. “This organization set itself the task of fighting pedophilia and extremism on the internet, mostly by hands of the so called ‘cyber-warriors’ [*kiberdružinniki*], who provoke and expose pedophiles, and report about contentious websites to the law-enforcement bodies,” underlines Russian scholar Mihail Suslov (2015, 13).

The ROC has a leading position among religious communities involved in online communication; Muslim activity is not as expanded. The biggest and most influential Muslim digital resource in Russia is the Internet portal Islam.ru, whose main goal is to protect the interests of traditional Muslims, as well as popularize the works of traditional Islamic values. It launched the first daily Islamic news feed and opened 13 thematic sections along with a full-fledged English version of the site. Beside news, Islam.ru publishes analytical articles, religious texts (in particular, prayers) and provides psychological, legal and theological advisory. The resource has pages on all popular social networks through which feedback from readers is maintained. Islam.ru opened the possibility to become a member of the Muslim community virtually. “People become Muslims because of their convictions and sincere faith. On the site, they can leave their data in order to inform the world about their decision,” said the chief editor of the Islam.ru Rinat Muhamedov (Luchenko 2008). There is a button “I accept Islam” on the Islam.ru website; pressing it is equal to publicly pronouncing the formula “There is no God but Allah, and



Mohammed is His prophet.” In addition to Islam.ru there are some independent Muslim socio-political channels, such as “Voice of Islam,” “Russian Islamist”, as well as educational projects.

Jewish, Catholic and Protestant digital resources are focused mostly *ad intra*, serving local communities and those who show some interest in them. Together with other non-traditional religious media and networks, they are marginal and less visible in the Russian public sphere in comparison to the dominant Orthodox and Muslim religious communities.

The only major television project for Russian Protestants is “Television of Good News,” which began as part of the global Trinity Broadcasting Network (TBN) and now is positioning itself as an independent public broadcaster. Without any doubt, this is the biggest Protestant media resource that broadcasts via satellites and cable networks. Protestant radio “Teos” lost its frequency and is now a fully Internet-based station. Nevertheless, it is developing, inviting interesting presenters, such as Orthodox journalist Sergej Hudiev and a number of others, trying to be interesting and relevant to a wide range of audiences, not only for Protestants. Newspaper “Mirt” is a serious newspaper for ministers and parishioners, publishing reflections and sermons, sometimes not understandable to non-Protestants. There are also a number of successful printed media outlets outside Moscow and Saint Petersburg: newspapers in Yaroslavl, Penza, Yoshkar-Ola, Voronezh, Vladivostok, Irkutsk, and other cities of Russia. Among the Internet portals the leading project is Protestant.ru that presents a good example of successful migration from a printed newspaper to web portal. The press secretary of the Union of Christians of Evangelical Faith (Pentecostals) in Russia Anton Kruglikov pointed out two major visible trends in Protestant media: (1) to move content from printed media to digital platforms and (2) to address the general public, not only those who already are Protestants.

Generally speaking, there are several problematic areas in religious digital media:

1. *Subordination of journalism to public relations (PR)*. Many of the employees of religious media in Russia find themselves serving the religious institutions in terms of public relations and advertising much more than following journalistic standards. Both the employers and the employees do not find such a situation strange.
2. *Out of touch with mission and target audience*. Digital religious media fall into the trap of thinking that their structure would be “media for all,” but in reality, they find themselves with an unclear mission and target audience.
3. *Populism and primitivism*. In order to be closer to common people, digital religious media sometimes pursue populism through primitivism of the message. Such a simplification creates a distorted image of the religious reality and also “corrupts” the religious view of cultural and social issues in Russia.

4. *Conflict of formats*. Digital religious media lack a language that is clear and understandable for the general public. In many cases, because of the language secular journalists have the impression that religious world is strange and hard to cover, therefore it is underexposed and finds itself *ad marginem* of the national media system.
5. *Religious media as the "ghetto"*. Religious media still do not realize the need to be part of social dialogue. Meanwhile, media and digital culture is increasingly becoming a space of public life and cognition.
6. *Lack of professionalism is not understood as a problem*. The lack or total absence of professionalism in religious media often is not considered to be something inappropriate.
7. *Religious media are still run mostly by enthusiasts*. In many cases the editorial staff's enthusiasm does not receive any moral (and more material) support and understanding from the hierarchy of religious organizations, and that makes synergetic strategic planning and systematic work hardly possible.

So, from a religious perspective there are evident problems with news production, channeling, transmitting, broadcasting, with interaction and understanding; therefore, the voices of religious leaders are hardly heard in society (for more on digital journalism beyond religion, see Chap. 9).

### 11.5 SACRED AND PROFANE: DIGITAL REMAPPING

In the Russian digital sphere, there are two major contextual challenges for Durkheim's *sacred-profane* dichotomy (Durkheim 1915, 47): the enforced atheization during the Communist time and, after it, the religious revival in the context of secularization. Digitalization speeds up the remapping of the social space with sacred and profane markers: some profane objects and social practices have been sacralized, while some traditional religious ceremonies and sacred objects have been profanized. Digitalization can also lead to re-sacralization, to the creation of new sacred objects, new mysteries, and new explanations for events of supernatural origin.

The last two decades of the digital era have been a time of continuous sacred-profane remapping in Russia. Russian feminist punk rock group "Pussy Riot" staged a performance in Moscow's Cathedral of Christ the Savior in February 2012, which was stopped by church security guards. Online video sharing was essential for Pussy Riot's performance to reach an audience and create the scandal it created. Six months after, three members of Pussy Riot were convicted of hooliganism motivated by religious hatred and sentenced to two years imprisonment. Different ecclesiastics reactions followed the "punk-prayer" by Pussy Riot. Archpriest Vsevolod Chaplin appealed to "criminal sanctions for everyone, who affronts the faithful sense," while at the same time deacon Andrei Kuraev commented on the event on his LiveJournal in the opposite way: "If I were a sacristan of the Cathedral I would feed them with

pancakes, give a cup of mead to each of them and invite them to come round for a confession. And if I were an old layman, I would pinch them a bit at parting ... Just to make wise” (Kuraev 2012).

The more recent debate on “Matilda,” a film directed by the Russian filmmaker Aleksei Uchitel, which tells the story of a romance between the future Tsar Nicholas II, canonized by the Russian Orthodox church in 2000, and Matilda Kshesinskaya, a teenage prima ballerina at the Mariinsky theatre in St. Petersburg, is a good example of the “sacralization” trend in the Russian public sphere and how it is supported by media. Radical Russian Orthodox movements warned that “cinemas will burn” if Matilda was screened, because the film portrays the “holy tsar” in love scenes. In response to the threats, the largest network of cinemas in Russia in September 2017 refused to screen the film because of safety reasons. Various other spontaneous, grass-roots public initiatives in Russia (e.g. icons of Stalin painted with the nimbus as a saint, protests against digitalization in order to avoid the “number of devil” appearing in the documents) are not in line either with Church teaching or with government intentions, but widely covered by media, inspiring the sacralization of, for example Stalin or Ivan IV Terrible.

Another example—heavily rooted in digital media support—is the process of “sacralization” of Epiphany bathing (ice swimming). Ice swimming has been practiced in Russia for centuries and some historians suggest that the practice was a popular pagan tradition. Every year on Epiphany (January 19 in Russia), Russian Orthodox believers are plunged into a blessed section of frozen water three times in remembrance of Jesus’ baptism in the river Jordan by John the Baptist. In 2019, almost 460 thousand people took part in the Epiphany bath in Moscow, and over 2.4 million in Russia (for comparison—in 2018: 150 thousand in Moscow and over 1.8 million in the entire country). Russian President Vladimir Putin traditionally, year-by-year, attends a religious service and also participates in Epiphany bathing. Even the US ambassador to the Russian Federation John Huntsman, a Mormon by faith, took part in Epiphany bathing in 2018 and called this ritual “the great Russian tradition.” The Moscow authorities published on the Mayor’s website the “rules of baptismal bathing,” which did not contain a word about the religious character of the act. And the mayor of the city of Yaroslavl, with the words “you are Orthodox people”, convincingly asked the officials to lead the bathing. Generally speaking, Epiphany bathing has become a huge media event covered by all the major media in Russia and abroad—covered as religious tradition, as something all Russian Orthodox Christians are called to do, as a ritual blessed by the Church.

In fact, many Russian Orthodox bishops and priest condemned this ritual and called on believers not to take part in it and invited them to attend Epiphany liturgy instead. Bishop Evtikhy of Domodedovo put forward four reasons for this: (1) ice swimming is dangerous for the health, it contradicts the Gospel and therefore it is a sin; (2) bathing is a profanation of the sacred—blessed water; (3) bathing is not traditional for the Russian Orthodox Church and (4) it strengthens not faith, but superstitions (Evtikhy (Kurochkin) 2019). Such a

negative approach to Epiphany bathing was evident in previous centuries. “Bathing violates the sanctity and contradicts to the spirit of true Christianity; therefore, it cannot be tolerated and must be condemned,” wrote priest Sergij Bulgakov in the end of nineteenth century (Bulgakov 1913).

This opinion is low profiled both by media and state authorities and therefore not heard in the public sphere. Both media and politicians gain symbolic capital during Epiphany bathing ignoring the position of priests and bishops who have never been in fact proclaimed loudly “*ex cathedra*,” and therefore the ROC’s ecclesial approach to Epiphany bathing is not clear and understandable for the general public in Russia.

As Kseniya Luchenko mentioned, high-quality Church-related discussions are conducted not in mainstream media, but predominantly in digital social networks. “The answer to that question is closely linked to the analysis of dialogue culture in Russian society as a whole. Social institutions and mechanisms that are supposed to ensure and sustain that dialogue are overwhelmingly out of order. However, the need to discuss, share experiences and monitor publications is still there. And social networks make it possible,” the Russian scholar suggested (Luchenko 2015, 130). Almost all of the largest Orthodox websites have pages on social networks, such as VKontakte, Odnoklassniki and Facebook. On these social networks there are special pages of ecclesiastics, groups connected to parishes, with Orthodox public associations or churches.

The analysis of the self-expressions and discussions on religious topics in the digital platforms shows that young Russians, in matters of belief/disbelief, rely mainly on their own experience and the experience of other people (family and friends), and not on faith, authority or tradition, as would be expected (Khroul 2015). The most convincing is the socio-historical explanation for this phenomenon: the Russian tradition of faith that was consistently eradicated over a fairly long period of time. Minimizing appeals to faith, tradition and authority is a “birthmark” of Russian history, which can be described in terms of “post-atheism trauma.”

Paradoxically, the Internet users in their self-expression make evident their mostly positive attitudes towards God and predominantly negative attitudes towards Orthodox Christianity and Russian Orthodox Church. The social and political activity of the ROC faces more criticism than Orthodox Christianity as a religion: for example, “ROC proposal to impose a dress code for the people of Russia,” “ROC proposes to create a criminal penalty for heresy.” This suggestion may be proven not only quantitatively but also qualitatively, with the rhetoric of users’ voices: “ROC is a business project”; “ROC, in most cases do not care about people, but about the godless government,” “I love the Orthodox religion and Orthodox culture, myself, am an Orthodox man, but terribly hate ROC.” The arguments of those who are in favor of ROC and defend it are mostly rooted in ethnic and geopolitical discourse: “I am Russian and therefore I am an Orthodox. It is natural”; “ROC is an integral part of the thousand-year history of Russia, she has always supported our morals and I will always be with her, as the rest of the true believers.”

In 2012, a content analysis study of Russian digital Internet communication texts found observable “traces” of mainstream media publications (predominantly TV) against so-called non-traditional religious organizations (Khroul 2016). Consider, for example, some opinions on Jehovah Witnesses’ (JW) activities published on the website *lovehate.ru*: “According to news shows, journalists covered how some sect engaged in raping children”; “Recently in the news on TV it was said that a 50-year-old man, a Jehovah’s Witness, set himself on fire. He considered himself a great sinner who had allegedly had to wash away his sins. Thus, we see what this sect leads us to”; “This is a false religion, which is no good and kills a person (religiously destructive sect)”; “This is the most vile of sects, posing as Christianity. In fact, what we have is a simple case of Freemasons.” The analysis of the texts makes visible two important things: (1) behavioral attitudes of intolerance with respect to the JW, and (2) the willingness of people to take tough repressive measures against JW from the state. In sum, this “explosive mixture” is already provoking a request to the authorities, as in the case of aggravating state–religious relations or the case for a need to find another “enemy”. It can become a “trigger” for negative measures taken not only against the JW but also against other so-called non-traditional religions, who at the current juncture come across as an easy target. Indeed, JW were banned in Russia in April 2017 by the decision on the Supreme Court, and in February 2019 the Russian court for the first time found a Jehovah’s Witness, Danish national Dennis Christensen, guilty of extremism and sentenced him to 6 years behind bars (Russian Court 2019).

From a journalistic perspective, there is a visible problem of journalistic autonomy. According to recent studies, journalists in Russia do not enjoy autonomy because of their political and economic dependence. Secondly, the challenge of objectivity is apparent, which leads to a poor and stereotyped coverage of religious life in secular media. Agenda-setting process in media is not ethical-oriented: the main players are mostly focused not on the audience or on public interest, but on political subordination and commercial profit, therefore moral issues are secondary. Therefore, religious media are not able to change the content management: “infotainment” and “advertainment” oriented media decision makers do not seem to be concerned with fitting their products into even secular moral norms, so religious norms as more strict are ever more ignored.

## 11.6 CHALLENGES OF DIGITALIZATION IN RELIGIOUS PERSPECTIVE

For religions in Russia, all visible and invisible challenges and threats of digital communication—non-hierarchical structure, lack of authority, dogmatic corruption, information noise, fake news dissemination—seemed to be not so dangerous in comparison with their advantages and benefits and therefore manageable. Therefore, the concerns of Russian religious leaders with regard

to digital technologies are mostly (with some rare exceptions) focused on misuses of them in particular cases (the spread of heresies, online pornography, gaming addiction, playing Pokémon Go in church, etc.).

Nevertheless, there are visible “grassroots” protests among Russian Orthodox fundamentalists against digitalization in general. According to these fundamentalists, digitalization in the context of religion is not limited to its technological side, as it is always a threat. Moreover, for some ultraconservative Russian Orthodox Christians the “digital” as such has ontologically negative connotations related to “the number of the beast” and the process of the digitalization is seen as a visible sign of the Apocalypse, the end of the world.

Therefore, digitalization was accompanied with protests against, for example, the “barcode” or “666” digits in the passport numbers of some Orthodox believers. Paradoxically, the campaign against individual tax numbers (INN) in 2000 became the first civil action of a religious nature in Russia, in which the Internet was used as a tool of influence, the main mean of information exchange. Individual tax number opponents using digital platforms and channels brought this topic onto the agenda of mainstream media and of church–state relations (Luchenko 2008). The movements against electronic control and globalization processes are widely using one of the main tools of globalization—the Internet. While widely rumored, these views still are marginal in Russian media and public sphere.

Various semi-pagan cults and self-proclaimed “prophets,” who previously were not known beyond the regions of their activity, nowadays cover the entire territory of the country, thanks to digital network channels. In 2008, the case of the so-called Penza hermits—a group of believers who reject the foundations of modern society and the state and spent more than half a year, having closed themselves in a dugout in the Penza region, became widely known (New Cult 2008). The spread of myths about the “sanctity” of Ivan the Terrible, Grigori Rasputin, Russian Emperor Pavel I, and voices demanding their canonization by the ROC would not be so successful without digital networks. Moreover, some informal groups that hold completely different views and have completely different goals can act in the digital space on behalf of the Orthodox or Muslim, Jewish, Catholic, Protestant communities. The general shift of these movements toward greater radicalism seems to be consequent; since the center of social and political discussion in the digital world is shifting towards oppositional radical structures, it is easier to act on the Internet, exaggerating their ideology.

Yet, the biggest concern in terms of social security in the digital space is raised by radical extremist networking. After Twitter closed more than 300 thousand accounts on suspicion of spreading extremist ideology in 2015, the followers of the so-called Islamic state (IS) became more embittered on Telegram messenger (total number of users exceeded 100 million). Telegram officials informed that they suppressed activities related to extremism in public channels, but do not monitor private chats (encrypted and secret). On November 18, 2015, Telegram announced the blocking of 78 public channels,

connected with the IS extremist group (banned in Russia). Fundamentalists used 12 languages for digital extremist propaganda. After that case, the FSB (*Federal'náá služba bezopasnosti*, Federal Security Service) head Aleksandr Bortnikov considered the possibility of restricting Russians' access to Telegram (RBC 2015), but this initiative was not implemented at that time.

At the same time religious organizations use various digital channels of mass communication with missionary goals, as well as to maintain the integrity of the religious community and its development, to ensure the necessary information exchange in modern conditions. For the ROC, one of the main functions of the Internet is an electronic document management system that allows its structures and administrative units to more effectively coordinate their activities.

Despite the use of tablets and smartphones in order to follow the worships and using digital TV for live transmissions of religious events (some of which also became media events), and despite being involved more actively in web-based content production and consumption, for many Russians the core of religious practices still remains based on interpersonal communication.

## 11.7 CONCLUSION

In spring 2020, the reactions of Russia's various religions communities to the coronavirus pandemic were noticeably different, once more confirming the diversity of practices sketched in this chapter. While most places of worship were closed or switched to online services, some bishops in the Russian Orthodox Church insisted they would not stop in-person services or the tradition of kissing icons. Another traditional ritual in times of emergency took place in Moscow on 3 April: Patriarch Kirill took a miraculous icon of Maria, Mother of God, and made a round trip through Moscow, praying to save the city from the coronavirus.

Digitalization had a tremendous impact on religions practices during the pandemic as believers got a chance to participate in worships digitally at a distance. For example, Catholic masses all over Russia were broadcast online. Moreover, in the opinion of the Russian Orthodox Church, even the sacrament of confession became possible online. If a person wants to confess during self-isolation because of the coronavirus, then "in exceptional circumstances they can confess by phone or Skype," said Metropolitan Hilarion, the head of the ROC Synodal Department for External Church Relations (RIA Novosti 2020).

The use of religious apps has brought about a diverse range of religious practices (e.g. confession by smartphone) that often fall outside traditional thinking, yet the rituals performed with these apps are felt to be authentic (Scott 2016). The digital network structure also frees users from the need to integrate into strict hierarchical systems and rigorously participate in rituals—that is, from important elements of institutionalized religions. In the wake of the turn from religiosity to spirituality, user practices have become increasingly diverse, sometimes deviating from church (in the case of Christianity)



doctrines. Moreover, the individualization of religious practices leads to a situation in which church authorities lose their status as the final ethical and dogmatic referee.

Opening new channels and platforms for information flows, the digital era created opportunities and challenges both for religious institutions (new formats, genres, packages for preaching and communications) and for individual religiosity (variety of information sources, shift from interpersonal to digitally mediated communication). As this chapter has shown, digital technologies as a shaping force make religious life more transparent (challenging hierarchical information filters and church secrets), more liquid (after centuries of stability), and more ambivalent and pluralistic in terms of values and practices.

## REFERENCES

- Bulgakov, S. 1913. *Nastol'náá kniga dlá svášenno-cerkovno-služítelej: sbornik svedenij, kasaišsibá preimúšestvenno praktičeskoj deátel'nosti otčestvennogo duhovenstva* [Handbook for Priests: A Collection of Information Relating Mainly to the Practical Activities of the National Clergy]. Kiev: tipografiá Kievsko-Pečorskoj Uspenskoj lavry.
- Campbell, H. 2005. Spiritualizing the Internet: Uncovering Discourses and Narratives of Religious Internet Use. *Online-Heidelberg Journal of Religions on the Internet* 1 (1). <http://archiv.ub.uni-heidelberg.de/volltextserver/volltexte/2005/5824/pdf/Campbell4a.pdf>.
- Constitution of the Russian Federation. 1991. <http://www.constitution.ru/en/10003000-01.htm>.
- Danilova, A. 2011. The Russian Orthodox Church and the New Media // Religion and New Media in the Age of Convergence. Moscow: MSU, Journalism Faculty.
- Durkheim, E. 1915. *The Elementary Forms of the Religious Life*. Translated from French by Joseph Ward Swain. London: George Allen and Unwin Ltd.
- Engström, M. 2014. Contemporary Russian Messianism and New Russian Foreign Policy. *Contemporary Security Policy* 35 (3): 356–379.
- Evtikhy (Kurochkin). 2019. Počemu á protiv krešenskikh kupanij [Why I Am against Epiphany Bathing]. Accessed March 18, 2019. <https://www.pravmir.ru/pochemu-ya-protiv-kreshenskikh-kupanij/>.
- Furman, D., and K. Kaariainen. 2006. *Religioznost' v Rossii v 90-e gody XX – načala XXI veka* [Religiosity in Russia in 1990s and 2000s]. Moscow: OGNI TD.
- Gosduma. 2015. The State Duma of the Russian Federation Called on Journalists Not to Incite Religious Hatred. RIA Novosti, January 23. <https://ria.ru/20150123/1043989903.html>.
- Habermas, J. 1989. The Public Sphere: An Encyclopedia Article. In *Critical Theory and Society. A Reader*, ed. Stephen E. Bronner and Douglas Kellner, 136–142. New York: Routledge.
- Holland, C. 2000. Online-religion/Religion-online and Virtual Communitas. In *Religion on the Internet: Research Prospects and Promises*, ed. J.K. Hadden and D.E. Cowan, 205–233. New York: JAI Press.
- Hjarvard, S. 2008. The Mediatization of Religion: A Theory of the Media as Agents of Religious Change. *Northern Lights* 6 (1): 9–26.

- Kashinskaja, L., M. Lukina, and L. Resnianskaja, eds. 2002. *Religiá v informacionnom pole rossijskih SMI*. Moscow: MGU.
- Khroul, V. 2012. *Religion and Media in Russia: Functional and Ethical Perspectives*. Lambert Academic Publishing.
- . 2015. The Religious Identity of Russian Internet Users: Attitudes Towards God and Russian Orthodox Church. *Digital Icons: Studies in Russian, Eurasian and Central European New Media* 14: 133–143.
- . 2016. Hate Speech in the Internet Communication on Jehovah's Witnesses in Russia. *Kultura – Media – Teologia* 25: 9–18.
- Kuraev, A. 2012. Maslenica v Hrame Hrista Spasitelá [Maslenitsa in Christ the Savior Cathedral]. Accessed March 18, 2019. <http://diak-kuraev.livejournal.com/285875.html>.
- Luchenko, K. 2008. Internet i religioznye komunikatsii v Rossii [Internet and Religious Communications in Russia]. *Mediascope* 1. Accessed March 18, 2019. <http://mediascope.ru/node/32/>.
- . 2015. Orthodox Online Media on Runet: History of Development and Current State of Affairs. *Digital Icons: Studies in Russian, Eurasian and Central European New Media* 14 (2015): 123–132.
- Lundby, K., ed. 2014. *Mediatization of Communication*. Handbook of Communication Sciences, vol. 21. Berlin: Walter de Gruyter.
- New Cult Uncovered in Penza. 2008. *Russia Today*, July 31. Accessed March 18, 2019. <https://www.rt.com/news/new-cult-uncovered-in-penza/>.
- Obščestvennoe mnenie. 2013. Ežegodnik Levada-centra [Public Opinion—2013. Annual Book of Levada-Center]. <http://www.levada.ru/books/obshchestvennoe-mnenie-2013>.
- Papkova, I. 2011. *The Orthodox Church and Russian Politics*. New York: Oxford University Press.
- RBC. 2015. V Telegram naščitali okolo 40 kanalov “Islamskogo gosudarstva” [Around 40 Islamic State Channels were Counted on Telegram], November 19. [https://www.rbc.ru/technology\\_and\\_media/19/11/2015/564da7ab9a7947d3bd5cea31](https://www.rbc.ru/technology_and_media/19/11/2015/564da7ab9a7947d3bd5cea31).
- Religious Belief and National Belonging in Central and Eastern Europe. 2017. Pew Research Center. May 10. Accessed August 1, 2019. <https://www.pewforum.org/2017/05/10/religious-belief-and-national-belonging-in-central-and-eastern-europe/>
- RIA Novosti. 2020. Arhieriej: v period samoizoláčii ispovedovat' sá možno i po telefonu [Bishop: During Self-isolation One Can Confess Also by Telephone]. April 2. <https://ria.ru/20200402/1569513724.html>.
- Russian Court Convicts Danish Jehovah's Witness to 6 Years for “Extremism”. 2019. *Moscow Times*, February 6. Accessed March 18, 2019. <https://www.themoscow-times.com/2019/02/06/russia-convicts-danish-jehovahs-witness-6-years-for-extremism-reports-a64417>.
- Scott, S. 2016. Algorithmic Absolution—The Case of Catholic Confessional Apps. *Online—Heidelberg Journal of Religions on the Interne* 11. <https://heiup.uni-heidelberg.de/journals/index.php/religions/article/view/23634/17360>.
- Sibireva, O. 2020. Freedom of Conscience in Russia: Restrictions and Challenges in 2019. SOVA Center for Information and Analysis, March 19, 2020. [https://www.sova-center.ru/en/religion/publications/2020/03/d42209/#\\_ftn4](https://www.sova-center.ru/en/religion/publications/2020/03/d42209/#_ftn4).
- Simons, G., and D. Westerlund. 2015. *Religion, Politics and Nation-building in Post-communist Countries*. Farnham: Ashgate Publishing.

- Suslov, M. 2015. The Medium for Demonic Energies: ‘Digital Anxiety’ in the Russian Orthodox Church. *Digital Icons: Studies in Russian, Eurasian and Central European New Media* 14 (2015): 2–25.
- Suslov, M., M. Engström, and G. Simons. 2015. Digital Orthodoxy: Mediating Post-secularity in Russia. *Digital Icons: Studies in Russian, Eurasian and Central European New Media* 14 (2015): i–xi.
- Thomas, G. 2015. The Fluidity of Religious Forms and Their Attractiveness for Audiovisual Communicative Media. In *Mediatization of Religion: Historical and Functional Perspectives*. Media and Religion Book Series 3, ed. V. Khroul, Lomonosov Moscow State University 9–29.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Doing Gender Online: Digital Spaces for Identity Politics

*Olga Andreevskikh and Marianna Muravyeva*

## 12.1 INTRODUCTION

In contemporary Russia, online discourses on gender reflect the complex legacies of the Soviet and post-Soviet attitudes and approaches to masculinity and femininity. These complexities are defined by the seemingly contradictory combination of Russia's cultural matrifocality (i.e. reliance on women to run households in the absence or less significant presence of men in family life) and patriarchal social order (Kon 1995). They have also been affected by the new gender identities which evolved during the temporary liberation of Russian society in the 1990s. The appearance of the concept of “sexual freedom” in the post-Soviet Russia, as well as the critical rethinking of the Soviet gender roles—the “emasculated” men (Kay 2006) and the desexualized “masculinized” women under the “double burden” (Stella and Nartova 2015, 37)—led to the emergence of new gender contracts. These included the “housewife” and the “sponsored contract”—a type of relationship between wealthy men and women where the former sponsor the latter by paying their bills and offering gifts in return for sexual and romantic encounters (Zdravomyslova and Tyomkina 2007; Pilkington 1996; Stella 2015), as well as the new aesthetics and ideology of “*glamour*” (“glamor”) (Goscilo and Strukov 2011). The emergence of grass-roots feminist and LGBTQ (lesbian, gay, bisexual, transgender, queer) rights

---

O. Andreevskikh  
University of Leeds, Leeds, UK

M. Muravyeva (✉)  
University of Helsinki, Helsinki, Finland  
e-mail: [marianna.muravyeva@helsinki.fi](mailto:marianna.muravyeva@helsinki.fi)

movements led to the rise of the visibility of new types of masculinity and femininity in public discourses—that is, queer, gay, lesbian and transgender identities. The shift from the command to mixed economy and the overall democratization of the public sphere, in general, also led to the increase in women’s involvement in various forms of civic activism, which in Russia tends to be historically associated with maternal care (Salmenniemi 2008).

While Russian women explored new opportunities and fought the challenges of the new capitalist society, Russian men appeared to be even deeper impacted by the radical social shifts and especially by the economic and political turmoil of the 1990s. The dramatic changes in Russian masculinities are rooted in the Soviet gender order which consisted in men being deprived of the patriarchal status in the family by the state patriarch. In the post-Soviet times, the paradox of masculinity (Kaganovsky 2008, 4) was complicated by men losing their positions on the economic and political arenas to women, as well as by the rise of nationalism and militarism in socio-political life (Yusupova 2018; Sremac and Ganzevoort 2015). Russian men faced a new crisis of masculinity, this time being deprived of their professional dignity and achievements (Goscilo and Hashamova 2010; Kay 2006).

The current discourses on gender, despite the post-Soviet socioeconomic changes, continue to maintain the patriarchal matrifocal dichotomy, which in its turn affects the digital construction of gender. The Internet is seen as a predominantly male activity (Huppertz 2012), monopolized by men as part of gendered masculine capital (Bourdieu 2001), that is, a patriarchal digital space. Early internet scholars while explaining the relative absence of women online pointed to how the World Wide Web (WWW) was constituted dominantly as a “white male playground” (Green and Adam 2001). They made evident how men took over discussions online, even when they were directly related to women and their gendered experiences. Other scholars often hailed “cyberspace” as an arena where individuals could escape social shackles of their biological gender. In their vision, digital technologies facilitated bodily transcendence, catalyzed new ways of engaging in gender politics and provided new contexts whereby individuals could reconstruct their identity free from bodily stereotypes (Castells 2010; Plant 2000). Contemporary researchers take this discussion to a different level by looking at the Internet and related digital technologies (such as social networks and online platforms) as material actors that perform important tasks within dynamic settings, that is, a form of digital work that creates, maintains and transforms human institutions alongside new information technologies (IT) uses (Arvidsson and Foka 2015). These approaches are particularly relevant to booming digitalization in Russia, where women and men go online to perform a material-discursive translation of digital technologies and their cultural use to enable and constrain certain activities, roles, and identities (Hodder 2012). In other words, women and men take their materiality with them into cyberspace, which often becomes further oppressive rather than liberating.

Thus, mirroring the gendered discourses on masculine and feminine roles and patterns of behavior, digital media spaces impose similar restrictions and expectations on female users as those experienced by women in their offline activities. Therefore, female activists operating online tend to be seen as transgressing the accepted gendered behavioral norms solely by the fact of their leadership in digital media. When engaging in any interaction or activity online, women are expected to employ their feminine emotional capital in socially acceptable ways (i.e. providing emotional labor for the benefit of others), and the failure to do so tends to cause disapproval and criticism. At the same time, digital spaces attract female users interested in civic activism, which on the contrary is seen as non-transgressive. This paradox creates a complex environment for individual users and for virtual communities engaged in constructing alternative gendered identities online, both feminine and masculine.

This chapter offers an analysis of how the World Wide Web and digital technologies influence gender identity politics in contemporary Russian society. We look at the ways Russians construct gender online, how their practices become means of resistance and activism, and how they adapt and shape digital technologies to perform their gender identities and communicate with the State in the situation of increasing surveillance and control of material and cyberspaces.

## 12.2 CONSTRUCTING GENDER ONLINE

One of the responses to the post-Soviet crisis of masculinity and the emerging feminist movements that are perceived as a direct personal threat by some Russian men has been a rise in radical anti-feminist and masculinist movements operating primarily in online, digital spaces. Fuelled by the state-sponsored ideology of “traditional values,” misogyny became a part of any online debate (see also Lokot 2019). New versions of and new views on masculinity have been shaping up, with the gendered masculine identity being rethought through the opposition to “woman” as the “other” and reimagined in a world where women would not exist at all or would play a less prominent social role. These new views on masculinity can take relatively harmless forms, such as Internet memes. For example, since approximately 2012 there has circulated a popular meme “We don’t need chans” (“*tân ne nužny*”).<sup>1</sup> It was first applied by fans of Japanese anime, hence the use of the Japanese suffix *ちゃん*, Eng. “chan” / Rus. “tân” (a form of reference to children, female family members and female friends), but soon gained viral popularity on Runet. Oftentimes the new masculine identities are not only openly misogynist but are borderline extremist: one such example is the radical misogynist online community MD (*Mužskoe dviženie*, Masculine Movement; over 34,000 followers in March 2020).<sup>2</sup> The community’s motto is “We are not fighting against women—we are fighting for men’s rights.” The MD public accepts female members provided they do not post any content or comment, that is, have no voice, and the content circulated by the public consists mostly of misogynist hate speech and discussions of what is perceived by the public community members as violation of men’s rights.

Radical masculine movements existing as offline groups or online communities are by no means a specific feature of Russian society—in this respect, Russia is fully included into the global trends of anti-feminist backlashes. Another example of Russia following the global developments in terms of renegotiating gender roles and gender (in)equality is the popularity of the extremist movement of incels,<sup>3</sup> or “involuntary celibates,” which started in 1997 as an online community where its members shared their life experiences. It soon developed into a radical anti-feminist misogynist movement and is currently being spread across the globe, including Russia.<sup>4</sup> There, one of the best-known incels is probably Aleksej Podnebesnyj (aka Alex Undersky)—a Nizhny Novgorod-based anarchist and civic rights activist notorious for his misogynist social media posts calling for the end of women’s rule, which he refers to as “vaginocapitalism,” and for physical violence against women and, especially, feminists. As a result of Podnebesnyj’s activity on social media, in December 2019 a court case was started to investigate into the man’s extremist rhetoric against women.<sup>5</sup>

The accessibility of various social media platforms has enabled Russian men to explore their gendered identities through the construction of online hoped-for selves (Bouvier 2018) and outside of the agendas of grassroots movements. For these alternative masculinities the visual representations of gendered identities are particularly important, and picture- and video-based platforms—Instagram, TikTok and YouTube—have become a primary digital space for expressing those alternative masculinities (Kudaibergenova 2019). For example, the October 2019 ratings of top-twenty Instagram accounts and TikTok bloggers showed that the number two position in the rating was taken by the blogger Sima (@alexmymymy; over 31.1 million followers in March 2020) with almost three million followers.<sup>6</sup> Young and bold, Sima experiments with camp visuality, representing a queer take on masculinity, for example, through the use of make-up and feminine clothes.<sup>7</sup>

The popularity of bloggers like Sima is not a one-off success but rather a social media trend, with openly gay queer bloggers like Andrei Petrov attracting thousands of subscribers (in March 2020, Petrov’s YouTube channel had 1.05 million subscribers).<sup>8</sup> Petrov, who identifies as a gay cisgender man and uses his channel primarily to offer advice on beauty products, make-up trends and fashion, positions himself not only as a beauty and lifestyle blogger but also as a spokesman for the LGBTQ communities. Thus, on November 27, 2019, alongside five other openly gay celebrities and public figures, he participated in the YouTube TV show “*Ostorožno, Sobčak!*” (Beware of Sobchak!) hosted by the oppositional pro-LGBTQ celebrity politician Kseniya Sobchak. The episode was called “Coming-outs, gay-lobby and banning of propaganda: six gays and Sobchak” and was devoted to a range of issues connected with LGBTQ rights in Russia.<sup>9</sup>

The examples of Sima and Andrei Petrov demonstrate that Russian social media have become a relatively safe digital space for constructing transgressive non-heteronormative masculinities as far as adult audiences are involved. The online practices applied by Russian women also include transgressive patterns



of gendered behavior, which is consistent with the emergence of new gendered identities over the post-Soviet decades and which reflects the ongoing renegotiations of gender inequality and the relationships within the binary dichotomy “men–women.” For example, resisting the neo-conservative socio-political turn which took place in Russian public discourses throughout the 2000s and 2010s (Cucciola 2017), women have been challenging the imposed gender stereotypes about women’s primary social roles being those of mother and wife. On the Russian social networking site VKontakte, public online communities like “*Šast’e materinstva*” (Ze joy of motherhood; over 75,000 subscribers in March 2020)<sup>10</sup> and “*Šast’e byt’ ženoj*” (Ze joy of being a wife; over 28,000 followers in March 2020)<sup>11</sup> aim to disclose the truth about the challenges, difficulties and obstacles women face when performing the “traditional” gender roles, including domestic violence, mental health problems, financial struggles and broken relationships. Female inclusivity bloggers on platforms like Instagram, for example, Eleni (@loukoumh; over 65,500 followers in March 2020) or Ekaterina (@ekaterinaxiii; over 23,900 followers in March 2020), share digital images representing body-positive non-stereotypical concepts of female physicality and beauty.<sup>12</sup> Feminist bloggers like Tatyana Nikonova (@nikonova.online; over 243,000 subscribers on Instagram in March 2020), who is active across various social media platforms—Telegram, VKontakte, Facebook and Instagram—tackle various aspects of female sexuality and desire, offering open and honest advice on a range of issues, from choosing a sex toy to resolving the problem of sexual incompatibility between partners.

These insta-gender practices represent non-violent resistance or quiet activism women have been employing in the past decade to carve out their online space.

### 12.3 DIGITAL SERVICES FOR (WO)MEN: CREATING GENDER-SPECIFIC SPACES

Challenging gender binaries and traditional gender roles is also achieved by translating socio-economic materiality into digital spaces. With the rapid digitalization of services Russian state has been offering, women move online to perform their femininities and “traditional” roles of motherhood by using digital services to take care of their health, diet, body politics and, even, protect themselves from abuse. Women first organized around internet or web forums that served as an online discussion/message boards specialized around certain themes. Eventually those evolved into full time web resources and communities for women to exchange experiences and get help and information. Forums such as [www.myjulia.ru](http://www.myjulia.ru) (launched in 2008) or [www.woman.ru](http://www.woman.ru) (based on internet magazine launched in 2016) cater to different groups of women and cover a wide range of topics on health, beauty, personal relations, intimacy, family and sex. More specialized forums include [www.baby.ru](http://www.baby.ru) (launched in 2009) or [www.materinstvo.ru](http://www.materinstvo.ru) (in existence since 1999) that provide health and

educational advice for expecting and experienced mothers, but also provide a discussion space for women. While these online spaces are tagged by scholars as “traditional” (Gnedash 2012), they can also be viewed as a site of quiet activism (Pottinger 2017) where women manage and practice their femininity the way they see appropriate to them.

Women have also quickly learnt the advantages of digital citizenship, that is, using state-provided digital platforms to improve their wellbeing. One of these digital platforms—omnipotent *Gosuslugi* (Public services portal)—offers a range of services to make women’s lives better. Thus, everyone can make an appointment with health services, that is often important for women with small children, that they could do it from home and not call or go in person. Another service—enrolling children into kindergarten or school—is supposed to remove obstacles for disadvantaged families and make the procedure more transparent. While these services are positioned as gender-neutral—any one of the parents could use them—in reality it is still women who are tasked with everything related to motherhood and family obligations. Therefore, women not only become active digital citizens, they also are the ones who provide a feedback to the state to make these technologies better (see also Vivienne et al. 2016).

The IT industry has recently moved to create gender-specific apps to gain additional markets and better appeal to the user. In this move, gender dynamic remained essentially the same and even has been further re-enforced by pushing women to use more health apps (such as mHealth, dieting, yoga, fitness and other apps) and reproductive apps (such as baby.ru app to monitor pregnancy and breastfeeding or time-factor app to monitor monthly periods, both created by men). This distribution of apps promotes the healthy female subject who is embodied in three types of subject positions: (1) Barbie; (2) Earth goddess, and (3) entrepreneur. These themes fix White, middle-class, skinny, young, and fertile female bodies as the standards for health. Women are encouraged to achieve these bodies through practices of self-surveillance, disclosure, and self-advocacy, which are encouraged and normalized through routine use of apps. Thus, apps allow women to actively participate in choosing traditional subject positions, revealing the postfeminist sensibilities of this form of technology-based embodiment (Doshi 2018). At the same time, maternity apps help women self-survey their reproduction and claim autonomy by avoiding medical professionals for frequent check-ups.

By contrast, male apps reproduce masculinity, healthy male body, sexuality and grooming. In Russia, app market is especially full of barber and other grooming apps (such as Muzhikipro app) that claim to turn men into “real men.” Other apps such as Yourbro app are reinforcing heterosexual male identity by exploiting porn and female body. Sex is central for new digital technologies. Dating apps occupy a significant segment of Runet: alongside international Tinder and Grindr apps, Russia developed their own dating services such as Rambler dating app or the newly produced by VKontakte’s owners the Lovina app. Scholars suggest that while hetero-apps have a power to reinforce gender

stereotypes and heteronormativity, they empower and compromise women at the same time (Solovyeva and Logunova 2018; Chan 2018). Women receive opportunities to challenge traditional feminine behavior as chaste by arranging multiple and anonymous dating as well as sharing their experiences about dating apps (as above-mentioned blogger Tatyana Nikonova does). At the same time, they put themselves in a position of criticism and vulnerability.

Women's safety has become a part of Russian public discourse, thanks to massive online campaigns and activism, which we will look in detail in the following sections. The app market responded by creating safe apps (such as *Between Us* from Vodafone) for women that allow to share locations, make fake calls, and push the emergency button. The feminist non-governmental organization (NGO) [Nasiliu.net](http://Nasiliu.net) (Stop violence), that has a very prominent presence online and provides services to survivors of gender-based violence, created their own app ([bit.ly/NasiliuNetIOS](http://bit.ly/NasiliuNetIOS) for IOS and [bit.ly/NasiliuNetAndroid](http://bit.ly/NasiliuNetAndroid) for Android), which has the complete information regarding shelters, crisis centers, legal aid and other useful information for women, but mostly importantly has an SOS button that allows to alert people who the user trusts about danger at home and on the street. The developers hope that the app radically contributes to women's wellbeing.<sup>13</sup>

Assessing women's presence online, cyberfeminist theoretical framework offers to look at it as an "alliance" or "connection" between women and technology by exploring the intersection between gender identity, culture and technology (Mohanty and Samantaray 2017). Digital space liberates women and challenges binary gender order by its very process of transgressing material reality into digital one. Women increasingly use online and social networking for activism and mobilization in ways that were not possible before. One of those ways is to make women visible via *feminitivy* (feminitives)—feminine gender counterparts of all lexical terms denoting professional occupations used by Russian feminists to fight against the invisibility of professional women in public discourses (Guzaerova et al. 2018). In linguistics, the category of gender includes grammatical, lexical, referential and social gender (Hellinger and Motschenbacher 2015, 6), and the fact that Russian is a language with a grammatical gender means that all nouns fall into a gender category—masculine, feminine or neuter. Masculine and feminine gender nouns are unmistakably recognized by Russian speakers as referring to the social categories of femininity and masculinity. Although most terms denoting occupations have both masculine and feminine forms, quite a few nouns do not have a feminine counterpart, which aggravates the already existing issue of higher frequency of masculine–male expressions in Russian public communication (Hellinger and Bussmann 2001, 261). In the 2000s, to overcome the "androcentric perspective" of the Russian language (Hellinger and Bussmann 2001, 270), feminist activists started introducing into their online communication new feminine counterparts of masculine nouns formed with the suffix "k" and feminine gender ending "a," for example, "doktor—*doktorka*." These words were used in cases when the Russian lexicon did not have a feminitive to refer to a female

professional: for example, “*avtorka*” (authoress), “*redaktorka*” (editoress), “*direktorka*” (directoress). Throughout the 2000s and early 2010s, discussions about effectiveness and urgency of feminities were mainly conducted within the Russian feminist movement, primarily online but also in offline spaces. As of the late 2010s, these debates have entered mainstream public discourses, both in digital and offline spaces, and have polarized Russian society into supporters of such linguistic visibility for women and opponents, who are worried about the purity of the Russian language affected by feminist linguistic innovations.

## 12.4 WOMEN’S AND QUEER ONLINE ACTIVISM

When it comes to challenging and transgressing patriarchal discourses on women’s gendered behavior and social roles, digital media offer Russian women invaluable opportunities for activism. In the same way that digital media have impacted politics in general, transforming top-down political hierarchies into participatory networks (Dartnell 2006), social movements and the notion of social activism have also evolved in the Internet era. Protest voices (Coudry 2010) have been amplified by social media campaigns (Jenkins et al. 2016; Kaun 2017) and citizen journalists generating amateur media-content on social media have come to be considered a reliable and trustworthy source of information (Bewabi and Bossio 2014). Since their appearance in the global media landscape, social networking sites, or social media, have evolved from focusing on “bonding social capital,” that is, social bonds within a family or a small local or ethnic community, to “bridging social capital” by providing links across ethnic groups or between various communities and “linking social capital” by offering a new means of communication between political elites and the general public and between different social classes (Flew 2014, 66–67). Social networks have become an integral part and a valuable tool of participatory media cultures across the globe (Flew 2014, 77–78). Like other internet resources, social networking sites can be viewed as dynamic horizontal communication spaces (Youngs 2013, 176), which, due to the shared internet tools’ characteristics of multiplicity and interactivity, are often perceived as resources with “radical liberatory potential” (Curran et al. 2012, 151).

Despite Runet being prone to state surveillance and political monitoring (Uldam 2018), its users nevertheless enjoy a high level of participation and autonomy (Curran et al. 2012, 164), which is especially high on social media platforms. Taking political and social protest to social networking sites provides activists with wider opportunities for contacting like-minded people and promoting individually framed agendas. Social networking sites thus afford a means of coordinating and boosting collective action of various social movements: “collective actions are also becoming more inclusive, that is, they encourage participation of those who would not want to commit to the interpretations of a formal group and who would traditionally not be the target of organizational outreach efforts” (Schumann 2015, 55). Although the digital

divide still has a gendered dimension, in that women have suffered from inequalities in terms of access to the internet and other ICT (Ross and Byerly 2004, 187), the Internet has also enabled a considerable empowerment for women through cyberpolitics and cyberfeminism (Ross and Byerly 2004, 197–198). This is especially so for women involved in grassroots and community groups, whose activism increasingly takes place on the internet (Ross and Byerly 2004, 200). Internet-based activism has become vital for feminist activists and activist groups promoting the rights of lesbian, bisexual and transgender (LBT) women (Brown et al. 2017; Serano 2013).

Online feminist activism in Russia is developing fast and evolving consistently, comprising a variety of platforms and employing various strategies, among them—those of emotional capital and of “do it yourself” (DIY) brand identity (Turner 2010). Social media platforms offer Russian feminists such important tools as opportunities for transgressing patriarchal discourses, creating safe digital spaces in the form of emotional communities, and managing their own online identity as personal celebrity or influencer brands. On the other hand, activism performed online entails potential threats in the form of cyberbullying. For example, the case of the 2019 “*Lushgate*” campaign in support of prominent Russian feminist and lesbian activist Bella Rapoport, introduced into media discourses a debate on what kinds of online emotional expressions are acceptable for a woman. In March 2019, in an Instagram story Bella expressed her disappointment in the Lush handmade cosmetics brand which claims to be pro-feminist but failed to extend its support to her, that is, rejected her offer to collaborate. This made Bella a subject of cyberbullying across various social media platforms: she received hate mail via direct messaging on Instagram; Twitter users (both personal and corporate accounts) started a flashmob making a ridicule of Bella’s correspondence with Lush; the activist received hateful and threatening comments and messages on her personal Facebook page.<sup>14</sup> The cyberbullying was further promoted by multiple online media and mainstream media. The emotions shared in the Instagram story were interpreted as a transgression of socially acceptable feminine emotional boundaries by an overdemanding and self-absorbed feminist: a “good” woman does not use her emotions to demand benefits for herself but uses them to provide benefits for others. The example of “*Lushgate*” is only one of the numerous cases where Russian feminist activists faced a backlash of complex societal responses to their transgressive emotional expression and gendered behavior while performing their activism online.

Hashtag campaign mobilizations work to make women’s and feminist voices heard in situations of aggressive misogyny. Similar to emotional management, hashtags provide a form of active and quick mobilization as an immediate response to abusive actions. A hashtag, created as a means of structuring content in social networks, is increasingly used to attract attention to social and political issues and events. After its emergence, hashtag campaigns were considered mainly in the context of protests against government actions and decisions. Nowadays more and more attention is being attracted to hashtag

campaigns, which are against existing social practices, behavior and norms. In these cases, the protest is addressed not so much to the state as it is to power in a broader sense. Such campaigns often take the form of discursive activism that was described by Shaw (2012) as “speech or texts that seek to challenge opposing discourses.” Here the issue of participants’ choice of discursive strategies might be raised (Arbatskaya 2019).

Russian feminists and activists have started using hashtags increasingly after a very powerful “flashmob” *#yaNeBoyusSkazati/t* (in Ukrainian and Russian, respectively; “I am not afraid to tell”) started by a Ukrainian activist, Anastasiya Melnichenko, in Summer 2016. In response to a Facebook post blaming women for becoming victims of rape, Melnichenko shared her own story of sexual assault with the hashtag. The post went viral across Ukrainian- and Russian-language social media: hundreds of women shared their own stories of sexual assault and sexual harassment at work. In the first two months alone, there were 12,282 original posts and over 16 million views (Aripova and Johnson 2018). Following the success of *#yaNeBoyusSkazati/t*, other hashtag campaigns followed: *#etoNePovodUbit'* (*#ItIsNotAReasonToKill*) in 2018 and *#yaBoyusMuzhchin* (*#IAmAfraidOfMen*) in 2019. All of them represent an example of participants’ attempt to challenge patriarchy by sharing stories of abuse that women are not supposed to talk about. By articulating trauma and translating it into narratives, these campaigns also provided therapeutic effect as well as solidarity and space for sharing.

At same time, there is plenty of online resistance to feminist and women’s activism. Conservative social movement organizations (SMO) utilize online spaces for their own brand of activism to claim legitimacy by supporting traditional values that include stereotypical gender roles, the heteronormative family, protection of the family, and attacking anyone who says different. They efficiently use tools that are similar to those used by feminist activists: exclusive online spaces and hashtags as a response to what they see as a threat to “authentic Russia.” The SMOs such as *Sorok Sorokov* ([www.sorokorokov.ru](http://www.sorokorokov.ru)) and All-Russian Parental Resistance (RVS, [www.rvs.su](http://www.rvs.su)) have very visible online presence by conducting aggressive mobilization campaigns and organizing fake media events. Their media is de-personified by using the pronoun “we”; they rarely mention any representatives by names, instead hiding behind webpages and hashtags. Their most recent campaign is resistance to passing prevention of domestic violence law in the Russian Federal Assembly (Russian parliament). Not only they started an abusive and aggressive media campaign against the law and its authors (all women), they also mobilized online using hashtag *#zaSemyu* (*#ProFamily*) to encourage their supporters to participate in an online discussion of the draft at the Council of Federation (upper house of the parliament) webpage.



## 12.5 CONCLUSION

Mirroring the complex discourse on gender roles and gender equality in contemporary Russian society, digital spaces have evolved into a battleground for new gender politics and identities. Early cyberfeminists and activists considered those spaces safe, safer than actual public spaces for protest (see, e.g., Rolleston Collective 2014), which has resulted in the existence of a wide and diverse variety of online communities, activist public accounts, and personal blogs with a solid potential to influence and shape offline debates on the feminism, non-heteronormative identities, men's and women's rights. However, the example of Runet together with other "nets" suggests that people take their politics and their materiality to virtual spaces that, in turn, are becoming even more dangerous due to illusion of safety. Cases of cyberstalking, cyberbullying, and simple online campaigns calling to "deal" with feminist and LGBTQ+ activists make us revisit the concept of cyberspace. In Russia, the situation is further aggravated by selective but tight state-imposed control and censorship over internet as well as state's official patriarchal discourse.

Yet, the development of gendered online practices, tools and strategies point to an emergence of mosaic virtual reality in which multiple identities debate and negotiate but remain fluid in its discursivity. Russian feminist and anti-feminist and anti-gender conflict online mirrors a global backlash against feminism in digital media. Online spaces and digital platforms reproduce materiality of "real-life" conflict with serious political consequences. In Russia, gender politics online and offline indicates the debates and negotiations important for constructing identities in situations when freedom of expression can be limited. Russians use online and digital platforms as a strategy to communicate their difference to the State and to their fellow Russians.

## NOTES

1. Source: <https://memepedia.ru/tyan-ne-nuzhny-tnn/>, accessed December 4, 2019.
2. Source: <https://vk.com/mensrights>, accessed March 27, 2020.
3. See, for example, "Not as ironic as I imagined: the incels spokesman on why he is renouncing them," by the *Guardian*, published on July 19, 2018, <https://www.theguardian.com/world/2018/jun/19/incels-why-jack-peterson-left-elliott-rodger>, accessed December 13, 2019.
4. See, for example, "*Tán ne nužny: agressivnye devstvenniki stanovátsá novymi terroristami* [We don't need chans: aggressive virgins become new terrorists]," by Russian TV channel NTV, published on November 16, 2019, <https://www.ntv.ru/novosti/2255780/>, accessed December 13, 2019.
5. See "Women must provide sex for everyone: The most famous incel of Russia to be trialed for the ideas of vaginocapitalism," by the Russian newspaper *Komsomol'skaá pravda*, published on December 12, 2019, accessed December 13, 2019. <https://www.kp.ru/daily/27062.5/4130982/>.
6. Source: <https://www.mlg.ru/ratings/>, accessed November 27, 2019.



7. Source: <https://www.tiktok.com/@alexmymymy>, accessed March 27, 2020.
8. Source: [https://www.youtube.com/channel/UC0UK9eo\\_m6v4DDGX-QKQf5Ww](https://www.youtube.com/channel/UC0UK9eo_m6v4DDGX-QKQf5Ww), accessed December 4, 2019.
9. The episode is available at <https://www.youtube.com/watch?v=ksdptbnbu8c>, accessed December 13, 2019.
10. Source: [https://vk.com/zaiki\\_luzhaiki](https://vk.com/zaiki_luzhaiki), accessed March 27, 2020.
11. Source: [https://vk.com/prelesti\\_braka](https://vk.com/prelesti_braka), accessed March 27, 2020.
12. “9 inclusivity activists from Russia you need to follow on Instagram,” by the *Calvert Journal*, published on March 9, 2019, <https://www.calvertjournal.com/articles/show/11058/9-inclusivity-activists-from-russia-you-need-to-follow-on-instagram>, accessed December 13, 2019.
13. Source: <https://nasiliu.net/nuzhna-pomoshh/mobilnoe-prilozhenie/>.
14. See, for example, “Lush vs. activist: a serious debate about feminism born out of a Twitter craze,” by the RT’s multimedia project *Russia Beyond*, published on March 26, 2019, <https://www.rbth.com/lifestyle/330157-lush-vs-activist-serious-debate-feminism>, accessed December 13, 2019.

## REFERENCES

- Arbatskaya, Elena. 2019. Discursive Activism in the Russian Feminist Hashtag Campaign: The # ItIsNotAReasonToKill Case. *Russian Journal of Communication* 11 (3): 253–273.
- Aripova, Feruza, and Janet Elise Johnson. 2018. The Ukrainian-Russian Virtual Flashmob Against Sexual Assault. *The Journal of Social Policy Studies* 16: 487–500.
- Arvidsson, Viktor, and Anna Foka. 2015. Digital Gender: Perspective, Phenomena, Practice. *First Monday* 20. <https://firstmonday.org/ojs/index.php/fm/article/view/5930/4430>; <https://doi.org/10.5210/fm.v20i4.5930>.
- Bewabi, Saba, and Diana Bossio, eds. 2014. *Social Media and the Politics of Reportage. The “Arab Spring”*. London: Palgrave Macmillan.
- Bourdieu, Pierre. 2001. *Masculine Domination*. Stanford: Stanford University Press.
- Bouvier, Gwen, ed. 2018. *Discourse and Social Media*. London and New York: Routledge.
- Brown, Melissa, Rashwan Ray, Ed Summers, and Neil Fraistat. 2017. #SayHerName: A Case Study of Intersectional Social Media Activism. *Ethnic and Racial Studies* 40 (11): 1831–1846. <https://doi.org/10.1080/01419870.2017.1334934>.
- Castells, Manuel. 2010. *The Power of Identity*. 2nd ed. Malden, MA: Wiley-Blackwell.
- Chan, Lik Sam. 2018. Liberating or Disciplining? A Technofeminist Analysis of the Use of Dating Apps Among Women in Urban China. *In Communication, Culture and Critique* 11 (2): 298–314. <https://doi.org/10.1093/ccc/tcy004>.
- Collective, Roestone. 2014. Safe Space: Towards a Reconceptualization. *Antipode* 46 (5): 1346–1365.
- Couldry, Nick. 2010. *Why Voice Matters. Culture and Politics After Neoliberalism*. Los Angeles: SAGE.
- Cucciola, Mario, ed. 2017. *State and Political Discourse in Russia*. Rome: Reset-Dialogues on Civilizations.
- Curran, James, Natalie Fenton, and Des Freedman. 2012. *Misunderstanding the Internet*. London and New York: Routledge.

- Dartnell, Michael Y. 2006. *Insurgency Online. Web Activism and Global Conflict*. Toronto and Buffalo: University of Toronto Press.
- Doshi, Marissa J. 2018. Barbies, Goddesses, and Entrepreneurs: Discourses of Gendered Digital Embodiment in Women's Health Apps. *Women's Studies in Communication* 41 (2): 183–203.
- Flew, Terry, ed. 2014. *New Media*. South Melbourne and Victoria: Oxford University Press.
- Gnedash, A.A. 2012. Ženske soobšestva v onlajn-prostranstve: 'Režimy vidivosti' v publičnoj politike RF. *Ūžno-rosijskij žurnal social'nyh nauk*, no. 1: 97–103. <https://cyberleninka.ru/article/n/zhenskie-soobschestva-v-online-prostranstve-rezhimy-vidivosti-v-publichnoy-politike-rf>.
- Goscilo, Helen, and Yana Hashamova, eds. 2010. *Cinepaternity. Fathers and Sons in Soviet and Post-Soviet Film*. Bloomington and Indianapolis: Indiana University Press.
- Goscilo, Helen, and Vlad Strukov, eds. 2011. *Celebrity and Glamour in Contemporary Russia. Shocking chic*. London and New York: Routledge.
- Green, Eileen, and Alison Adam, eds. 2001. *Virtual Gender: Technology, Consumption, and Identity*. London: Routledge.
- Guzaerova, Regina, Drahomira Sabolová, and Vera A. Kosova. 2018. Russian Feminine Nouns with Suffix -k(a) in the Modern Mediaspace. *Revista San Gergorio*, no. 25: 36–43. <https://dialnet.unirioja.es/servlet/articulo?codigo=6841019>.
- Hellinger, Marlis, and Hadmund Bussmann, eds. 2001. *Gender Across Language: The Linguistic Representation of Women and Men, Volume 1*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Hellinger, Marlis, and Heiko Motschenbacher, eds. 2015. *Gender Across Languages 4*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Hodder, Ian. 2012. *Entangled: An Archaeology of the Relationships Between Humans and Things*. Malden, MA: Wiley-Blackwell.
- Huppatz, Kate. 2012. *Gender Capital at Work. Intersections of Femininity, Masculinity, Class and Occupation*. Basingstoke: Palgrave Macmillan.
- Jenkins, Henry, Sangita Shresthova, Liana Gamber-Thompson, Neta Kligler-Vilenchik, and Arely M. Zimmerman. 2016. *By Any Media Necessary. The New Youth Activism*. New York: New York University Press.
- Kaganovsky, Liliya. 2008. *How the Soviet Man was Unmade. Cultural Fantasy and Male Subjectivity Under Stalin*. Pittsburgh: University of Pittsburgh Press.
- Kaun, Anne. 2017. 'Our Time to Act Has Come': Desynchronization, Social Media Time and Protest Movements. *Media, Culture & Society* 39 (4): 469–486. <https://doi.org/10.1177/0163443716646178>.
- Kay, Rebecca. 2006. *Men in Contemporary Russia: The Fallen Heroes of Post-Soviet Change?* London: Ashgate.
- Kon, Igor. 1995. *The Sexual Revolution in Russia. From the Age of the Czars to Today*. New York: The Free Press.
- Kudaibergenova, Diana T. 2019. The Body Global and the Body Traditional: A Digital Ethnography of Instagram and Nationalism in Kazakhstan and Russia. *Central Asian Survey* 38 (3): 363–380. <https://doi.org/10.1080/02634937.2019.1650718>.
- Lokot, Tetyana. 2019. Affective Resistance Against Online Misogyny and Homophobia on the RuNet. In *Gender Hate Online*, ed. Debbie Ging and Eugenia Siapera, 213–232. Cham: Palgrave Macmillan.

- Mohanty, J.R., and Swati Samantaray. 2017. Cyber Feminism: Unleashing Women Power Through Technology. *Rupkatha Journal of Interdisciplinary Studies in Humanities* 9 (2): 328–336.
- Pilkington, Hilary, ed. 1996. *Gender and Identity in Contemporary Russia*. London, and New York: Routledge.
- Plant, Sadie. 2000. On the Matrix: Cyberfeminist Simulations. In *The Cybercultures Reader*, ed. David Bell and Barbara M. Kennedy, 325–336. New York: Routledge.
- Pottinger, Laura. 2017. Planting the Seeds of a Quiet Activism. *Area* 49 (2): 215–222.
- Ross, Karen, and Carolyn M. Byerly, eds. 2004. *Women and Media. International Perspectives*. Oxford: Blackwell Publishing.
- Salmenniemi, Suvi. 2008. *Democratization and Gender in Contemporary Russia*. London and New York: Routledge.
- Schumann, Sandy. 2015. *How the Internet Shapes Collective Actions*. Basingstoke: Palgrave Macmillan.
- Serano, Julia. 2013. *Excluded. Making Feminism and Queer Movements More Inclusive*. Berkeley, CA: Seal Press.
- Shaw, Francis, 2012. The Politics of Blogs: Theories of Discursive Activism Online. *Media International Australia* 142: 41–49.
- Solovyeva, Olga, and Olga Logunova. 2018. Self-Presentation Strategies Among Tinder Users: Gender Differences in Russia. In *International Conference on Digital Transformation and Global Society*, 474–482. Cham: Springer.
- Sremac, Srdjan, and Ruard Ganzevoort, eds. 2015. *Religious and Sexual Nationalisms in Central and Eastern Europe. Gods, Gays and Governments*. Leiden and Boston: Brill.
- Stella, Francesca. 2015. *Lesbian Lives in Soviet and Post-Soviet Russia: Post/Socialism and Gendered Sexualities*. Basingstoke and New York: Palgrave Macmillan.
- Stella, Francesca, and Nadya Nartova. 2015. Sexuality, Nationalism and Biopolitics in Putin's Russia. In *Sexuality, Citizenship and Belonging: Trans/National and Intersectional Perspectives*, ed. Francesca Stella, Yvette Taylor, Tracey Reynolds, and Antoine Rogers. London: Routledge.
- Turner, Graeme. 2010. *Ordinary People and the Media. The Demotic Turn*. Los Angeles: SAGE.
- Uldam, Julie. 2018. Social Media Visibility: Challenges to Activism. *Media, Culture & Society* 40 (1): 41–58. <https://doi.org/10.1177/0163443717704997>.
- Vivienne, S., A. McCosker, and A. Johns. 2016. Digital Citizenship as Fluid Interface: Between Control, Contest and Culture. In *Negotiating Digital Citizenship: Control, Contest and Culture*, ed. A. McCosker, S. Vivienne, and A. Johns, 1–17. Rowman and Littlefield International.
- Youngs, Gillian. ed. 2013. *Digital world: connectivity, creativity and rights*. London-New York: Routledge.
- Yusupova, Marina. 2018. Between Militarism and Antimilitarism: 'Masculine' Choice in Post-Soviet Russia. In *Gender and Choice After Socialism*, ed. L. Lynne Attwood, Elisabeth Schimpfössl, and Marina Yusupova. Cham: Palgrave Macmillan.
- Zdravomyslova, Elena, and Anna Tyomkina, eds. 2007. *Rossiiskij gendernyj porjadok: sociologičeskij podhod* [Russian Gender Order: A Sociological Approach]. Saint Petersburg: European University Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Digitalization of Consumption in Russia: Online Platforms, Regulations and Consumer Behavior

*Olga Gurova and Daria Morozova*

## 13.1 INTRODUCTION

Digital consumption is a complex field that can be defined as “online retail, marketing approaches, or seen as an expanding field of technological platforms and mobile applications that advance various forms of production, distribution, and consumption” (Ruckenstein 2017, 562). Within Russian studies, this is a still emerging field. At the moment, scholarship on digital consumption in Russia is quite limited, although think tanks and marketing companies have been following the situation closely and provide the most up-to-date data.

In this chapter, we focus on two main topics within digital consumption which have emerged from the two main areas of scholarly interest: online shopping and the sharing economy. Research into online shopping has examined e-commerce business models (Doern and Fey 2006), barriers and drivers of e-commerce (Daviy et al. 2018; Daviy and Rebiazina 2015), effects of national culture on e-commerce acceptance (Kim et al. 2016), and the emergence of one of the biggest Russian e-retailers—Ozon (Hawk 2002). The sharing economy has been studied from the point of view of its barriers and drivers in acceptance of services (Rebiazina et al. 2018) and of socio-cultural meanings of particular sharing platforms, such as the gift-giving platform [DaruDar.org](#) (Polukhina and Strelnikova 2014; Strelnikova and Polukhina 2014; Polukhina and Strelnikova 2015; Bocharova and Echevskaya 2014; Ivanenko et al. 2014).

---

In alphabetical order, contribution of both authors is equal.

---

O. Gurova (✉) • D. Morozova  
Aalborg University, Aalborg, Denmark  
e-mail: [gurova@hum.aau.dk](mailto:gurova@hum.aau.dk); [morozova@hum.aau.dk](mailto:morozova@hum.aau.dk)

Whereas marketing researchers mainly apply online quantitative surveys, sociologists utilize qualitative methods—in particular, netnography or online ethnographic research (Kozinets 2010) complemented by face-to-face interviews. Of the marketing companies and think tanks providing data, we are drawing upon studies conducted by Morgan Stanley (2018), PayPal Inc. (2018), Russian Association of Internet Trade Companies (2017) and Data Insight (2014, 2017, 2018, 2019). Therefore, for this chapter we utilized three types of data: academic articles on the subject; research produced by think tanks and marketing companies; and media publications to identify the current trends in digital consumption in Russia.

The first section of the chapter focuses on e-commerce, while the second section sheds light on the sharing economy. Each section contains definitions and brief theoretical concepts related to the phenomena and concrete empirical examples taken from the context of digital consumption in Russia. In conclusion, we summarize the findings and suggest directions for future research.

### 13.2 E-COMMERCE, M-COMMERCE AND ONLINE SHOPPING

Digital transformations of retail and shopping are linked to the shift from offline to online retail (e-commerce, m-commerce) and various forms of their co-existence. E-commerce is broadly defined as “using the internet to sell products and services” (Doern and Fey 2006, 315). M-commerce refers to purchases made from mobile devices such as smartphones and tablets. These transformations are enabled by the emergence of digital solutions for retailers and consumers, which transform the traditional offline retail and shape the everyday shopping experience. Digital transformations of retail and shopping occur in the context of a broader transition to the “service economy,” where retailers act as “integrators of services” based on knowledge-intensive service innovations (Pantano and Gandini 2018, 1). This results in a new concept of retail that “overcomes the traditional physical boundaries of the store ... to foster the growth of new forms of commerce strongly based on the usage of technologies such as online and mobile for shopping” (ibid., 1).

Digital transformations of retail in Russia can be approached through the concept of “liquid retail,” that is an open metaphor that helps to problematize the dynamics of retail, with the purpose of shedding light on the current circumstances in which retail stakeholders and consumers navigate the accelerated transformations (de Kervenoael et al. 2018, 417–418). Following this framework, in this section we pay attention to macro-level changes of market transformations, normative shifts, techno-economic (infra)structures and meso- and micro-level activities of multiple actors (new retail formations, changing consumption practices, etc.) (ibid., 418) in order to reflect upon how digitalization has changed retail and shopping in Russia.

### 13.2.1 *E-commerce, M-commerce*

E-commerce has been acknowledged as one of the fastest growing markets in Russia. According to Statista, e-commerce is expected to show an annual growth of 7.5% in the coming years (Ecommerce Foundation 2018). In 2018, it was considered to be in an emerging state, with high potential for contributing to the development of the Russian national economy (Daviy and Rebiazina 2015, 4). As for m-commerce, statistics (Ipiev 2018) show the share of mobile payments by July 2018 rose by 11% compared to the first half of the year. At the same time, the number of payments from desktop computers dropped by 4%. Nevertheless, 55% of e-commerce purchases were made from desktop and 45% from smartphones and tablets. Yet, the number is expected to change in favor of m-commerce, since retailers actively continue developing shopping apps and platforms for mobile devices (ibid.). In addition, shopping with an emphasis on “social commerce”—meaning shopping on social media—has become a noticeable phenomenon (Pantano and Gandini 2018, 2). According to research conducted by Yandex.Kassa and Data Insight, near 39 million Russians made purchases in various peer-to-peer platforms, such as social media and messengers (Yandex Kassa and Data Insight 2018).

There was virtually no e-commerce in Russia prior to 1998 (Hawk 2002, 702). It started to gain popularity after the Russian financial crisis of 1998, which forced many people to become self-employed and turned out to be a catalyst for entrepreneurial activities. The crisis was one of the reasons for the companies to start to operate more efficiently and develop e-solutions. As for other factors, such as internet penetration rate and access to computers, Hawk (2002, 703) mentions that the Internet usage in Russia was still very low at the time. Interestingly, the majority of those who accessed Internet did it at work (57%), whereas only 27% accessed it at home. Between 1998 and 1999, the number of dot-com companies in Russia grew from 50 to 400 (Doern and Fey 2006, 317). Compared to the developed economies of the United States, Canada and Western Europe at this time, in developing countries such as Russia, India and countries in Latin America, e-commerce was miniscule (Hawk 2004, 181).

In the years that followed, e-commerce developed rapidly and continued to boom in Russia up till 2013, though its growth slowed down during the financial crisis of 2008–2009 (Daviy and Rebiazina 2015, 9). As for the situation in the second half of the 2010s, the economic recession followed by the Crimea crisis in 2014 had a significant effect on the development of Russian e-commerce (Sadyki 2017, 1). On the one hand, these changes included a drop in gross domestic product (GDP), a decrease of buying capacity of consumers, and increased political risks for Western companies to operate in Russia due to sanctions and counter-sanctions implemented after the annexation of Crimea. At the same time, these changes helped push Russian companies into the e-commerce market (ibid., 1).



According to global ratings, Russia lags behind in terms of Internet penetration (23rd in 2017), reaching an estimated 62% in 2018 (Ecommerce Foundation 2018, 13), with the majority of users concentrated in Moscow and major cities (Sadyki 2017, 2). The country was placed 99th in the United Nation's logistical performance list, 35th in its ease of doing business list, and 35th in e-government Index (*ibid.*, 14). In addition, there is a substantial gap in regional development across the country (Morgan Stanley 2018, 3). Considering these barriers, the Russian e-commerce market is characterized as under-developed and fragmented—for instance, the four major e-commerce players in Russia account for 27% of the market compared to 63% in the United States and 84% in China (*ibid.*, 6), it is regionally highly disproportionate, and dominated by cash payments and poor quality of service, especially delivery (Sadyki 2017).

One of the drivers of e-commerce has been the increase of the quality and availability of Internet connections in Russian regions. There have been significant improvements in the services provided by the Post of Russia—the main operator of delivery services. Further, the development of numerous services, online payment platforms, digital signatures and a general increase in trust towards these types of tools contribute to e-commerce progress (*ibid.*, 11).

#### *13.2.1.1 Russian E-commerce Retailers*

Digitalization and the development of e-commerce and m-commerce have affected retailers of all sizes, from large- to small-scale. According to a report by Morgan Stanley (2018, 1), “Russia is the last major emerging market without a dominant online retailer.” In the same report, it is estimated that the Russian e-commerce market will reach 31 billion dollars by 2020. In 2018, the two most influential actors on the Russian market were Yandex (the largest technological company specializing in internet products and services, including the search engine Yandex.ru) and the Mail.ru Group (a major technological company, operating the most popular Russian social networking sites VKontakte.ru, Odnoklassniki.ru and “Moj mir” [My.mail.ru]). Yandex joined forces with the largest state-owned bank, Sberbank, in order to create “a leading e-commerce ecosystem” based on the existing Yandex marketplace (Henni 2017). On the other hand, Mail.ru Group partnered with Chinese retail giant Alibaba, the owner of [Aliexpress.com](http://Aliexpress.com)—the most popular online platform in Russia—to develop a “one-stop platform for social communication, gaming and shopping” (Henni 2018).

Meanwhile, some of the largest international e-commerce retailers have not been successful in their attempt to conquer the Russian market. [eBay.com](http://eBay.com) entered Russia in 2011, but so far has failed to gain significant traction. [JD.com](http://JD.com)—China's second largest e-commerce player—left Russia in 2016 after just one year of operating. Among the main challenges, the executives of the company listed cross-border logistics and high marketing expenses (Sun 2018). In 2018, a number of German e-commerce giants such as Otto, Quelle and Westwing ceased their activity in Russia due to reduced purchasing power of

consumers and significant revenue drops after converting funds from rubles into euros (East-West Digital News 2018). This latter phenomenon resulted from serious Russian ruble depreciation after 2014 (Urbanovsky 2015). The notable exception is [Aliexpress.com](https://www.aliexpress.com), which has been one of the leaders on Russian online market, and will be discussed in more detail in the section on cross-border shopping.

At the same time, digital transformations in retail gave a significant boost to Russian small-scale innovative companies and startups—for instance, in fashion. Online retailing has various benefits: it allows these companies to reduce costs associated with launching and operating their businesses; it gives the opportunity to be discovered by consumers more quickly; it allows using new business models and flexibility in adjusting to fluctuations of the market; it gives data and instruments for the immediate analysis of consumer behavior; and, it also provides tools for immediate interaction with consumers through videos, blogs, messages. In a separate study, we noticed a “boost” in small-scale fashion businesses (Gurova and Morozova 2018) when startup companies created formal and informal businesses through social networks, using Instagram or VKontakte as sales channels.

#### *13.2.1.2 New Retail Platforms*

Digital transformations have globally led to the development of online retail platforms (Daviy and Rebiyazina 2015; Doern and Fey 2006). The fastest-growing and most highly valued e-commerce platform in Russia has been Wildberries.ru, an online platform for selling clothes, accessories, home goods and so on. The platform operates in Russia, Belarus, Kazakhstan and Kyrgyzstan and has plans to enter the European Union market in 2019, starting with Poland (Ganzhur 2019). By 2019, Wildberries was ranked number one among online fashion retailers worldwide with the highest traffic volume, followed by H&M and Zara (Popova 2019). Wildberries’ revenues have soared, showing 85% growth in the first three months of 2019 compared to the same period in 2018 (Intellinews 2019). Their fastest growing segments have been electronics, office equipment, gardening equipment and kitchen equipment (Kommersant 2019). The co-owner of Wildberries, Vladislav Bakalchuk, admits that their strongest advantages are the wide network of delivery points across Russia (with fewer couriers and the possibility to try clothes on the spot), and its commission model in which a supplier/producer pays commission for every sale (Kommersant 2018b). In 2018, the second and third most popular e-commerce sites in Russia were Ozon.ru (one of the leading multcategory retailer, offering goods in about 20 categories, including electronics, household appliances, clothes, food, Digital Versatile Discs (DVDs) and Citilink.ru (an online platform that positions itself as an electronics discounter), respectively (Kommersant 2018c).

Online retail platforms in Russia can be classified into seven categories (see Table 13.1). We have taken mostly examples from the list of top-100 companies in e-commerce in Russia in 2018 (Data Insight 2019) to illustrate the

**Table 13.1** Classification of online retail platforms in Russia

<i>No</i>	<i>E-Format</i>	<i>Description</i>	<i>Examples</i>
1	Price-based online platforms	Platforms that have price as the main strategy for differentiation	Citilink.ru (electronics discounter); KupiVIP.ru (discounter of premium garment brands); Kuponator.ru (discount coupon service)
2	Community forming retailers	Platforms that incorporate community activities, such as forums and discussion boards	Exist.ru, the biggest retailer for car parts, has forum on their website
3	Online marketplaces/ shopping centers	Platforms that offer an online retail infrastructure for other companies to sell their products	Wildberries.ru sells clothes, accessories, home goods; Ozon.ru offers goods in about 20 categories
4	Online department stores	Online equivalent of brick-and-mortar department stores	Tsum.ru, an online shop of Central Universal Department store in Moscow
5	“Category killers”	Price-aggressive retailers selling a certain category of commodities	LaModa.ru sells clothes and shoes; Mvideo.ru for electronics; Petrovich.ru for homeware; Online pharmacy Apteka.ru
6	Niche retailers	Websites selling a particular type of goods	Yamaguchi.ru offers massage equipment
7	Micro-retailers of social commerce	Micro-companies that operate predominantly on social media	Roseville.ru is a fashion brand selling women’s clothes (being a micro-brand, it is not included in the top-100 list)

platform types. According to the top-100 list, among the top-10 companies are online marketplaces (Wildberries.ru, Ozon.ru), price-based online platforms (Citilink.ru) and “category killers”<sup>1</sup> (Mvideo.ru for electronics, Lamoda.ru for clothes and Petrovich.ru for homeware).

In 2011, Darrell K. Rigby wrote about “omnichannel retailing.” It refers to retailers who are “able to interact with customers through countless channels—websites, physical stores, kiosks, direct mail and catalogs, call centers, social media, mobile devices, gaming consoles, televisions, networked appliances, home services, and more” (Rigby 2011). Retailing in Russia is developing in the direction of omnichannel retailing. Therefore, in addition to traditional brick-and-mortar stores, companies operate as click-and-mortar stores, merging offline and online formats of retail in different ways. For instance, the online retailer Lamoda expressed its interest in opening the first offline store (Kommersant 2018a). Offline supermarket Perekrestok voiced a plan to launch a “shop-window” on its website where consumers can order commodities unavailable in the stores and offered by partners, and then pick them up in one of Perekrestok stores (Ishchenko 2018). Clothing online retailer Aizel.ru opened a pop-up store in the Moscow department store Atrium in 2018 (Utesheva 2018). However, managing director of KupiVIP.ru

Miroslav Zubačevskij noted that although the omnichannel retailing is emerging in Russia, it is still underdeveloped and faces various problems, ranging from inconvenient Information Technologies (IT) solutions to logistical issues (Fashion United 2014).

An important dimension of the service economy is the consumer experience. Therefore, retailers use technological solutions to address the needs of retail and shopping as part of the experience economy (Pine and Gilmore 1998). Retailers actively use online tools to enhance the experiential dimension of shopping online with videos, blogs and other interactive formats. For example, Bonprix.ru has a rubric “Fashion and life” featuring news on current fashion trends and useful advice regarding fashion and lifestyle from editors and bloggers. At the same time, they aim at individualization or customization of experience and products. For instance, Vsemayki.ru offers to “construct” your own T-shirt with chosen print, while Holodilnik.ru offers the option of customizing the color and adding exclusive decorations on fridges, washing machines and dishwashers.

Augmented reality and virtual reality technologies are emerging trends in retailing, also in Russia (Utesheva 2018). For instance, the Russian company Mirow (Mirow.ru) developed a touch-screen mirror that is able to identify which items are taken to a fitting room and to give personal recommendations on what is usually bought or may suit with those items. It can also provide information about discounts, including personalized ones, offer to call a consultant or ask for a different size/color of an outfit, and to arrange the issuing of a loyalty card. Another technological solution, Tardis (Tardis3d.ru), helps to identify a customer’s clothing size and the way outfit will look on their figure with the use of a selfie and a short questionnaire. There is also a startup, Fittin.ru, that has developed virtual fitting for shoes. Although these solutions are to a large extent at the experimental stage, they are in line with global trends, with a direction towards “smart retail”—that is, the search for digital solutions for retail and further consumer acceptance of these solutions (Dacko 2017).

### 13.2.2 *The Profile of Online Consumers*

Between 2011 and 2014, major indicators of e-commerce in Russia were on the rise, including the quantity of orders and average purchase amount, followed by a substantial drop in 2015 (Data Insight 2017). However, after a steep decline, Russian e-commerce started to slowly recover, and as of 2018, 65% of the country’s online users have shopped online at least once (Data Insight 2018).

As far as socio-economic characteristics are concerned, between 2011 and 2017, Russian trends remained typical for developing countries, where factors such as place of living, income and education have more influence on internet activity than physical access to the Internet (van Deursen et al. 2011; van Dijk and Hacker 2003). On the other hand, immaterial resources such as the knowledge of a foreign language and availability of free time may also influence the

popularity of the online shopping (Firsova 2013, 47). Interestingly, gender did not show significant correlation with the frequency of online shopping for a household; therefore, the researchers suggest that inconspicuousness of the online shopping (e.g. sitting alone in front of the monitor versus trying new clothes in public) might challenge the stereotype of shopping as a female prerogative (ibid., 48).

The leading group of online consumers, particularly in terms of frequency, are residents of megacities with higher education and above-average incomes, while poorer rural dwellers with less than 10 years' education lag behind (Data Insight 2018). However, experts (Morgan Stanley 2018; Data Insight 2018) assume that the current growth in online purchases is driven by shoppers from small towns, residential communities and villages who are quite price-conscious and have embraced online shopping in search of a better deal. Additionally, Russian rural settlements are now better equipped with pick-up points for online orders: in January 2017, 76% of all delivery points were located in small towns and villages (Data Insight 2017). Furthermore, over half of Russian internet users make online payments and transfers: this indicator equals 61%, 55% and 44% for large cities, middle-sized cities and the rest of the settlements, respectively (Data Insight 2018).

For a long time, the three most popular categories for online shopping have been electronics, clothing and household appliances (Russian Association of Internet Trade Companies 2017, 27). In addition to material goods, the popularity of food delivery, airplane and train tickets, as well as online games' complement products have been on the rise (Data Insight 2014).

The results of a poll by BrandMonitor, published in 2018, revealed that a large proportion of Russians (63%) may mistake fakes for luxury brand items (Tishina 2018). An even larger proportion (84%) of Russians shopping online is actually quite eager to buy counterfeit items (usually A-brands such as Apple or Louis Vuitton). As a rule, these consumers have some previous experience of buying fake goods through traditional channels, and afterwards turn to the Internet as an information tool. Frequently, the webpages of original brands are used as points of reference for comparing how well the fakes match the descriptions and pictures of authentic products (Radon 2012).

Overall, some pieces of research conclude (Firsova 2013) that the growth of the Internet shopping has great potential in Russia, but it will primarily spread among people who have previous positive experience of using the web.

### 13.2.3 *Online Cross-Border Shopping*

Online retail gives a boost to cross-border shopping. In 2015, 30% of the Russian population made at least one online cross-border purchase in the course of the year (Ecommerce Foundation 2018). In 2017, 30% of Russian online shoppers were limiting their purchases to the domestic market, whereas 56% were shopping both domestically and abroad, with 14% abroad only (PayPal Inc. 2018). These shares of cross-border online purchases are among the highest in Eastern Europe (ibid.)

Clothes and accessories, smartphones/tablets (including related items), home appliances and electronics have been the most popular goods for ordering abroad (Yandex 2016). The main reasons for cross-border shopping among Russians have been better deals in terms of price, wider selection of goods, and access to brands unavailable in Russia (ibid.), which corresponds to the cross-border shopping drivers among US (Invesp Consulting 2016) and European Union (EU) shoppers (Hunter and Wilson 2015, 25), but differs from Chinese consumers who are primarily interested in certified and authentic goods (Zhang 2018).

China has been the most popular country for cross-border orders in Russia, accounting for up to 80% of all cross-border sales in 2016 (East-West Digital News 2017, 22). In 2014, AliExpress became the number one e-commerce platform in Russia, and since then has been offering, in addition to Chinese goods, products made in Russia with a same-day delivery option and purchases on credit (ibid.). Honoring Russian-language consumers, the platform has translated its web-interface into Russian. However, the automated translation is not always grammatically correct or smooth and the descriptions of consumer goods can look like a mere collection of words without proper declension, for instance, “new fashion print design Russian crime tattoo.” AliExpress gave rise to social media groups educating Russian consumers on how to navigate online shopping with the retailer. The company takes the Russian market seriously; it experimented with 3D virtual stores, and Russia is the only country where the retailer has tested an opportunity to enter the brick-and-mortar market (Vedomosti 2017).

### *13.2.3.1 Regulation of the Online Cross-Border Shopping*

As online shopping has experienced noticeable growth, so did the number of purchases from the foreign platforms that have to cross the Russian border. As a result, substantial legislative changes regarding cross-border shopping have been introduced recently. From 2019, individual purchases with a price over 500 euro and weight over 25 kg are charged with a customs duty. Previously, customs duties were levied on individual purchases totaling over 1000 euro and weight over 31 kg per month. From 2020, this tax border for online cross-border shopping will be further decreased to 200 euro. Therefore, the direction of changes is to increase taxation by lowering the threshold amount for tax-free purchases. Experts and entrepreneurs working in the e-commerce sector question the necessity of this measure and predict low revenues to the government (Russia Business Today 2018; Skuratova 2018). As of 2018, very few online purchases in Russia exceed 200 euro, and 90% of receipts on the most popular platform AliExpress amount to less than 100 euro (Russia Business Today 2018).

Another legislative measure related to e-commerce offered by the Russian Association of Internet Trade Companies is the introduction of value added tax (VAT) for foreign online retailers. Arguing for this measure, the association maintains that the taxation of foreign e-retailers will support domestic

companies (Russian Association of Internet Trade Companies 2018, for more, see also Chap. 5). In addition, in its statement, the Russian Association of Internet Trade Companies stresses that customs duties will bring approximately 300 million rubles to the Russian budget in the first three years (*ibid.*). The measure has so far been postponed due to a lack of agreement on how the procedure should be carried out technically. For instance, it is unclear if VAT should be paid by the online platform (e.g. AliExpress) or by the retailers that are using the platform. Experts argue that additional taxes may force international e-commerce platforms to cease their activity in Russia: equaling just 0.7%, Russia is a minor player for world-wide operating companies (TASS 2017); additionally, foreign retailers are already paying operational costs, commission for using the online platforms and service providers' fees (Kommersant 2018d).

It is unclear how the new measures will influence cross-border shopping. In the meantime, new platforms are emerging to address the needs of the consumers. For instance, there is a Russian platform called [Tudatuda.com](http://Tudatuda.com) where consumers gather to find someone who could buy and pass on a needed item from abroad. This service has been helpful for those looking for medicines or brands that are unavailable in Russia or when local prices are too high (for example, Apple products).<sup>2</sup> Such practices suggest that, even if affected by the changes in legislation, cross-border shopping will most likely adjust, for instance, by changing its format.

### 13.3 ONLINE EXCHANGES: SHARING ECONOMY AND COLLABORATIVE CONSUMPTION

The sharing economy and collaborative consumption refers to the collective use of consumer goods (Botsman and Rogers 2010) enhanced by the development of online platforms. Collaborative consumption is “the peer-to-peer-based activity of obtaining, giving, or sharing access to goods and services, coordinated through community-based online services” (Hamari et al. 2015, 2047). The sharing economy is an environment in which goods and services are offered and consumed through community-based platforms (*ibid.*). The terms can be used interchangeably since there are common denominators, namely, the mediating role of new digital technologies connecting various actors and modes of transfer. One position even argues that collaborative consumption is embedded in the sharing economy and is one of its forms (Wahlen and Laamanen 2017); therefore, we use sharing economy as the term embracing both categories.

The sharing economy has been triggered by the development of the Internet and technology, thanks to which it is easier to create trust between strangers (Botsman and Rogers 2010). There are other reasons for the growing popularity of the sharing economy, including economic (saving money), environmental (reducing ecological footprint) and social (expanding one's networks) (Schor and Thompson 2014).



In Russia, the sharing economy has become a noticeable phenomenon for several reasons. On the one hand, its wider emergence has been linked to the financial crisis of 2008–2009, after which the use of many sharing services was a coping strategy of dealing with economic hardship. For example, the gift-giving platform [DaruDar.org](http://DaruDar.org), launched in 2008, where participants make gifts of different daily objects such as books, children’s products, furniture, home-ware and others in exchange for gratitude, illustrate this copying strategy (Polukhina and Strelnikova 2014, 90). The boost of the sharing economy can also be seen as a result of people searching for alternative income sources during cutbacks. For instance, [Youdo.com](http://Youdo.com), launched in 2012, matches people who need minor home services (repair, cleaning) with service providers. The platform has been positioned as an opportunity to earn extra money in one’s spare time and, on the other side, solve a home problem for a lower price.

At the same time, the proliferation of the sharing economy was connected to deeper socio-cultural processes in Russia. The first decade of 2010s was associated with the relative growth of well-being of Russian citizens (“fat noughties”); hence, there were people whose financial conditions made them eligible to participate in various lending and gift-giving activities. Bocharova and Echevskaya (2014, 102–103) have noticed that some people join DaruDar due to surplus rather than need, and share the motive of contributing to the common good. This is evidence of the fact that the increase in well-being resulted in a shift towards post-materialist values, among which are contributions to the common good, self-expression and environmentalism (Polukhina and Strelnikova 2014, 88). In addition, participation in the sharing economy has become a form of “consumer solidarity.” Since the sharing economy is often a part of the “informal economy” (Polukhina and Strelnikova 2015), not directly regulated by the government (Polukhina and Strelnikova 2014, 87) and existing along with the “formal economy” of online stores (*ibid.*), it serves as a horizontal grassroots form of solidarity.

### 13.3.1 *Types of Sharing Economy*

Treapăt et al. (2018) distinguish between three types of sharing economy. The first one is paying for the benefit of using some good without purchasing it, such as the platform [Rentmania.com](http://Rentmania.com) launched in Moscow in 2013. The most popular items for borrowing (Shlyahov 2017) are children’s goods; sports equipment; various gadgets, including laptops, game consoles, and scooters; and evening and carnival garments. Goods for borrowing are provided by both companies and individuals. The former dominates in the sector of larger sports equipment, such as treadmills, that is characterized by seasonal demand. The latter usually prevail in children’s and gadget sectors. There are particular types of goods, such as photo booths, 3D printers or cotton-candy machines that are borrowed along with service providers. In addition, a separate and influential group of goods providers is comprised by winter downshifters, that is, people who leave Moscow for the whole winter period and are willing to offer

long-term lease of their items. Rentmania was initially launched as a start-up with venture capital for Moscow region only. Without any interest from the consumers, it ceased its operation in Russia. In 2018, the owners of the platform decided to move from Russia to the United States due to lack of investors and, consequently, limited perspectives in the home market (Mihajlova 2018).

Another type of sharing economy is the redistribution of used and no longer needed goods to the users who need them, for money or exchange (Treapăt et al. 2018). In Russia, the most popular platform of this type is Avito.ru founded in 2007 by Jonas Nordlander and now owned by a South-African company, Naspers Holding. At first, Avito was a platform for re-selling various everyday products such as clothes, cutlery, furniture etc., but eventually it has expanded into other directions, including recruiting, real-estate and short-term property leases. According to a Mediascope study (Ishunkina 2018), Avito's audience reached 4.3 million people aged between 12 and 64 years old by September 2018; and over 14 million Russians used Avito monthly with average 8 minutes per day (as of January 2019) (Mediascope 2019).

Finally, the third type of sharing economy is sharing of lifestyles, meaning that in addition to tangible goods, something intangible such as space and time is offered (Treapăt et al. 2018). Various rental services such as [AirBnb.com](https://www.airbnb.com) are part of this type. In Russia, AirBnb has become popular in Moscow and St. Petersburg, showing growth of 121% between 2015 and 2016 (Egorova 2017). A research on AirBnb offers on Tverskaya street (Moscow) shows that Muscovites are renting out the rooms (40–72 euro per night) or apartments (69–120 euro per night) in Stalinist high rises for prices 5–15 times lower than hotels (from 499 euro per night) on the same street (Treapăt et al. 2018). On the one hand, Russians have been known for a quite conservative and protective attitude towards their homes (“My home is my castle”); on the other, they have a long history of sharing homes (*kommunalka*, a communal apartment in the former Soviet Union, typically shared by several families). According to Treapăt et al. (2018), apart from extra income, people in Russia are opening their flats in order to gain new experiences, broaden horizons and practice languages.

### 13.3.2 *Participants of Sharing Economy*

The proliferation of the Internet in Russia facilitated the increasing number of participants of various sharing services, spread far beyond one's close circle and scaling the sharing economy to a higher level. According to a survey by Data Insight (2018), 394 million transactions equaling 591 million rubles (7.5 million EUR) within the Russian sharing economy were conducted in 2018. The most popular product category was clothing and boots with an average transaction price of 1950 rub (25 EUR). The second and third most popular categories were electronics and real estate. The online platform Avito was the most popular source for offering one's goods (65% of users), followed by a similar platform Youla.ru (39%), social networking site VKontakte (33%), and Instagram (9%) (ibid.).

In relation to socio-demographic characteristics of consumers, researchers (Rebiazina et al. 2018) have shown that people aged 18–35 and those who live in big cities across Russia used services based on the sharing economy several times per month, while older generations (35–60) and residents of smaller towns used the services once every few months. The majority of consumers had an average income. The findings showed no difference between them and high-income consumers. Representatives of both income groups used services several times per month. Therefore, the sharing economy is a mostly urban phenomenon and is not linked to a particular income group. It has been found that the most popular services are Uber, GetTaxi and Avito, which were each used by 4 out of 10 people (ibid., 394).

### 13.3.3 *Drivers and Barriers of Sharing Economy*

Scholars have studied drivers and barriers of the use of services based on the sharing economy in Russia (Rebiazina et al. 2018). Regarding drivers, they found that although 60% of consumers trust the services and are ready to rent things, only 30% want to rent out their own things due to risks related to sharing, personal safety and hygiene. They also found that the ownership of things is considered to be a symbol of status; therefore, some consumers prefer to own things rather than to rent them (ibid., 394). Among the drivers, the following were named: (1) interest, comfort and utility towards/of sharing services, (2) recommendations of reference group (family, friends), (3) ecological and environmental benefits and (4) ease of use. Barriers included: (1) risks associated with participation in sharing, (2) additional efforts caused by participation (time, financial costs) and (3) preferences for ownership as an indicator of higher status. The researchers concluded that the companies willing to build their business in the Russian market should take into consideration the meaning of ownership as a status symbol and the necessity of building credibility and trust; a major issue in an emerging market (ibid., 397).

Noticeably, the researchers mostly discussed the sharing economy as a new phenomenon, coming from the Western countries and sometimes appearing in contradiction to the Russian mentality with its inclination towards ownership as opposed to renting. However, it is interesting how this new sharing economy co-exists with older practices of exchange and sharing coming from socialist times and familiar to the older generations.

## 13.4 CONCLUSION

In this chapter, we have approached digital consumption as consumption mediated by the Internet and analyzed it at the level of market transformations, changes in technological infrastructures and solutions, and in the activities of multiple actors (governments, retailers, professional associations, IT companies, consumers). To study digital consumption thus meant to focus on market developments (retail formats, retail culture, business models for companies),

regulatory aspects (laws and regulations provided by the governments), technological infrastructure (possibilities and limitations of platforms, IT solutions for data collection and analysis, IT solutions to address the needs of retailing and consumers) and consumer behavior (patterns, objects and channels of purchases, online brand communities).

The key impact of digitalization of consumption in Russia by the end of the 2010s has been the swift growth of online retail platforms, such as Wildberries.ru, which have a potential of becoming global, as well as of sharing economy platforms, such as Avito.ru and Darudar.ru, aimed at arranging circulation of goods and services. Offline retail has evolved in the direction of omnichannel retailing by developing various online solutions aimed at creating, on the one hand, fast and convenient and, on the other hand, an immersive and customized experience of shopping. The interest of government in regulating this quickly developing sphere, in cross-border shopping, particularly with its potential tax revenues, has been observed. In terms of consumer practices, Russian consumers have been turning to online shopping in search of lower prices and a wider selection of goods. Initially, the average Russian online consumer's income and education were higher than the country's average, but recently, consumers from lower income and education and from older age groups have been joining the practice of online shopping more actively. This last development might be attributed, on the one hand, to personal income drops, but, on the other, to positive developments in terms of delivery or pick-up options across the country as well as saturation of internet connectivity. Overall, online shopping trends in Russia are heterogeneous due to sharp differences across Russian regions.

We would like to suggest the following directions for future research: first, more ethnographic research on the shifts in the culture of consumption caused by digital transformations is needed. Here, such diverse topics as online brand and consumer communities, how they function and shape consumer identities, and new forms of socialities and consumer behavior in Russia are to be addressed. Another research direction could be based on the use of big data on consumption in cooperation with companies. This might help in studying the peculiarities of consumer behavior of various social groups depending on different factors, such as time and place. A third direction could be a more nuanced research on tastes, social characteristics and consumer practices, based on data collected by retailers. The researchers could also take a holistic approach by examining the intersection of various aspects of digital consumption, for example overlaps between regulations, retail, technology, and consumer behavior.

## NOTES

1. This is the term from the professional field of retailing.
2. According to a Deutsche Bank report on prices in 2017, Russia was the third most expensive country for buying an iPhone 7. The report was accessed on December 2018. <https://www.finews.ch/images/download/Mapping.the.worlds.prices.2017.pdf>.

## REFERENCES

- Bocharova, Ekaterina, and Olga Echevskaya. 2014. Kollaborativnoe potreblenie v sovremennoj Rossii: formy organizacii praktiki i vključeniiâ na primere soobščestva Darudar [Collaborative Consumption in Contemporary Russia: Forms of Organization and Practices of Engagement Using Community Darudar as an Example]. *Labirint*, no. 2: 97–108.
- Botsman, Rachel, and Roo Rogers. 2010. *What's Mine Is Yours: The Rise of Collaborative Consumption*. New York: HarperCollins books.
- Dacko, Scott G. 2017. Enabling Smart Retail Settings via Mobile Augmented Reality Shopping Apps. *Technological Forecasting and Social Change* 124: 243–256.
- Data Insight. 2014. Internet-torgovlá v Rossii 2014. Godovoj otčet [Online Retail in Russia 2014, Annual Report]. [http://www.datainsight.ru/files/DI\\_InSales\\_PayU-Ecommerce2014.pdf](http://www.datainsight.ru/files/DI_InSales_PayU-Ecommerce2014.pdf). Accessed 28 Dec 2018.
- . 2017. Internet-torgovlá v Rossii 2017. Cifry i fakty [Online Retail in Russia 2017, Figures and Facts]. [https://s3.amazonaws.com/wix-anyfile/r2G6e4KrT-TqbdH7EU7iB\\_Data%20Insight\\_2017\\_Электронная%20торговля.pdf](https://s3.amazonaws.com/wix-anyfile/r2G6e4KrT-TqbdH7EU7iB_Data%20Insight_2017_Электронная%20торговля.pdf). Accessed 28 Dec 2018.
- . 2018. Socseti, messendžery, sajty ob"âvlenij i sharing economy kak kanaly prodaz̄ [Social Media Platforms, Messengers, Classifieds Sites and Sharing Economy as Retail Channels]. <http://www.datainsight.ru/sites/default/files/DI-SocCommerce-YandexKassa.pdf>. Accessed 28 Dec 2018.
- . 2019. Rejting top-100 krupnejših internet-magazinov Rossii (za 2018 god) [Top-100 Largest Internet Retailers in Russia (as for 2018)]. <http://datainsight.ru/top100/>. c
- Daviy Anna O., and Vera A. Rebiazina. 2015. *Investigating Barriers and Drivers of the E-commerce Market in Russia*. Working Paper. Series: Management. National Research University – Higher School of Economics.
- Daviy, Anna O., Vera A. Rebiazina, and Maria M. Smirnova. 2018. Bar'ery i drajvery pri soveršenii internet-pokupok v Rossii: rezul'taty èmpiričeskogo issledovaniâ. *Vestnik SPBGU. Menedžment* 17 (1): 69–98.
- van Deursen, Alexander J.M., Jan A.G.M. Dijk, and Oscar Peters. 2011. Rethinking Internet Skills: The Contribution of Gender, Age, Education, Internet Experience, an Hour Online to Medium- and Content-Related Internet Skills. *Poetics* 39: 125–144.
- van Dijk, Johannes A.G.M., and Kenneth Hacker. 2003. The Digital Divide as a Complex and Dynamic Phenomenon. *The Information Society* 19: 315–327.
- Doern, Rachel R., and Carl F. Fey. 2006. E-commerce Developments and Strategies for Value Creation: The Case of Russia. *Journal of World Business* 41: 315–327.
- East-West Digital News. 2017. E-commerce in Russia, What International Merchants, Service Providers and Entrepreneurs Must Know to Succeed in Fast-Changing Market. <http://www.ewdn.com/files/etail.pdf>. Accessed 28 Dec 2018.
- . 2018. <http://www.ewdn.com/2018/10/03/german-retailers-leave-russia/>. Accessed 28 Dec 2018.
- Ecommerce Foundation. 2018. Ecommerce Report Russia 2018. <https://www.ecommercewiki.org/reports>. Accessed 28 Dec 2018.
- Egorova, Kira. 2017. Russians Weather the Crisis by Sharing Beds, Rides and Jobs. *Russia Beyond*. January 30, 2017. <https://www.rbth.com/business/2017/01/30/>

- [russians-weather-the-crisis-by-sharing-beds-rides-and-jobs\\_692088](#). Accessed 28 Dec 2018.
- Fashion United. 2014. KupiVIP: Lúdi stali pokupat' ne men'she, a umnee [KupiVip: People Have Started to Buy Less, But Smarter]. *Fashion United*, September 9, 2014. <https://fashionunited.ru/v1/fashion/kupivip-ludi-stali-pokupat-umnee>. Accessed 8 Sept 2018.
- Firsova, Natalya. 2013. Prediktory innovacionnyh potrebitel'skih praktik: osvoenie internet-šopinga v rossijskih domohozájstvah [The Predictors of Innovative Consumers' Practices: Uptaking the Internet Shopping Within the Russian Households]. *Journal of Economic Sociology* 14 (4): 27–57.
- Ganzhur, Elena. 2019. Bakal'čuk protiv Bezosa: Wildberries budet konkurirovat' s Amazon [Bakalchuk vs Bezos. Wildberries Will Compete Against Amazon]. *Forbes*, March 14, 2019. <https://www.forbes.ru/biznes/373327-bakalchuk-protiv-bezosa-gde-wildberries-budet-konkurirovat-s-amazon>. Accessed 14 Apr 2019.
- Gurova, Olga, and Daria Morozova. 2018. Creative Precarity? Young Fashion Designers as Entrepreneurs in Russia. *Cultural Studies* 32 (5): 704–726.
- Hamari, Juho, Sjöklint Mimmi, and Antti Ukkonen. 2015. The Sharing Economy: Why People Participate in Collaborative Consumption. *Journal of the Association for Information Science and Technology* 67 (9): 2047–2059.
- Hawk, Stephen. 2002. The Development of Russian E-commerce: The Case of Ozon. *Management Decision* 40 (7): 702–709.
- . 2004. A Comparison of B2C E-Commerce in Developing Countries. *Electronic Commerce Research* 4 (3): 181–199.
- Henni, Adrien. 2017. Yandex and Sberbank Agree \$500 Million JV Plan to Create “Leading E-commerce Ecosystem” in Russia. <http://www.ewdn.com/2017/08/09/yandex-and-sberbank-agree-500-million-jv-plan-to-create-leading-e-commerce-ecosystem-in-russia/>. Accessed 28 Dec 2018.
- . 2018. Alibaba and Mail.Ru Group Join Forces to Create “One-Stop Platform for Social, Communications, Gaming and Shopping.” <http://www.ewdn.com/2018/09/11/alibaba-and-mail-ru-group-join-forces-to-create-one-stop-platform-for-social-communications-gaming-and-shopping/>. Accessed 28 Dec 2018.
- Hunter, Julie, and Mark Wilson. 2015. Cross-Border Online Shopping Within the EU. Learning from Consumer Experience. <http://www.anec.eu/attachments/ANEC-RT-2015-SERV-005.pdf>. Accessed 28 Dec 2018.
- Intellinews. 2019. Russian Online Retailer Wildberries posts 85% Growth in Revenues in 1Q19. Intellinews. *Intellinews*, April 12, 2019. <https://www.intellinews.com/russian-online-retailer-wildberries-posts-85-growth-in-revenues-in-1q19-159491/?source=russia>. Accessed 13 Apr 2019.
- Invespro Consulting. 2016. Cross Border Shopping: Statistics and Trends. <https://www.invespro.com/blog/cross-border-shopping/>. Accessed 28 Dec 2018.
- Ipiev, Temir. 2018. Trendy rossijskih e-commerce platežej – 2018: analiz statistiki [Trends of Russian E-commerce Payments: Analysis of Statistics]. <https://vc.ru/trade/45487-trendy-rossijskih-e-commerce-platezhey-2018-analiz-statistiki>. Accessed 25 Dec 2018.
- Ishchenko, Natalia. 2018. ‘Perekrestok’ zapuskaet ‘marketplejs’ dlá onlajn-magazina [Perekrestok Launches a ‘Marketplace’ for an Online Shop]. *Vedomosti*, July 20, 2018. <https://www.vedomosti.ru/business/articles/2018/07/20/776030-perkrestok-zapuskaet>. Accessed 24 Apr 2018.



- Ishunkina, Inessa. 2018. Digital. Auditoriã Internetã [Digital. The Internet Audience]. Mediascope. [https://mediascope.net/upload/iblock/ea0/RIW2018\\_I-Ishunkina\\_21.11.2018.pdf](https://mediascope.net/upload/iblock/ea0/RIW2018_I-Ishunkina_21.11.2018.pdf). Accessed 13 Apr 2019.
- Ivanenko, E.A., M.A. Koretskaja, and E.V. Savenkova. 2014. Darudar. Žizn' i smert' prostyh vešej [Darudar. Life and Death of Ordinary Things]. In *Sila prostyh vešej*, ed. S.A. Lishaeva, 137–161. St. Petersburg: Aleteja.
- de Kervenoael, Ronan, Domen Bajde, and Alexandre Schwob. 2018. Liquid Retail: Cultural Perspectives on Marketplace Transformation. *Consumption, Markets and Culture* 21 (5): 417–422.
- Kim, Eungkyu, Roman Urunov, and Hyungjoon Kim. 2016. The Effects of National Culture Values on Consumer Acceptance of E-commerce: Online Shoppers in Russia. *Procedia Computer Science* 91: 966–970.
- Kommersant. 2018a. Lamoda primerãet magaziny. Internet-ritejler vyhodit oflajn [Lamoda Is Testing Out the Shops. Internet Retailer Is Heading Offline]. *Kommersant*, June 29, 2018. <https://www.kommersant.ru/doc/3670886>. Accessed 27 Dec 2018.
- . 2018b. 'Neinteresno prodavat' to, što proizvedeno v Kitae i zavozitsã bez nalogov.' Sovladelec Wildberries Vladislav Bakal'čuk ob osobennostãh rossijskogo onlajn-ritejla ['It's Boring to Sell Goods Made in China and Delivered Across the Border Untaxed.' Co-owner of Wildberries Vladislav Bakalchuk on Peculiarities of Russian Online-Retail]. *Kommersant*, March 30, 2018. <https://www.kommersant.ru/doc/3586567>. Accessed 28 Dec 2018.
- . 2018c. U Wildberries srabotal assortiment. Internet ritejler otčitalsã o rekordnoj vyručke [Assortment Has Been Pivotal in Wildberries' Case. The Internet Retailer Has Reported on Record-Breaking Profits]. *Kommersant*, December 17, 2018. <https://www.kommersant.ru/doc/3833542>. Accessed 28 Dec 2018.
- . 2018d. V posylki podkladyvaüt NDS [VAT Is Being Added into the Parcels]. *Kommersant*, September 15, 2018. <https://www.kommersant.ru/doc/3410434>. Accessed 22 Dec 2018.
- . 2019. Wildberries v I kvartale 2019 goda uveličil oborot na 85% [Wildberries Show 85% Turnover Growth in the I Quarter of 2019]. *Kommersant*, April 11, 2019. <https://www.kommersant.ru/doc/3940021>. Accessed 13 Apr 2019.
- Kozinets, Robert V. 2010. *Netnography: Doing Ethnographic Research Online*. Sage Publications.
- Mediascope. 2019. Statistics on Avito.ru. January 2019. <https://webindex.mediascope.net/report?byGeo=1&byDevice=3&byDevice=1&byDevice=2&byMonth=201901&cid=15828>. Accessed 12 Apr 2019.
- Mihajlova, Lera. 2018. Glava servisa arendy vešej Rentmania rešil perevesti servis v SŠA iz-za otsutstviã finansirovaniã v Rossii [The Head of Renting Service Rentmania Has Decided to Move the Service the USA Due to Lack of Financing in Russia]. <https://vc.ru/flood/38904-glava-servisa-arendy-veshchey-rentmania-reshil-perevezti-servis-v-ssha-iz-za-otsutstviya-finansirovaniya-v-rossii>. Accessed 28 Dec 2018.
- Morgan Stanley. 2018. Russia Ecommerce: Last But Not Least. <https://www.cemat-russia.ru/news/expo/6890-morgan-stanley-ru-ecom-internet.pdf>. Accessed 28 Dec 2018.
- Pantano, Eleonora, and Alessandro Gandini. 2018. Shopping as a 'Networked Experience': An Emerging Framework in the Retail Industry. *International Journal of Retail & Distribution Management* 46 (7): 690–704. <https://doi.org/10.1108/ijrdm-01-2018-0024>.



- PayPal Inc. 2018. Cross Border Consumer Research 2018. [https://www.paypalobjects.com/digitalassets/c/website/marketing/global/shared/global/media-resources/documents/PayPal\\_Insights\\_2018\\_Global\\_Report.pdf](https://www.paypalobjects.com/digitalassets/c/website/marketing/global/shared/global/media-resources/documents/PayPal_Insights_2018_Global_Report.pdf). Accessed 28 Dec 2018.
- Pine, Joseph B., and James Gilmore. 1998. Welcome to the Experience Economy. *Harvard Business Review*, July–August, 1998. <https://hbr.org/1998/07/welcome-to-the-experience-economy>.
- Polukhina, Elizaveta, and Anna Strelnikova. 2014. Ot izbytočnogo potrebleniâ – k ustojčivomu: fenomen onlajn daroobmena [From Overconsumption to Sustainable Consumption: The Phenomena of Free Online Exchange]. *Labirint*, no. 2: 87–96.
- . 2015. Exploring the Russian Online Gift-Exchange Communities: The Results of Netnographic Approach. *The Qualitative Report* 20 (12): 2041–2049.
- Popova, Natalia. 2019. Wildberries obošel H&M i Zara po ob’emu trafika sredi onlajn-magazinov odeždy [Wildberries Outpaced H&M and Zara in the Amount of Traffic Volume Among Fashion Online Retailers]. *FashionUnited*, March 12. <https://fashionunited.ru/novostee/receytil/wildberries-oboshel-h-m-i-zara-po-obemu-trafika-sredi-onlajn-magazinov-odezhdy/2019031225091?fbclid=IwAR0siPJGk5Mwujtrhs7PqG53sKVb0nbpD7bVuqQiGgupu0aQyOX5909gPg>. Accessed 13 Apr 2019.
- Radon, Anita. 2012. Counterfeit Luxury Goods Online: An Investigation of Consumer Perceptions. *International Journal of Marketing Studies* 4 (2): 74–79.
- Rebiazina, Vera A., Anastasia V. Shalaeva, and Maria Smirnova. 2018. Do Russian Consumers Understand and Accept the Sharing Economy as a New Digital Business Model? In *Digital Transformation and Global Society*, 386–401. Cham: Springer.
- Rigby, Darrell K. 2011. The Future of Shopping. *Harvard Business Review*, December 2011. <https://hbr.org/2011/12/the-future-of-shopping>.
- Ruckenstein, Minna. 2017. Digital Consumption. In *Routledge Handbook on Consumption*, ed. Margit Keller, Bente Halkier, Terhi-Anna Wilska, and Monica Truninger, 562–575. London and New York: Routledge.
- Russia Business Today. 2018. Russian Government Postpones Duty Threshold Cut for Internet Shopping Abroad. <https://russiabusinesstoday.com/economy/russian-government-postpones-duty-threshold-cut-for-internet-shopping-abroad/>. Accessed 29 Dec 2018.
- Russian Association of Internet Trade Companies. 2017. Ecommerce Report Russia 2017. [https://www.arvato.com/content/dam/arvato/documents/reports/russia-reports/Russian\\_Country%20Report\\_2017\\_%20Ecommerce%20Foundation\\_Arvato\\_Rus.pdf](https://www.arvato.com/content/dam/arvato/documents/reports/russia-reports/Russian_Country%20Report_2017_%20Ecommerce%20Foundation_Arvato_Rus.pdf). Accessed 28 Dec 2018.
- . 2018. Position of RAITC on the Issue of Cross-Border Shopping. <http://www.akit.ru/position-cross-border/>. Accessed 22 Dec 2018.
- Sadyki, Marina. 2017. *National Report on E-commerce Development in Russia*. Working Paper. Series: Inclusive and Sustainable Industrial Development. United Nations Industrial Development Organisation. <https://www.unido.org/api/opentext/documents/download/9920890/unido-file-9920890>. Accessed 28 Dec 2018.
- Schor, Juliet B., and Craig J. Thompson. 2014. *Sustainable Lifestyles and the Quest for Plenitude: Case Studies of the New Economy*. New Haven, CT: Yale University Press.
- Shlyahov, Evgenij. 2017. Velikaâ šeringovaâ revolúciâ: kak zarabotat’ na arende vsego na svete [Great Sharing Revolution: How to Earn on Renting Out Everything Possible]. *Biz360*, January 16, 2017. <https://biz360.ru/materials/kak-zarabotat-na-arende-vsego-na-svete/>. Accessed 28 Dec 2018.

- Skuratova, Maria. 2018. Dan' gosudarstvu: čem opasny ograničeniâ na onlajn pokupki [Custom Duties to the Government: What Is Dangerous About Limits on Online Purchases]. *RBC*, September 17, 2018. [https://www.rbc.ru/opinions/own\\_business/17/09/2018/5b9b9f8b9a7947cb8f082b1e](https://www.rbc.ru/opinions/own_business/17/09/2018/5b9b9f8b9a7947cb8f082b1e). Accessed 22 Dec 2018.
- Strelnikova, Anna, and Elizaveta Polukhina. 2014. 'Čto takoe povsemestnoe darenie? Èto maksimal'noe doverie drug k drugu': osobennosti social'nogo porâdka v virtual'nyh soobšestvah daroobmena ['What Is All-Around Gift-Giving? This Is the Highest Possible Level of Trust Towards Each Other': Peculiarities of Social Order in Virtual Gift-Giving Communities]. *Inter*, no. 7: 7–21.
- Sun, Leo. 2018. JD.com Could Be Returning to Russia. *The Motley Fool*, July 5, 2018. <https://www.fool.com/investing/2018/07/05/jdcom-could-be-returning-to-russia.aspx>. Accessed 28 Dec 2018.
- TASS. 2017. 'Nalog na Alièkspress': smogut li rossiâne delat' pokupki v onlajn magazinah? ['Aliexpress Tax': Will Russians Be Able to Shop in Online Shops?]. *TASS*, September 15, 2017. <https://tass.ru/ekonomika/4564606>. Accessed 27 Dec 2018.
- Tishina, Yuliya. 2018. Onlajn pokupki ne otliačûtsâ original'nost'û: rossiâne vybiraût kontrafakt [Online Purchases Are Not Quite Unconventional: Russians Do Choose Pirate Goods]. *Kommersant*, December 21, 2018. <https://www.kommersant.ru/doc/3836670>. Accessed 14 Apr 2019.
- Treapăt, Laurentiu-Mihai, Gheorghiu Anda, and Marina Ochkovskaya. 2018. A Synthesis of the Sharing Economy in Romania and Russia. In *Knowledge Management in the Sharing Economy*, ed. Elena-Mădălina Vătămănescu and Florina Pinzaru, 57–73. Springer: Cham.
- Urbanovsky, Tomas. 2015. Factors Behind the Russian Ruble Depreciation. *Procedia Economics and Finance* 26: 242–248. [https://doi.org/10.1016/S2212-5671\(15\)00827-8](https://doi.org/10.1016/S2212-5671(15)00827-8).
- Utesheva, Galina. 2018. Offlajn ne umret: 4 trenda radikal'noj transformacii šopinga [Offline Will Not Die: 4 Trends of Shopping Radical Transformation]. *FashionUnited*, December 24, 2018. <https://fashionunited.ru/novostee/reetyeil/offlajn-ne-umret-4-trenda-radikalnoj-transformatsii-šopinga/2018122424347>. Accessed 25 Dec 2018.
- Vedomosti. 2017. Alièxpress oprobuet v Rossii magaziny virtual'noj real'nosti [Aliexpress Will Test Virtual Reality Stores in Russia]. *Vedomosti*, November 3, 2017. <https://www.vedomosti.ru/business/news/2017/11/03/740493-aliexpress-virtualnoi-realnosti>. Accessed 27 Dec 2018.
- Wahlen, Stefan, and Mikko Laamanen. 2017. Collaborative Consumption and Sharing Economy. In *Routledge Handbook on Consumption*, ed. Margit Keller, Bente Halkier, Terhi-Anna Wilska, and Monica Truninger, 94–105. London and New York: Routledge.
- Yandex. 2016. Razvitie rozničnoj online trgovli v Rossii [Development of Online Retail Trade in Russia]. *Yandex*, December 21, 2016. [https://yandex.ru/company/researches/2016/ya\\_market\\_gfk](https://yandex.ru/company/researches/2016/ya_market_gfk). Accessed 28 Dec 2018.
- Yandex Kassa and Data Insight. 2018. Pokupki i pokupateli v social'nyh setâh, messendžerah, na sajtah ob"âvlenij i šaring economy [Purchases and Customers in Social Media, Messengers, on Classified Sites and Sharing Economy]. *Yandex Kassa and Data Insight*, December 26, 2018. <http://datainsight.ru/sites/default/files/ecombehavior.pdf>. Accessed 12 Jan 2019.
- Zhang, Autumn. 2018. 2017 Statistics on the Chinese Cross-Border Online Shopping Market. *Eggplantdigital*, May 21, 2018. <https://eggplantdigital.cn/2017-statistics-on-the-chinese-cross-border-online-shopping-market/>. Accessed 28 Dec 2018.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Digital Art: A Sourcebook of Ideas for Conceptualizing New Practices, Networks and Modes of Self-Expression

*Vlad Strukov*

## 14.1 INTRODUCTION

Computer-enabled, digital technologies have altered the ways in which art is produced, experienced and thought of. For example, in the 1990s, European, North American and Russian art museums and galleries developed multi-media products—Compact Discs (CD), Compact Disc Read-Only Memory (CD-ROM) and Digital Versatile Discs (DVD)—featuring images of artworks from their permanent collections along with critical commentary. The user was now able to appreciate works of art on the computer screen, and not just in the space of the gallery or in a book format. The user was also able to modify the image of an artwork, or add it to their personal web page, thus emerging as an “active” consumer of art. In the same period, online galleries appeared on the internet, competing with established institutions. In the early 2000s, online galleries emerged. For example, the Olga Gallery ([abcgallery.com](http://abcgallery.com)) was set up by teenage brothers Yury and Sergej Mataev, who published catalogued works by famous artists, first Russian and later world masters. Their clandestine gallery became an important teaching tool for those in the field of Russian Studies and Arts, providing easy access to quality reproductions of artworks.

At the same time, large museums started to provide online tours of their galleries. The Russian State Hermitage Museum was a pioneer of innovative virtual tours. On the one hand, the museum allows users anywhere in the

---

V. Strukov (✉)  
University of Leeds, Leeds, UK  
e-mail: [v.strukov@leeds.ac.uk](mailto:v.strukov@leeds.ac.uk)

world to experience its galleries online. On the other, visitors to the museum in St. Petersburg can watch 3D movies and become a witness of historical events that had taken place in the Winter Palace. These experiments with virtual reality occurred at the same time as the production of Aleksander Sokurov's 2002 *Russkij kovčeg* (Russian Ark), a movie that was shot entirely in the Winter Palace of the Russian State Hermitage Museum on 23 December 2001 using a single-take single 96-minute Steadicam sequence shot. Russian Ark has become a digital artwork itself insofar as it had challenged existing theories of film and audio-visual presentation, paving the way for experiments with digital filmmaking in Hollywood and elsewhere.<sup>1</sup> With the rise of social media in the late 2000s, museums started to use digital technologies, including providing immersive experiences so that the visitor can enjoy art across different platforms. Garage Museum of Contemporary Art in Moscow leads the way in terms of using digital technologies in its various inclusivity and access programs such as those for deaf people and people with visual impairments.

Digital technologies have changed the ways in which museums and galleries operate, including the kind of objects and practices they acquire for their collections. The debates about what constitutes art and how to collect, curate and exhibit it are ongoing. Digital art is commonly understood as a form of art produced, distributed and appreciated with the help of digital technologies. For the purpose of this chapter, I limit this definition to that kind of art which exists exclusively in the digital form. For example, an installation featuring objects, photographs and a digital component such as digital animation has been eliminated from my consideration. This process of elimination is not discriminatory but empowering because it makes one wonder about some principal notions helping us understand the nature and purpose of art. For example, is the digital a new medium or a new form of expression? Is digital just another way to say "contemporary"? Does the digital convey new forms of subjectivity or does it translate existing issues into a new "language"?<sup>2</sup>

The discussion is based on the analysis of specific works of art, archival work, interviews with artists<sup>3</sup> and critical assessment of exhibitions, biennales and festivals of contemporary art. The discussion is organized around two nodes: (a) historical and artistic contexts and (b) the scope and dynamics of Russian digital art. In the first instance, the chapter traces the evolution of digital technologies, artistic practice and cultural and aesthetic transformations. In the second instance, the chapter supplies a conceptualization of a diverse range of artworks around the notion of image transformation thanks to new digital technologies. All artworks and images discussed in the chapter are available in the public domain and so can be easily found in Wikimedia commons and other sites.

## 14.2 RE-STRUCTURING THE IMAGE:<sup>4</sup> OLGA TOBRELUTS AND THE DIGITAL COLLAGE (THE 1980s AND THE 1990s)

Computer technologies and digital literacy is one of the key components to a successful economy. This was recognized at the state level already in the 1980s under Soviet late socialism. Articulated as an imperative to develop new means of automation, the policy of digitalization was at the core of Mikhail Gorbachev's *perestroika*, which translates into English as "re-structuring." It aimed to supply new, more efficient means to carry out planning and management for the Soviet economy. It encompassed a development of a few generations of computers and computational technologies and a development of a workforce capable of operating complex machinery and running computer programs. These goals were achieved thanks to professional training made available to school pupils, students and those already in employment through re-training programs. As a result, in the late 1980s there was a large supply of engineers and other personnel involved in the production and maintenance of computer technologies (for more on digital education, see Chap. 10).

They were involved—often anonymously—in early experiments with computer art, or the application of algorithms capable of producing or copying artworks. These included, for instance, images rendering famous works of art with the help of zeros and ones, thus visualizing the computer code. Other experiments involved visualizations of mathematical formulas such as fractals. These are curves or geometrical figures, each part of which has the same statistical character as the whole. Fractals are used in modeling of complex structures such as snowflakes. On one level, digitally produced images incorporating and imitating fractal laws had the characteristics of geometrical patterns and so appeared as decorative elements. On another, they were works of art insofar as they enquired about the laws of the physical world and their mathematical representations, making references to abstract art and its predecessors such as the Russian avant-garde. These artworks were exchanged freely among communities of technical intelligentsia who experimented with computer technologies, moving beyond their utilitarianism and producing artworks. In doing so they embraced Gorbachev's neoliberal reforms such as unregulated information exchanges, the privatization of national resources and self-sufficiency.

The growing availability of computers and the emergence of new software such as Adobe Illustrator meant that artists started experimenting with digital art. In the early 1990s, in St. Petersburg, a young artist Olga Tobreluts (b. 1970) joined the art scene after making friends with Timur Novikov, an influential art manager and curator. At that time Novikov was pre-occupied with hangings made of different kinds of fabric and decorated with appliqués. These were "textile collages" aimed at re-organizing space in novel ways. Later he presented his ideas in the form of a theoretical treatise in which he called his visual experiments "*perekompoziciá*" (re-compositions), a term which designates re-modeling and re-structuring of space (Andreeva 2007). Tobreluts

responded to Novikov's ideas by making digital collages. She learned computer graphics and 3D modeling while on a visit to Berlin. On her return to St. Petersburg, she produced a series of images that featured digital "re-compositions." They were shown at exhibitions in Russia, the United Kingdom (UK), the United States of America (USA) and other countries, securing Tobreluts the title of "a leading Russian digital artist" (Geusa 2013).

The novelty of her work was in the realistic "effect": instead of rejecting perceptive realism of classical art, Tobreluts utilized it to query the status of image and illusion in the digital era. For example, her project *Models* from the late 1990s consisted of re-interpretations of classical art for the digital era. Tobreluts followed the conventions of traditional portraiture by choosing a "head and shoulders," full face or three-quarter view, and depicting her subjects with a thoughtful expression of face. She enhanced the conventionality of her portraits by using sculptural elements available from antiquity. At the same time, Tobreluts challenged the viewers' perceptions by applying bright colors and making use of symbols from popular culture, for example, the Lacoste fashion brand. Ultimately, the artist enquired about the value of art and individual expression in the era of digital reproduction. Here originality stems from a "re-composition" of elements, not from new elements. Tobreluts "re-structures" the artistic canon and the image itself by accentuating the composite quality of culture and memory. She conceives of digital art as a new medium, and by employing classical imagery she re-inscribes the digital into art history.

In an interview published in 1995, Tobreluts defines digital art in the following way. "First, the work is composed of different pieces. Then it is transferred from the computer to a compact disc (CD). Then a negative is printed, and then a photograph is printed ... The computer is a stupid machine. It is just a metal box that can do nothing unless it is instructed to do something"<sup>5</sup> (Sharandak 1995). Tobreluts describes different stages in the production of a digital artwork whereby the digital is materialized, that is, different manipulations are used to present the digital as an object. Different stages in the production of the artwork refer to the process of layering employed in image editing programs such as Photoshop. Another artist—Natalia Kamenetskaia from Moscow—described the same process in an interview in 2011. Speaking of her digital collage titled *St. Sebastian* and produced in 1993 with the help of Photoshop, she notes that "St. Sebastian is a multi-layered, poly-semantic figure which brings together images and characters from classical and contemporary art" (Strukov 2011, 123–124).<sup>6</sup>

In her 1995 interview, Tobreluts conceived of computers and digital technologies as a new medium. She compared them to "a new kind of brush which is just more convenient to use"<sup>7</sup> (Sharandak 1995). In my interview with her in 2017, Tobreluts spoke about "*cifrovaâ èstetika*" (digital visuality), or a particular way of thinking about the world, not just representing it artistically (interview with the author, 2017). In other words, in twenty years Tobreluts' understanding of computers and digitality has evolved from one which considers the digital as a more efficient medium to one which utilizes the digital to



construct new worlds. The change in her thinking is manifested artistically: from using the digital to re-structure the image in the 1990s, in the 2010s she turned to using the medium of painting to reveal the nature and dynamics of the digital. I argue that this reversal of her artistic focus reveals the transformations propelled by the greater use of digital technologies in the present-day society.

In the early 1990s, Tobreluts, Kamenetskaia and other artists centered on the image as a key component of artistic expression. Their attempts to “re-structure” the image using digital technologies resulted in a new understanding of artistic originality and authorship. Like their predecessors such as Marcel Duchamp, Andy Warhol and Ilya Kabakov, Russian digital artists queried art as an autonomous sphere of production. They continued to challenge the notion of the artistic genius by engaging with technologies that they could not fully control.<sup>8</sup> Kamenetskaia acknowledges that “the computer was an unpredictable thing that would generate unplanned, unexpected results. Working with a computer was a mystical process” (Strukov 2011, 122–123). On one level, Kamenetskaia ascribes some degree of authorship to the machine which, in her view, is responsible for the outcome without discernible human intention. Like Dadaists, she embraces chance as a stimulus to expression in the work of art. Like Pollock, who practiced the technique called “Action Painting,” which relied on chance, she is interested in random connections generated by the computer software.

On another level, Kamenetskaia re-claims ownership of art as a collective enterprise, thus opposing the long-standing tradition of perceiving art as a result of individual expression, or Romantic genius. She reminisces (Strukov 2011) that in the early 1990s she did not own her own computer and made use of her friends’ computers, for example, of a computer that belonged to Irina Sandomirskaia, now a professor of Russian Studies at Södertörn University. Kamenetskaia would spend hours working on her computer at night. According to the artist, it was more than borrowing some tools from a friend; rather it was a collective enterprise insofar as they wanted to achieve something new in their work, namely, to open to the global community. Kamenetskaia recalls Sandomirskaia saying that “by learning how to use the computer we can show to the western world that we are part of it. The computer was a language in which all modern people communicated but Russians not yet” (Strukov 2011, 123).<sup>9</sup> In this regard, Kamenetskaia and her friend, perhaps unknowingly, rehearsed the vision of global solidarity originally articulated by Sergei Eisenstein for the medium of film. For him, film would be a universal language, one that does not require translation, which would unite people of the world (2007 [1934]).

### 14.3 RE-WIRING THE EAST: OLIA LIALINA AND NET.ART (THE 1990s)

These ideas of shared knowledge, collective authorship and international solidarity were at the core of an artistic movement known as net.art. The main members of the movement were Vuk Ćosić, [Jodi.org](#), Heath Bunting, Aleksei Shulgin and Olia Lialina, based in countries that just a few years ago were separated by the Iron Curtain. To achieve a new post-Cold War commonality, they formed an artistic collective, defining their art as “net.art,” or “internet art.” Though they wished to explore similar political and social concerns, from the aesthetic standpoint their works were very different. Net.art is a synonym of “internet art.” According to Shulgin, who allegedly coined the term, net.art stemmed from “conjoined phrases in an email bungled by a technical glitch (a morass of alphanumeric junk, its only legible term net.art)” (Greene 2004, 12). The term has been used in the title of various exhibitions celebrating internet art. It covers a wide range of artistic practices that use the internet as its main medium.

One of the most celebrated net.artists is Olia Lialina (b. 1971). She is widely recognized for developing the internet as a medium for artistic expression and storytelling. For example, her network-based artwork *My Boyfriend Came Back from the War* (1996) tells the story of a young woman and man who have been separated by war. To a Russian user, Lialina makes a reference to the first Chechen War, which had devastated the newly founded Russian Federation (RF); to other users, she speaks of a universal situation. The lovers attempt to engage in a conversation, but they find it difficult. It is not entirely clear whether they are communicating in the “real” or online world; the boundaries between spaces, lines of communication and identities are constantly blurred, creating a Chekhov-style drama of misunderstanding. Unlike other examples of net.art, *My Boyfriend Came Back from the War* is directly involved with the user’s emotions. In fact, the work reflects on what constitutes expression, meaning and emotion on the internet. In many ways, it anticipated the conflicts and dramas of social media which are to appear a decade later.

*My Boyfriend Came Back from the War* makes use of interactive hypertext storytelling. The work consists of nested frames with black and white web pages and grainy GIF images that show human faces and objects. Lialina conceives of the internet as a space where the boundaries between words and images, and between connections and emotions, are erased. Each element is an arena of action, reflection and observation. When clicking hyperlinks in the work, the frame splits into smaller frames and the user reveals a nonlinear story about the couple. The story takes on a number of routes but eventually it leads to the point where the screen becomes a mosaic of empty black frames. They stand for emotional emptiness, a breakdown in communication and impossibility of genuine dialogue in the modern world (for more on hypertext, see Chap. 15).

On one level, the squares and frames make a reference to the film strip, that is, a roll of frames. The grainy black-and-white images and intertitles evoke early silent movies. Like Eisenstein, Lialina is interested in montage as a means to construct meaning on the internet. On another level, the work reveals the potentialities of the internet as a new medium, particularly the role of the user in assembling data and constructing meaning. Without the user, the frames and images in *My Boyfriend Came Back from the War* would remain static. With the user's involvement they become animated. Here, reading the story is a ludic experience insofar as the user is guided but not directed to act, thus producing new connections and exploring new spheres of meaning. The user begins to wonder about their role and about the impact of their actions: are they there to observe an intimate conversation between a man and a woman? Are they responsible for the breakup of communication?

*My Boyfriend Came Back from the War* was displayed in Lialina's online gallery, which was one of the first internet-based galleries in the world. Nowadays artists employ the internet to produce, showcase and distribute their work, with many artists boasting profiles on numerous platforms. What Lialina has been interested in is the exploration of the possibilities of the new medium, on the one hand, and, on the other, the challenges of preserving early internet art and culture for future generations. With many programs now obsolete, how can a user experience the internet of the 1990s? Particularly, how can they feel the joy of connecting with someone they do not know in another country? This seems banal in the present-day world, but in the early 1990s with the world just emerging from the Cold War, being able to communicate directly with someone from another country was an extraordinary experience. What net.artists did in that period was to re-wire Europe and re-connect the world in new ways that would be free of government controls, ideological blocks and national, racial and gender stereotypes. *My Boyfriend Came Back from the War* is a record of this kind of aspiration of the post-Cold War Europe.

In her pioneering net.art, Lialina poses a number of important questions. The ethical ones are: what is the nature of communication? How does the internet change communication? What is privacy? How can we be intimate when there is no privacy? And the aesthetic questions are: what is duration on the internet? How do users define time? Does the digital have its own ontology? What kind of visuality and visibility does the digital supply? Is it possible to conserve the digital? In other words, *My Boyfriend Came Back from the War* and Lialina's other works are about knowledge and its calibrations and misnomers, about the scale and trajectory of communication and performance, and about the difference between connectivity and community. Lialina's works are simultaneously contextual—they exist within a specific technological and social context—and universal as they speak of global issues and assert universal values.

#### 14.4 MINI AND MAXI: GLOBAL VISIONS FROM OLEG KUYAEV AND AES+F (THE 2000s AND 2010s)

While Lialina's works are significant from the standpoint of art criticism, history of communication and theory of the internet and the digital, they remain marginal from the standpoint of popular cultural industries and global consumption. Who were the artists who made digital art popular? Conversely, how did artists respond to the rise of popular use of digital technologies? What effects did the changes in technologies have on the aesthetics, distribution and significance of digital art? In this section, I aim to answer these questions by addressing two interrelated concerns. The first is the role of individual artists in the development of the cultural industry with its digital segment. The second is the transnational realm of Russian culture in general and digital art in particular. Indeed, my analysis of the works by Tobreluts and Lialina indicates that Russian digital art has been international from its inception. Here I wish to emphasize that it has always occupied a transnational domain. For instance, Tobreluts' collages signify the process of symbolic layering of culture in the era of globalization. She mixes tropes and forms stemming from different periods and contexts, and, following the imperial tradition of artistic expression such as the classical architecture of St. Petersburg, what makes her works Russian is the radical appropriation of seemingly un-Russian symbols. She reveals subjectivity through renouncing identities, or, to be precise, by demonstrating their constructed nature. With Lialina, transnational social networks define the processes of articulation and dissemination of her art. She works with artists based in other countries, and she makes art which is possible thanks to the actions of users located anywhere in the world. Lialina's interest in specificity and universalism points to the effects of global communication networks which, on the one hand, allow us to connect to anyone anywhere and, on the other, keep us trapped in our information bubbles. In addition, I argue that individual artists, not government-funded or corporate initiatives, are responsible for the emergence of cultural industry and digital economy in the RF.

The developments occurred at different levels and through employment of sundry strategies. Here I reflect on two of these, which I coded using the terms "mini" and "maxi." The former stands for a particular sense of intimacy, personal space, reflexivity and a steer toward abstraction (see the discussion of Lialina's works above). The latter signifies an infatuation with popular culture, spectacle and a steer toward figuration. To showcase the latter, I first investigate the work of Oleg Kuvaev before turning to the art collective known as AES+F (the name is initials of the artists Tatiana Arzamasova, Lev Evzovich, Evgenii Sviatskii and Vladimir Fridkes). Kuvaev's work characterizes the tendencies of the early 2000s while AES+F address the concern of the late 2000s and early 2010s.

In 2001, Kuvaev (b. 1967 in St. Petersburg) founded a small animation studio called Mult.ru and started promoting *Masyanya*, a series of short clips about the adventures of a young girl called Masyanya who lives in St. Petersburg

with her boyfriend. Kuvaev worked with Macromedia Flash to produce films that were distributed over the network using viral marketing. Macromedia Flash uses vector technology to produce layered imagery. It appears quite simple—geometric lines, bright colors, lack of shading, and so on, but this simplicity, or rather naivety, was the key to success. In a few years, and in spite of Kuvaev being involved in a legal battle over his brand,<sup>10</sup> *Masyanya* was the most popular phenomenon on the Russian language internet, linking communities in the RF, Europe, Israel, North America and elsewhere. Some describe the 2000s as “Putin’s Russia” due to the rise of the new form of governance associated with the figure of the president (see, for example, Wegren 2018). I argue that the 2000s were “*Masyanya*’s Russia” (see Strukov 2004 for full analysis) because Kuvaev and his *Masyanya* transformed the ways in which people communicated online, and gave rise to digital economy (for more, see Chap. 4).<sup>11</sup>

Kuvaev employs caustic humor and depicts *Masyanya*’s absurd behavior while reflecting on the struggles of the young generation of Russians who had been affected by neoliberal reforms. Visually, *Masyanya* is an example of naïve, or primitive, art, that is, art that (looks as if it) was produced by non-professional artists. Elsewhere (Strukov 2004), I called *Masyanya* “a visual anecdote,” meaning that the series functions as a digital form of joke-telling which has traditionally characterized Russian culture. Indeed, *Masyanya* has the qualities of humorous GIFs and memes, making it an alternative to commercial, mainstream culture. It is also a good example of how niche digital art may become popular. On the one hand, *Masyanya* resisted the dominance of Hollywood<sup>12</sup> with its specific visual language and symbolic economy. On the other, it constructed its own alternative form of globalization based on principles of free labor, pirating and sharing. These practices have become commodified and commercialized since the emergence of Western social media giants such as Facebook and Instagram. *Masyanya* spoke of community, intimacy and honest conversation before they became catch phrases in the new global digital economy.

AES+F are also interested in the effects of digital globalization on local communities. Their award-winning multi-channel digital video installation *Allegoria Sacra* (2011–2013) shows some passengers stranded in an international airport. The location alludes to Arthur Hailey’s eponymous novel which has been hugely popular in Russia. It represents a global community stuck in some kind of temporal warp. The title of the video is of course a reference to Giovanni Bellini’s painting (1490–1500) which represents the purgatory. Their artwork speaks of limbo and of the intemporality of the internet where everything is available forever and yet changes and disappears all the time. AES+F present a series of biblical figures, mythological creatures, cyborgs, clones and so on who are transposed into the eternal realm of Bellini’s painting. Like Tobreluts, AES+F adopt classical forms for the digital environment when, for example, the Saracen-Muslim is transformed into a group of refugees and St. Sebastian turns into a young, shirtless traveler, hitchhiking his way through

tropical countries. Yet, AES+F's artwork is more of an allegory of the contemporary life than a postmodern reinterpretation of Bellini's painting.

*Allegoria Sacra* weaves complex global issues such as the refugee crisis, global warming, identity politics, and gender and sexuality into visually rich metaphors. The group conceives of the digital as the element that holds the global society together. However, it is not clear whether this hold is a genuine bond or, in fact, a form of captivity. Like Lialina, AES+F are concerned with the issues of identity, privacy, freedom and choice. *Allegoria Sacra* reflects on human condition from a Russian yet global perspective. This global vision is accounted for by the artwork outreach—it has been shown at art venues all over the world—and it is encoded aesthetically through the use of a multi-screen projection which creates an extraordinary spectacle of performance and immediacy such as the slow digitally enhanced movement of characters and objects against the pulsating background. The massive scale of the project—the digital maxi—is also a reflection on the spectacularity of the digital, its omnipresence and panopticism. If Kuvaev ignited Russian digital economy by supplying a product that speaks of intimacy, community and commonality, AES+F showcase the might of this digital economy as they orchestrate a global show of connectivity and (mis)communication. All the artists address ethical and aesthetic questions posed by Lialina a decade ago, which suggests that these questions remain unanswered. This leads me to enquire about the legacy of digital art experiments in the RF.

#### 14.5 THE DIGITAL ARCHIVE: CYLAND AND CYFEST (THE 2000S AND THE 2010S)

After early experiments since the late 1980s, in the 2010s digital art has become a mainstay of Russian contemporary art scene. For example, there are art galleries that specialize in showing digital art, such as the Multimedia Art Museum headed by the diva of the Russian art scene Olga Sviblova and the Solyanka Art Gallery, which hires young curators to stage shows. Both are located in the center of Moscow and both are sponsored by the government. However, if the Multimedia Art Museum puts on big exhibitions showing blockbusters such as AES+F's *Allegoria Sacra*, for which the Museum gets sponsorship from Russian oil and gas monopolies, the Solyanka Art Gallery is a small space, hidden away from the tourist crowds and specializing in edgy, intellectually challenging exhibitions of international artists and artists from Russian regions. In addition to art spaces, there are numerous mergers—art and fashion as well as art and technology spaces—which include digital art in their programs. For example, Art Play Design Centre in Moscow stages immersive digital shows that enable the visitors to interact with artworks and digital environments.<sup>13</sup> This type of exhibition does not engage with innovative technologies and complex issues; however, they do attract wider audiences to museum spaces, thus promoting digital art generally. Another example would be the use of digital art in popular

culture, such as 3D projections and immersive videos during live concerts of the Ukrainian-born Russian singer Svetlana Loboda.

The burning issue facing Russian cultural managers is not the promotion of digital art but its preservation. Indeed, how does one conserve pieces produced using obsolete technologies like Lialina's *My Boyfriend Came Back from the War*? And how does one ensure that the Russian public, especially in Russian regions, remains aware of advances in digital art nationally and internationally? While these issues are being acknowledged in the professional community (Biryukova 2018), more work is needed in this direction. At present, no Russian national (federal) museum of digital art exists, and principal museums do not list digital art as their priority area in terms of acquisition. Three major institutions—the Hermitage, the Tretyakov Gallery and the Russian Museum—have departments specializing in contemporary art but acquisition of digital art is still very rare. This reveals a gap between artistic practices and the cultural economy whereby there is a perceived lack of national strategy in terms of promotion and preservation of digital art. For example, digital art does not feature in the nationally funded government-led program of digitalization of Russian economy introduced by President Dmitry Medvedev, and discussions in the Russian government and parliament tend to focus on digital literacy, which is, in fact, a re-hash of Gorbachev's policies of perestroika (“greater automation and greater efficiency”), and on digital security, which is in actual terms a string of legislation limiting freedoms of communication on the internet.

As a result, the arena of preservation of digital art has been occupied by private initiatives. One of the most influential ones is Cyland Media Lab. Founded in 2007, Cyland is a non-profit organization dedicated to digital art and broadly the intersection of art and technology through exhibitions, a collection of art, and educational programming. Overall, Cyland aims to connect emerging and established artists, educate how to use creative technology and foster innovation in new technologies (<http://cyland.org/lab/about/>). Co-founded by Marina Koldobskaja and Anna Frants, Cyland is sponsored by Frants, who, in addition to being a philanthropist, is an internationally renowned multi-media artist specializing in interactive art installations. Cyland collaborates with museums such as the Hermitage and the Chelsea Art Museum (New York, USA), but it has an ambition to build a museum of its own. For a decade Cyland has been building an online collection of artworks. Divided into a video archive and a sound archive, Cyland's online collection is a comprehensive survey of Russian and international art (over 100 individual artists and groups from the RF). Video and sound are understood as a means to categorize works, whereas in actual terms, the collection, managed by Viktoria Ilushkina, features video art, experimental films, computer graphics, 3D animation and so on. The collection reveals the technological, platform and genre diversity of what is understood as digital art.

In addition to an online collection, Cyland is committed to promoting digital art nationally and internationally through Cyfest. Running since 2008, Cyfest is an annual festival celebrating digital and new media art. The main part



of the festival takes place at different venues in St. Petersburg, and some parts of the festival at exhibitions in partner institutions in London, New York, Venice and other places. As with similar festivals in other countries, Cyland festivals are themed; for example, in 2019 the theme was “ID,” and in the previous year it was “Digital Cloudness.” These themes refer to pressing social, political and aesthetic concerns in the contemporary world. Unlike *Ars Electronica* in Linz, Austria, Cyfest is a much more focused enterprise with a commitment to experimentation and community building, and not city branding and industry collaborations. Cyfest remains the principal platform for showcasing experimental digital art in the RF. The legacy of Cyland and similar initiatives is to be assessed in future research.

On the one hand, Russian digital art is frequently presented at international art festivals such as Cyfest. On the other, a national museum or archive of computer-based and digital art is to be formed. This is highly unusual for a country obsessed with museums and museification. In fact, digital artworks are still to be included in permanent collections of existing museums such as the Russian Museum in St. Petersburg and the Tretyakov Gallery in Moscow. Similarly, a history and a theory of Russian digital art and new media art are to be written. In this context of research possibilities and probabilities, an essential history of Russian art allows for an in-depth understanding of the development of internet technologies in the RF (for more on types of digital archives, see Chaps. 20 and 21).

Nowadays the internet is a mundane thing and users are more likely to speak of specific platforms such as VKontakte or Twitter. In the mid-1990s the internet was a novel phenomenon which relied on the user’s advanced technical knowledge and produced an important effect of instantaneous connectivity in a world where people still used landline telephone connections, faxes and telegrams to communicate with each other. Indeed, instantaneity of communication and production of online social networks were two focal points of net. artists. They employed a variety of techniques some of which would be considered dubious by present-day users, such as fake websites, spam mails and unsolicited distribution of information. Their purpose was to explore networked modes of communication and interplays of exchanges. They understood collaborative and cooperative work differently whereby they frequently delegated the production of the artwork to the user, not just to other members of the artistic community. Ultimately, they aimed at working across national borders, building a digital utopia for the next generation of artists. For many contemporary Russian artists, the digital remains an arena of utopian possibilities to be explored.

## NOTES

1. For an in-depth discussion of the film, see Strukov (2009).
2. On new media as a form of language, see Manovich (2001).

3. Some of these interviews were published in *Studies in Russian, Eurasian and Central European New Media*; see, for example, Strukov (2011).
4. In Russian, “*perestroika obraza*.”
5. [“*Vse montiruetsâ iz raznykh kusočkov. Potom s komp’ûtera peregonâetsâ na lazernyj disk. Pečataetsâ negativ, potom s nego—fotografiâ.... Komp’ûtery—oni že glupye mašiny. Metalličeskaâ korobka, kotoraâ ničego ne možet sdelat’*”, *esli ty ej ne skážeš’, čto nužno sdelat’*].
6. [“*Svâtoj Sebast’ân—èto rabota, v kotoroj sloi kul’tury naslaivaûtsâ. Èto mnogoslonyj, mnogoznačnyj obraz, gde soedinâûtsâ personaži sovremennogo iskusstva—sovremennogo tomu vremeni—i klassičeskie personaži*].
7. [“*Ran’she, kogda pisali kartiny, poâvlâlas’ novaâ kist’, bolee udobnaâ, nikto že ot nee ne otkazyvalsâ*].
8. [“*Komp’ûter byl dlâ menâ istočnikom raznykh neožidannykh vêsěj, kogda ty čto-to delaeš’, i polučaetsâ nezaplanirovannyj rezul’tat. Kakoj-to mističeskij daže process*].
9. [“*Â pomnû frazu Sandomirskoj, kogda poâvilsâ komp’ûter, čto èto očen’ važno, čto my osvoili komp’ûter, potomu čto my pokazali zapadnomu miru, čto my čast’ ètogo mira. To est’ èto byl âzyk, na kotorom govorili vse sovremennye lûdi, a Rossiâ èse net*].
10. Kuvaev lost and in the end decided to emigrate to Israel.
11. In retrospect, it is possible to interpret Masyanya’s adventures as a parody on Putin, who is also from St. Petersburg.
12. See Norris (2012) and Strukov (2016) on the dominance of Hollywood aesthetics on Russian culture in the 2000s.
13. See, for example, their recent Samskara project (<https://www.samskara.pro/> [10.10.2019]).

## REFERENCES

- Andreeva, Ekaterina. 2007. *Postmodernizm: Iskusstvo vtoroj poloviny 20-go — načala 21-go veka* [Postmodernism: The Art of the Second Half of the 20th – Beginning of the 21st Century]. Moscow: Azbuka-Klassika.
- Biryukova, Svetlana. 2018. Iskusstvo interakcii [The Art of Interaction]. Accessed October 10, 2019. <http://rusmuseum.ru/pavilions-mikhailovsky-castle/news/march-27-from-16-00-to-21-00-at-the-media-centre-held-a-roundtable-on-the-art-of-interaction-from-co/>.
- Eisenstein, Sergei. 2007. *Writings, 1934–1947*. London: I.B.Tauris.
- Geusa, Antonio. 2013. Russian Digital Artist Olga Tobreluts Gets Huge Retrospective at Moscow Museum of Modern Art. *Phaidon*. Accessed October 10, 2019. <https://uk.phaidon.com/agenda/art/articles/2013/january/25/russian-digital-artist-olga-tobreluts-gets-huge-retrospective-at-moscow-museum-of-modern-art/>.
- Greene, Rachel. 2004. *Internet Art*. London: Thames & Hudson.
- Manovich, Lev. 2001. *The Language of New Media*. Cambridge, MA: The MIT Press.
- Norris, Stephen. 2012. *Blockbuster History in the New Russia: Movies, Memory, and Patriotism*. Bloomsbury: Indiana University Press.
- Sharandak Natalya. 1995. Komp’ûternyj akademizm [Computer Academism]. Accessed October 10, 2019. <http://www.owl.ru/win/books/sharandak/11.htm>.

- Strukov, Vlad. 2004. Masiania, or Reimagining the Self in the Cyberspace of Rusnet. *The Slavic and East European Journal* 48 (3): 438–461.
- . 2009. A Journey through Time: Alexander Sokurov’s Russian Ark and Theories of Mimesis. In *Realism and the Audiovisual Media*, ed. Lucia Nagib and C. Mello, 119–132. London: Palgrave Macmillan.
- . 2011. Digital Sebastian: Interview with Natalia Kamenetskaia. *Studies in Russian, Eurasian and Central European New Media* 6: 121–127.
- . 2016. *Contemporary Russian Cinema: Symbols of a New Era*. Edinburgh: Edinburgh University Press.
- Wegren, Stephen. 2018. *Putin’s Russia: Past Imperfect, Future Uncertain*. Lanham: Rowman & Littlefield.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# From Samizdat to New Sincerity. Digital Literature on the Russian-Language Internet

*Henrike Schmidt*

## 15.1 INTRODUCTION. THE HYBRID NATURE OF DIGITAL LITERATURE

No clear-cut definition exists to describe digital literature, which is characterized by its hybrid nature and which borders on the fields of information technology, media art, media activism and computer games. Katherine Hayles speaks about “new horizons for the literary,” emphasizing that digital literature transgresses a restricted understanding of literature, in the sense of discrete texts and established literary devices (Hayles 2008, 4). Scott Rettberg alternatively uses the term “electronic literature” (2016, 2019; see also Tabbi 2018; O’Sullivan 2019), underlining also its inherent hybridity and the difficulty, if not impossibility, of working with fixed genre definitions. Electronic literature in his view stands at the crossroads of literary practice and critique and is characterized “by the approach rather than content” (2016, 166). In the following, digital literature is understood accordingly as an umbrella term for “literary practices in digital and networked environments.” Exemplary manifestations of digital literature are animated poetry, text generators producing poems relying on algorithms and hyperfiction, that is, stories told in a digressive, interactive way by using hyperlinks: icons, graphics or text that link to another document or website. For the later phase of increasingly mobile devices, since approximately the early 2010s, one might think of “locative literature” (ibid., 170), in which smartphone apps guide readers through story worlds at real locations,

---

H. Schmidt (✉)  
Freie Universität Berlin, Berlin, Germany  
e-mail: [schmidth@zedat.fu-berlin.de](mailto:schmidth@zedat.fu-berlin.de)

thus enabling a temporal-physical immersion into the narration. Whether digitized texts, that is, works previously published in print and later converted into digital format, can be classified as digital literature is a topic of debate (Bouchardon 2017, 3). What unites all appearances of literature or “the literary” in the digital sphere is the fact that they are computer-processed and thus rely on code. The literary texts, which the readers perceive on the surface of the computer screen, are secondary, products of the underlying primary text of the computer code. They tend to either hide their computer-generated nature, which we can call media in-transparency, or display it openly, exposing the texts’ mediated nature.

Literary practices on the Russian-language Internet (Runet), ranging from online libraries to Facebook life-writing, have become as of 2019 an established theme in Slavic Literary Studies, which analyzes how such phenomena relate to historical developments in Russian literature. Autobiographical blogging may, for example, be researched in connection to Fyodor Dostoevsky’s *Dnevnik pisatelâ* (A Writer’s Diary, 1873–1881). These practices constitute an integral part of the broader sphere of Digital Russia Studies, which investigates the interaction between the different segments of culture, politics and economy, for example, the use of literary memes for political campaigning. The newly evolving discipline of Global Russian Studies (Platt 2019) tackles, in turn, questions of transnationally dispersed communities beyond traditional understandings of exile or diaspora, which are important for analyzing Russian-language writing-scapes and reading-scapes. Concurrently, digital literature on the Runet is being integrated into the wider context of Global Digital Studies, including literary aspects (Rettberg 2019; O’Sullivan 2019; Tabbi 2018). For some time now, this field has been opening itself up to non-English/non-Latin alphabet based case studies in order to overcome its Western-centricism (Russell and Echchaibi 2009).

Runet literary studies do not differ theoretically or methodologically from global approaches. But they do offer interesting insights into the specific correlations between literary and socio-political evolution in a given national/cultural context. This is particularly significant for the first phase of Runet development in the early 1990s, when, after the dissolution of the Soviet Union, political transformation and media “revolution” coincided. But it is also relevant to the politicized media environment that has established itself after almost two decades of Vladimir Putin’s executive rule as President and Prime Minister. This environment is characterized by a return to vertical power structures, neo-imperialist tendencies and new identity politics (for more on history of Runet, see Chap. 16).

The early manifestations of both Information and Communication Technologies (ICT) and Computer-Mediated Communication (CMC) have been a global inspiration in terms of their potentially democratizing impact, with democratization understood here as access to publication technology, and not primarily in a values sense (Jenkins 2006, 241). Hypertext appeared as the embodiment of a new epistemological system or the realization of long dreamt

of literary utopias: the ultimate library—non-linear story telling. In global theory on CMC, researchers coined concepts including the “wreader” (George Landow 2006) and “the prosumer” (Jenkins 2006). “Wreaders” co-create meaning in collaborative literary projects; “prosumers” in today’s participatory cultures consume and produce at the same time.

Due to the high technical and financial barriers to Internet access in the early phase of Internet development (i.e., from the late 1980s to the mid-1990s), the first users were mostly scientists, programmers at research institutions or academics at universities. As the Internet became more widespread and commoditized, the technically advanced pioneers and “fathers of the Runet” were replaced by a mass of unsophisticated enthusiasts. With each new succeeding generation, the ways in which digital technologies are used are changing, including of course in the field of culture and literature. Editing, copying, sharing and commenting are gradually replacing the creation of “genuine” content. The associated discourses range from the concept of an emancipating collective vernacular creativity, as a continuation of traditional folklore in modern garb, to critical interpretations in the sense of an emerging “prosumer capitalism” (Beck-Pristed 2020, 418). This process is taking place in Russia in analogy to global dynamics. As regards the Runet, such global phenomena and terminology are sometimes embedded into national cultural contexts. A characteristic example of this is the study of amateur creativity, a global phenomenon on the Internet, stimulated by the web’s easy-to-use publication technologies (Vadde 2017). In Runet contexts, however, amateur culture tends to be “nationalized,” that is, explained with an emphasis on historical or cultural traditions. Both protagonists and researchers contextualize amateur writing with reference to the historical phenomenon of Soviet *samizdat* literature (Gorny 2006, 197; Kuznecov 2004). *Samizdat* literally means “self-publishing” (from the Russian “*sam*” = “self” and “*izdat*” = “to publish”) and relates to a highly elaborated, clandestine publication system of works that were subject to political censorship.

The present chapter continues with a further clarification of terminology. It then offers a survey of the main “genres” of digital literature, ranging from hypertext to blogging. The conclusion outlines main research trends and future desiderata.

## 15.2 LITERARY PRACTICES/LITERARY FACTS ON THE RUNET: DEFINITIONS AND APPROACHES

Roughly two decades have passed since Computer-Mediated Communication was broadly implemented worldwide. In this period, a complex terminology has been elaborated—and continually deconstructed—that distinguishes (1) digitized from (2) digital and from (3) Internet (networked; *setevaiâ*) literature (Bouchardon 2017, 3; Gendolla et al. 2010). According to this approach, *digitized literature* denotes previously published print texts, digitized to achieve

broader or different dissemination. A reproduction of a historic poetry collection in one of the numerous online libraries would be a typical example. Digitized literature is differentiated from *born digital* materials, texts originally written on a computer and which do not have a paper substrate (Hayles 2008, 3). *Digital literature*, in turn, designates works that rely aesthetically on distinct features of ICT, such as the inclusion of hyperlinks, multimedia or animation. The hyperfiction poem *V metro* (In the Subway) by Sergey Vlasov and Georgy Zherdev (together with Aleksey Dobkin 2001), offering multiple possibilities to navigate through a set of stories, may serve here as illustration. *Internet or networked literature* is closely related to digital literature. It also relies naturally on code and is embedded into hyperlinked CMC environments, but its conceptual core is concerned with communication practices (sharing, commenting, liking) and is characterized by the Internet's volatile and often very large communities. An example would be the *virtual personae* that were popular on the Runet in the late 1990s to early 2000s. A *virtual persona* is "a fictitious personality, established by a person or group of people which creates semiotic artifacts" (Gorny 2006, 194). *Virtual personae*, or, respectively, their "authors," use communication forums and websites as a playground for identity games including gender swapping (ibid., 208).

As Scott Rettberg underlines, in electronic or digital literature the individual literary work often is less important than the "exploratory engagement" (2016, 166–167) with contemporary computer technology. Consequently, "toolmaking and platform development" should be considered to be an integral part of it. The latter can be specially generated creative environments. But any existing, even commercial, platforms can also be subjected to poetic uses. On Twitter, for example, literary quotations from an author or on a specific topic can be posted, individually selected or automatically processed relying on algorithmic procedures. Such (semi-)automated forms of poetic meaning production sometimes play with the principle of chance, in a continuation of aleatory avant-garde practices. There exist numerous Twitter accounts of historical and contemporary Russian writers or celebrities—some real, some fictional. Since 2012, to name but one, the Russian exiled poet and Nobel Prize winner Joseph Brodsky (†1996) has a Twitter account (@brodsky\_joseph), which is followed by around 350,000 users. The "authors" standing behind such *virtual personae* often remain (semi-)anonymous, masked by their pseudonyms. This demonstrates how specific "genres" or usage patterns can migrate from one technical environment like forums or blogs to another (Twitter).

At the same time, since ICT increasingly infiltrates everyday life and routines, including writing and reading, clear-cut distinctions between on and offline become obsolete. Consequently, the concept of a *post-digital* or *post-Internet* literature has evolved. Post-Internet literature refers to texts that have been produced online but re/turn to paper (Hayles 2008, 159). A typical example would be a blog that is subsequently printed in book form as a kind of sequel, as was the case with the popular online diary written by scriptwriter and novelist Yevgeni Grishkovetz (*Izbrannyye zapisi*, Selected posts, 2014).



The growing amalgamation of the on and offline spheres poses challenges to fixed definitions. More flexible approaches (re)gain significance, for example, the concept of “remediation.” Bolter and Grusin ([1999] 2000) introduced the term in order to describe the multiple processes of media transformation to which any content is subjected. Often labeled “revolutions,” these do not, at least in most cases, lead to the extinction of the previous forms but rather to convergence. Concerning the digital sphere, remediations run in opposite directions and on two tracks: from analogue to digital—from print books to digitized manuscripts—and from digital to analogue—from Twitter posts to poetry collections.

Another approach, which relies not on definition but rather on function, revitalizes Russian Formalist theory, particularly in its concepts of “*literaturnost*” (literariness) and the “*literaturnyj fakt*” (literary fact) as developed by Viktor Shklovsky and others in the early twentieth century (“Russian Formalism,” in Buchanan 2018). “Literariness” is understood as an aesthetic quality (function), which exists not only in literary texts proper but characterizes (online) communication in a broader sense. “Literary facts” are, by contrast, features of non-literary communication (in the present case of digital culture, for example, encodings, media formats or colloquial styles), which in turn affect literary practices.

The chapter follows the typology sketched above, using the concepts of *digitized*, *digital*, *networked* and *post-digital* literature as a rough grid. In so doing, it avoids normative judgments such as the one arguing that digitized literature, as a “simple” remediation, is less culturally significant than experiments with hypertext or critical explorations of code.

### 15.3 THE RUSSIAN-LANGUAGE INTERNET (RUNET): HORIZONTAL VERSUS VERTICAL COMMUNICATION PATTERNS

The term “Runet” as an object of scientific inquiry is not less elusive than “digital literature.” On the global Internet, reading and writing audiences cannot be clearly differentiated according to territorial, national, ethnic or language criteria, despite recent trends toward a re-emerging national sovereignty in the digital sphere (for more, see Chap. 2). This is especially true for Russian contexts, with the existence of a large diaspora and the new “global Russians,” constantly moving between their native country and the second homes they have chosen across the world. Grigory Chkhartishvili, also known as Boris Akunin, an author of sophisticated historical detective fiction, is a good example for a global Russian writer: he lives in Europe, in London and Northern France, and continues to influence Russian prose as well as the socio-political discourse via his Facebook account. In this chapter, the term Runet designates the *Russian-language* Internet accordingly. Where applicable, it will be distinguished from the Internet in the Russian Federation, for example, with regard to the discussion of legislation and regulation.

In early conceptualizations, the structural horizontality of the Internet was hailed as a technological embodiment of postmodernist concepts, including de-hierarchization and non-linearity (ignoring the fact that the technology was actually developed as part of United States [US] military programs). In the contexts of the Runet, this had two major implications. Firstly, on a philosophical level, media change and the regime change of perestroika seemed to coincide with the metaphor of the horizontal, denoting the non-hierarchical, whether this be realized in political democratization or in digressive narration. Secondly, on a pragmatic level, the abolition of censorship put an end to the hunger for books of the late 1980s, an appetite that was immediately satisfied on the Internet—at least for those who could access it.

The Runet of the early and mid-1990s was a marginal phenomenon, with less than one percent of the Russian population online. The early adopters were either members of the technological elite working at scientific institutions or living abroad, mostly in the US, Israel or Germany. The foremost implication of this was that early literary communication on the Runet took place using the Latin alphabet, as Cyrillic encodings did not yet exist. This new communication environment stimulated linguistic creativity, including the systematic use of obscenities, traditionally named *mat*. A later offspring of this linguistic inventiveness is the so-called *padonki* slang (*padonki* translates as “scumbags”), which relies on the principle of distorted phonetic transcription. Questions of coding thus turned into literary facts. The *padonki* movement produced, besides an immense corpus of texts that partly can be considered a form of Internet folklore, also its own platform, which was very popular in the 2000s ([udaff.com](http://udaff.com); see Goriunova 2012).

The year 1998 put an end to the Runet’s marginal status. Paradoxically, this was a consequence of the severe financial crisis, which was accompanied by galloping inflation. The new medium demonstrated that broader user groups could efficiently exploit it, by monitoring ruble exchange rates in real time. Consequently, money and politics entered the scene, and with them “professional” literature. The opening of the Reading Room (*Žurnal’nyj zal*) in the decisive year of 1998 symbolized and embodied the arrival of established canons and authorities in the digital sphere. The Reading Room represented the *tolstye žurnaly* (thick journals) and published excerpts or whole issues of renowned journals such as *Novyj mir* (New World) free of charge. The thick journals have been a peculiarity of Russian reading culture since the late eighteenth century. They publish both literary works and literary criticism and exemplify what is alleged to be Russian literature’s exceptional significance, a literature that fulfills not only aesthetic but also ethical and political functions in a public sphere curbed by censorship. As such, they contribute to the essentialist and literature-centric view of Russia as a “reading country.” In the perestroika era, their popularity rocketed as they took part in political and social transformation. By the 1990s, however, these journals were ailing, due to an overall tendency of de-canonization and because of the economic problems in disseminating their content to more peripheral regions. Paradoxically, the

Runet provided a remedy against diminishing circulation and influence while simultaneously representing a diametrically opposed attitude of non-hierarchical literary communication.

With the overall growth of the Internet—jumping from 2 percent of the population in the late 1990s to 40–50 percent in the 2010s, and finally catching up in comparison to average global Internet penetration, reaching 76 percent in 2019—state institutions also arrived on the Runet. A legal framework was elaborated for the previously largely unregulated sphere of the Internet within the Russian Federation. Of major significance for literary issues are copyright regulations, implemented in the course of Russia’s accession to the World Trading Organization (WTO) in 2012. But legislative measures also include the registration of popular literary blogs under the category of mass media and the blocking of individual works or whole websites for allegedly propagating pornography and pedophilia or “extremism” (for more, see Chap. 5). The ban of the popular instant messenger Telegram in 2018, for example, met with resistance on the part of young users, in particular, and attracted a lot of attention abroad. Experts differentiate between first, second and third generations of Internet control, with the latter embracing repressive methods *and* so-called positive content, that is, cultural narratives, used to disseminate pro-regime information and values (Deibert et al. 2010, 7). It is writers like the prose author and TV journalist Sergey Minaev who contribute to such content creation in the first place. In his successful novel *Media Sapiens. Povest’ o tret’em sroke* (Media Sapiens. The Story of the Third Term, 2007), Minaev creates an influential picture of oppositional media as manipulated and corrupt. This needs to be contrasted with protest movements against electoral fraud and against vertical power structures since the 2010s. These movements rely massively on online mobilization, and, by so doing, they challenge official Internet policies (for more on digital activism, see Chap. 8). Literary practices on the Runet thus take place in a highly politicized environment. The trope of horizontality, ascribed to the new medium of communication in the post-perestroika period, was superseded by the metaphor of the “*vertikal’ vlasti*” (power vertical, Ryazanova-Clarke 2009) as a description of the political system of the Putin era.

## 15.4 LITERARY PRACTICES ON THE RUNET: LIBRARIES AND LIFE-WRITING

### 15.4.1 *Digitized Literature: Forming the Canon from Below*

Online libraries figured among the first literary projects of the Runet. Born out of the hunger for books in the post-perestroika era, they made previously censored texts available. These first digital libraries were personal text collections, intended to be shared with like-minded readers. Their initiators belonged to the technical intelligentsia. Typical examples are EEL (*Publičnaá èlektronnaá*

*biblioteka Evgeniâ Peskina*, Eugene’s Electronic Library, 1992–1998) and the Moshkov library lib.ru (1994), named after its initiator, the programmer Maksim Moshkov. The latter refused the title of librarian, describing himself instead as a mere “doorman” (Mjør 2014, 217). Readers digitized literary works they wanted to see on the virtual shelves and submitted them for electronic publication. The library reflected an eclectic mix of individual tastes and previously marginalized genres, ranging from religious and esoteric texts to science and cyberfiction writing.

With the arrivals of the “professionals” onto the field of play, academically trained literary critics and philologists, new library projects emerged. The RVB (*Russkaâ virtual’naâ biblioteka*, Russian Virtual Library, 1999) offered literary works in accordance with academic standards while modifying the canon by including *samizdat* poetry. The FÈB (*Fundamental’naâ èlektronnaâ biblioteka ‘Russkaâ literatura i fol’klor’*, Fundamental Digital Library of Russian Literature and Folklore) was the first online library partly financed by state money and affiliated to pre-digital academic institutions, in this case the Gorky Institute of World Literature. The FÈB reproduced the literary canon of pre-revolutionary Russia in authoritative digital editions, partly relying on Soviet scholarship and thus implicitly its norms (Mjør 2014, 223).

All libraries provided popular communication forums and metamorphosed from text repositories into social networks in their own rights. For the Runet as a global reading-scape, embracing remote Russian regions and the global diaspora, the online libraries represented a much-needed source of information. At the same time, through their functioning as social networks, they turned into “source[s] of identification” (Mjør 2014, 219). In addition to the troika of the renowned Runet libraries, there exists a multiplicity of smaller, less conceptual, but not less popular, online libraries, where books—especially contemporary prose—can be downloaded for free, in part still illegally. Peter Shillingsburg has called such amateur libraries the “dank cellar” of the Internet, worth consideration as an expression of canon formation from below (Shillingsburg 2006, 138).

An abrupt change in the history of these book repositories occurred in 2004, when the Moshkov library was sued for copyright violations. As a reaction to the trial, the “readers’ librarian” changed his publication policies. New entries in the library were restricted to texts available in the public domain. In addition, Moshkov initiated a platform associated to the library where authors can publish their texts themselves. Named *Samizdat*, the nomenclature refers to discourses about the Runet as an extension of unofficial Soviet publication practices, as detailed above.

A decade later, in 2014, the first large-scale state-financed digital library project was initiated: NÈB (*Nacional’naâ èlektronnaâ biblioteka*, National Electronic Library). The NÈB unites the digital collections of a multiplicity of Russian libraries. It is oriented toward the professional reader. Contemporary fiction protected by copyright is not publicly available but can be accessed from the electronically equipped reading rooms of participating institutions.

Thus, 2004 was a watershed year, marked by the Moshkov trial and the gradual implementation of regulations covering authorial rights. Alongside this, a commercial sector for literary content evolved. This process was stimulated technologically, by the availability of mobile devices—including smartphones, tablets and e-book readers—and specific e-book formats—epub, fb2—which detached reading from stationary computers. Only one year later, the Litres.ru e-book store, originally a network of smaller online libraries, started its activities on a pay-per-download basis. In the years following it established itself as market leader, actively opposing “pirated” resources. Other providers of legal literary content followed, offering different distribution systems. In 2007, Kroogi (Circles), a sharing platform for music, art and literature, went online, based on a pay-what-you-want strategy. Kroogi also offers crowdfunding models. A little later, in 2010, Bookmate was founded as a Freemium service. Users pay a monthly fee to access copyrighted content, which consisted of roughly 800,000 literary texts and audio books as of 2018. In order to structure the abounding wealth of content and to work with their audiences, all of the named e-book services provide multiple communication forums and incentive systems. They arrange editors’ and readers’ recommendations, rankings and awards, incorporating functions that were previously distributed among different institutions (online libraries, magazines, awards).

As a result, a functioning e-book market has emerged, accounting in 2018 for five to seven percent of the book market as a whole (Federal’noe agentstvo 2018, 57): compare with thirty percent in the US. Pay-per-download, subscription and sharing models co-exist. Nevertheless, as of 2018, about half of all e-books were being downloaded illegally using torrents and social networks or being read free of charge from online libraries (Anuryev n.d., 6). Among electronic bestsellers, genre fiction dominates: romantic fiction, detective novels and sci-fi. A significant tendency is the growing popularity of audio books. A more crucial trend still is the dynamically evolving segment of self-publishing, similar to the development in the US, where, as of the late 2010s, one-third of all e-books are indie productions. The company Rideró is the market leader in the field of self-publishing in Russia. But all of the big players in the field of legal e-book content offer self-publishing services. LitRes characteristically named it Samizdat, referring—as Moshkov had done before it—to the Soviet reading and publishing tradition discussed above but stripping the term of any political significance.

A noteworthy number of Russian authors agree to flexible publication models, which combine free access for on-screen reading with payment models for downloads, for example, Internet-savvy writers like Viktor Pelevin or Boris Akunin. Genres that have no market value are broadly accessible on the Runet. The main trends in contemporary poetry are represented free of charge on websites and journals such as *Vavilon*, *Text Only*, and *Novaâ kamera braneniâ* (New Storage Room).

### 15.4.2 *Hypertext Digressions and Media Criticism*

In their nascent phases, the Internet in general—and hypertext in particular—stirred multiple utopian visions. For literature proper, these were dreams of the ideal library or the emancipation of narration from the yoke of linearity, inspired by the short stories *Library of Babel* (1941) and *The Garden of Forking Paths* (1941), by Argentinian writer Jorge Luis Borges. The Runet's literary pioneers also soon explored hypertext as a possibility for new writing modes, for example, in the collective poetry project *Sad rashodāsihsā bokku* (The Garden of Forking Haiku, 1997; Roman Leibov/Dmitry Manin), paying homage to Borges as the global icon of pre-hypertext digressive narration. They were well acquainted, too, with the hypertext experiments in what was, at the time, the dominant player in digital literature: texts by American authors, including Michael Joyce's *afternoon, a story* (1987).

While the utopia of the Internet as a library was realized spontaneously, fueled by the late Soviet hunger for books, hyperfiction remained restricted to a small number of experiments. These Russian explorations of hyperfiction often critically reflected on rampant hypertext euphoria. Thus, media artist Alexei Shulgin in his manifesto *Art, Power and Communication* (1996) dismantled hyperlinking as a simulation of interactivity, while behind the screens the author held even more subtle powers than previously for manipulating readers (for more, see Chap. 14). Another such epistemological critique of hypertext is articulated in the cyberfiction of postmodernist writer Pelevin, the chronicler of digital culture in Russia, for example, in his short stories *Princ Gosplana* (Prince of Central Planning, 1992) or *Akiko* (2003). Skepticism about hypertext is partly motivated by (auto)biographical experiences of the advanced manipulative techniques of Soviet totalitarianism.

Iconic works of hyperfiction are Roman Leibov co-authorship of *Roman*, which would translate into English as *Novel*, though no English translation has been published to date (1995–1996; programmer Dmitriy Manin), and Olia Lialina's *My Boyfriend Came Back from the War* (1996; for more, see Chap. 14). Leibov *Roman* is a conceptual experiment with the im/possibilities of turning readers into co/writers. The title has a trifold meaning, denoting the genre (novel), the style (romance) and the first name of its author, including an allusion to the Roman alphabet, in which the text was written, due to the lack of Cyrillic web encodings at the time. Its core consists of a short text fragment, a juvenile love story with an open end. Readers were invited to send in alternative versions. A dozen author-readers produced around two hundred pages of text. After a year of organic growth, the text became unreadable and Leibov stopped the experiment, which from the beginning was intended as a philological critique of hypertext theory.

An immersive version of a multimedia, animated hypertext is presented by the creative collective consisting of Sergey Vlasov (text), Georgy Zherdev (concept/animation) and Aleksey Dobkin (photography). *V metro* (In the Subway) organizes its fragmentary text as a Moscow metro map, with readers



“entering” and “leaving” it with the help of hyperlinks. Media theoretician Roberto Simanowski describes such creative cooperation among authors, artists and programmers as a new “artes mechanicae” (Simanowski 2002, 148).

Runet hyperfiction is of interest today for reasons of literary history rather than formal innovation. The tireless innovator Akunin continues to experiment with digressive narration, for example, in his novel *Kvest* (Quest, 2009), designed as a game and supplemented by its own interactive website. Animation and code work in the sense of aesthetic explorations of computer code are less frequent still. An example of critical work with code is Aleksroma’s digitized version of the novel *Idiot* (The Idiot, 1868), by Fyodor Dostoevsky, rearranged as a news ticker (2001). “Reading” the text would take 24 hours and is intentionally inconvenient. Aleksroma’s animated version of *The Idiot* underlines how disrespectful remediation can flash out the specific gains and losses that a text can be effected by, in its transfer from analogue to digital format. It thus functions as a multimedia critique of euphoria about technology.

### 15.4.3 *Bottom-Up Creativity: Amateur Literature, Fan Fiction, kreatiff*

While hypertext and the concept of the “wreader” were soon criticized as simulating rather than stimulating interactivity (Simanowski 2002, 66–68), amateur literature and fan fiction blossomed worldwide. “Amateur” is not a clearly definable term in literary theory. Instead, it should be viewed as one part of the cultural battles between “professionals” and “dilettantes” (Vadde 2017). The potentially democratizing effects of easy to use digital publication technologies provoke a redistribution of symbolical capital between established institutions, which act as gatekeepers, and newcomers. Practices and discourses on the Runet do not differ much from similar dynamics on a global scale, although two areas of divergence are worthy of discussion. Firstly, the terrain of Russian literature has traditionally been characterized by a strong orientation around canon and authority, a result of the long periods of strong state interference into culture. On the one hand, this intensifies the quarrels between “amateurs” and “professionals.” On the other, amateur culture is not by default critical vis-à-vis the canon but rather reproduces it by (re)cycling its “masterpieces.” Secondly, self-publishing is terminologically and historically linked to the phenomenon of Soviet *samizdat*, as elaborated earlier. However, literary critic Dmitry Kuz’min (1999) stresses instead the differences between a politically motivated *samizdat* of the Soviet type and today’s media-stimulated self-publishing activities: the existence of an informal but strong quality control in the former.

Since 2000, the largest self-publishing portals on the Runet have been the twin portals stihi.ru for poetry and proza.ru for prose genres. Hundreds of thousands of authors have published literally millions of texts on both portals. Publication on these privately initiated platforms is free of charge. These immense text repositories are structured with the help of editors’ and readers’



recommendations. Stihi.ru and proza.ru regularly organize literary awards in order to motivate and promote their authors. Some, such as the Heritage Award (*Nasledie*), express a patriotic agenda. Texts published on stihi.ru and proza.ru adhere to the category of *born digitals*, in having no paper substrate and in not being primarily intended for print publication. But these platforms, as with most semi-professional content providers, also offer self-publishing as print on demand, for a small charge. This illustrates the tendency of digital literature to move into a post-Internet sphere. Self-publishing reveals itself to be a lucrative market.

Fan fiction, in comparison to amateur literature, is closely tied to the narrative worlds of novels or film sagas such as *The Lord of the Rings* (1937–1949) by J.R.R. Tolkien or the *Harry Potter* saga by J.K. Rowling (1997–2007). Media theoreticians such as Marie-Laure Ryan and Thon (2014) attribute higher immersive potential to fan fiction than they do to hyperfiction. In fan fiction, the reader turns into a writer herself—fan fiction writers are mostly women—and is able to expand or change the narrative. Amateur and fan fiction have generated commercially very successful authors, including E.L. James (Erika Leonard) with her erotic novel sequence *Fifty Shades of Grey* (2011–2017). Disregarding these economic success stories, the majority of its adepts perceive amateur and fan fiction as a basically non-commercial activity, the last realm of “pure” creativity. Fan fiction, as amateur literature, represents the *born digital* text type. While the technology to print it does of course exist, both protagonists and researchers often perceive it as not transferable to paper, due to its high embeddedness in the specific communication environments (Samutina 2017).

Russian fan fiction, at [ficbook.net](http://ficbook.net), for example, does not differ structurally from analogous writing worldwide. Harry Potter fiction, to name just one of the most popular fan fiction universes globally, also has its share of Russian users (*ibid.*). It is generally fantasy and sci-fi with their complex story worlds that generate the most impressive amounts of fan fiction. Thus, the narrative universe of the Strugatsky Brothers (Arkady and Boris), who dominated the genre in the late Soviet era, stimulate a lot of Russian fan fiction, as do contemporary sci-fi and cyberfiction writers like Dmitry Glukhovskiy (*Metro* series, 2002–2015) or Sergey Lukyanenko (*Dozory/The Watch* sequence, 1998–2018), both of whom started as indie or fan fiction writers.

Further phenomena relating to participatory culture are Internet memes and “netlore”—Internet folklore. The term “meme” developed out of Richard Dawkins’ contested theories of cultural evolution and describes micro-narratives that spread across media. Memes typically include not only linguistic or literary components but also visual ones. In contrast to amateur or fan fiction, memes are often created anonymously, moving them closer to the pole of folklore production. In Russian contexts, they are sometimes associated with *lubok*, popular prints that circulated in pre-revolutionary Russia.

Moreover, the new concept of *kreatiff* appeared on the Runet in the 2000s, designating non-commercial cultural creation that is located in the intersection

between amateur fiction, fan fiction and netlore. The term is a linguistic distortion of the English word “creative.” The most popular *kreatiff* has been the Preved-Medved meme. Its narrative core consists of an erotic scene, with a bear (in Russian: *medved'*) surprising a couple having sex in the woods by saying “hello” to them (in Russian: *privet*). The picture is taken from the US artist John Lurie and its English text is “translated” into *padonki*. At the time the meme was created, the bear motif referred implicitly as well to President Dmitry Medvedev. The meme combines allusions to traditional Russian folklore (the bear motif in fairy tales), counter-cultural linguistic creativity and political humor. Both *padonki* jargon and the Preved-Medved meme function as literary facts in the Russian formalist sense: they both influence literary writing. Thus, postmodernist writer Pelevin titled his chat-novel *Šlem užasa. Kreatiffo Tese i Minotavre* (The Helmet of Horror: The Myth of Theseus and the Minotaur, 2005), a *kreatiff*.

#### 15.4.4 *Blogging: Non-literariness and New Sincerity*

Around the year 2000, global Internet culture witnessed the paradigm shift from web 1.0 to web 2.0. This shift was characterized by a move from individual homepages to standardized blogging and social media platforms. On the Runet, writers' homepages as the central location, where the author's *persona* was constructed, became outdated. Previously, this is where Akunin had played his games of self-mystification, related to the hero of his series of historical mystery novels, Fandorin (1998–2018). Pelevin hid as much behind his fan community as he did behind his trademark sunglasses. The queen of crime fiction, Aleksandra Marinina, invited readers to virtually and visually inspect her writing desk. But such self-staging always remained embedded in these authors' respective narrative text-worlds. The communication format of the blog, by way of contrast, with the timeline as the main organizational principle, pulled the author back to the front of the stage, after their role had been marginalized by hypertext theory. Writing on the Internet became increasingly autobiographical.

Blogging was one of the most popular forms of online activity on the Runet from 2002 until 2017. The beginning is clearly marked by a typesetting blog entry by Leibov, who had already “invented” Russian hyperfiction. This triggered a blogging boom. A significant number of Russian authors engaged in intensive blogging, in close interaction with their geographically dispersed readership: Akunin, a literary Internet explorer in all senses of the word; Grishkovetz, playwright and author of neo-sentimental prose; Lukyanenko, prominent sci-fi and cyberfiction writer; and Tatyana Tolstaya, author of sophisticated post-mythological prose. But blogging also offered possibilities to previously less known writers. These included the prolific essayist Linor Goralik (snorapp) or the poetry performer Vera Polozkova (vero4ka). While the early era of literary activities on the Runet had been predominantly male—with the exception of renowned figures such as media artist Lialina—women

writers have caught up since the 2000s. Polozkova has published her blog poetry in book format (*Nepoèmanie*, an untranslatable neologism playing with the Russian word for “misunderstanding,” 2008) and produces carefully staged poetry clips. Her example illustrates the tendency toward post-Internet literature, with digital literature reverting to paper and, at the same time, a trend to a remediated orality.

Participant and observer Yevgeni Gorny portrays the Russian blogosphere as a playground for virtual identities (Gorny 2006). Literary scholar Ellen Rutten takes a different standpoint, highlighting the seemingly paradoxical fact that Russian writers are attracted by blogging specifically because it is perceived as a non-literary activity (Rutten 2017). From this perspective, it is precisely the quality of the blog as an informal communication channel, again, a literary fact in the Russian formalist sense, which has enabled non-polished, everyday language to refresh literary communication. Russian literary blogging stands symptomatically for a broader tendency, moving from postmodernist irony toward “new sincerity” (ibid.).

Runet blogging is characterized by the peculiarity that it was closely linked to one specific blog provider: the US-based [LiveJournal.com](http://LiveJournal.com) (LJ). The brand name was even translated into Russian as *Živoj Žurnal*, meaning “the lively journal.” Blog researcher Gorny relies on cultural psychology to explain this: LJ nurtured the integration of individual blogs into a wider community by offering specific technological features. This process chimed with the allegedly collectivist psychology of Russian society (Gorny 2006, 253). Others contextualize this development in political terms (Howanitz 2020, 4–5): the strong emergence of blogging coincided with a wave of control of the Runet. The fact that LJ servers were based physically in the US was experienced as a protection from surveillance at home. The end of the LJ era was directly linked to these issues. In 2017, LJ moved its servers to Russian Federation territory to comply with Russian data location laws (for more, see Chap. 5). Parallel to this, the company changed its terms and conditions, prohibiting “political agitation.” Bloggers interpreted this as kowtow before Russian authorities. Prominent authors deleted their *Živoj Žurnal* accounts en masse.

#### 15.4.5 *Social Networks: Life-Writing, Public Expression and “Prosumer Capitalism”*

For the social network services (SNS) in a narrower sense, the Internet in Russia shows peculiarities comparable to those evident in Runet blogging. Besides Facebook as the globally dominant actor, local social media platforms have grown up: Odnoklassniki (“Classmates,” founded 2006) and VKontakte, which is known and branded as VK (“In Contact,” also founded 2006). The latter has since then outmatched both its local as well as its US-based competitors. A multiplicity of literary activities thrive on VKontakte, from reading clubs to Russian authors connecting directly with their audiences. VK is also

used to circulate creative content, often still illegally, and enforces authorial rights regulations less rigorously than its global competitors.

Although smaller in terms of user numbers in Russia, the social media giant Facebook is especially popular among writers and public intellectuals. It was the exodus from LiveJournal that led authors to Facebook in the first place: Tolstaya (206,000 followers) and Akunin (250,000 followers) are among the most prominent to date. Social media profiles of Russian authors, be they on VKontakte or on Facebook, intensify the trend toward autobiographical or life-writing. Writers stage their author personalities in direct interaction with the audiences (autoheterobiography; Lüdeker 2012, 147). Strategies are diverse. Tolstaya presents herself as a private person, mixing personal photographs with invitations to her readings. Akunin retains elements of self-mystifying identity play. His username is a combination of pseudonym and surname, Akunin Chkhartishvili. He uses Facebook as an efficient channel to promote his work in cooperation with e-book stores. Concurrently, he continues to participate in political debate, representing the Putin-critical wing among the Russian intelligentsia. At the other end of the political spectrum stands the prominent patriotic writer Zakhar Prilepin (98,000 followers). Prilepin, who rose to fame through his novel about the Chechen War (*Patologii*, *The Pathologies*, 2005), comments on literary culture in contemporary Russia but also reports from the armed conflict between the Ukraine and Donbass secessionists, who are supported by Russia.

Hence, not only do social media profiles by Russian writers function as autobiographical life-writing, they are also part of the composition of the Runet as a deformed but effective public sphere in an otherwise tightly controlled media landscape. They exemplify the formation of global reading-scapes, which are united by language and partly shared collective experience but are also undermined by new ethnic, cultural, national or political affiliations.

In addition to Facebook and its Russian analogues, SNS encompass a variety of other platforms, each of which is characterized by distinct features, operating as literary facts and fostering specific literary usages. Twitter has been used for political mobilization, but the brevity of its messages also promotes the emergence of poetic miniatures. Despite this fact, Russian-language Twitter and Instagram poetry have yet to produce literary celebrities comparable to Indian-born Canadian poet Rupi Kaur. Instant messaging apps are also used for literary purposes. Despite the blocking of the aforementioned popular messenger Telegram, numerous literary channels are active there. As with other SNS, the forms of use are wide ranging. Professional translators or publishers offer glimpses into their work, and addicted readers give personal book recommendations. The “Chekhov writes” channel (@chekhovpishet, initiated by Yevgeni Pekach, about 16,000 subscribers), on the other hand, is an example of projects that closely integrate literature into the lives of readers. Subscribers regularly receive (historical) letters from the famous innovator of Russian prose from the beginning of the twentieth century, Anton Chekhov, via their Telegram account. In contrast to “locative literature,” there is not a spatial but

a temporal immersion. Historical and contemporary reading contexts are fused and contrasted. For the future, Pekach and the editorial team plan to use bots, software applications that execute automatic tasks, to process Chekhov's letters according to search keywords. The final vision is the creation of a virtual "Anton Chekhov" dialogue partner relying on artificial intelligence technology. On YouTube, and its local equivalent Rutube, spoken-word artists and poets circulate recordings of poetry readings or produce poetry clips, fostering a newly mediated orality. One especially popular example of this was occasional poetry by writer and journalist Dmitry Bykov in the early 2010s, who fittingly named his literary project *Citizen Poet* (an allusion to Nikolay Nekrasov's famous political poem "*Poët i graždanin*," *The Poet and the Citizen*, 1856). In a serialized form, Bykov commented on daily politics in traditionally rhymed verses, which were performed by renowned actor Mikhail Yefremov (producer: Andrey Vasilyev).

SNS are not restricted to life-writing, literary experiments and political communication by writers proper but have also stimulated the emergence of huge reading communities (Livelib.ru being the Russian equivalent to Amazon's Goodreads). Brigitte Beck-Pristed presents a case study of such "social reading," understood as "sharing reading experiences through user-generated book comments, reviews, readers' rankings and recommendations" (Beck-Pristed 2020, 407). She shows how reading in digital environments is returned to its "haptic, bodily experience" by being staged as a sporting challenge (reading marathon) on the one hand and as individual quality time on the other. Photographs of the "good old paper books" are posted on the social reading platforms, which show readers relaxing lazily with a steaming teacup in their hands (420–422). These reading networks have market power and popularize authors beyond the established institutions of literary criticism (Vadde 2017). From a more critical point of view, readers are doubly exploited in terms of "prosumer capitalism," stresses Beck-Pristed (2020): they produce unpaid content and are the object of targeted advertising.

## 15.5 FIELDS OF RESEARCH: TOWARD MIXED METHODS

Runet literary studies rely on terminology and concepts developed in global Internet theory—remediation, convergence, participatory culture, transmedia story telling—but also incorporate approaches from Russian Formalism, including the notions of non-/literariness and the literary fact. Especially in the Runet's early years, the mid-1990s, researchers made sense of the new medium by embedding it into local reading traditions (the *samizdat* narrative). Such cultural "domestications" of the new global medium were partly essentializing, ascribing to it seemingly inherent characteristics of Russian culture (literature-centrism and collectivism). There is a strong tendency to personalize the (literary) history of the Runet by focusing on pioneering protagonists (Gorny 2006). Given the especially high percentage of male forerunners, feminist narratives of developments have only recently begun to appear (Ratilainen et al.

2019). The same is true for studies that focus on the significance of digital literature for the regions and ethnic minorities. More attention has been paid to transnational Russian-language reading-scapes (Stahl 2018). Dirk Uffelmann (2014) discusses aspects of Russian cyber-imperialism, with Russian being the lingua franca for users in ex-Soviet countries. Rutten et al. (2013) focus on “web wars” concerning disputed events of twentieth century and contemporary history, which are fueled by and feed into literary narratives. Complementary to such large-scale approaches, a multiplicity of specialized studies exist, which focus on protagonists (Gorny 2006), institutions (Mjør 2014), genres (Coati 2012; Schmidt 2014) and discourses (Rutten 2017).

Concerning methodology, qualitative approaches, including hermeneutic or formalist readings (literary devices, genres patterns), have the upper hand. Quantitative approaches are applied in Digital Humanities and Russian and East European Studies (DHREES) at Yale University (Marijeta Bozovic) and the Digital Humanities in the Slavic Field research association. Natalia Samutina (2017) in her analysis of Russian fan fiction employs long-term participant observation. First exemplary case studies use quantitative methods (topic modeling, literary network analysis; Howanitz 2020). Challenges for future research lie: in combining quantitative and qualitative research (mixed methods); in documentation and archiving; in feminist renderings of Runet literature; in case studies of translocal and transnational Russian-language reading-scapes; and in a further integration into the discipline of Global Russian Studies, highlighting similarities as well as autonomous developments while avoiding essentialization and exoticization.

## 15.6 CONCLUSIONS: CONTENT OUTPLAYS CODE

Literary practices on the Russian-language Internet are, as we would expect, a phenomenon of “glocalization.” The term is a portmanteau of globalization and localization, introduced in the 1990s by renowned sociologists such as Roland Robertson and Zygmunt Bauman, in order to describe overlapping global and local dynamics in an increasingly networked world. With the ever-growing popularity of worldwide SNS and the dominance of global Internet companies such as Amazon and Google, which influence the literary field with game-changing publication and digitization technology, the Runet integrates structurally and functionally more closely into global reading cultures and trends as “New Sincerity” (Rutten 2017). That said, and while the dynamics on the Russian e-book market in the late 2010s are comparable to those in the US (while starting from a lower total level of sales), its local market leaders like LitRes or Rideró outsell Amazon. The appropriation of LiveJournal for specifically Russian-language blogging needs also illustrates how global Internet brands can become “localized.”

Supposedly specific features of Runet literature are located on the level of cultural discourse—for example, self-publishing as *samizdat*—rather than on the level of the textual artifacts themselves. But Runet literary studies show that



genre patterns may differ as regards socio-political dynamics. Thus, content creation was partly more influential than coding experiments, in contrast to what Scott Rettberg states in his approach to electronic literature (2016, 166). This does not mean that content and code (form) should be seen as unrelated but that code is perceived as “transparent” (neutral in terms of meaning) by both the authors and the readers. Such content orientation on the Runet is a consequence of the pronounced needs to communicate that a literature in transition contained. The early Runet filled the gaps in the post-perestroika literary infrastructure and generated textual riches, which amaze readers until today. Against the background of Russian official culture’s strongly normative orientation, and in light of new identity politics, the digital arena continuously renegotiates norms (Lunde and Paulsen 2009). The remarkable activity of renowned writers on the Runet therefore is less a consequence of the persistent myth of Russian “literature-centrism” and rather more the result of highly politicized reading environments. Cultural change is often generated outside the literary field in the narrow sense, overlapping with net art, media activism, computer games or linguistic evolution, for example, *padonki* slang.

The outlined overview of literary practices on the Russian-language Internet shows that digital literature in the narrower sense, from hypertext to code experiments, and changes in literary communication due to alternative distribution channels of digitized literature are closely intertwined. The case of the Runet encourages rethinking overly rigid definitions of digital or electronic literature (Gendolla et al. 2010; Rettberg 2016; see O’Sullivan 2019, 26–38), which tend to exclude digitized texts or post-Internet literature.

## REFERENCES

- Aleksroma [Aleksandr Romadanov]. 2001. F. Dostoevsky, Idiot. *Setevaa slovesnost'* [Web Literature], May 4. <http://www.litera.ru/slova/alexroma/idiot.htm>. Copy in *Internet Archive*. <http://web.archive.org/web/20040123170327/http://www.russianwriter.com/flash/idiot/idiot.swf>.
- Anuryev, Sergey. n.d. Rynok elektronnyh knig. Tendencii razvitiâ i predvaritel'nye rezul'taty 2017 goda [The e-book Market. Trends in Development and Preliminary Results]. Accessed January 9, 2019. <https://bookunion.ru/upload/iblock/f83/f833d535c0995e199fb6206c056d03eb.pdf>.
- Beck-Pristed, Brigitte. 2020. Social Reading in Contemporary Russia. In *Reading Russia. A History of Reading in Modern Russia*, ed. Damiano Rebecchini and Raffaella Vassena, vol. 3, 407–432. Milano: Ledizioni.
- Bolter, Jay David, and Richard Grusin. [1999] 2000. *Remediation: Understanding New Media*. 3. print. Cambridge, MA: MIT Press.
- Bouchardon, Serge. 2017. Towards a Tension-Based Definition of Digital Literature. *Journal of Creative Writing Studies* 2 (1), article 6: 1–13. <https://scholarworks.rit.edu/jcws/vol2/iss1/6>.
- Buchanan, Ian. 2018. Russian Formalism. In *A Dictionary of Critical Theory*, ed. Ian Buchanan. Oxford: Oxford University Press. Current Online Version. <https://doi.org/10.1093/acref/9780198794790.001.0001>.



- Coati, Elisa. 2012. *Russian Readers and Writers in the Twenty-First Century: The Internet as a Meeting Point*. PhD diss., The University of Manchester. ProQuest Dissertations Publishing. <http://search.proquest.com/docview/1774233751/?pq-origsite=primo>.
- Deibert, Rafal, John Palfrey, Rafal Rohozinski, and Jonathan Zittrain, eds. 2010. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge, MA: Cambridge University Press.
- Federal'noe agentstvo 2018 = Federal'noe agentstvo po pecati i massovym kommunikaciám [Federal Agency on Media and Mass Communications]. 2018. Knižnyj ryok Rossii. Sostoánie, tendencii i perspektivy razvitiá v 2017 godu [The Russian Book Market. Status, Tendencies and Outlooks for the Year 2017]. Moscow. <http://www.fapmc.ru/rospechat/activities/reports/2018/pechat2.html>.
- Gendolla, Peter, Jürgen Schäfer, and Roberto Simanowski, eds. 2010. *Reading Moving Letters: Digital Literature in Research and Teaching. A Handbook*. Bielefeld: transcript Verlag. <https://doi.org/10.14361/9783839411308>.
- Goriunova, Olga. 2012. *Art Platforms and Cultural Production on the Internet*. London: Routledge.
- Gorny, Eugene. 2006. *A Creative History of the Russian Internet*. PhD diss., Goldsmiths College, University of London. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.1099&rep=rep1&type=pdf>.
- Hayles, Katherine. 2008. *Electronic Literature: New Horizons for the Literary*. Notre Dame: University of Notre Dame Press.
- Howanitz, Gernot. 2020. *Leben weben: (auto-)biographische Praktiken russischer Autorinnen und Autoren im Internet* [Weaving Life: (Auto)biographical Practices of Russian Authors on the Internet]. Bielefeld: transcript. <https://doi.org/10.14361/9783839451328>.
- Jenkins, Henry. 2006. *Convergence Culture: Where Old and New Media Collide*. New York: New York University Press.
- Kuz'min, Dmitry. 1999. Kratkij katehizis russkogo literaturnogo Interneta [A Short Catechism of the Russian Literary Internet]. Inostrannaá literatura [Foreign Literature] 10. Cited from Setevaá slovesnost' [Web Literature]. Accessed February 22, 2019. <https://www.netslova.ru/kuzmin/kuzm-inlit.html>.
- Kuznecov, Sergej. 2004. *Ošup'pyaá slona (Zametki po istorii russkogo Interneta)* [The Elephant in the Dark (Notes on the History of the Russian Internet)]. Moscow: Novoe literaturnoe obozrenie.
- Landow, George. 2006. *Hypertext 3.0: Critical Theory and New Media in an Era of Globalization*. 3rd ed. Baltimore: Johns Hopkins University Press.
- Leibov, Roman, and Dmitry Manin. 1995–1996. Roman [Novel]. [http://www.cs.ut.ee/~roman\\_1/hyperfiction/](http://www.cs.ut.ee/~roman_1/hyperfiction/).
- Lüdeker, Gerhard. 2012. Identität als Selbstverwirklichungsprogramm: Zu den autobiographischen Konstruktionen auf Facebook [Identity as Self-realization: Autobiographical Constructions on Facebook]. In *Narrative Genres im Internet: theoretische Bezugsrahmen, Mediengattungstypologie und Funktionen* [Narrative Genres on the Internet: Theoretical Framework, Typology of Media Genres and Functions], ed. Ansgar Nünning, Jan Rupp, Rebecca Hagelmoser, and Jonas-Ivo Meyer. Trier: WVT.
- Lunde, Ingunn, and Martin Paulsen, eds. 2009. *From Poets to Padonki: Linguistic Authority and Norm Negotiation in Modern Russian Culture*. Bergen: University of Bergen.

- Mjør, Kåre Johan. 2014. Digitizing Everything? Online Libraries on the Runet. In *Digital Russia: The Language, Culture, and Politics of New Media Communication*, ed. Michael S. Gorham, Ingunn Lunde, and Martin Paulsen, 215–230. London: Routledge.
- O’Sullivan, James. 2019. *Towards a Digital Poetics: Electronic Literature & Literary Games*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-11310-0>.
- Platt, Kevin M.F., ed. 2019. *Global Russian Cultures*. Madison: University of Wisconsin Press.
- Ratilainen, Sara, Mariëlle Wijermars, and Justin Wilmes, eds. 2019. Re-framing Women and Technology in Global Digital Spaces. *Digital Icons. Studies in Russian, Eurasian and Central European New Media* 19: i–iii. <https://www.digitalicons.org/issue19/editorial-issue-19>.
- Rettberg, Scott. 2016. Electronic Literature as Digital Humanities. In *A New Companion to Digital Humanities*, ed. Susan Schreibman, Raymond George Siemens, and John Unsworth, 127–136. Chichester: Wiley-Blackwell.
- . 2019. *Electronic Literature*. Cambridge: Polity.
- Russell, Adrienne, and Nabil Echchaibi, eds. 2009. *International Blogging: Identity, Politics and Networked Publics*. New York: Lang.
- Rutten, Ellen. 2017. *Sincerity after Communism: A Cultural History*. New Haven and London: Yale University Press.
- Rutten, Ellen, Julie Fedor, and Vera Zvereva, eds. 2013. *Memory, Conflict and New Media: Web Wars in Post-socialist States*. London: Routledge.
- Ryan, Marie-Laure, and Jan-Noël Thon, eds. 2014. *Storyworlds across Media: Toward a Media-Conscious Narratology*. Lincoln: University of Nebraska Press.
- Ryazanova-Clarke, Lara. 2009. How Upright Is the Vertical? Ideological Norm Negotiation in Russian Media Discourse. In *From Poets to Padonki: Linguistic Authority and Norm Negotiation in Modern Russian Culture*, ed. Ingunn Lunde and Martin Paulsen, 288–314. Bergen: University of Bergen.
- Samutina, Natalia. 2017. Emotional Landscapes of Reading: Fan Fiction in the Context of Contemporary Reading Practices. *International Journal of Cultural Studies* 20 (3): 253–269. <https://doi.org/10.1177/1367877916628238>.
- Schmidt, Henrike. 2014. Russian Literature on the Internet: From Hypertext to Fairy Tale. In *Digital Russia: The Language, Culture, and Politics of New Media Communication*, ed. Michael Gorham, Ingunn Lunde, and Martin Paulsen, 177–193. London: Routledge.
- Shillingsburg, Peter. 2006. *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge: Cambridge University Press.
- Shulgin, Alexei. 1996. Art, Power, and Communication. *nettime*, October 7. <https://www.nettime.org/Lists-Archives/nettime-l-9610/msg00036.html>.
- Simanowski, Roberto. 2002. *Interfictions. Vom Schreiben im Netz* [Interfictions. On Web Writing]. Frankfurt a.M.: Suhrkamp.
- Stahl, Henrieke. 2018. Subject and Space in Political Viral Poetry Concerning the Russia-Ukraine-Crisis. In *People and Cultures in Motion: Environment, Space, Subject and Humanities. Proceedings of an Interdisciplinary Conference*, ed. Ralf Hertel and Christian Soffel, 89–96. Taipei: Chengda University Press.
- Tabbi, Joseph, ed. 2018. *The Bloomsbury Handbook of Electronic Literature*. London et al.: Bloomsbury.

- Uffelmann, Dirk. 2014. Is There a Russian Cyber Empire? In *Digital Russia: The Language, Culture, and Politics of New Media Communication*, ed. Michael Gorham, Ingunn Lunde, and Martin Paulsen, 266–284. London: Routledge.
- Vadde, Aarthi. 2017. Amateur Creativity: Contemporary Literature and the Digital Publishing Scene. *New Literary History* 48 (1): 27–51. <https://doi.org/10.1353/nlh.2017.0001>.
- Vlasov, Sergey, Georgy Zherdev, and Aleksey Dobkin. 2001. V metro (i snaruži). Nablúdeniâ [In the Metro (and Outside). Observations]. Setevaâ slovesnost' [Web Literature], October 30. <http://www.netslova.ru/vlasov/metro/index.html>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Run Runet Runaway: The Transformation of the Russian Internet as a Cultural-Historical Object

*Gregory Asmolov and Polina Kolozaridi*

## 16.1 INTRODUCTION

Unlike some other national segments of the World Wide Web, the Russian Internet has a name of its own: it is often called Runet. One may ask why there is a need for a special term focused on one country and the Internet in that country. The question, however, is even more complicated, since we face two simultaneously important designations when working with the Internet and Russia: the first is *Runet* and the second is the *Internet in Russia*. If we explore the Russian Internet, are we exploring Runet or the Internet in Russia? Is this merely a matter of language, since the Russian language is typically considered one of the features designating Runet? How can we distinguish between these two concepts and what are the methodological consequences of this distinction? The Internet in Russia seems to be a wider concept, but a clear one. For

---

The original version of this chapter was revised. A correction to this chapter can be found at [https://doi.org/10.1007/978-3-030-42855-6\\_33](https://doi.org/10.1007/978-3-030-42855-6_33)

Gregory Asmolov is Leverhulme Early Career Fellow at the Russia Institute. Polina Kolozaridi is the coordinator of the Russia-based Club of Internet and Society Enthusiasts.

---

G. Asmolov (✉)  
King's College London, London, UK  
e-mail: [gregory.asmolov@kcl.ac.uk](mailto:gregory.asmolov@kcl.ac.uk)

P. Kolozaridi  
Higher School of Economics (HSE University), Moscow, Russia

instance, if we speak about the “Internet of things” in Russia, this is an element of the Internet in Russia, but it is doubtful if it can be called part of Runet. The latter is usually seen as a socio-cultural space or a segment of the Internet. Both terms designate the Internet as a place, something that has borders intended both to include and to exclude (Markham 1998).

In our previous study, we argued that Runet is the object of continuous construction by a variety of actors, including technological, political, cultural, business and media elites, and that changes in the process of construction are associated with the dynamics of power relations between these actors (Asmolov and Kolozaridi 2017). However, following these dynamics is not sufficient to identify the boundaries of Runet as an object or to distinguish it from the Internet in Russia or from the World Wide Web. This chapter is based on historical analysis and aims to offer a conceptual framework for this distinction and to illustrate how this framework can be applied in order to deepen our understanding of Runet as an object. The purpose of the chapter is to explore the history of Russian Internet development in the context of the tension between different approaches to understanding the Internet at the country level.

We focus here on two key properties of Runet: it is historically sensitive and it is multidimensional. The historicity of Runet highlights the fact that what has been developing is not only the content of the object (e.g. what happened with Runet) but the object itself (what Runet is). In this sense, our history has an ambivalent position, since it is both a history of the construction of an object and a historical description of various events related to this object. Therefore, when telling the story of Runet we should constantly question whether our story is still taking place within the boundaries of Runet or whether perhaps it is already the story of something else, for instance, of the Internet in Russia.

Following the dynamics of the historical process, not just as an ongoing chain of events but as the evolution of an object, requires a framework for following changes in the object. Previously, we identified five stages of Internet change in Russia (Asmolov and Kolozaridi 2017). Here we seek to advance this approach by replacing the linear structure of periodization with a framework that approaches Runet as a multidimensional socio-technical object with a number of vectors that are ongoing through continuous change.

## 16.2 RUNET AS AN OBJECT: THEORETICAL AND HISTORICAL APPROACHES

The Internet in Russia is older than Runet. The history of the Russian Internet, at least as a concept, starts many years before the collapse of the Union of Soviet Socialist Republics (USSR). The conceptual origins of the Internet in Russia have been linked to information networks and cybernetics development as part of the Soviet planned economy (Gerovitch 2002). Peters (2016, 4) explores the failure of the early development of a nationwide Soviet computer

network (the All-State Automated System), which was inspired by “a utopian vision of [a] distinctly state socialist information society.”

There are a number of other events that could be considered as the starting point of the Internet in Russia, including the first instance of modem-based communication between the Kurchatov Institute and the University of Helsinki in August 1990, the foundation of the first Soviet Internet service provider or the registration of the domain zone. The Soviet domain zone .su was established on September 19, 1990, while the .ru zone traditionally associated with Runet was registered on April 7, 1994. The word Runet, however, only appeared later. A number of sources argue that it was first used in 1996 by Raf Aslanbeyli, a journalist living at that time in Israel (Lihachev 2015).

Researchers argue that the “Internet in Russia” and the “Russian Internet” form a “complex matrix of overlapping areas and distinct segments, producing constant fractions” (Schmidt et al. 2006, 130). Schmidt and Teubener (2006, 14) highlight how the notion of Runet as a dedicated term for a specific segment of cyber space “has almost no analogue in Western languages.” They point out that the boundaries of Runet rely on a variety of factors, including “language, technology, territory, cultural norms, traditions or values and political power” (Schmidt and Teubener 2006, 14). Deibert and Rohozinski (2010, 19) highlight how Runet relies mostly on digital platforms that “are modelled on services available in the United States and the English-speaking world, but are completely separate, independent, and only available in Russian.”<sup>1</sup>

That said, the significance of the distinction between Runet and the World Wide Web is also questioned. For instance, according to Bowles (2006, 30), the “differences between the RuNet and the rest of the Internet have gradually been dropping away” while “RuNet is simply another backwater of the Internet, fenced in by a language barrier and sometimes subject to mystification by loyal denizens, but not essentially different.” Recent literature presents an understanding of Runet based on its perception by the Russian state. According to Nocetti (2015), the Russian authorities conceive of “cyberspace as a territory with virtual borders corresponding to physical state borders, and wishes to see the remit of international laws extended to the internet space, thereby reaffirming the principles of sovereignty and non-intervention.” Building on this argument, Ristolainen (2017, 8) proposes that “RuNet—the Russian segment of the Internet—is considered an extension of the existing territory in the Russian ‘information space.’”

Runet is not the only national segment of cyberspace in the former USSR, and not the final chain in the hierarchy of segmentation of cyberspace. The idea of a national segment of the Internet, as discursively manifested through a dedicated name, can also be seen in Kazakhstan (Kaznet), Ukraine (Uanet), Belarus (Bynet) and other states (Shklovski and Struthers 2010). There are also socio-cultural online spaces in some of the Russian regions. For instance, Tonet was the name for a city-based network in Tomsk, Chuvashitet is the title given to the Internet associated with users from Chuvashia, while Tatnet is described as the “Internet for Tatars and in the Tatar language” (Sibgatullin 2009). So

Runet is not the only “net” in Russia, or in the Russian language, and it is not the same as the Internet in Russia in general.

The following section offers a conceptual framework that allows us to resolve some of the challenges for the conceptualization of Runet as an object of investigation.

### 16.3 RUNET AS A RUNAWAY OBJECT

As argued above, Runet cannot be reduced to the experience of a shared language (Bowles 2006). Some early approaches addressed it in terms of socio-political phenomena seen in the USSR. For instance, comparing the role of Runet to a “Soviet Kitchen” (Popkova 2014, 98) would suggest that Runet should be explored as a new type of public sphere “where people can get together and freely discuss and identify societal problems” (Habermas 1991, 398). Another notion taken from the Soviet Union, that of *samizdat*, presents Runet as a space for the independent generation and distribution of content (for more, see Chap. 15). In both cases, the conceptualization of Runet builds on the antagonism between an authoritarian state and users seeking new, uncontrolled spaces of freedom. Drawing on Bakhtin, Gorny (2007) seeks to go beyond the political conceptualization of Runet and to address it as an alternative socio-cultural space that deconstructs traditional cultural hierarchies, offering space for the flourishing of new identities and alternative ways of living. Runet can also be addressed as a space that allows the emergence of a Russian network society (Castells and Kiselyova 2003).

Previously we have argued that Runet can be explored by following the changes in Internet elites and in the dominant/alternative Internet imaginaries (Mansell 2012) promoted by different actors (Asmolov and Kolozaridi 2017). This approach highlights how Runet cannot be defined as a static entity or as a set of technological properties. It requires a conceptualization drawing on a historical perspective that allows us to capture the dynamics of continuous change. In this sense, historical description is not the purpose of our investigation but a method that allows us to deal with the complexity of its object.

Traditional concepts of the social construction of technology have a limited capacity to address large-scale and constantly developing socio-technical objects. Following Giddens (2000) and Engeström (2008), we would argue that these types of objects can be considered as “runaway objects”—objects that are constantly shaped by the forces of both technological development and social construction. According to Engeström (2008, 227), a “runaway object” is a large-scale, complex object which is “pervasive and [whose] boundaries are hard to draw” and “poorly under anyone’s control and have far-reaching, unexpected side effects”. Runaway objects are not artifacts in a traditional sense but are constantly addressed, shaped and changed by the activities of numerous actors, while every event may create a contradiction between different actors and potentially lead to a new chain of events. Runet as an object has constantly created new challenges, new opportunities and “alternative ways of



living” (Mansell 2002, 408) for various types of actors. It has thus also triggered some actors to address these changes.

Our analysis presents Runet as a “net,” opposing it to the Internet as a single network spreading all over the world. As Kevin Driscoll and Camille Paloque-Berges emphasize, taking this “net” into account helps us to avoid a simplification of the Internet as solely a technology and to conceptualize its socio-technical role. “Nets” are various and highly dependent on the historical and cultural context, while “the Internet” remains a global phenomenon (Driscoll and Paloque-Berges 2017).

#### 16.4 THE VECTORS OF RUNET DEVELOPMENT: DEFINING RUNET AS AN OBJECT IN A CULTURAL-HISTORICAL CONTEXT

The description of Runet as a runaway object requires us to approach Runet as a multi-vector object and to follow its historical development in terms of each different vector. A runaway object is developed through the activity of a variety of actors, including not only political, cultural, media and business actors but also developers and everyday users. Accordingly, these sets of relationships between different actors can be seen in terms of each vector. The vectors are interrelated, however, and distinguishing between the actors allows us to conceptualize the complexity of Runet as a multidimensional and complex runaway object. Our analysis of the vectors relied on a thematic analysis of media sources and on the research literature on Runet.

We have chosen to distinguish the following five vectors of Runet development: the technological vector, the cultural vector, the media vector, the user and everyday life vector, and the political vector. This selection does not necessarily mean that these are the only vectors that could be followed or that there is no place for alternative descriptions. For instance, one may argue that there is a need to follow an “economic vector”; however, we have not addressed this as a distinct vector since the manifestation of economic power can be seen in all the vectors, as can the manifestation of political power.

The *technological vector* is concerned with the development of the hardware and software that Runet relies on, including fiber cables, domains and their registers, various online platforms and the infrastructure of surveillance. The technological question is concerned with the identification of the most popular online Runet platforms, including search engines, social networks and blogospheres. It examines the extent to which Runet relies on local or foreign platforms and follows the changes in dominant platforms. This vector is particularly concerned with forces of technological development and with who controls the technological segments of Runet.

The *cultural vector* allows an exploration of the role of Runet as a space of cultural development. On the one hand, it examines whether Runet offers a space for alternative and underground cultures that were not able to find a proper place in traditional offline space or participatory cultures (Jenkins

2006). On the other hand, it examines different manifestations of traditional and mainstream culture, how these fight to establish their presence in Runet and the relationships between underground and mainstream.

The *media vector* addresses the role of Runet as a space for media development. It explores how new types of media platforms shape the news consumption and production of the Runet audience and examines the extent to which online media have been able to set the agenda and frame different types of events. It is particularly concerned with the relationship between the new online and traditional offline media. It also explores how power relations are manifested in changes in the structure of ownership, different modes of censorship and various forms of state-sponsored regulation.

The role of technologies substantially changes during the transition from usage by a minority of early adopters to when new technologies become domesticated (Silverstone 2002). The *user vector* follows how Runet became a part of everyday life in almost every sphere for a wide spectrum of the population. It explores how Runet has configured its users and the functions of Runet in everyday life. This includes an analysis of the changing popularity of platforms, sociological data on Runet usage in different time periods and the mapping of new forms of social interaction and community building. It also addresses various forms of facilitation of user activity in order to address different types of everyday life issues and crisis situations. A distinct sub-topic of this vector is the role of Runet in the lives of children and teens.

The *political vector* follows the role of Runet in the political life of Russia. It encompasses approaching Runet as a public sphere, the role of Runet in political mobilization and the role of Runet in the empowerment of the state, including new technologies of surveillance and crowd control. In this sense, this vector follows the tension between the different imaginaries of Runet as an alternative political space, a space of political discussion and mobilization as well as the securitization and sovereignization trends on Runet that seem to make it one more sphere of the state's political influence and an additional set of technologies of political power.

## 16.5 THE HISTORY OF RUNET THROUGH FIVE VECTORS

### 16.5.1 *The Technological Vector: From Enthusiasts to Corporations*

Some of the technological origins of Runet relate to the development of informational systems for communication, scientific purposes and the advancement of the planned economy in the Soviet Union (Gerovitch 2002; Peters 2016). The experience of early Internet usage could be connected to that of earlier computer-based network systems like Usenet and, later, Bulletin Board Systems (BBS) and FidoNet<sup>2</sup> (Driscoll 2016). However, the development of FidoNet and BBS differed from that of the Internet in terms of both technology and social organization: “Unlike the Internet, which in the United States was the preserve of academic and military institutions up to the early 1990s, FidoNet

has been more the preserve of talented computerphiles, run on a purely non-commercial, anyone-can-join basis” (Rohozinski 1999).

The early development of Runet can be linked to the continuous development of the Internet in Russia, but, as mentioned, there are different approaches to what can be considered its starting point. For instance, Kuznetsov (2004) identifies two events as the starting points of the Russian Internet: the registration of the .su domain and the creation of the Relcom/Demos computer network. In this sense, the development of the technology that offered an infrastructure for Runet was driven by scientists and programmers together with businessmen who identified the commercial potential of the Internet.

From a relatively early stage, the Russian security services interfered in the development of the new informational system. A number of scholars highlight, however, how KGB (*Komitet gosudarstvennoj bezopasnosti*, Committee for State Security) apparently had no capacity to control the electronic flow of information in the first phase of Runet development, and specifically around the political events that triggered the final collapse of the USSR (Konradova 2016). The systematic surveillance of Internet-based communication started with the implementation of SORM-2 (*Sistema tehničeskikh sredstv dlâ obespečeniâ funkcij operativno-razysknyh meropriâtij-2*, System for Operative Investigative Activities-2) in 1998, when all telecommunications operators were required to integrate this into their communication hardware.

In addition to cables and hardware, the technical aspects of Runet relied on the development of various types of online services. The Russian search engines Aport (1996), Rambler (1996) and Yandex (1997) were founded before Google. A social network, Odnoklassniki, was launched in March 2006 and followed by VKontakte in January 2007. The most popular e-mail services were offered by Mail.ru and Yandex. Russian blogging relied mostly on an American platform, LiveJournal, which was subsequently sold to a Russian company, Sup Media, in 2007. Since then, Yandex and Mail.ru have become the two major Russian Internet giants, while VKontakte dominates the social networks market. However, the dominance of Russian online platforms has not excluded Western platforms. Google, YouTube, Facebook, Twitter and Instagram have continued to be popular destinations for the users of Runet (for more on social networks, see Chap. 19).

One of the ongoing developments of Runet within the technological vector is the change in the structure of ownership of the major online platforms. Gold stock in Yandex was purchased by Sberbank in 2009. Most of the platforms, including Mail.ru (Mail.ru Group has been controlled by Alisher Usmanov since 2015), Odnoklassniki (owned by Mail.ru Group), LiveJournal (since 2013 a part of Rambler, owned by Aleksandr Mamut) and VKontakte (owned by the Mail.ru Group since 2014), came under the control of oligarchs alleged to have close ties with the Kremlin. The founder of VKontakte, Pavel Durov, was forced to sell his share of the company in 2014. At the same time, the Russian authorities increased the scale of regulation of the activity of foreign Internet companies including Facebook, Google and Twitter. Russian law

required these companies to keep the private data of Russian citizens on servers located in Russia. LinkedIn did not comply and was banned. Other major Western platforms such as Twitter and Facebook have also not complied, but in 2020 they remain accessible in Russia.

Efforts to increase state control can also be seen at the infrastructural level. The introduction of the Cyrillic .рф domain in 2010, actively supported by the Russian authorities, afforded new technical opportunities for the russification of the Internet in Russia. In 2017, the Kremlin required Russian Information Technologies (IT) entrepreneurs to focus locally at the expense of the global market in order to be independent of foreign influences (Budnitsky and Jia 2018, 607). A number of initiatives promoted a vision of Runet as a “sovereign Internet” (Asmolv 2010; Kukkola and Ristolainen 2018). In 2019, this vision led to a law requiring the development of an independent infrastructure for the Russian Internet that would enable it to continue functioning while relying solely on Russian servers. Increasing control over technological infrastructure and software can also be seen at the policy level. Strategic documents from the late 1990s promote the idea that “our” technologies, produced and used in Russia, were treated by the state as a “social good” while global technologies were considered a threat (Shubenkova and Kolozaridi 2016).

### 16.5.2 *The Cultural Vector: From Alternative to Mainstream*

The first popular websites on Runet included an online library (lib.ru) and online competitions for writers and poets. Since the early 1990s, Runet has been rapidly occupied by artists, journalists and members of the academic community, who have not only shared their work but also actively participated in the construction of the new space. Roman Leibov, a semiotics scholar from Tartu, Estonia, is considered to have been the first Russian-language blogger on LiveJournal. These writers and scholars considered Runet a laboratory for cultural experiments such as collaborative production and hypertext. A range of online projects crossed national boundaries and offered a common space of cultural production for people in former USSR countries and for emigrants all over the world, including in the United States (US), Europe and Israel.

One of the first web design studios that actively contributed to designing the early Runet space was launched by Artemy Lebedev in October 1995. A special space was also offered for the production and sharing of humor, which had played an oppositional role in Russian culture. The list of the most popular websites included at that time anekdot.ru, created by Dmitry Verner. Later, the web project Lurkmore.to, launched by David Homak in 2007, sought to offer an encyclopedia of memes illustrating the underground culture of Runet.

At the beginning of the 2000s, LiveJournal became the most popular platform among Russian cultural elites and, as highlighted by Alexanyan (2013), could be considered a unique mix of blogging and social networking. Initially, the option to create a blog on LiveJournal was by invitation only. This type of model ensured the elitist nature of the LiveJournal community. In 2002,

however, the invitation-only requirement was cancelled, and LiveJournal opened its gates to the growing community of Runet. In 2010 Harvard-based researchers identified this cultural cluster as still one of the biggest clusters in the Russian blogosphere, although it was less dominant by comparison with the public affairs cluster (Alexanyan et al. 2010). Later, the first Russian social networks, VKontakte and Odnoklassniki, contributed to a shift from content-generation toward social networking among friends as a dominant form of activity of Russian Internet users.

The shift from Runet as a space of alternative culture to a mainstream domain could be seen in a number of aspects. Firstly, the Russian social network VKontakte offered not only an option for communication but also a limitless and unregulated environment for sharing any type of music and video content. Accordingly, despite copyright laws, any type of cultural content could be found online. Later, VKontakte started to comply with some of the copyright laws; however, it remained one of the major music and video hosts on Runet. The increasing role of mainstream culture is associated with the increasing dominance of content created for the traditional media. For instance, the most popular YouTube accounts among Russian audiences are *KVN (Klub veselyh i nahodčivyh*, a Russian humor show,), with 4 million subscribers, and a talk show, *The Evening Urgant*, with 2.7 million viewers. The most popular Russian account on Instagram belongs to a pop-singer, Olga Buzova, who has 14 million followers (Lebedev 2018).

At the same time both YouTube and Instagram are key sites for new celebrities competing with traditional media content, such as videobloggers, beauty bloggers and musicians. However, these phenomena are rarely treated as specifically characteristic of Runet, since they partly belong to a global culture of micro-celebrities, various youth scenes (Omelchenko 2019) or particular genres. They use the Russian language, but it is arguable whether they share that sense of commonality which was so important for the Runet culture of the 1990s and 2000s.

### 16.5.3 *The Media Vector: From Alternative Media to State Control*

The first time Runet was able to play a substantial role as an alternative form of media was during the coup attempt against Gorbachev in 1991. While Soviet TV was broadcasting the ballet Swan Lake, Relcom allowed geeks and scientists to break the information blockade through UseNet groups and inform the Western audience about what was happening (Konradova 2016). The first Russian media websites appeared a few years later, when early adopters started to occupy the Runet space. The first news website, *Vechernij Internet* (Evening Internet), launched by Anton Nosik in 1996, covered mostly news concerning Runet.<sup>3</sup> As pointed out by Kuznetsov, “The Russian Internet was so small at that time, that the appearance of any new page was an event” (Kuznetsov 2004).

The first website of an offline newspaper was launched in spring 1995 by *Učitelskaâ gazeta* (Teachers’ Newspaper). However, very soon Runet was

offering a space for the development of new media organizations. These included Vesti.ru, Gazeta.ru and Lenta.ru. While the democratization of the Russian media sphere was led by traditional media in the 1990s, the Internet took a lead as a major liberal media domain in the 2000s. Under the new president, Vladimir Putin, who took office in 2000, the Russian state succeeded within a short time in taking control of the major TV channels from the oligarchs Berezovsky and Gusinsky, while online media remained relatively independent. Although Vesti.ru was taken under the control of the Russian national TV channel, Lenta.ru and Gazeta.ru were considered among the most popular independent online sources for about another ten years.

Social media also started to play an increasing role in shaping the Russian media environment. The rise of blogs, citizen journalism and groups on VKontakte can be seen as important factors that challenged the control of the traditional Russian media. Many traditional journalists also started using blogs to develop their personal professional brands, to share unedited content and to have direct communication with their audience. Other types of actors also contributed to the transformation of the Russian online media system. An increasing number of newsmakers, including politicians (such as President Dmitry Medvedev), experts and celebrities, started using blogs and social networks, which now could often be considered a source of first-hand information.

Social media activists and opposition politicians also contributed to the development of Runet as a media sphere. These activists launched online investigations that were able to set the news agenda and make an impact on traditional media. This included securing investigations of police corruption as well as helping to hold high-ranking businessmen and officials accountable for their misdeeds, as in the case of a car accident involving the vice president of the Lukoil oil company in February 2010. That said, Toepfl (2011) points out that the traditional Russian political elites learned how to manage public outrage and restructure it to serve their own political goals.

During parliamentary and presidential elections in 2011–2012 the Russian online media played a central role in exposing the scale of fraud and in covering the protests. Following the protests, the Russian authorities started to increase their control over and pressure on online media. Some, like Grani.ru, were blocked. The editorial teams of two leading news websites, Gazeta.ru and Lenta.ru, were changed and some former members of the Lenta.ru team moved to Latvia to found a new website, Meduza.io, in 2014. At that time LiveJournal also lost its political function while most of the influential media bloggers moved to standalone platforms or to social networks. Opposition sources also became less visible in the Yandex News aggregator following political pressure from the Kremlin (Soldatov and Borogan 2015).

Alexanyan has argued that in the 2000s Runet gave rise to a different type of imagined community of Russian citizens, distinguishing between “Internet Russia and TV Russia” (Alexanyan 2013, 161). However, as a result of state media regulation, the Russian authorities increased their control over the Runet media sphere. Only a few liberal online media outlets, including



NovayaGazeta.ru, Meduza.io, Ekho Moskv ( <https://echo.msk.ru/> ) and the *TV Rain (Dožď)* channel (tvrain.ru), remained active. Facebook also continued to play some role, whereas a new digital platform, the messaging app Telegram, assumed increasing importance for the circulation of political rumors through anonymous channels. While on the one hand the Russian authorities made a failed attempt to ban Telegram in 2018 for non-compliance with anti-terrorist legislation, on the other hand it was also being actively used by the Kremlin for various types of political media manipulation through popular anonymous political channels (Rubin and Badanin 2018). While the Runet media sphere lost its oppositional power as an alternative media environment, it still offered a diversity of media voices and genres, although since 2014 it has started to be dominated by state-affiliated platforms (e.g. Lenta.ru, which changed its ownership, Yandex News, RIA Novosti, KP.ru and Izvestia.ru) and the Russian authorities gained more control over agenda-setting and the framing of political events. At the same time, some opposition content moved to non-Russian platforms, such as in the case of the popular YouTube video channels of opposition leader Alexei Navalny and TV presenter Yury Dud, as well as of the independent political channels on Telegram (for more on digital journalism, see Chap. 9).

#### 16.5.4 *The User Vector: From Elites to Everyday Usage*

In the 1990s and the first part of the 2000s, the Internet was used actively by a minority of Russian citizens. The major trend, however, that changed the profile of the Russian user was the gradual increase in the number of Internet users in Russia. This could be seen in terms of both the regions covered by the Internet and the frequency of usage. The socio-economic groups that had had limited access to the Internet during the first years of Runet became active users. This happened as a result of the reduction in costs of Internet access and the broader availability of computers and mobile phones.

In 2017 Russia had more than 107 million Internet users (more than 76% of the Russian population) and the number of users aged between 10 and 55 was more than the TV audience. The growth in the number of users was linked to the increase in instrumental usage of Runet. According to Nisbet et al. (2015), the most popular usage of the Russian Internet included: “search for information for personal usage”; “communicating in social networks”; “reading national news”; “e-mail correspondence,” and “downloading and listening to/viewing of music and video.” These types of usage are related to the increasing popularity of a number of websites, including Avito.ru (online sales), weather forecasts (Gismeteo.ru) and Head Hunter [hh.ru] (recruitment). The ratings of most popular Russian websites are constantly changing, while the top placings are not only dominated by media, social networks, e-mail services and search engines but also determined by trends in digital consumption and online education. The rankings of statistically most-visited websites among Russia users can be seen at radar.yandex.ru and top1000-ru.hotlog.ru.



In 2020 VKontakte remains one of the most popular websites, offering not only social networking but also various forms of entertainment including movies, music and pornography (Ostrovsky 2019). VKontakte also offers a platform for the development of communities of different kinds, from vibrant youth culture to intellectual clubs, wives of prisoners and street-food testers in small towns. An additional sector that fulfills instrumental functions and addresses the needs of Russian citizens includes state-related services offered through the e-governance portal Gosuslugi. The increasing scope of instrumental functions is also manifested through rapid growth in online banking services and online payment systems.

We may also find evidence of how digital platforms afford Russian users an opportunity to address everyday life challenges. This is related to various forms of crowdsourcing, as a digitally mediated form of mobilization of resources to address different goals. One of the groups of digital platforms that allow users to be mobilized around everyday life issues consists of civic applications (Ermoshina 2014). Runet has offered a rich diversity of platforms of this type, from the mapping of potholes (the Rosyama.ru project, initiated by Navalny in 2010) to RosZKH.ru and Zalivaet.SPB.ru, which map the failure of local authorities to fix buildings and local infrastructure. Charity platforms like [pomogi.org](http://pomogi.org) and [TakiDela.ru](http://TakiDela.ru) raise awareness of individuals needing various kinds of help and allow users' financial resources to be mobilized to address these problems.

The Internet has also played a substantial role in the case of various emergencies, where it has not only offered independent sources of information but also allowed people to take part in response. One of the most significant cases of digitally mediated civic mobilization was the response to wildfires in 2010 (Asmolov 2013b). Some of these projects support continuous engagement to save people's lives. For instance, the Liza Alert platform allows people to be mobilized for search and rescue operations when elderly people and children become lost in Russian forests.

The Russian authorities also seek to develop platforms to engage users and harness crowd resources. State-affiliated initiatives for the engagement of citizens in decision-making, such as the Active Citizen project ([ag.mos.ru](http://ag.mos.ru)) launched by the mayor of Moscow, have been criticized for offering "a semblance of openness and participation, while in practice neutralizing citizens' activity and exerting control over them" (Asmolov 2018).

The user vector, perhaps, is the sphere where the contrast between Runet and the Internet in Russia is most visible. This is where the Russian Internet continuously becomes an instrument of the "uses and gratifications" (Katz et al. 1973) of a majority of Russian citizens. Here, we also see how the change in the demography of Russian Internet users, specifically the increase in the number of users among older generations and in more remote areas of Russia, is associated with the change in the role of Runet. The instrumental usage of the Russian Internet also makes it more similar to the Internet in other countries.

### 16.5.5 *The Political Vector: From Democratic Promise to Digital Sovereignty*

During the 1990s politicians started slowly to explore the new political technologies. In March 1996, Yabloko was the first Russian political party to open a website. However, Runet is sometimes considered to be a space for opposition political actors, various types of movements and individuals that have had no affiliation with traditional political organizations. In 1999, Putin—then prime minister—held his first meeting with leaders of Runet. Despite some pressure from a minister of communication, Mikhail Lesin, to introduce some form of Internet regulation, Putin opposed Lesin’s proposal. He stated: “We are not going to look for a balance between freedom and regulation. We will always choose freedom” (Soldatov and Borogan 2015).

The elections of 2000 were the first where the Internet started to play a significant role. A new type of political consultant with the Internet as an area of expertise appeared. This group included such people as Gleb Pavlovsky, a founder of the Fund for Effective Politics (FEP). FEP was the first organization to release public opinion polls online. During the first two terms of President Putin the authorities did not actively interfere in the online space, although a number of legislative initiatives for the regulation of communication were introduced. Meanwhile some liberal governors like Oleg Chirkunov and Nikita Belykh started to experiment with the online space by managing LiveJournal blogs. In 2008 Dmitry Medvedev became president and started a campaign of popularization of open data and e-government. Medvedev visited the head office of Twitter in California, where he opened an account and wrote his first Tweet.

At the same time, in the late 2000s, Runet displayed a “growing use of digital platforms in social mobilization and civic action” (Alexanyan et al. 2012). This political mobilization was not necessarily associated with any political organization but rather with “issue-based campaign[s]” initiated by Internet users (Alexanyan et al. 2012). At the same time, some leaders started to develop their political capital online, without affiliation with any political party. One example of the new generation of Internet-enabled leaders was Alexei Navalny, who gained popularity via his blog on LiveJournal, where he published his investigations into corruption. Later, when LiveJournal came under the control of pro-Kremlin owners, Navalny launched a standalone website, Navalny.ru, as well as actively using YouTube, Twitter, Facebook and Telegram.

That said, according to Fossato (2009), “The state remained the main mobilizing agent.” She argues that Runet operates “as a device to spread and share information, but largely among closed clusters of like-minded users who are seldom able or willing to cooperate.” In 2010, contradicting his previous positive assessments of the Internet, Putin stated that it was well known that 50 percent of online content was pornography. Since then, one can see the domination of the state’s discourse on the role of the Internet as a dangerous technology and a threat to socio-political stability that has to be regulated. The

major examination of the political role of Runet, however, took place around the parliamentary and presidential elections in winter 2011–2012.

During the parliamentary elections of 2012, social networks, crowdsourcing platforms and dedicated websites were employed to monitor electoral fraud (Oates 2013). At the same time, the Russian authorities launched the WebVybory2012 ([webvybory2012.ru](http://webvybory2012.ru)) operation to cover 95,000 polling stations with two web cameras for each station and offer online live broadcasting of the vote and the counting process. The project sought to prove that the Russian elections were transparent and legitimate. Despite the efforts of the Russian authorities to protect the legitimacy of the elections, independent monitoring efforts and online media challenged the results. The parliamentary elections were followed by a wave of protests facilitated via social networks.

The electoral cycle of 2011–2012 provided a momentum for accelerated political innovation (Asmolov 2013a) and specifically for new forms of digitally enabled horizontal mobilization of protests. This included the development of crowdsourcing platforms for election monitoring ([Kartanarusheniy.ru](http://Kartanarusheniy.ru)), using social networks including Facebook for large-scale mobilization, and the development of dedicated digital tools for the organization of distributed protests (e.g. in the case of the White Circle protest, where a website, [Feb26.ru](http://Feb26.ru), supported self-organization, enabling people to create a live chain around the center of Moscow). Digital political innovation also offered new ways of collecting data on the scale of arrests and of offering assistance to people who were detained. The wave of political innovation continued after the elections. During the Moscow mayoral election in 2013 Navalny's team was able to develop online tools to mobilize support despite the lack of coverage in the traditional media. Eventually Navalny received 27 percent of the vote, which was considered an unexpected success. Later Dmitry Gudkov developed so-called "political Uber" to simplify voting for the most liberal politician at a neighborhood level. However, this success never went beyond local level.

Following the electoral cycle of 2011–2012, the authorities identified the political threat associated with Runet, through the challenge to the legitimacy of elections or the capacity to facilitate large-scale political action. Klyueva (2016) argues that "[T]he successes of the protest movement initiated a government crackdown on the Russian Internet and social media." She concludes that "the pro-government actors were able to monopolize and control the public sphere with their issues and messages" (Klyueva 2016, 4674). Gunitsky (2015, 50) suggests that the case of Runet illustrates a "shift from contestation to co-optation" of social media (for more on social networks and politics, see Chap. 30).

The third Putin presidency (2012–2018) started with a series of restrictive laws. The Yarovaya package obliged Internet Service Providers (ISP) to store their information about user activity for a long time. The state also supported groups of cyber guards who search for prohibited content online and report it to the authorities. At the same time a new generation of pro-Kremlin digital-savvy politicians started to play an increasingly significant role online (for

instance, spokesperson of the Russian Ministry of Foreign Affairs, Maria Zaharova). Some experts started talking about building a “great Russian fire-wall” (Kulikova 2014). The process of actually doing this, however, would be substantially different from that of its Chinese predecessor.

Taking control of Runet required a multidimensional operation that addressed content, technological infrastructure, the structure of ownership of major Internet platforms, shaping the perception of the Internet among Russian citizens and creating a legal environment to support various forms of repressive measures. This took the form of *sovereignization*—that is, the type and scale of control over online space became more and more like the control exercised over offline space (Nocetti 2015). Another notion that applied to the state’s approach to Runet was *fragmentation* (Kolozaridi 2019) or what is sometimes called *balkanization* (Kulikova 2014). Another trend seen in the most recent history of Runet development is the increasing *securitization* of the Russian online space. The online sphere became a major domain in the context of international conflict, which included not only cyberattacks and the use of trolls and bots as a part of state-sponsored propaganda but also the mobilization of users’ resources to support various aspects of warfare. These tendencies were visible in the conflict between Russian and Ukraine (2014–2016) (Asmolov 2019).

The increasing role of regulation and approval of new sovereignization also led to the emergence of a new wave of “digital resistance.” The first wave of protests in April 2018, with about 12,000 participants, addressed the efforts of the Russian authorities to ban Telegram, which led to the blocking of hundreds of other websites as “collateral damage.” The second wave of protests “against the isolation of Runet,” with about 15,000 participants, was triggered by the approval of the “Internet sovereignization” law and took place in March 2019. The new restrictions of sovereignization have been addressed by proliferation of Virtual Private Network (VPN) services and other circumvention tools. In August 2019 Telegram chats and chatbots became a major tool for the coordination of protests after a ban on the participation of opposition candidates in local Moscow elections (for more on digital politics, see Chap. 2).

## 16.6 CONCLUSION

This vector-based historical overview of Runet allows us to identify some important properties of Runet as an object that has been developed as an alternative socio-political and cultural space. First, all the vectors seem to be inter-related. The major trend that can be seen in all the vectors is the increasing conflict between understanding Runet as an alternative phenomenon with its own rules influencing the outer social world and treating it like other entities that follow the offline cultural and political order. This conflict is manifested in the increasing efforts of state institutions to impose various forms of regulation on the online networked environment. This regulation seems to be aimed at

restricting Runet as a construct with a distinct cultural and socio-political role (as seen from the first stages of Runet development), while also offering more space for the Internet in Russia as an instrumental construct that serves a broad spectrum of needs of Russian citizens, from digital consumption to e-government services. Most recent digital innovations offer a broad range of new services and contribute to the development of the Internet in Russia, but it is debatable whether these can be considered part of the continuous development of Runet as a socio-political and cultural object.

The notion of a runaway object highlights the fact that objects are shaped by the continuous activity of a variety of actors who do not necessarily agree about what the object should look like. That said, their activity is still driven by a shared vision of the object to be constituted as a distinct entity with its own boundaries. All the vectors described here demonstrate that the early development of the Russian Internet was driven by various imaginaries of Runet as a socio-cultural project and an alternative political space. It seems, however, that the increase in the number of users, the change of policy on Information and Communication Technologies (ICT) development and the increase in various forms of regulations and other trends not only drastically changed Runet but gradually decreased its salience as an object of participatory socio-political construction.

What continued was the development of the Internet as an advanced form of communication infrastructure in modern society that supports various aspects of people's lives as well as being used by governments as a tool of political influence. However, the decline of Runet is not necessarily an outcome of political Internet regulation but also of a range of socio-technical processes related to the development of the Internet, its accessibility and functions. Moreover, one may argue that the political regulation of the Internet in Russia in fact contributes to the continuation of Runet, since the act of regulation reinforces the boundaries of the object regulated.

We are not necessarily arguing, in imitation of Fukuyama, that Runet is at the end of its history. However, a historical consideration of the Russian Internet seems to suggest a major shift. The main outcome of the trends identified through this historical analysis of five vectors is not increasing state control of Runet but a gradual replacement of Runet by the whole Internet in Russia. That said, Runet and the Internet in Russia continue to co-exist. One may argue that the latent resources of Runet could still be mobilized and take center stage in Russian cyberspace.

## NOTES

1. For example, in the cases of Yandex, which can be considered as the "Russian Google," or of VKontakte, which can be considered as the "Russian Facebook."
2. FidoNet is a worldwide computer network used for communication between bulletin board systems (BBSes).
3. The online archive of the project is available at: <http://www.gagin.ru/vi/>

## REFERENCES

- Alexanyan, K. 2013. *The Map and the Territory: Russian Social Media Networks and Society*. Doctoral Thesis, Columbia University, NY. <https://academiccommons.columbia.edu/doi/10.7916/D8XP7C49>.
- Alexanyan, K., Barash, V., Etling, B., Faris, R., Gasser, U., Kelly, J., Palfrey Jr., J.G., and Roberts, H. 2012. Exploring Russian cyberspace: Digitally-mediated Collective Action and the Networked Public Sphere. *Berkman Center Research Publication No. 2012-2*.
- Alexanyan, K., Etling, B., Kelly, J., Faris, R., Palfrey, J.G. and U. Gasser. 2010. *Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization*. Cambridge, MA: Berkman Center.
- Asmolov, G. 2010. Russia: From ‘Sovereign Democracy’ to ‘Sovereign Internet’? *Global Voices Online*, June 13, 2010. <https://globalvoices.org/2010/06/13/russia-from-sovereign-democracy-to-sovereign-internet/>.
- . 2013a. Dynamics of Innovation and the Balance of Power in Russia. In *State Power 2.0 Authoritarian Entrenchment and Political Engagement Worldwide*, ed. M.M. Hussain and P.N. Howard. Farnham: Ashgate.
- . 2013b. Natural Disasters and Alternative Modes of Governance: The Role of Social Networks and Crowdsourcing Platforms in Russia. In *Bits and Atoms Information and Communication Technology in Areas of Limited Statehood*, ed. S. Livingston and G. Walter-Drop. Oxford: Oxford University Press.
- . 2018. Vertical Crowdsourcing (Russia). In *The Global Encyclopedia of Informality: Towards an Understanding of Social & Cultural Complexity 2*, eds. A. Ledeneva, A. Bailey, S. Barron, and E. Teague. 463–467. London: UCL Press. [http://in-formality.com/wiki/index.php?title=Vertical\\_Crowdsourcing\\_\(Russia\)](http://in-formality.com/wiki/index.php?title=Vertical_Crowdsourcing_(Russia)).
- . 2019. The Effects of Participatory Propaganda: From Socialization to Internalization of Conflicts. *Journal of Design and Science* 6: 20. <https://doi.org/10.21428/7808da6b.833c9940>.
- Asmolov, G., and P. Kolozaridi. 2017. The Imaginaries of RuNet: The Change of the Elites and the Construction of Online Space. *Russian Politics* 2: 54–79.
- Bowles, A. 2006. The Changing Face of the RuNet. In *Control + Shift. Public and Private Usages of the Russian Internet*, ed. H. Schmidt, K. Teubener, and N. Konradova. Norderstedt: Books on Demand.
- Budnitsky, S., and L. Jia. 2018. Branding Internet Sovereignty: Digital Media and the Chinese-Russian Cyberalliance. *European Journal of Cultural Studies* 21 (5): 594–613.
- Castells, M., and E. Kiselyova. 2003. *The Collapse of the Soviet Union. The View from the Information Society*. Los Angeles: Figueroa Press/USC Bookstore.
- Deibert, R., and R. Rohozinski. 2010. Control and Subversion in Russian Cyberspace. In *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*, ed. Ronald Deibert et al. Cambridge: The MIT Press.
- Driscoll, K. 2016. Social Media’s Dial-up Roots. *IEEE Spectrum* 53 (11): 54–60.
- Driscoll, K., and C. Paloque-Berges. 2017. Searching for Missing ‘Net Histories’. *Internet Histories* 1 (1–2): 47–59.
- Engeström, Y. 2008. *From Teams to Knots: Studies of Collaboration and Learning at Work*. New York, NY: Cambridge University Press.



- Ermoshina, K. 2014. Democracy as Pothole Repair: Civic Applications and Cyber-empowerment in Russia. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 8 (3): Article 4.
- Etling, B., K. Alexanyan, J. Kelly, R. Faris, J.G. Palfrey, and U. Gasser. 2010. *Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization*. Cambridge, MA: Berkman Center.
- Fossato, F. 2009. Discussion: Is Runet the Last Adaptation Tool? *Russian Cyberspace* 1 (1): 8.
- Gerovitch, S. 2002. *From Newspeak to Cyberspeak: A History of Soviet Cybernetics*. Cambridge, MA: MIT Press.
- Giddens, A. 2000. *Runaway World: How Globalization Is Reshaping Our Lives*. New York: Routledge.
- Gorny E. 2007, October 18. The Russian Internet: Between Kitchen-Table Talks and the Public Sphere. *Art Margins*. <http://www.artmargins.com/index.php/2-articles/145-the-russian-internet-between-kitchen-tabletalks-and-the-public-sphere>.
- Gunitsky, S. 2015. Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability. *Perspectives on Politics* 13 (01): 42–54.
- Habermas, Jurgen. 1991. The Public Sphere. In *Rethinking Popular Culture: Contemporary Perspectives in Cultural Studies*, ed. C. Mukerji and M. Schudson. Berkeley: University of California Press.
- Jenkins, H. 2006. *Convergence Culture: Where Old and New Media Collide*. New York: New York University Press.
- Katz, E., J.G. Blumler, and M. Gurevitch. 1973. Uses and Gratifications Research. *The Public Opinion Quarterly* 37 (4): 509–523.
- Klyueva, A. 2016. Taming Online Political Engagement in Russia: Disempowered Publics, Empowered State and Challenges of the Fully Functioning Society. *International Journal of Communication* 10: 20.
- Kolozaridi, P. 2019, forthcoming. Internet in Fragments: How Can We Think Over It.
- Konradova, N. 2016. A Putsch in Usenet. A hall of unfulfilled fame, *OpenDemocracy.net*. <https://www.opendemocracy.net/od-russia/natalia-konradova/putch-usenet>. Accessed Aug 16.
- Kukkola, J., and M. Ristolainen. 2018. Projected Territoriality: A Case Study of the Infrastructure of Russian 'Digital Borders'. *Journal of Information Warfare* 17 (2).
- Kulikova, A. 2014, May 19. The 'Balkanisation' of Russia's Internet. *OpenDemocracy.net*. <https://www.opendemocracy.net/od-russia/alexandra-kulikova/%E2%80%98balkanisation%E2%80%99-of-russia%E2%80%99s-internet>.
- Kuznetsov, S. 2004. *Osypyaâ slona: Zametki po istorii russkogo Interneta [Feeling an Elephant: Notes on the History of the Russian Internet]*. Moscow: Novoe literaturnoe obozrenie.
- Lebedev, Pavel. 2018. Cifrovaa èpoha. Zaterânnnye v socsetâh [The Digital Age. Lost in Social Networks]. IPSOS Views. <https://www.ipsos.com/ru-ru/cifrovaya-epokha-zateryannnye-v-socsetyakh>.
- Lihachev, N. 2015. 21-letie Runeta [21 Years of Runet]. *TJournal*. <http://tjournal.ru/paper/runet-21>.
- Mansell, R. 2002. From Digital Divides to Digital Entitlements in Knowledge Societies. *Current Sociology* 50 (3): 407–426.
- . 2012. *Imagining the Internet: Communication, Innovation, and Governance*. Oxford: Oxford University Press.



- Markham, A.N. 1998. *Life Online: Researching Real Experience in Virtual Space*. Lanham, MD: Rowman Altamira.
- Nisbet, E. 2015. *Benchmarking Public Demand: Russia's Appetite for Internet Control*. Philadelphia: Center for Global Communications Studies. <http://www.global.asc.upenn.edu/publications/benchmarking-public-demand-russias-appetite-for-internet-control/>.
- Nocetti, J. 2015. Contest and Conquest: Russia and Global Internet Governance. *International Affairs* 91 (1): 111–130.
- Oates, S. 2013. *Revolution Stalled: The Political Limits of the Internet in the Post-Soviet Sphere*. Oxford: Oxford University Press.
- Omelchenko, E.L. 2019. Is the Russian Case of the Transformation of Youth Cultural Practices Unique? *Monitoring of Public Opinion: Economic and Social Changes* 1: 3–25.
- Ostrovsky, A. 2019, March 7. Russians Are Shunning State-controlled TV for YouTube. *The Economist*. <https://www.economist.com/europe/2019/03/09/russians-are-shunning-statecontrolled-tv-for-youtube>.
- Peters, B.J. 2016. *How Not to Network a Nation. The Uneasy History of the Soviet Internet*. Boston, MA: MIT Press.
- Popkova, A. 2014. Political Criticism from the Soviet Kitchen to the Russian Internet: A Comparative Analysis of Russian Media Coverage of the December 2011 Election Protests. *Journal of Communication Inquiry* 38 (2): 95–112.
- Ristolainen, M. 2017. Should 'RuNet 2020' Be Taken Seriously? Contradictory Views About Cybersecurity Between Russia and the West. *Journal of Information Warfare* 16 (4): 21.
- Rohozinski, R. 1999. Mapping Russian Cyberspace: Perspectives on Democracy and the Net. *United Nations Research Institute for Social Development (UNRISD), Discussion Paper 115*. New York: United Nations.
- Rubin, M., and Badanin, R. 2018. Telega Iz Kremlâ [The cart from the Kremlin]. Proekt.Media. <https://www.proekt.media/narrative/telegram-kanaly/>.
- Schmidt, H. 2014. Russian Literature on the Internet. Digital Russia: The Language, Culture and Politics of New Media Communication. In *Digital Russia: The Language, Culture, and Politics of New Media Communication*, ed. M. Gorham, I. Lunde, and M. Paulsen, 177–194. London and New York: Routledge.
- Schmidt, H., and K. Teubener. 2006. Our RuNet. Cultural Identity and Media Usage. In *Control+ Shift. Public and Private Usages of the Russian Internet*, ed. H. Schmidt, K. Teubener, and N. Konradova, 14–20. Norderstedt: Books on Demand.
- Schmidt, H., K. Teubener, and N. Zurawski. 2006. Virtual (Re) unification? Diasporic Cultures on the Russian Internet. In *Control + Shift. Public and Private Usages of the Russian Internet*, ed. H. Schmidt, K. Teubener, and N. Konradova. Norderstedt: Books on Demand.
- Shklovski, I., and D.M. Struthers. 2010. Of States and Borders on the Internet: The Role of Domain Name Extensions in Expressions of Nationalism Online in Kazakhstan. *Policy and Internet* 2 (4): 107–129.
- Shubenkova, A.Y., and P.V. Kolozaridi. 2016. The Internet as a Subject of Social Policy in the Official Discourse of Russia: A Benefit or a Threat? *Journal of Researches of Social Policy* 14 (1): 39–54.
- Sibgatullin, A.A. 2009. Tatarskij Internet [Tatarian Internet]. Medina. [http://www.idmedina.ru/books/history\\_culture/?1802](http://www.idmedina.ru/books/history_culture/?1802).

- Silverstone, R. 2002. Complicity and Collusion in the Mediation of Everyday Life. *New Literary History* 33: 745–764.
- Soldatov, A., and I. Borogan. 2015. *The Red Web: The Struggle Between Russia's Digital Dictators and the New Online Revolutionaries*. New York: Public Affairs.
- Toepfl, F. 2011. Managing Public Outrage: Power, Scandal, and New Media in Contemporary Russia. *New Media & Society* 13 (8): 1301–1319.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



PART II

---

Digital Sources and Methods



## Corpora in Text-Based Russian Studies

*Mikhail Kopotev, Arto Mustajoki,  
and Anastasia Bonch-Osmolovskaya*

### 17.1 INTRODUCTION

This chapter focuses on textual data that are collected for a specific purpose, which are usually referred to as *corpora*. Scholars use corpora when they examine existing instances of a certain phenomenon or to conduct systematic quantitative analyses of occurrences, which in turn reflect habits, attitudes, opinions, or trends. For these contexts, it is extremely useful to combine different approaches. For example, a linguist might analyze the frequency of a certain buzzword, whereas a scholar in the political, cultural, or sociological sciences might attempt to explain the change in language usage from the data in question. This handbook is no exception: the reader will find several chapters (for additional information, see Chaps. 26, 23, 29 and 24) that are either primarily or secondarily based on Russian textual data.

Russian text-based studies represent a well-established area of science, unknown in part to Western readers due to the language barrier. However, this

---

M. Kopotev (✉)

Higher School of Economics (HSE University), Saint Petersburg, Russia  
e-mail: [mkopotev@hse.ru](mailto:mkopotev@hse.ru)

A. Mustajoki

Higher School of Economics (HSE University), Moscow, Russia

University of Helsinki, Helsinki, Finland

e-mail: [arto.mustajoki@helsinki.fi](mailto:arto.mustajoki@helsinki.fi)

A. Bonch-Osmolovskaya

Higher School of Economics (HSE University), Moscow, Russia  
e-mail: [abonch@hse.ru](mailto:abonch@hse.ru)

should not overshadow the existence of well-developed tools and promising results (Dobrushina 2007; Mustajoki and Pussinen 2008; Plungian 2009). Naturally, scholars in linguistics have made the most visible progress in corpus studies, offering a wide spectrum of data (described in Sect. 17.3 of this chapter) and a range of corpus-based methods that are reflected in recent publications (Plungian 2009; Plungian and Shestakova 2014; Zobotkina 2015; Lyashevskaya 2016; Kopotev et al. 2018).

In the chapter, we describe existing textual resources in Russian, from available online sites to DIY (“do-it-yourself”) corpora, with a special focus on two of the most significant examples: the Russian National Corpus and the Integrum database. Finally, in the last section, we present two cases of corpus-based analysis: the first investigates the collective mnemonic patterns for names of decades in Soviet and post-Soviet history and the second concerns political trends in modern Russia.

T. McEnery and A. Wilson (1996, 24) offer the following definition of a corpus:

Corpus in modern linguistics, in contrast to being simply any body of text, might more accurately be described as a *finite-sized body of machine-readable* text, sampled in order to be *maximally representative* of the language variety under consideration. (Italics added)

Three features of this definition need to be highlighted as they constitute the quality criteria for any corpus data. The first is that it is finite-sized. This means that the number of tokens is known so the user can apply various statistics to the data, ranging from simple frequency rankings to sophisticated neuronal algorithms. The second quality is that it is in a machine-readable format that allows users to conduct quick searches within an unlimited amount of data, from Tolstoy’s masterpieces to ordinary texts available on the internet. The third quality is maximal representativeness, which makes it possible to draw conclusions from a finite number of examples on the infinity of a language or its variety. In this sense, the usage of corpora in the humanities makes it similar to a hard science, meaning that the results are calculable and replicable, and thus able to be tested.

## 17.2 THE WEB AS A CORPUS

The emergence of search engines such as Yahoo, and later Google, has made it possible to explore the World Wide Web and its expanding massive number of sites. This development has given rise to new verbs such as “googling” (meaning to search on [google.com](http://google.com)) and “yandexing” (to search on [yandex.ru](http://yandex.ru)). The Russian part of the global internet is often referred to as the Runet (for additional information, see Chap. 16). This includes not only sites under the country code’s top-level domain. RU but every site

available in the Russian language. Runet had a six percent share of all internet sites for 2018, putting it in second place after advanced English (see Usage 2019). However, a clear differentiation should be made between search engines that index websites and corpora. Search engines that index websites allow users to make searches, whereas corpora constitute data, the results of which are controlled and replicable.

Whichever commercial search engine is used, it is primarily intended to deliver information that includes, first and foremost, marketing material that targets specific consumer groups. One can, of course, use the internet for information mining but the results may be scientifically unreliable without additional verification. Information from data mining tends to contain drivel attributable to varying spelling norms, scanning errors, fluctuation in internet communication, and so forth. As Adam Kilgarriff observes:

[L]ike Borges's Library of Babel, [the internet] contains duplicates, near duplicates, documents pointing to duplicates that may not be there, and documents that claim to be duplicates but are not. (Kilgarriff 2001, 342)

A simple internet search yields *a priori* unknown results, which are usable only if they are task-specific and the researcher is cognizant of all the limitations. Even then, using the internet as a source is fraught with serious risks. Among the most serious is the fact that users do not control the data they search and they do not control the search engines they use (see Bozdag 2013; Flaxman et al. 2016).

It is difficult to conduct data-based research without texts that are *reliable* and *accessible*. By reliable, we mean texts that are consistently of high quality, and by accessible, we refer to texts that are easily obtainable. A general caveat with regard to the data that are available online is that the smaller the text and the more unique its contents, the more reliable the source should be. If the features of an individual text are not crucially important, then any potential noise in the data can be ignored, at least to some extent. A large amount of noisy data may nonetheless be used effectively to study general tendencies in the language variety under consideration. For example, a noise would be caused by errors related to a source, as in mixing Latin and Cyrillic letters after Optical Character Recognition (OCR) processing, and these are dissolved in the total mass of data.

Electronic texts that are available on the internet fall into one of three, uneven, categories: the majority are insufficiently prepared (e.g., a source is not reliable), error-filled (e.g., inaccurately digitized), or non-authorized (such as a doubtful copyright status). A smaller amount of textual data, with more attention given to their quality, can be further categorized as non-linguistic collections, or "electronic libraries," and linguistically oriented collections, or "linguistic corpora." Naturally, the distinction between non-linguistic and linguistic data is somewhat vague and depends heavily on the task at hand, the

main difference being whether or not the data are linguistically annotated, that is, enriched with linguistic information.

### 17.3 ELECTRONIC LIBRARIES

Collections of texts are not corpora in the strict sense of the term. However, large text collections have a wide circulation in digital studies and are reliable resources for Russian studies. The largest of these collections on Runet are Moshkov's Library ([www.lib.ru](http://www.lib.ru)) and Librusec ([www.lib.rus.ec](http://www.lib.rus.ec)).<sup>1</sup> Access to both sites is free and includes massive collections of fictional and non-fictional Russian texts. Furthermore, both could serve as good initial sources for big-data studies in Russian digital humanities (for more, see Chap. 29).

When the research objective is to analyze literary masterpieces, the sources need to be more carefully selected. In this context, Runet has three useful websites that aim to provide high-quality data. The first is the Fundamental Electronic Library of "Russian Literature and Folklore" ([www.feb-web.ru](http://www.feb-web.ru)), which is a fast-developing collection of belles lettres that follows the strict guidelines of academic publications, enriched with commentaries and an extended reference apparatus. The website contains fiction from the eighteenth to the twentieth century as well as Old Russian literature and folklore. The second resource is the Russian Virtual Library ([www.rvb.ru](http://www.rvb.ru)). The content, principles, and developers of this collection partly overlap with the Fundamental Electronic Library, although the latter focuses more on published Russian texts from the eighteenth century, the *fin de siècle*, and from Soviet underground poetry. The third resource, [lib.pushkinskijdom.ru](http://lib.pushkinskijdom.ru), is maintained by the Institute of Russian Literature (RAS, also known as *Pushkinskij dom*). This site provides access to thousands of texts from the ninth to the twentieth century. These consist mainly of fiction and poetry, but also memoirs, critical reviews, and critical bibliographies. A true gem of the collection is the library of Old Russian literature, which includes most of the surviving ancient texts and their Russian translations.

### 17.4 LINGUISTIC CORPORA<sup>2</sup>

While the aforementioned sources are sufficient for many researchers, linguists require resources that are specifically designed for their analyses of language phenomena. These are referred to as "linguistic corpora," which means that the entries are enriched with specific linguistic information. Some examples of this are tokenization, lemmatization, part-of-speech tagging, and syntactic relations. This detailed information enables scholars who are more interested in the linguistic content of the texts to search in sources that are more directly oriented to linguistic information.

The dawn of computer-assisted research in the Russian language occurred at the turn of the twenty-first century, which was shortly after the emergence of



resources specifically designed to meet the needs of linguistics scholars, namely linguistic corpora. Russian corpus linguistics is currently a highly developed branch of linguistic studies and is well represented in national computational linguistic landscapes (see the “Dialogue” conferences at [www.dialog-21.ru/en](http://www.dialog-21.ru/en)) and in international collaboration (see, e.g., Erjavec et al. 2010; Nivre et al. 2018). The following extensive “big data” resources were made available from the beginning, presented below in an ascending order of tokens:

- the Araneum Russicum corpora of 1.2 billion tokens (Benko 2014);
- the ruWac: the Russian portion of the project “The Web as a Corpus” of 1.3 billion tokens (Sharoff and Nivre 2011);
- the Taiga corpus of 5 billion tokens (Shavrina and Shapovalova 2017);
- ruTenTen of 14.5 billion tokens, a member of the commercial TenTen corpus family (Jakubíček et al. 2013);
- General Internet Corpus of Russian of 19.8 billion tokens (GICR; see Belikov et al. 2013).

The above list of corpora and resources is by no means comprehensive, and many smaller, more specific and more deeply annotated corpora are available for academic use (see the catalogue at [www.ruscorpora.ru/new/corpora-other.html](http://www.ruscorpora.ru/new/corpora-other.html)). There are also various historical and parallel corpora, as well as corpora that are not publicly available, which are beyond the scope of this chapter (see reviews in Mitrenina 2014; Mikhailov and Cooper 2016; Kopotev et al. 2018). Nonetheless, in many cases, the best available option is to create a task-specific corpus.

A do-it-yourself (DIY) corpus eliminates many issues caused by raw internet data, such as repetition, disproportion, and babelization (language mixture). Many special tools have been developed to create DIY corpora, typically referred to as a “concordancer” or “corpus manager” (see [https://en.wikipedia.org/wiki/Corpus\\_manager](https://en.wikipedia.org/wiki/Corpus_manager)). Researchers can use these programs to look up contexts, construct lists of keywords or frequencies, analyze word co-occurrences, and determine the distribution of words across texts or topics. A reliable option that is available to scholars is the commercial Sketch Engine service and its non-commercial version, No Sketch Engine ([www.sketchengine.eu/nosketch-engine](http://www.sketchengine.eu/nosketch-engine)). The service includes many specific linguistic tools that are available upon registration.

#### 17.4.1 *The Russian National Corpus ([www.ruscorpora.ru](http://www.ruscorpora.ru))*

A national corpus of any language, the acme of linguistic resources, is characterized by two fundamental features. First, it is essential that the corpus represent the entire language in question. This means that it should contain all types of communication, both written and oral, in all genres, from the belletristic to the dialectal, and represent all historical periods, from antiquity to the present. Second, it should be maximally balanced insofar as the text types in the corpus

correspond to their proportion of usage in real-life communication to the extent that it is feasible, taking into consideration aspects such as data availability and legal restrictions.

A national corpus makes it possible to conduct a wide range of linguistic analyses into the language for which it is available. As the creators of the Russian National Corpus (hereafter RNC) explain:

[Electronic] libraries are not well suited to academic work on the nature of language; they tend to focus on the content of texts rather than their language properties, while the creators of the Corpus recognize the importance of literary or scientific value of the texts, but see them as a secondary feature. Unlike an electronic library, the National Corpus is not a collection of texts which are deemed “interesting” or “useful” of themselves; the texts in the Corpus are interesting and useful for the study of language. Such texts might include not only great works of literature, but also works of a “secondary” writer, or a transcription of an ordinary conversation. (<http://www.ruscorpora.ru/en/corpora-intro.html>)

Since the RNC became available in 2004, it has developed into a functional and extensively annotated resource. Today, in terms of its size and scientific value, it is comparable to the American, British, Czech, and Polish national corpora. The core collection of the RNC includes manually selected samples of written and spoken texts. Those samples represent various genres, such as fiction, drama, memoirs, news and literary criticism, popular non-fiction and textbooks, religious and technical texts, business and jurisprudence papers, and texts on daily life. The samples include texts that were not initially intended for publication.

Any national corpus by definition is large and multifaceted. At the time of writing, all subcorpora and spin-off projects available on the [ruscorpora.ru](http://www.ruscorpora.ru) site comprise 600 million tokens. Table 17.1 lists the detailed statistics on the main

**Table 17.1** Russian National Corpus: texts by subcorpora

<i>Subcorpora</i>	<i>Number of texts</i>	<i>Number of sentences</i>	<i>Number of tokens</i>	<i>% of tokens</i>
The main subcorpus	76,882	17,574,752	209,198,275	57.3
The news-media subcorpus	181,175	8,553,495	113,292,003	31.0
The dialectal subcorpus	197	20,273	194,283	0.1
The educational subcorpus	229	65,666	664,751	0.2
The parallel subcorpus	370	1,609,609	24,022,437	6.6
The poetry subcorpus	41,448	638,861	6,738,474	1.8
The oral subcorpus	3034	1,604,626	10,122,579	2.8
The multimodal subcorpus	31,741	148,619	648,576	0.2
<i>In total:</i>	<i>335,076</i>	<i>30,215,901</i>	<i>364,881,378</i>	<i>100</i>

Source: <http://www.ruscorpora.ru/corpora-stat.html>. The English translation is ours

**Table 17.2** Russian National Corpus: texts by creation date (the main subcorpus only)

<i>Periods</i>	<i>Number of texts</i>	<i>Number of sentences</i>	<i>Number of tokens</i>	<i>% of tokens</i>
1701–1750	298	27,090	590,541	0.3
1751–1800	979	176,207	2,981,803	1.4
1801–1850	1098	704,678	10,380,375	4.8
1851–1900	2063	2,366,209	31,761,447	14.7
1901–1950	26,325	4,646,823	53,445,536	24.7
1951–2000	14,486	6,172,190	67,252,763	31.0
2001–2010	31,491	4,094,011	50,231,677	23.2
<i>In total:</i>	<i>76,740</i>	<i>18,187,208</i>	<i>216,644,142</i>	<i>100</i>

Source: <http://www.ruscorpora.ru/corpora-stat.html>. The English translation is ours

parts of the collections; Table 17.2 provides additional details on the core collection. The represented time periods vary due to the availability of the digitized sources of the particular period.

All the subcorpora are lemmatized, which occurs when all forms of a word are arranged under a headword as in dictionary form, called *lemma*, and annotated both morphologically and syntactically. Some of the subcorpora are also analyzed semantically (grouped in lexical classes according to the meaning) and derivationally (grouped by word formation). The crowning touches of this monumental resource are its diverse rich metadata and sophisticated search options, such as multiword expressions, tag repetition in adjacent tokens, and stress marking.

The site also hosts several spin-off projects of which the most interesting is the Old Russian subcorpus (Pichhadze 2005), which includes original Old Russian texts (such as chronicles and Novgorodian birch-bark letters) as well as translations from Greek texts (e.g., *The Romance of Alexander*, Flavius Josephus's *Books of the History of the Jewish War against the Romans*) and South Slavic texts, rewritten in Old Russian (e.g., *Izbornik* [Miscellany] of 1076). Other notable projects are the SynTagRus corpus (Boguslavsky et al. 2000), which is manually annotated with syntactic dependency and lexical function markups, and the FrameBank (Lyashevskaya and Kashkin 2015), which is annotated with semantic roles. To the best of our knowledge, the RNC is also the only resource that includes a corpus of Russian poetry, which allows searches by meter and rhyme of poetic texts from the eighteenth century to the present (Grishina et al. 2009).

#### 17.4.1.1 Case Study: Tracking Collective Memory Through “Decade Constructions”<sup>23</sup>

The study of collective memory is a strong interdisciplinary field that concentrates on the exploration of collective mnemonic concepts. The aim is to analyze how and why people and society think about and collect the events of their mutual past. This research objective has drawn the attention of historians,

scholars of cultural studies, and anthropologists. However, this has almost never been addressed by linguists, despite the generally acknowledged importance of language as a key translator of culture (Lotman 2009; Koselek 2004). Attempts to explore the Russian collective memory through corpus analysis have been made by Bonch-Osmolovskaya (2018) and Götzelmann et al. (2019). The former analysis focuses on the constructions, which include a word-denoted decade preceded by an epithet, such as *libie devânostye* (wild nineties), *zolyte pâtidêsâtye* (golden fifties), and *groznye tridcatye* (terrible thirties). We refer to them hereafter as *decade constructions*. The basic assumption is that these constructions reflect the mnemonic patterns of each decade in Soviet and post-Soviet history; hence, their linguistic analysis makes it possible to reconstruct patterns of collective memory.

The data obtained from the Russian National Corpus have been re-organized so that the final dataset had a total of 242 sentences with decade constructions, which refer to the period from the 1920s until the 1990s. A non-trivial semantic feature of this construction is that the ordinal, such as *dvadcatye* (twenties), refers to a timespan that does not fully coincide with a corresponding decade. A timespan is perceived as a featured historical period, with specific connotations, expressed by an adjective and shared between a speaker and an audience. As Zerubavel (2003, 31) observes, the corpus analysis of decade constructions reveals a non-even distribution of historical periods so that “hills and valleys” appear in the collective memory. Some decades seem to be salient and prominent mnemonic concepts, whereas others remain almost forgotten.

Frequency analyses of the examples have their own methodological specificity. Most corpus methods focus on the most frequent entries, and those that are statistically non-significant are typically not considered. In this case, however, even a unique entry should not be neglected and must be included in the analysis, as the adjective still refers to a shared collective concept that can only be understood if this association occurs. Figure 17.1 displays the overall frequency distribution of the construction for each decade. The radar-chart values for each decade correspond to the mean value for all constructions. Table 17.3 presents the number of constructions that occur in the RNC for each ordinal.

It is clear from both Fig. 17.1 and Table 17.3 that the distribution is not even. Naïve chronology covers almost all of the decades in the twentieth century, but some are more important (the 1930s and 1990s). Some decades are rarely referred to (the 1950s and 1980s), which means that they do not form a mnemonic pattern and barely exist in the collective memory. The 1990s, which was the turbulent period of post-Soviet political and economic transition and a time of intensive and highly emotional social reflection, display the highest frequency, whereas the 1950s and the 1980s represent the lowest, which is less than the overall mean. These two periods coincide with the end of two historical epochs: Stalin’s reign of terror and Brezhnev’s era of stagnation. One might speculate that they do not form a holistic mnemonic pattern because they are more likely to represent a rupture between the preceding and subsequent decades.

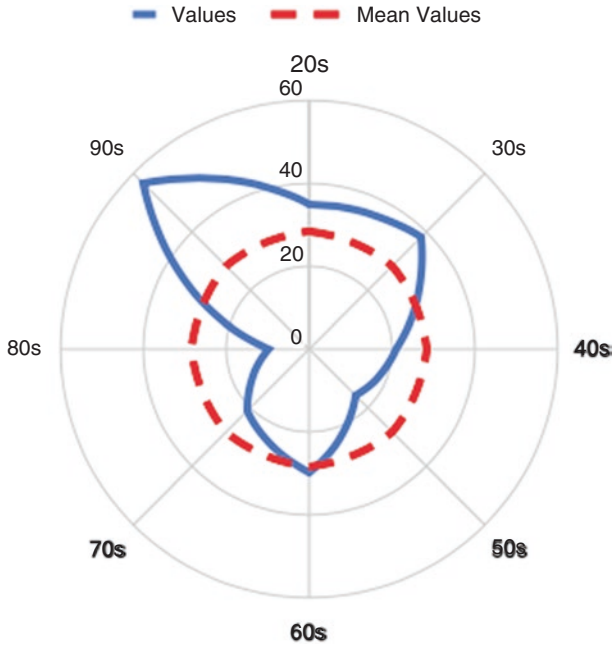


Fig. 17.1 Frequency of adjective decade constructions for each decade

Table 17.3 Frequency of adjective decade constructions for each decade

Decades	Construction frequency
1920s	35
1930s	38
1940s	21
1950s	16
1960s	30
1970s	21
1980s	10
1990s	57

Our research continues with the analysis of 117 adjectives, which are used with the ordinals in question and fall into several semantic classes. The first three classes are united by the semantics of direct or indirect emotional assessment toward an ordinal. The epithet basically defines the decade as a separate cultural phenomenon with specific symbolic meaning; the epithet also contains a built-in assessment of the epoch by the speakers. These are adjectives that refer to real-world attributes that are characteristic of the historical period, such as *ateističeskie dvadcatye* (atheistic twenties), *stilážnye päťidesätje* (dandy fifties), and *banditskie devänostje* (gangster nineties). Another major class

comprises adjectives of positive or negative assessment, which include metaphorical expressions, such as *libie devânostye* (wild nineties). There are also adjectives that emphasize the prominence of the decade that cannot be classified as either positive or negative, such as *nepovtorimye devânostye* (unique nineties) and *rokovye sorokovye* (fatal forties). Two more adjectival classes are connected by spatial or geographical references, such as *sovetskie semidesâtye* (Soviet seventies) and *moskovskie šestidesâtye* (Moscow sixties), or by temporal references of which the most frequent are *rannie* (early) and *pozdnie* (late). One might expect the latter two to reflect a common characteristic of any decade, but this is not the case because their distribution across the decades is uneven (see Fig. 17.2): the concept “early/late” is not selected randomly but corresponds to micro-historical patterns. Hence, the “early thirties” is a period that precedes the Great Terror, which is not referred to as the “late thirties” because it has its own name. On the other hand, the “late fifties” and “early sixties” combined constitute the conceptual memory of *the Khrushchev Thaw* (Rus. *ottepel'*).

As Fig. 17.1 indicates, “the nineties” is the most frequently occurring nomination in the dataset and it represents a very special case of collective memory modeling. Approximately 70 percent of all “nineties” examples contain attributes of either a positive or negative assessment. The most common is *libie devânostye* (wild nineties), which occurs 14 times (30%). However, on 10 of those occasions, the adjective *libie* (wild) is enclosed within quotation marks, which makes the whole pattern more complex. One might assume that the speaker uses quotation marks to refer not to the collective memory but to the

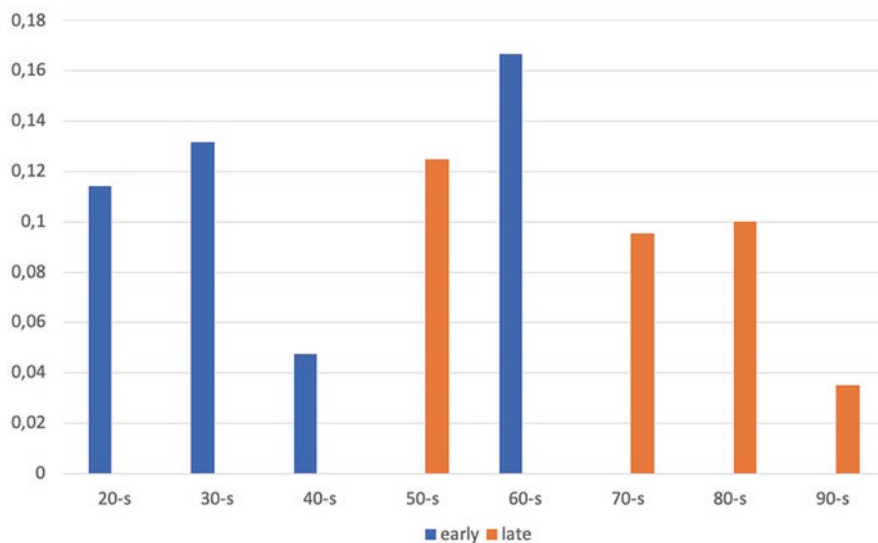


Fig. 17.2 Distribution of *rannie* (early) and *pozdnie* (late) in decade constructions

preceding contextual usage of the expression specifically adopted by those in power. This is where the process of lexicalization begins and initial conceptual semantics fade. This is even more obvious when examining the Newspaper subcorpus within the RNC (about 133 million tokens from 2001 to 2014). The “nineties” constructions again constitute the dominant majority, comprising approximately 50 percent of all the examples, of which 30 percent is *libie devánostye* (wild nineties). However, the marked difference in distribution demonstrates that the collective memory of the post-Soviet nineties was formed later in the noughties when it became a phrasal cliché through the perpetual repetition of *libie devánostye* (wild nineties) in the media. Figure 17.3 presents the rapidly increasing frequency of the “wild” nineties compared to all other adjectives followed by the ordinal; “wild” becomes nearly dominant from 2008 to present. Having become a fixed-word combination, *libie devánostye* (wild nineties) no longer triggers collective memory but is instead a meme, a semantically bleached language sign that has nothing in common with the concept of “wildness and chaos,” which is something that could be associated with the period in question.

The case study presented above demonstrates the potential usefulness of relatively small datasets in collecting promising historical observations on “memory landscapes” by using linguistic corpora. Although the dataset is too small to apply standard statistical measures, the qualitative analysis of symbolic value provides an alternative basis for interpretation, which is based on evidence rather than statistics. There is no single occurrence of the construction nor is a single use of the adjectives random because they are all bricks in the construction of a controversial and multifaceted collective memory. What is of significance here is the reliability of the data: it is a corpus that is balanced in

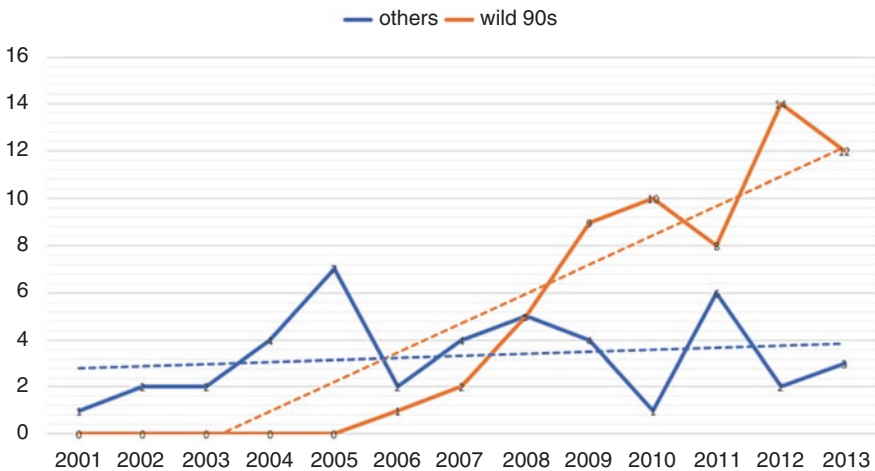


Fig. 17.3 Frequencies of *libie devánostye* (wild nineties) compared to all adjectives attested in the construction (2001–2013, the Newspaper subcorpus)



terms of both genre and timespan. Its morphological mark-up also allows the user to search not only for a word but also for a moving context that yields insights that are otherwise inaccessible.

#### 17.4.2 *Integrum* ([www.integrumworld.com](http://www.integrumworld.com))

Although it is not a corpus in the strict sense, the Integrum database of Russian media has features that render it extremely useful for research purposes in comparison with both linguistic corpora and biased raw Internet data (for a comparison, see Mustajoki 2006; Plungian 2006). The service is not free, but libraries and universities throughout the world provide access to it online.

The main benefit of Integrum is that it covers almost all newspapers and magazines published in Russia from the beginning of the 1990s. Thus, users have easy access to the full texts of metropolitan media publications, such as *Izvestia* and *Komsomolskaya pravda*, as well as to far more remote and thus difficult to obtain media including *Vesti respubliki* (Grozny, Chechnya), *Večernij Murmansk*, and *Saratovskaya panorama*. Dozens of Russian-language newspapers published outside Russia are likewise available, including *Evropa-Èkspress* (Berlin), *Karavan* (Kazakhstan), and *Minskij kur'er* (Belarus). Complementing the printed media, Integrum also includes a wide variety of data from radio and television broadcasts, online media, news agencies, and legislation. A total of approximately 200 million texts are available, which means many more than 50 billion running words.

A researcher can find some of the materials available in Integrum elsewhere on the Internet. Yet what makes Integrum invaluable is the thorough categorization of the data. Within the categories, users can search for further sources of interest simply by clicking on a given list of resources. This option is especially useful for those who are interested in examining different opinions on political issues, such as pension reforms throughout Russia, or in comparing regional differences in attitudes, such as how foreign powers are perceived in the eastern part of Siberia versus attitudes that prevail in the capital region.

The data in Integrum are not deeply morphologically annotated, but the search options are diverse nonetheless. To make searches, users can utilize tokens (word forms), lemmas (words), or parts of words (using wildcards). It is possible to determine the distance between the searched words, that is, how far apart they are to be included in the results. For example, the query [*modernizac\* :3 Rossi\**] returns all contexts in which all forms of the words occur within one to three words of each other. In addition, a brief excerpt and the full text are provided for the examples found. Researchers may also conduct more sophisticated searches to create macros that enable them to more precisely pinpoint the passages they find most interesting and useful. For anyone with a limited command of Russian, one available option is to make a quick automatic English translation in the search box. Thus, a look-up value, such as “digital Russia,” returns texts containing corresponding Russian words highlighted in Russian-language articles.

#### 17.4.2.1 Case Study: Political Buzzwords in Russian<sup>4</sup>

Integrum is intended primarily for business people, journalists, and scholars who are interested in Russian society and politics, and the economy, but it can also be used effectively in linguistic studies, such as to determine how people use the Russian language (see Mustajoki and Pussinen 2006, 2008). Below we present a case that demonstrates the use of Integrum in interdisciplinary research to examine attitudes toward the modernization process of the Russian media, with special reference to events that made modernization impossible to achieve.

Although “modernization,” or *modernizaciâ* in Russian, has a colloquial usage, its appropriation by Dmitry Medvedev’s administration made it a buzzword that is identifiable as a marker in certain types of political discourse. This word became a central concept in Medvedev’s political program during his presidential term (2008–2012). Thus, *modernizaciâ* has both political and economic connotations and has continued to be associated with Medvedev and his politics.

In their study of media texts on modernization, Laine and Mustajoki (2017) concentrated on the period from December 31, 2000, to December 31, 2012, because it covers the rise and fall in usage of this notion in Russian media discourse. Within that timeframe, 94,500 occurrences of the word *modernizaciâ* in all its forms were detected in 350 national Russian newspapers (see Fig. 17.4).

A preliminary investigation of the examples revealed that discussion related to the concept was frequent, but that the overall attitude was rather skeptical.

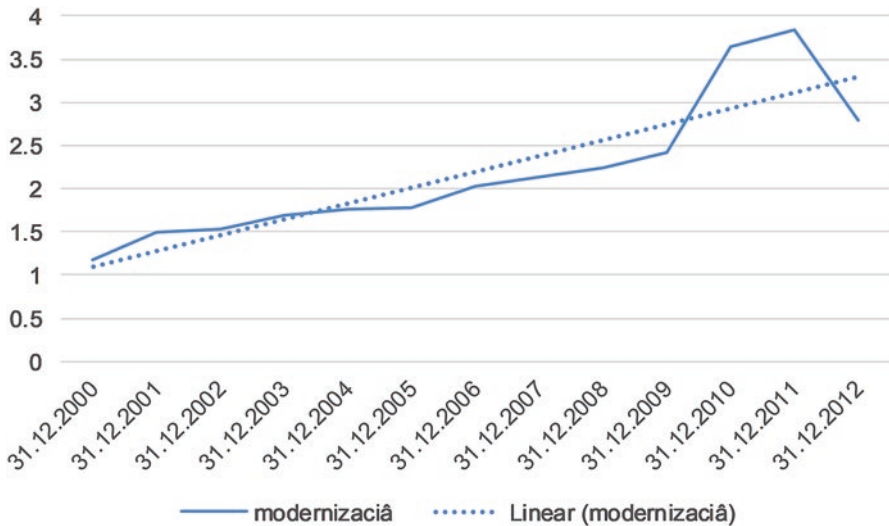


Fig. 17.4 The relative frequency (%) of *modernizaciâ* (modernization) occurring in texts from Russian national newspapers (Source: Integrum, Dec. 31, 2000—Dec. 31, 2012)

Many writers welcomed the modernization process per se, but they expected it to fail as did all previous attempts at reform. They insisted that reform could only succeed if *X* were to take place, *X* being something specific that should be undertaken, as in the following example:

*Bez ètoj mobil'nosti nevozmožna modernizaciâ strany, a značit, gosudarstvu pri-detsâ pojti na strukturnye izmeneniâ v èlitah.* (RBC, July 1, 2008)

Without this mobility, modernization of the country is impossible, which means that the state will have to go for structural changes in the elites. (RBC, July 1, 2008)

Our observation corresponds to that of Juri Prokhorov and Iosif Sternin (2006, 67–68), who claimed that Russians tend to typically adopt reasoning based on a “single-explanation” in their public representations of themselves. These sociolinguists examined the cliché that Russians search for a centralized solution for all problems and put their trust in quick and simple resolutions for complex problems. According to Prokhorov and Sternin, what lies behind this stereotype is the historically grounded, left-leaning reasoning that responsibility for everything rests with *oni* (in Russian, “they”; here, “the ones with power”). This responsibility pertains to not only the country’s prosperity but also the well-being of the nation. “They” may be personalized, as a czar or a president, or it may be an abstract concept referring to those who have power. The implicit belief underlying this attitude is that the solution lies outside and above, not with the people themselves, whereas “they”—the ones with the power—have the opportunity, the capability, to make life better in Russia.

Laine and Mustajoki (2017) used the multistage cascade search technique to explore that line of argument more deeply as it applies to the concept of modernization. As a first step, all contexts of all forms of the word *modernizaciâ* were extracted. Thereafter, only contexts that referred to the modernization of the whole country were considered further, rather than those that related to a specific sector, such as transportation, education, or the army. To achieve this, they introduced additional search criteria: contextual conditions, which restrict the context to all-Russian modernization, for example, *modernizaciâ + Rossii* (modernization of Russia) or *modernizaciâ strany* (modernization of the country). More detailed restrictions were applied during the next step—finding the “single-explanation” argument. This means that certain expressions had to be attested in a nearby context within the same sentence, such as [*modernizaciâ*] *vozmožna, tol'ko esli* ([modernization] is possible only if) or [*dlja modernizacii*] *neobhodimo* ([for modernization,] it is necessary to). The corpus was restricted to the news media, which excluded scientific articles, official documents, and historical texts. In total, approximately 100 contexts were subject to further detailed analysis.

To summarize, according to the results by Laine and Mustajoki, the factors that obstruct modernization fall into several categories: (a) economic (such as a low level of investment in industry, raw-material dependency and a lack of

“civilized” competition); (b) scientific and educational (the country should create the “necessary” environment for young scientists and “normal” conditions for specialist education in order to avoid a national brain drain); and (c) political (controversial opinions such as “Under this rule, modernization is impossible” and “Only Putin would have the ability to modernize the country”; “The party in power, United Russia, can ensure the success of modernization”).

The Russian word *importozamêsenie*, which is both difficult to pronounce and comprehend, means “import substitution” in a Russian-specific sense. A new phase of Russian political rhetoric began in 2012 when Putin embarked on his successive terms in the Kremlin. The context of both his third and fourth terms was that of empowered authoritarianism. After the annexation of Crimea, the European Union (EU), the United States (US) and some other countries imposed sanctions on Russia, and Russia enacted counter-sanctions on EU products (see Travin et al. 2020). In the changed political situation, President Putin introduced the new concept of *importozamêsenie* (import substitution), among other buzzwords. Its meteoric rise in the media is astonishing and comparable with that of “Russian modernization”; Fig. 17.5 illustrates how quickly its frequency increased in Russian media coverage from 2014 onward.

The “single-explanation” comments were again attested in the data after the new buzzword appeared. This time, the explanations tempering the effect of *importozamêsenie* (import substitution) included the competitiveness of Russian enterprises and a new attitude toward agriculture:

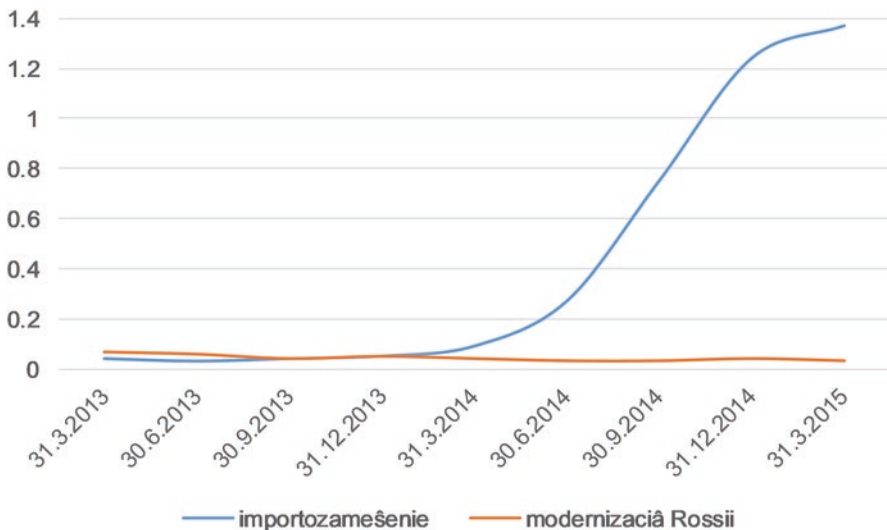


Fig. 17.5 The usage of *modernizaciâ Rossii* (modernization of Russia) in comparison to *importozamêsenie* (import substitution) (Source: Integrum, Russian National Media, 2013–2015)

*Vozmožno, ambicioznye plany ekspertov sel'hozotrasli po importozaměneniũ i sbudutsã, no tol'ko esli rynek tepličnyh ovõsej budet horošo udobren bankovskimi investiciãmi i gosudarstvennoj podderžkoj. (Rossijskaya gazeta, September 4, 2015)*

Perhaps the ambitious plans of experts in the agricultural sector for import substitution will come true, but only if the greenhouse vegetables market is well-fertilized with bank investments and government support.

*[S]trane nužno importozaměnenie, no ono vozmožno tol'ko pri nizkoj inflácii. (Sovetskaya Rossiya, November 20, 2014)*

[T]he country needs import substitution, but it is possible only with low inflation.

To summarize, the large-scale media data provided by Integrum revealed three major findings. First, a large amount of data distinctly reflect the extent to which awareness of the political agenda set by Russian leaders is spreading among people. The concepts of “modernization” and “import substitution” aroused interest, having been introduced by leaders and reproduced in the media. Second, a more detailed analysis revealed recurring attitudes toward the concepts: there were frequent occurrences of “single-explanation” reasoning concerning the possibilities of modernization and import substitution, which appears to be a recurrent argument in Russian media discourse. Third, a qualitative analysis made it possible to identify the reasons that were used in media discourse to prevent changes in Russia. A single reason was usually provided to explain the failure, be it economic, educational, or political.

## 17.5 CONCLUSION

Texts are the principle sources of analysis in various types of research. Large textual corpora are an excellent source for investigating diverse concepts and their reflection in the language and attitudes in a society. These types of studies need both statistical data and in-depth analysis, which the described resources have to offer. If a researcher is aware of how to use the available resources and conducts an investigation within the limits that the data impose, then the results are reliable and inspiring.

We have presented various textual resources that are available for Russian studies: the web as a corpus, electronic libraries, and linguistics corpora. Some of these are specifically designed for linguistic research, but the majority may be effectively utilized in wider text-based studies. We emphasized the two most significant resources in particular: the Russian National Corpus and the Integrum database. The case studies we presented utilized a basic corpus-informed analysis to illustrate the usefulness of both resources in the study of societal changes as they are reflected in the language.

## NOTES

1. The projects embrace the cause of promoting copyleft ideas, the free distribution of copies (<https://en.wikipedia.org/wiki/Copyleft>). Although many of the publications on these sites are no longer under copyright, there have been many accusations of copyright infringement.
2. This section is adapted from our previous review by Kopotev et al. (2018). Readers who are interested in the specific linguistic details are advised to consult that publication.
3. This section is based on Bonch-Osmolovskaya (2018), where more details are provided.
4. This section is based to some extent on Laine and Mustajoki (2017), where more details are provided.

## REFERENCES

- Belikov, Vladimir, Alexander Piperski, Vladimir Selegey, and Serge Sharoff. 2013. Big and Diverse Is Beautiful: A Large Corpus of Russian to Study Linguistic Variation. In *Proceedings of the 8th Web as Corpus Workshop (WAC-8)/International Conference on Corpus Linguistics*, Lancaster.
- Benko, Vladimir. 2014. Aranea: Yet Another Family of (Comparable) Web Corpora. In *International Conference on Text, Speech, and Dialogue*, 247–256. Cham: Springer.
- Boguslavsky, Igor, Svetlana Grigorieva, Nikolai Grigoriev, Leonid Kreidlin, and Nadezhda Frid. 2000. Dependency Treebank for Russian: Concept, Tools, Types of Information. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, 987–991. Saarbrücken.
- Bonch-Osmolovskaya, Anastasia. 2018. Imena vremeni: èpitety desàtiletij v Nacional'nom korpuse russkogo àzyka kak proekcià kul'turnoj pamàti [Names of Time: Epithets of Decades in the Russian National Corpus as a Projection of Cultural Memory]. *Shagi/Steps* 4 (3): 115–146.
- Bozdag, Engin. 2013. Bias in Algorithmic Filtering and Personalization. *Ethics and Information Technology* 15 (3): 209–227. <https://doi.org/10.1007/s10676-013-9321-6>.
- Dobrushina, Nina, ed. 2007. *Nacional'nyj korpus russkogo àzyka i problemy gumanitarnogo obrazovanià* [The Russian National Corpus and Issues in Humanitarian Education]. Moscow: Higher School of Economics.
- Erjavec, Tomaž, Ivan Deržanski, Dagmar Divjak, Anna Feldman, Mikhail Kopotev, Natalia Kotsyba, Cvetana Krstev, et al. 2010. *MULTEXT-East Non-commercial Lexicons 4.0*. Slovenian Language Resource Repository CLARIN.SI. Accessed June 1, 2019. <http://hdl.handle.net/11356/1042>.
- Flaxman, Seth, Sharad Goel, and Justin M. Rao. 2016. Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly* 80: 298–320. Accessed October 9, 2017. <https://doi.org/10.1093/poq/nfw006>.
- Götzelmann, Michael, Kirill Postoutenko, Olga Sabelfeld, and Willibald Steinmetz. 2019. The Historical Semantics of Temporal Comparisons through the Lens of Digital Humanities: Promises and Pitfalls. In print. Accessed December 15, 2019. [https://www.academia.edu/41122839/The\\_Historical\\_Semantics\\_of\\_Temporal\\_Comparisons\\_through\\_the\\_Lens\\_of\\_Digital\\_Humanities\\_Promises\\_and\\_Pitfalls](https://www.academia.edu/41122839/The_Historical_Semantics_of_Temporal_Comparisons_through_the_Lens_of_Digital_Humanities_Promises_and_Pitfalls).

- Grishina, E.A., K.M. Korchagin, V.A. Plungyan, and D.V. Sichinava. 2009. *Poëtičeskij korpus v ramkah NKRA: obšaa struktura i perspektivy ispol'zovaniâ* [The Poetic Corpus in RNC: Its Structure and Prospects of Use]. In *Nacional'nyj korpus russkogo âzyka: 2006–2008* [Russian National corpus: 2006—2008], ed. V.A. Plungyan, 71–113. Saint-Petersburg: Nestor-istoriâ.
- Jakubiček M. et al. 2013. The TenTen Corpus Family. In *7th International Corpus Linguistics Conference CL*. 125–127.
- Kilgarriff, Adam. 2001. Web as Corpus. *Proceedings of the Corpus Linguistics Conference (CL 2001)*. University Centre for Computer Research on Language Technical Paper Vol. 13, Special Issue, Lancaster University, 342–344. <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/kilgarri.pdf>.
- Kopotev, Mikhail, Olga Lyashevskaya, and Arto Mustajoki. 2018. Russian Challenges for Quantitative Research. In *Quantitative Approaches to the Russian Language*, ed. Mikhail Kopotev, Olga Lyashevskaya, and Arto Mustajoki, 3–29. Abingdon: Routledge.
- Koselek, R. 2004. *Futures Past: On the Semantics of Historical Time. Series: Studies in Contemporary German Social Thought*. New York: Columbia University Press.
- Laine, Veera, and Arto Mustajoki. 2017. Preconditions for Russian Modernisation: A Media Analysis. In *Philosophical and Cultural Interpretations of Russian Modernisation*, ed. Katja Lehtisaari and Arto Mustajoki, 175–190. Abingdon: Routledge.
- Lotman, Yu. 2009. *Culture and Explosion (Semiotics, Communication and Cognition)*. Translated by Wilma Clark and edited by Marina Grishakova. De Gruyter Mouton.
- Lyashevskaya, O. 2016. *Korpusnye instrumenty v grammatičeskikh issledovaniâh russkogo âzyka* [Corpus Tools in Grammatical Studies of the Russian Language]. Moscow: LRC Publishing House.
- Lyashevskaya, O., and E. Kashkin. 2015. FrameBank: A Database of Russian Lexical Constructions. In *International Conference on Analysis of Images, Social Networks and Texts*, 350–360. Cham: Springer.
- McEnery, Tony, and Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Mikhailov, Mikhail, and Robert Cooper. 2016. *Corpus Linguistics for Translation and Contrastive Studies: A Guide for Research*. Routledge.
- Mitrenina, Olga. 2014. The Corpora of Old and Middle Russian Texts as an Advanced Tool for Exploring an Extinguished Language. *Scrinium* 10 (1): 455–461.
- Mustajoki, Arto. 2006. The Integrum Database as a Powerful Tool in Research on Contemporary Russian. In *Integrum: točnye metody i gumanitarnye nauki*, ed. Galina Nikiporets-Takigava, 50–76. Moscow: Letnij sad.
- Mustajoki, A., and O. Pussinen. 2006. Počemu narodu mnogo, ili Novye nablūdeniâ nad upotrebleniem vtorogo roditel'nogo padeža v sovremennom russkom âzyke [Why *narodu mnogo*: New Observations on the Use of the Second Genitive in Russian]. In *Integrum: točnye metody i gumanitarnye nauki*, ed. G. Nikiporets-Takigava, 50–75. Moscow: Letnij sad.
- . 2008. Ob èkspansii glagol'noj pristavki PO- v sovremennom russkom âzyke [Expansion of the Prefix PO in the Contemporary Russian]. In *Instrumentarij rusitiki: korpusnye podbody* (= Slavica Helsingiensia 34), 247–275. Helsinki.
- Nivre, Joakim, Mitchell Abrams, Željko Agić. et al. 2018. *Universal Dependencies 2.3*, LINDAT/CLARIN Digital Library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <http://hdl.handle.net/11234/1-2895>.



- Pichhadze, A.A. 2005. *Korpus drevnerusskikh perevodov XI-XII vv. i izučenie perevodnoj knižnosti Drevnej Rusi* [The Corpus of Old-Russian Translations from 11–12 Centuries and Study of Translated Literature of Ancient Russia]. In *Nacional'nyj korpus russkogo ázyka: 2003–2005* [Russian National corpus: 2003–2005], ed. V.A. Plungian, 251–262. Moscow: Indrik.
- Plungian, V.A. 2006. 'Integrum' i Nacional'nyj korpus russkogo ázyka v lingvističeskikh issledovaniáh [Integrum and the Russian National Corpus in linguistic research]. In *Integrum: točnye metody i gumanitarnye nauki*, ed. G. Nikiporets-Takigava, 76–84. Moscow: Letnij sad.
- , ed. 2009. *Nacional'nyj korpus russkogo ázyka: 2006–2008* [Russian National Corpus: 2006–2008], 71–113. Saint-Petersburg: Nestor-istoriá.
- Plungian, V.A., and L. Shestakova. eds. 2014. *Korpusnyj analiz russkogo stiha* [Corpus Analysis of Russian Verse], 2. Moscow: Azbukovnik.
- Prokhorov, Y.A., and I.A. Sternin. 2006. *Russkie: kommunikativnoe povedenie* [Russian Communication Strategies]. Moscow: Flinta.
- Sharoff, Serge, and Joakim Nivre. 2011. The Proper Place of Men and Machines in Language Technology: Processing Russian Without any Linguistic Knowledge. In *Proceedings of Dialogue 2011, Russian Conference on Computational Linguistics*.
- Shavrina, T., and O. Shapovalova. 2017. To the Methodology of Corpus Construction for Machine Learning: Taiga Syntax Tree Corpus and Parser. In *Proceedings of "CORPORA-2017" International Conference*, Saint-Petersburg, 78–84.
- Travin, Dmitry, Vladimir Gel'man, and Otar Marganiya. 2020. *The Russian Path: Ideas, Interests, Institutions*. Stuttgart: ibidem Press.
- Usage. 2019. *Usage of Content Languages for Websites*. Accessed November 28, 2019. [https://w3techs.com/technologies/overview/content\\_language/all](https://w3techs.com/technologies/overview/content_language/all).
- Zabotkina, V.I., ed. 2015. *Metody kognitivnogo analiza semantiki slova. Komp'úterno-korpusnyj podhod* [Methods in Cognitive Analysis of Word Semantics: A Computational and Corpus-based Approach]. Moscow: Yazyki Slavyanskoi Kultury.
- Zerubavel, E. 2003. *Time Maps. Collective Memory and the Social Shape of the Past*. Chicago: The University of Chicago Press.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# RuThes Thesaurus for Natural Language Processing

*Natalia Loukachevitch and Boris Dobrov*

## 18.1 INTRODUCTION

In natural language processing (NLP) and information retrieval (IR), it is often useful to utilize various types of knowledge, including lexical knowledge about relations between words, their senses, domain-specific knowledge, and commonsense knowledge. The conventional way to represent this knowledge within NLP systems are the so-called thesauri (= thesauruses). In NLP and IR domains, a thesaurus is a language or terminological resource describing relations between lexical or terminological units in a formalized form (in form of links), which makes it possible to use such descriptions in computer text processing.

There exist two well-known paradigms of thesauri used in computer information systems. The first paradigm is information retrieval thesauri, designated for improving document search in information retrieval systems. The role of such thesauri in information retrieval was most significant during the 1960–1980s of the twentieth century. Currently, global search engines do not use manually created thesauri. Nevertheless, the importance of such resources continues to be quite high, because such thesauri are used in information services of large international organizations as a source of recommended keywords for document indexing and search. However, these thesauri are not intended for automatic procedures of indexing and information search (ISO-25964 2011; NISO 2005).

---

N. Loukachevitch (✉) • B. Dobrov  
Lomonosov Moscow State University, Moscow, Russia

Another paradigm of thesaurus-like resources is implemented in Princeton WordNet, created for the English language (Fellbaum 1998; Miller 1998). Since its appearance, WordNet has attracted a lot of attention of researchers and other specialists in natural language processing and information retrieval. WordNet-like thesauri (wordnets) have been initiated for many languages in the world (Vossen 1998; Bond and Foster 2013; Maziarz et al. 2016). In contrast to information retrieval thesauri, which are created for specific domains, wordnets usually represent the lexical system of a specific language in the form of sets of synonyms and relations between them.

As a detailed formalized description of the language lexical system, WordNet is used in numerous applications as a tool for automatic text processing, as a basis for generating new computational resources (e.g., ImageNet [Mishkin et al. 2017] or SentiWordNet [Baccianella et al. 2010]). But WordNet's structure is not convenient for describing the conceptual system of a broad domain because of WordNet's orientation to representing the lexical system of the language including parts of speech, lexical relations (synonyms, antonyms, derivation, etc.), and language registers (Loukachevitch and Dobrov 2014).

In this chapter, we describe the Russian thesaurus RuThes, which has been created as a tool for automatic document processing of contemporary news texts, newspaper articles, and legal texts to enable their search, categorization, clustering, and so on. In its structure, RuThes combines approaches for language and knowledge representation that are accepted in information retrieval thesauri and WordNet-like resources. The development of RuThes began more than 20 years ago. The thesaurus continues to be updated with novel concepts, words, senses, and multiword expressions, which represent the current state of the Russian language used in contemporary texts. RuThes stores knowledge about current social and political life in Russia, which can be described using the thesaurus' relations. We compare the RuThes structure with other thesaurus paradigms and provide several examples of recently introduced concepts.

The chapter is structured as follows. In Sect. 18.2 we describe the main methodologies for creating large thesauri for natural language processing and information retrieval. In Sect. 18.3, we discuss the approach to knowledge representation in the RuThes thesaurus. Section 18.4 is devoted to the description of current social and political concepts in RuThes.

## 18.2 THESAURI IN NLP AN IR

### 18.2.1 *WordNet Thesaurus and Wordnets*

The structure of Princeton University's WordNet (and other wordnets) is based on sets of synonyms—synsets. Most synsets are provided with a “gloss” explaining their meaning. If a word has several meanings, it is included into several synsets. Synset is considered by the authors as a representation of the lexicalized concept of the English language. The current WordNet (version 3.0) covers approximately 155,000 unique words and phrases, organized into

117,000 synsets. Each synset has relations with other synsets, such as hyponyms (more specific words), hyperonyms (more general words), meronyms (parts), holonyms (wholes), and others. The WordNet thesaurus includes the words of four parts of speech (nouns, adjectives, verbs, and adverbs) and is divided into four lexical nets according to these parts of speech. The synsets of each part of speech in WordNet have their own sets of relationships. Also, specific words in synsets can have their own lexical relations (antonyms, derivation). Princeton WordNet (Fellbaum 1998; Miller 1998) is freely available on the Internet (WordNet 2019), and on its basis thousands of experiments in the field of information retrieval and natural language processing were carried out (for more on linguistic resources, see Chaps. 29, 19 and 26).

Bond et al. (2016) noted that WordNet-like resources (wordnets) created for different languages, while preserving the basic structure of WordNet, can differ significantly from each other in terms of the inclusion of words and expressions in synsets, the use of semantic relations between synsets, and the interpretation of specific semantic relations. Also, in wordnets, approaches to the description of polysemy can vary considerably, which leads to a more fine or coarse system of representing the senses of ambiguous words. There may be different approaches to the inclusion of multiword expressions into wordnets.

Some features of the WordNet structure are not very convenient for describing the conceptual system of a specific domain. These features include sets of synonyms (synsets) as a basic unit of the thesaurus, the division into part-of-speech structures, lexical relations, and approaches to inclusion phrases. However, several attempts to create domain-specific wordnets (e.g., ArchiWordNet, Jur-WordNet) have been made (for a review, see Längen et al. 2008).

### 18.2.2 *Information Retrieval Thesauri*

Information retrieval thesauri are important instruments in information and library services; for years, they were used for representing the domain knowledge in information retrieval systems. International and national standards have been published in the 1980s and continue to be updated (ISO-25964 2011; NISO 2005; Dextre Clarke and Zeng 2012). There exist some very influential international thesauri such as EUROVOC—the thesaurus of the European Community (EUROVOC Thesaurus 1995), the UNBIS thesaurus of the United Nations (United Nations 1976), the Art and Architecture thesaurus (Art & Architecture Thesaurus Online 2018) and others.

Information retrieval thesauri are less known and utilized for NLP purposes because they are intended to be used only in manual or automated indexing by human indexers, according to the thesaurus standards (ISO-25964 2011; NISO 2005). However, the principles of describing broad and complex domains are important for comparison with the WordNet structure.

The main units of information retrieval thesauri are domain terms denoting domain concepts. Domain concepts can have several variants of text representation, which are considered as synonyms. Among synonyms, the most representative variant, called a descriptor or preferred term, is chosen. Other terms included in the thesaurus are called nonpreferred terms and used as auxiliary units helping to find preferred terms.

Every descriptor should be formulated unambiguously. If a clear and unambiguous descriptor cannot be formulated, the term that is taken as a descriptor is supplied with a relator (a short label) or comment. In standards, there are special guidelines for introducing multiword descriptors (NISO 2005). The set of the thesaurus descriptors should be sufficient to describe the topics of the absolute majority of the documents in the domain. To explain why such thesauri are not suited for use in automatic document processing, we would like to provide several examples from the EUROVOC thesaurus (EUROVOC 1995). EUROVOC is created for 23 languages of the European Union and therefore it does not include Russian, but this thesaurus is one of the most well-known resources and therefore its principles are important to consider.

To improve the domain representation for humans, the guidelines for the creation of information retrieval thesauri often recommend not to include certain kinds of terms in a thesaurus (infrequent terms, terms that are too specific, similar terms etc.; United Nations 2009). Relying on human indexers, traditional information retrieval thesauri try to limit the inclusion of ambiguous terms, which leads to problems in automatic document processing. In EUROVOC, for example, the single-word term *bank* is presented in only one sense; other senses are described in form of multiword terms (*sperm bank*, *data bank*, *blood bank*). Note, that in WordNet, the word *bank* has ten senses as a noun and eight senses as a verb. In the defense category, EUROVOC does not contain such terms as *soldier* or *military force*; only the descriptor *armed force* is presented.

The relations in information retrieval thesauri are quite different from WordNet-like lexical relations. Information retrieval thesauri have a small set of generalized relations, which are usually subdivided into two classes: hierarchical and associative. The most frequent type of hierarchical relations between preferred terms in information retrieval thesauri are the broader-narrow relations (BT and NT relations), comprising class-subclass, instance-class, and sometimes part-whole relationships. The associative relations convey various other types of domain-specific relations between concepts (related term (RT) relation). The standards and manuals on thesaurus development formulate principles for representing associative relations as the most significant ones (NISO 2005; Aitchinson and Gilchrist 1987).

The RT relations are considered to be symmetric, but looking at the existing thesauri, it is possible to see that this is not true in many cases. For example, in EUROVOC the *air transport* descriptor has RT relations with such descriptors as *air law*, *air traffic control*, and *aviation fuel*, which are much narrower than the *air transport* descriptor. This simple system of relations has been criticized

in many works (Tudhope et al. 2001) but it has an important advantage: it can be applied to any domain without additional efforts to develop the detailed set of domain-specific relations, which always is a very complex task.

In Russia, the most known information retrieval thesauri are developed in the Institute of Scientific Information of Russian Academy of Sciences (INION RAN). This institution publishes separate issues of thesauri on economics, sociology, linguistics, and so on, created according to the guidelines of international and national standards on thesaurus construction. These thesauri also cannot be used for automatic processing of document and news flows (Mdivani 2013).

## 18.3 RuThES STRUCTURE, UNITS, AND RELATIONS

### 18.3.1 *RuThes General Structure*

In the construction of RuThes, both popular paradigms for computer thesauri were used: concept-based units, a small set of relation types, and rules for including multiword expressions as in information retrieval thesauri; language-motivated units, detailed sets of synonyms, and description of ambiguous words as in wordnets. Also, some issues of ontology research—for example, concepts as main units, strictness of relation description, necessity for many-step inference—are accounted for (Guarino 1998, 2009).

RuThes is a hierarchical network of concepts. Each concept has a name, relations with other concepts, and a set of language expressions (words, phrases, terms) whose meanings correspond to the concept. The whole set of RuThes' concepts is subdivided into general lexicon and sociopolitical thesaurus. *General Lexicon* comprises general concepts and words that can be met in various specific domains such as *sozdanie* (creation), *udalit'* (remove), *uslovnye* (conditional). *Sociopolitical Thesaurus* contains thematically oriented lexemes and multiword expressions as well as domain-specific terms of the broad sociopolitical domain. The whole RuThes thesaurus includes more than 60,000 concepts and more than 200,000 Russian text entries (words and expressions). The published version of RuThes for use in noncommercial applications includes 110,000 text entries (RuThes 2019).

The *sociopolitical domain* is the domain of problems, relationships, and situations of the contemporary society (Loukachevitch and Dobrov 2015). Subdomains of the sociopolitical domain are themselves large domains such as economics, law, or international relations, each with its own terminology. However, the specific feature of the sociopolitical domain (and its subdomains) is that most domain terms are known to nonprofessionals. Here, in the sociopolitical domain, the general language and domain terminologies adjoin and mix with each other. At present, the RuThes sociopolitical thesaurus includes terminology from such domains as politics, elections, sociology, demography, social security, civil and criminal law, the court system, banking, security, economics (including macroeconomics, industry, agriculture, and transport), ecology, accidents, sports, culture, and others.

### 18.3.2 *RuThes Units*

The RuThes thesaurus is a hierarchy of concepts viewed as units of thought. A concept is associated with the set of language expressions that refer to it in texts. This approach is similar to approaches of traditional information retrieval construction (NISO 2005). In most cases, concepts should have denotational distinctions from related concepts. Such distinctions can be expressed in a specific set of relationships or associated language expressions: *text entries*.

Words and phrases whose meanings refer to the same concepts represented in the thesaurus are called ontological synonyms. Ontological synonyms can comprise sense-related words belonging to different parts of speech (i.e., *privatizaciâ* [privatization] vs. *privatizirovat'* [to privatize]); in contrast to traditional terminological resources and information retrieval thesauri that contain mainly nouns or noun phrases. A thesaurus for automatic document processing should contain various types of language units. Also, language expressions relating to different linguistic styles, technical terms, and lexical units can be presented as ontological synonyms related to the same concept. For example, the concept *Oil industry* has the following text entries: *neftânaâ promyšlennost'* (oil industry)—neutral, *neftânka*—slang, *nefteprom*—abbreviation. Compositional multiword expressions may be included into synonymic sets as well. Each concept should have a clear, univocal, and concise name. Such names often help to express and delimit the denotational scope of the concept. In addition, the concepts' names can be used in the analysis of the results of automatic document analysis, for example in visualization of trends or as cluster names.

Ontological synonyms, variants of lexical units, and technical terms (Nazarenko and Zargayouna 2009) are collected specially. After a concept has been introduced, an expert searches for all possible synonyms or orthographic variants, single words, and phrases that can be associated with it. These synonymic sets can also include multiple variants of the references to the same concept. For example, the concept *Obrana prirody* (Nature protection) is associated with almost 50 different text entries in Russian, for example *zašîta prirody* (defense of nature), *sohranenie prirody* (maintenance of nature), *zašîsat' prirodu* (to protect nature), *sohranât' prirodu* (to maintain nature), and others. These variants are useful to describe in the thesaurus because they directly refer to their concept. Besides, multiword term variants often contain ambiguous words within themselves. Thus, the inclusion of such term variants decreases the overall lexical ambiguity and facilitates disambiguation. All variants are collected during the analysis of real texts, usually news articles, legislative acts, or domain-specific documents.

In fact, the introduction of such a concept as *Nature protection* corresponds more to information retrieval thesauri than wordnets, because one of the important principles of WordNet-like resources is to include single words and lexicalized phrases into synsets (Bentivogli and Pianta 2004; Maziarz and Piasecki 2018). The phrase *nature protection* seems compositional, but the



concept *Nature protection* is significant for the contemporary life of the society and it has relations with other important concepts of the sociopolitical domain.

As can be seen, one of the difficult issues in developing application-oriented resources, such as wordnets or information retrieval thesauri is the inclusion of units (synsets or descriptors) based on the senses of multiword expressions, for example noun compounds (Bentivogli and Pianta 2004). Manuals and standards for information retrieval thesaurus development provide detailed principles for multiword term selection (NISO 2005; Aitchinson and Gilchrist 1987). In RuThes, the introduction of concepts based on multiword expressions is not restricted but encouraged if this concept adds some new information to the knowledge described in the thesaurus (Loukachevitch and Lashevich 2016).

### 18.3.3 *RuThes Relations*

Conceptual relations in the thesaurus may be utilized for several purposes, including query expansion in information retrieval, clustering related concepts mentioned in a text as a basis for better recognition of the main theme and subthemes in the document, and disambiguation of ambiguous terms and lexical units. Working with such a broad scope of concepts, we utilize a set of relations that can be applied to concepts in various domains, in contrast to domain-dependent relations.

RuThes has a small set of conceptual relations consisting of four main relations that describe the most important links of a concept. In fact, the current set of relations in the thesaurus is a more ontologically motivated variant of classic inter-descriptor relationships in information retrieval thesauri, which usually include hierarchical relations, such as broader term (BT) and narrower term (NT), and associative relations—related term (RT).

The first relation of RuThes is *the class-subclass relation* as it is treated in ontological approaches (Guarino 1998; Gangemi et al. 2003). To establish such relations, we apply tests similar to those used in ontology development. The tests are directed toward avoiding incorrect use of class-subclass relations and not mixing them up with other types of relations (such as type-role relation, class-instance relation), because errors in relation types degrade logical inference (Gangemi et al. 2003). The class-subclass relationship is considered as a transitive relation with the inheritance property.

The second relationship is *part-whole relation*, which is established using specific ontological restrictions (Gangemi et al. 2003). Our decision on part-whole relations is based on the following principles:

- Broad treatment of part-whole relations from the semantic point of view,
- Restriction of ontological subtypes of part-whole relations,
- Postulating the transitivity of part-whole relations.

Part-whole relations in RuThes comprise such relationships as parts of physical objects, territorial and geographical parts, process parts, and others (see examples in Table 18.1). Also, some other relationships are presented as part-whole relations in RuThes: an attribute and its bearer, a role or a participant in the situation (Winston et al. 1987, 27–28), entities and situations in the encompassing sphere of activity (Table 18.1).

In such a broad scope, part-whole relations described in RuThes are close to the so-called *internal relations* (parthood, constitution, quality inherence, and participation) as described by Guarino (2009). At the same time, part-whole relations in RuThes have a very important restriction (correlating with the information retrieval thesauri guidelines about the necessity to describe only inherent properties as hierarchical relations [NISO 2005]): a concept-part should be related to its whole during the normal existence of its instances: the so-called *ontological dependence*.

To analyze the ontological dependence between entities  $X$  and  $\mathcal{Y}$ , it is necessary to determine whether entity  $X$  can exist by itself or whether its existence depends on the existence of  $\mathcal{Y}$ . We describe the following types of dependent parts in RuThes:

**Table 18.1** Types and examples of part-whole relations in RuThes

<i>Type of relationship</i>	<i>Part</i>	<i>Whole</i>
Parts of physical objects	<i>starter dvigatelá</i> (motor starter) <i>kost'</i> (bone)	<i>dvigatel' vnutrennego sgoraniá</i> (internal combustion engine), <i>skelet</i> (skeleton)
Territorial and geographical parts	<i>oasis</i> (oasis) <i>izbiratel'nyj učastok</i> (electoral precinct)	<i>pustyná</i> (desert), <i>izbiratel'nyj okrug</i> (electoral district),
Process parts	<i>bankovskij sejf</i> (bank safe)— <i>izbiratel'naá tehnologiá</i> (electoral technology)	<i>bankovskoe braniliše</i> (bank vault) <i>predvybornaá kampaniá</i> (pre-election campaign)
Text and musical parts	<i>vvedenie</i> (text introduction) <i>muzykal'nyj interval</i> (musical interval)	<i>tekst</i> (text), <i>muzykal'naá kompoziciá</i> (musical composition)
Members	<i>člen političeskoj partii</i> (political party member) <i>deputat Gosudarstvennoj Dumy</i> (State Duma Deputy)	<i>političeskaá partiá</i> (political party), <i>Gosudarstvennaá Duma</i> (State Duma, the lower house of the Russian Parliament)
Substance as a part	<i>židkost' v organizme</i> (body fluids)	<i>telo</i> (body of living organism)
An attribute and its bearer	<i>skorost'</i> (speed) <i>glasnost' vyborov</i> (election publicity)	<i>dviženie</i> (movement), <i>vybory</i> (election)
Roles and participants in a situation	<i>investor</i> (investor) <i>igrok</i> (player)	<i>investirovanie</i> (investing), <i>igra</i> (game)
Entities and situations in the encompassing sphere of activity	<i>zavod</i> (industrial plant) <i>sportsmen</i> (sportsman)	<i>promyšlennost'</i> (industry), <i>sport</i> (sport)

- Inseparable part, which is a part that cannot exist without its whole, such as *oazis* (oasis)—*pustyná* (desert);
- Mandatory whole, when a part requires the existence of at least one entity described as a whole, such as *bankovskij sejf* (bank safe) and *bankovskoe hraniliše* (bank vault) (Guizzardi 2011).

Thus, we put existential constraints on the part-whole relations in RuThes. These constraints do not change the transitivity of part-whole relations if it was postulated. The inference mechanism can thereby utilize the transitivity of part-whole relations and rely on the chain of part-whole relations (Guizzardi 2011; Loukachevitch and Dobrov 2015).

The final types of relationships are *nonsymmetrical and symmetrical associations*, which are subdivided from the symmetric related term (RT) relation of conventional information retrieval thesauri. The nonsymmetrical associations are established on the basis of the ontological dependence of concepts. Symmetrical associations are described in the very restricted number of cases.

Associative relationships (RT relations) are quite common in information retrieval thesauri; they are established to provide additional links between descriptors for use in the indexing or retrieval of documents (NISO 2005). Such relations in information retrieval thesauri are always considered as symmetrical; however, many associative relations found in published thesauri demonstrate the evident absence of symmetry, for example *illness—disease prevention*, *illness—sick leave* (EUROVOC), et cetera. The first term in each pair is much more general than the other one.

Considering the problems involved in formalizing traditional information retrieval thesauri to adapt them to the contemporary level of ontological research, some authors propose changing the thesaurus's traditional system of relations to a formalized set of predicates and to provide axioms for such a set (Soergel et al. 2004). However, in creating such multidomain resources as RuThes, it is very difficult to find the universal set of semantic relations and apply them consistently. Therefore, we substituted the traditional thesaurus relation of symmetric association with another quite generalized relation, which can be applied in many various domains. We usually refer to this relation as a nonsymmetrical association,  $asc_1-asc_2$ . The definition of this relation is again based on a variant of ontological dependence, the so-called *external dependence* in ontological terms (Gangemi et al. 2003; Guarino 2009). This relation is established between two concepts  $c_1$  and  $c_2$  when two requirements are fulfilled:

- Neither class-subclass nor part-whole relations can be established between  $c_1$  and  $c_2$  in the thesaurus.
- The following assertion is true: “concept  $c_2$  exists” means “concept  $c_1$  exists” (necessarily existent entities are excluded from consideration).

**Table 18.2** Examples of conceptual dependence relations denoted as nonsymmetrical associations in RuThes

<i>Type of relationships</i>	<i>Main concept</i>	<i>Dependent concept</i>
Instrument—professional that uses this instrument	<i>skripka</i> (violin)	<i>skripač</i> (violinist)
Entity—branch of science that studies such entities	<i>životnoe</i> (animal)	<i>zoologíá</i> (zoology)
Entity and related entity	<i>serdce</i> (heart)	<i>kardiologíá</i> (cardiology)
Entity and actions that applied to these entities	<i>bagazh</i> (luggage)	<i>bagazháá karusel'</i> (luggage carousel)
Entity and its specific problems	<i>krov'</i> (blood)	<i>donorstvo krvi</i> (blood donation)
Entity and opposing entity or action	<i>eda</i> (food)	<i>žarka</i> (frying)
	<i>les</i> (forest)	<i>lesnoj požar</i> (forest fire)
	<i>serdce</i> (heart)	<i>bolezn' serdca</i> (heart disease)
	<i>virus</i> (virus)	<i>antivirus</i> (antivirus)

These two conditions mean that the concept  $c_2$  (dependent concept) externally depends on  $c_1$ :  $asc_1(c_2, c_1) = asc_2(c_1, c_2)$ . Table 18.2 presents some examples of conceptual relationships, where conceptual dependence can be seen.

Relations of ontological dependence are applicable to various domains; therefore, they are usually used in top-level ontologies (Gangemi et al. 2003). An additional advantage of using these relations in thesauri for automatic document processing is their usefulness for describing links between a concept based on the sense of a compositional multiword expression and concepts corresponding to the components of this multiword expression. As a result, a multiword-based concept (e.g., *Automobile racing*) is described as the dependent concept and its component concept (*Automobile*) as the main concept. This allows us to introduce concepts based on various types of multiword expressions and to establish their necessary relations.

## 18.4 DESCRIPTION OF SOCIAL AND POLITICAL CONCEPTS IN RUTHEs

The specific part of RuThes called Sociopolitical thesaurus provides detailed coverage of thematic lexical units and terms in the broad sociopolitical domain of contemporary written Russian (mainly news articles, laws, and official documents). The thesaurus was utilized in document-processing applications within information retrieval and information analytical systems (Loukachevitch and Dobrov 2015). Every project gave the opportunity to improve the descriptions of lexical senses, reveal useful expressions, and add domain terms of new subdomains of the sociopolitical domain, which, in turn, improved the description of related lexical senses.

Let us consider several examples of recently introduced concepts related to popular topics discussed in the Russian and international press and their descriptions in RuThes. Figure 18.1 represents the description of concepts

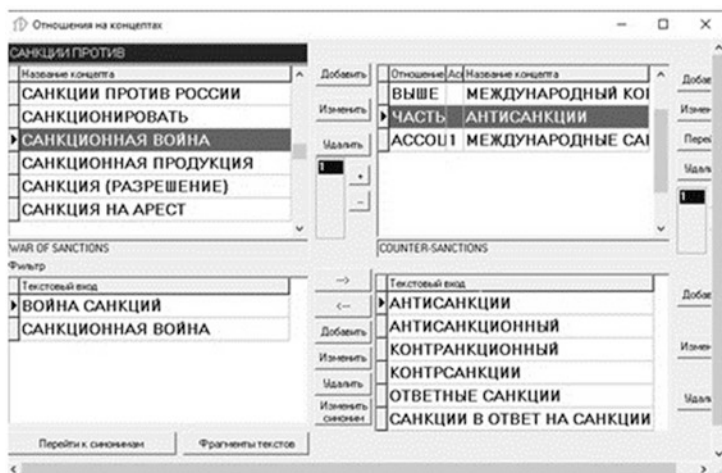


Fig. 18.1 Representation of the current international sanctions situation in the thesaurus form

related to sanctions: *Sankcii protiv Rossii* (Sanctions against Russia), *Sankcionnaâ vojna* (War of sanctions), *Sankcionnaâ produkciâ* (Products under sanctions). The upper-left form enumerates a list of concepts in alphabetical order.

The left-lower form shows Russian text entries for the concept *War of sanctions* such as *vojna sankcij* (war of sanctions) and *sankcionnaâ vojna* (sanctions war). The right-upper form presents the relations of the highlighted concept. Figure 18.1 shows the relation of the *War of sanctions* concept with such concepts as *Meždunarodnyj konflikt* (International conflict), *Antisankcii* (Counter-sanctions), and *Meždunarodnye sankcii* (International sanctions). In particular, the *War of sanctions* concept is described as dependent from the concept *International sanctions*, because it could not appear without this concept. The *Counter-sanctions* concept is described as a part of *War of sanctions*. The lower-right form shows Russian text entries for the related concept *Antisankcii* (Counter-sanctions). They include: nouns (*antisankcii*, *kontrsankcii*), noun groups (*otvetnye sankcii* [sanctions as an answer]), and adjectives (*antisankcionnyj*, *kontrsankcionnyj*).

After the pension reform in Russia was announced in 2018, new concepts *Predpensioner* (Person before retirement age) and *Predpensionnyj vozrast* (Before retirement age) were introduced in the thesaurus. These concepts appeared in Russian law to provide social security to some categories of the population in relation to raising the retirement age. The *Before retirement age* concept is described as a part (property) of *Person before retirement age* according the thesaurus guidelines. The concept *Person before retirement age* depends on the concepts *Pensioner* and *Pension system* because it requires their existence.

An innovation of the Russian transport law introduced yellow boxes on roads, a specific kind of road marks (box marking). In Russian, the concept is called *Vafel'naâ razmetka* (literally, *waffle marking*). The concept's set of Russian text entries includes word *vafel'nica*, which previously meant only “waffle iron,” a kitchen appliance for baking waffles. Therefore, the new sense of the word *vafel'nica* and new multiword expression *vafel'naâ razmetka* have been added into RuThes.

In recent years, cryptocurrencies were actively discussed. The corresponding concepts: *kriptovalûta* (cryptocurrency), *èlektronnyye den'gi* (electronic money), *Bitcoin*, *kriptomat* (Cryptocurrency ATM machine) have been introduced into the thesaurus.

Thus, RuThes provides detailed coverage of thematic lexical units and terms in the broad sociopolitical domain of contemporary written Russian (mainly news articles, laws, and official documents). The thesaurus can be used as a conceptual indexing tool in information analytical systems. RuThes can also be a useful instrument for developing knowledge-based categorization systems in conditions when a training collection for machine learning methods is absent and cannot be easily created. It is possible because the thesaurus contains thousands of words and expressions stored in a hierarchical structure, which can be used in the description of categories for automatic text categorization (Loukachevitch and Dobrov 2015).

## 18.5 RUTHEAS AS A SOURCE FOR RUSSIAN WORDNET

Despite the fact that RuThes is currently published for noncommercial use, people would like to have a large Russian wordnet. Therefore, a transforming procedure from the published version of RuThes (RuThes-lite) to the largest Russian WordNet (RuWordNet 2019) has been initiated. One of the most distinctive features of WordNet-like resources is their division into synset nets according to parts of speech. Therefore, all text entries of RuThes-lite were subdivided into three parts of speech: nouns (single nouns, noun groups, and preposition groups), verbs (single verbs and verb groups), adjectives (single adjectives and adjective groups). We have obtained 29,297 noun synsets, 12,865 adjective synsets, and 7636 verb synsets. The divided synsets were linked to each other with the relation of part-of-speech synonymy.

The hyponym-hypernym lexical relations (hyponymy shows the relationship between a generic term [hypernym] and a specific instance of it [hyponym]) were established between synsets of the same part of speech. These relations include direct hyponym-hypernym relations from RuThes-lite. In addition, the transitivity property of hyponym-hypernym relations was employed in cases when a specific synset did not contain a specific part of speech, but its parent and child had text entries of this part of speech. In such cases, the

hypernymy-hyponymy relation was established between the child and the parent of this synset.

Other RuThes relations were modified. The part-whole relations from RuThes were semi-automatically transferred and corrected according to traditions of WordNet-like without the expanded set of part-whole relations. Some part-whole relations were transformed to domain relations, for example *zavod* synset (industrial plant) is related to the domain *promyšlennost'* (industry) via the domain relation. The ontological dependence relations of RuThes were manually transformed to appropriate semantic relations such as antonyms, cause, entailment, and some others. RuWordNet is publicly available (RuWordNet 2019).

## 18.6 CONCLUSION

In this chapter, we described the RuThes thesaurus that was created as a linguistic and terminological resource for automatic document processing in Russian. In the construction of RuThes, both popular paradigms for computer thesauri were used: concept-based units, a small set of relation types, and rules for including multiword expression as in information retrieval thesauri; language-motivated units, detailed sets of synonyms, and description of ambiguous words as in wordnets. A large part of RuThes is devoted to the description of terms and concepts related to the current sociopolitical life in Russia and in the world—the so-called Sociopolitical thesaurus.

We have supported the development of RuThes for many years by introducing new concepts, representing new senses, and recording multiword expressions. In this chapter, we have showed some examples of representing newly appeared concepts related to important internal and international events. We demonstrated how we used the thesaurus' relation system for describing these concepts. Hence, we consider RuThes as a kind of formalized encyclopedia of social and political life of the contemporary society.

## REFERENCES

- Aitchinson, Jean, and Alan Gilchrist. 1987. *Thesaurus Construction: A Practical Manual*. London: Aslib.
- Art & Architecture Thesaurus Online. 2018. The J. Paul Getty Trust. <http://www.getty.edu/research/tools/vocabularies/aat/>.
- Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: an Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of Language Resources and Evaluation Conference LREC-2010* 10: 2200–2204.
- Bentivogli, Luisa, and Emanuele Pianta. 2004. Extending Wordnet with Syntagmatic Information. In *Proceedings of Second Global WordNet Conference*, 47–53. Brno, Czech Republic.



- Bond, Francis, and Ryan Foster. 2013. Linking and Extending an Open Multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* 1: 1352–1362.
- Bond, Francis, Piek Vossen, John McCrae, and Christiane Fellbaum. 2016. CILI: The Collaborative Interlingual Index. *Proceedings of the 8th Global WordNet Conference 2016 (GWC2016)*, 27–30.
- Dextre Clarke, Stella, and Marcia Lei Zeng. 2012. From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards Towards Interoperability and Data Modelling. *Information Standards Quarterly (ISQ)* 24 (1): 20–26
- EUROVOC, *Thesaurus*. 1995. Vol. 1–3/European Communities. Luxembourg: Office for Official Publications of the European Communities.
- Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Gangemi, Aldo, Roberto Navigli, and Paula Velardi. 2003. The OntoWordNet Project: Extension and Axiomatisation of Conceptual Relations in Wordnet. In *Proceedings of International Conference on Ontologies, Databases and Applications of Semantics (ODBASE)*, 820–838.
- Guarino, Nicola. 1998. Some Ontological Principles for Designing Upper Level Lexical Resources. In *Proceedings of the First International Conference on Language Resources and Evaluation*, 527–534. Granada, Spain.
- . 2009. The Ontological Level: Revisiting 30 Years of Knowledge Representation. In *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*, Lecture Notes in Computer Science, 5600, 52–67. Berlin and Heidelberg: Springer-Verlag.
- Guizzardi, Giancarlo. 2011. Ontological Foundations for Conceptual Part-Wholes Relation: The Case of Collectives and Their Parts. In *Advanced Information Systems Engineering: Proceedings of the 23rd International CAiSE*, (London, UK, June 20–24) Lecture Notes in Computer Science, 6741, 138–153. Berlin and Heidelberg: Springer-Verlag.
- INION RAN. Thesauri and Subject Headings. <http://inion.ru/resources/bazy-dannykh-inion-ran/>.
- International Standards Organization (ISO). 2011. *ISO 25964 Information and Documentation—Thesauri and Interoperability with Other Vocabularies*.
- Loukachevitch, Natalia, and Boris Dobrov. 2014. RuThes Linguistic Ontology Vs. Russian Wordnets. In *Proceedings of the Seventh Global WordNet Conference (GWC 2014)*, 154–162. Tartu, Estonia.
- . 2015. The Sociopolitical Thesaurus as a Resource for Automatic Document Processing in Russian. *Terminology. Special Issue Terminology across Languages and Domains* 21 (2): 238–263.
- Loukachevitch, Natalia, and German Lashevich. 2016. Multiword Expressions in Russian Thesauri RuThes and RuWordNet. In *Artificial Intelligence and Natural Language Conference (AINL)*, IEEE, 1–6.
- Lüngen, Harald, Claudia Kunze, Angelika Storrer, and Lothar Lemnitzer. 2008. Towards an Integrated OWL Model for Domain-Specific and General Language Wordnets. In *Proceedings of the 4th Global Wordnet Conference (GWC-2008)*, 281–296.
- Maziarz, Marek, and Maciej Piasecki. 2018. Towards Mapping Thesauri onto plWordNet. *Proceedings of Global Wordnet Conference GWC-2018*, 45–53.
- Maziarz, Marek, Maciej Piasecki, Eva Rudnicka, Stan Szpakowicz, and Paweł Kędzia. 2016. Plwordnet 3.0—a Comprehensive Lexical-Semantic Resource. In *Proceedings*

- of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2259–2268.
- Mdivani, Robert. 2013. Creating a Multilingual Thesaurus for the Social Sciences. Linguistic and Intercultural Problems. *Scientific and Technical Information Processing* 40 (3): 137–141.
- Miller, George. 1998. Nouns in WordNet. In *WordNet—An Electronic Lexical Database*, ed. Christiane Fellbaum, 23–47. Cambridge, MA: The MIT Press.
- Mishkin, Dmytro, Nikolay Sergievskiy, and Jiri Matas. 2017. Systematic Evaluation of Convolution Neural Network Advances on the Imagenet. *Computer Vision and Image Understanding* 161: 11–19.
- Nazarenko, Adeline, and Haifa Zargayouna. 2009. Evaluating Term Extraction. In *Proceedings of International Conference Recent Advances in Natural Language Processing (RANLP'09)*, 299–304. Hissar, Bulgaria.
- NISO. 2005. *Guidelines for the Construction, Format and Management of Monolingual Thesauri*, ANSI/NISO Z39.19. Bethesda, MD: NISO Press.
- RuThes. 2019. Thesaurus for Natural Language Processing in Russian. [http://www.labinform.ru/pub/ruthes/index\\_eng.htm](http://www.labinform.ru/pub/ruthes/index_eng.htm).
- RuWordNet. 2019. WordNet-like Thesaurus for Russian. <http://ruwordnet.ru/en>.
- Soergel, Dagobert, Boris Lauser, Anita Liang, Frehivot Fisseha, Johannes Keizer, and Stephen Katz. 2004. Reengineering Thesauri for New Applications: The AGROVOC Example. *Journal of Digital Information Article* 4 (4). Accessed January 3, 2019. <https://journals.tdl.org/jodi/index.php/jodi/article/view/112/111>.
- Tudhope, Douglas, Harith Alani, and Christopher Jones. 2001. Augmenting Thesaurus Relationships: Possibilities for Retrieval. *Journal of Digital Information* 1 (8). Available at [https://eprints.soton.ac.uk/254484/1/Tudhope\\_JoDI.pdf](https://eprints.soton.ac.uk/254484/1/Tudhope_JoDI.pdf) (Accessed 4 July 2020).
- United Nations. 1976. *UNBIS Thesaurus*, English edition. New York: Dag Hammarskjöld Library of the United Nations.
- . 2009. Guidelines for Analysis of UN Documents and Publications. UNBISnet–United Nations Dag Hammarskjöld Library. [http://www.un.org/Depts/dhl/unbisref\\_manual/indexpolicy/guidelines.htm](http://www.un.org/Depts/dhl/unbisref_manual/indexpolicy/guidelines.htm).
- Vossen, Piek. 1998. Introduction to EuroWordNet. In *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, 1–17. Netherlands: Springer.
- Winston, Morton, Roger Chaffin, and Douglas Herrmann. 1987. A Taxonomy of Part-Whole Relations. *Cognitive Science* 11 (4): 417–444.
- WordNet. 2019. <https://wordnet.princeton.edu/>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Social Media-based Research of Interpersonal and Group Communication in Russia

*Olessia Koltsova, Alexander Porshnev,  
and Yadviga Sinyavskaya*

## 19.1 INTRODUCTION

Social media and, in particular, social networking sites (SNS) have become an important source of research data both in Russia and worldwide which, correspondingly, has given rise to new research methods and approaches. Social media data serve as sources of two major types of data: first, the data about the “offline” reality, such as migration, electoral outcomes or mental disorders, and second, the data about human behavior within social media, which includes online self-presentation, networking, media consumption or purchasing behavior. Russian social media, unusual both in terms of their market configuration and data access opportunities, create a slim, but an interesting stream of research.

In this chapter, we critically review the most exemplary works in Russian social media studies. Our goal is to discuss strengths and weaknesses of different research designs and methods that are seldom reported in the papers focused on the research results. As of now, Russian social media studies can be classified along two major lines. The first line differentiates between studies using Russian SNSs as a source of data about human behavior in general, and those that aim at studying specifically Russian society or Russian-speaking community with the data from different social media, including the Russian SNSs.

The second line of differentiation is disciplinary, and we can single out three disciplines that have contributed most to Russian social media research:

---

O. Koltsova • A. Porshnev • Y. Sinyavskaya (✉)  
Higher School of Economics (HSE University), Saint Petersburg, Russia

psychology and health studies, sociology and political science. The latter has focused specifically on Russia, and especially on the relations between social media and protests. Sociology has addressed a wide range of topics, such as virtual demography, education and ethnic relations whose results reflect specific Russian context, but can often be extrapolated beyond the Russian society. Psychological research has mostly contributed to fundamental psychology by studying the relation of social media to such universal psychological phenomena as depression or personality traits. Health studies may be situated between psychology and sociology.

This review chapter focuses on the three above-mentioned disciplines. We find that they use a wide range of data types and data-collection techniques: self-reported data (from surveys and experiments collected off-line or on-line) and SNS activity user data (incl. user texts and their metadata, such as timestamps and geolocation, the data from group accounts, data on links between accounts and various external statistics). Methods of data analysis vary from traditional discourse analysis and classical statistics to social network analysis (SNA), supervised and unsupervised machine learning, and various combinations of those.

We should also note that Russian social media are actively used in computational linguistics. Though this research community works with the Russian language, it does not focus either on the Russian society or a broader Russian-speaking community, tackling such problems as optimization of information retrieval, text clustering and summarization, named entity recognition or automatic translation. It thus forms a distinct stream of literature addressed in this book (for more, see Chaps. 26, 19, 29, 25, 23 and 24) that we leave out in this review. We also omit the large and important topic of “research ethics is social media studies” as it is a subject for a separate contribution.

The rest of the chapter is structured as follows. First, we briefly introduce the context of social media development in Russia and show how it has resulted in their unique position within the global SNS landscape. The next three sections are devoted to the disciplinary overviews of political science, sociology, health and psychology. We conclude with summarizing both research opportunities and limitations created by social media.

## 19.2 SOCIAL MEDIA IN RUSSIA

Russian Internet landscape is unique in terms of its “home-grown” character. According to [Forbes.ru](#), Russian companies compete with global players nearly in all spheres of Internet business, including search engines, mailing services and social media ([Forbes.ru 2019](#)). Unlike in China, where the closed Internet ecosystem owes most of its success to the policy of technical, political and economic isolation known as the Great Chinese Firewall, Russian information technologies (IT) industry has until recently developed without any protectionist barriers.

**Table 19.1** SNS use in Russia according to media research (October 2018, in mil)

SNS	<i>Average daily reach (desktop) (Mediascope 2018)</i>	<i>Active users<sup>a</sup> per month (Brand Analytics 2018)</i>	<i>Monthly messages<sup>b</sup> (Brand Analytics 2018)</i>
Vkontakte	16.82	36.4	1096
Youtube	13.84	1.9	15.9
Odnoklassniki	8.22	15.8	364
Facebook	3.23	2.3	122
Instagram	2.70	23.7	304
LiveJournal	1.24	46	4.6
Twitter	NA	0.8	59.6
My World ("MojMir")	NA	0.099	7.2

<sup>a</sup>Active user: user who wrote at least one public message per month

<sup>b</sup>Message: any publicly available post—status, wall post or comment, post or comment in online group etc. The analysis did not include private messages

As a result, social media “diets” of Russian and Russian-language users are substantially different from global trends (see Table 19.1). Social networking site VKontakte (VK), a Russian replica of Facebook, has by far higher reach and especially higher user activity than all other SNSs, followed by popular, but much less active Odnoklassniki. Facebook (FB) in Russia is a niche network that attracts higher-educated audiences oriented towards international integration, business and, to some extent, more oppositional political views (e.g. Enikolopov et al. 2018). VK, on the other hand, has become a universal tool for everyday communication, practical task-solving and small business. A typical VK user consumes news from a few large entertainment and/or political public pages run by media organizations and also belongs to a multitude of smaller groups that include everything from school classes and self-help communities to pages of local businesses that use VK to promote their services.

Consequently, political and market pressures experienced by VK are to some extent different from those faced by Facebook. Unlike Facebook, VK has not been accused of illegitimate influence on elections, since it is widely accepted that electoral outcomes in Russia depend on very different things (Gel'man 2014). Combined with relatively low importance of privacy among the Russian population (Kisilevich et al. 2012), this until recently has been creating incentives for VK to privilege data sharing over privacy protection. As a result, the amount and diversity of data available through VK application programming interface (API) is incomparably higher than in FB, and thousands of business and research actors use it on a daily basis. It is this unique data availability that has made possible such large-scale virtual demography projects as Webcensus (Zamyatina and Yashunsky 2018) (see further below). Surprisingly, such opportunities have attracted nearly no attention from international scholars, which is why most VK-based research has been done by Russian researchers.

Another important trend is the fragmentation of the Russian-speaking online environment. For a while, VK was an integrating medium for

Russian-speakers on the Post-Soviet space, but this capacity has been severely hindered by the ban of VK in Ukraine in 2017 and the overall deterioration of Russia's relations with the rest of the world. It is plausible to expect that in the near future comparative studies that include Russia might be increasingly dominated by research based on global SNSs which will decrease the value of VK as a data source.

### 19.3 POLITICAL SCIENCE

Political science research on social media is the largest among the three mentioned disciplines. It focuses on a number of subfields, such as the role of social media for political protest and civil activity (Enikolopov et al. 2018; Koltsova and Selivanova 2019), mapping political discourses and agendas formed both by lay and professional SNS users, including media professionals and politicians (Bodrunova et al. 2018; Goncharov and Nechay 2018; Bulovsky 2019; Kelly et al. 2012; Koltsova and Koltcov 2013) or topic-specific political discussions (Filer and Fredheim 2016), and the newly emerged topic of SNS-channeled propaganda (Barash and Kelly 2012; Kelly et al. 2012; Stukal et al. 2017; Badawy et al. 2018; Sanovich 2017).

Studies based on discourse, agenda and discussion mapping are by their nature, mostly descriptive; and when based on manual text analysis only, usually not scalable. However, manual approach can be enhanced with automated text analysis, albeit in such case often at cost of its depth. Goncharov and Nechay (2018) benefit from applying such a combination to a collection of about 45 thousand tweets related to the anti-corruption protests organized by a prominent Russian oppositionist Alexei Navalny in spring 2017. To evaluate the mobilization potential of Twitter, they apply keyword-determined time-constrained sampling and then use topic modeling to reveal content-based clusters and social network analysis (SNA) to find link-based communities. They convincingly demonstrate the dominance of an oppositional and a loyalist meta-cluster in both partitions. An important methodological note of the authors is that hashtags seldom occur in their data, which questions the frequently used hashtag-based sampling in Twitter research. At the same time, the authors do not specify the type of links used, neither they explain how the two partitions of their data are related. Most importantly, they do not find an answer for their main question about mobilization effect of Twitter, since cluster analysis is a method suitable for descriptive, not for inferential investigation.

This limitation is shared by most SNS research based on clustering techniques and some research based on manual coding. Thus, Koltsova and Koltcov (2013) illustrate the growth of political topics at the expense of other topics in the Russian-language LiveJournal (LJ) top blogs on the eve of the Russian parliamentary elections in 2011 which, nevertheless, does not lead to any hypothesis testing. Bodrunova et al. (2018) go further by formulating the hypotheses about prevalence of certain media roles among professional authors of tweets discussing politicized violent events in four different countries,



including Russia. Tweets are classified manually; in principle, automatically clustered tweets might have equally been tested for prevalence, but no statistical procedures for doing so have been introduced in this research. The authors, nevertheless, offer a multitude of interesting details that help understand the structure of the political discussion in the four countries in a comparative perspective, which is still quite rare in quantitative media studies. In particular, they echo with Goncharov and Nechay (2018) in observing that Russian media on Twitter, unlike those of other countries, cluster along the pro-government—anti-government axis.

A successful attempt to do inferential research based on blog texts is presented by Bulovsky (2019). He fits a regression model to find out the difference between Twitter communication used by authoritarian and democratic political leaders across 144 countries, including Russia. He finds that the former have a significantly lower number of posts per day and a significantly smaller proportion of replies to other users.

Another difficulty with SNS research is a possible lack of context and the reliability of non-SNS-based data. Enikolopov et al. (Enikolopov et al. 2018) perform a most rigorous statistical inference to evaluate the influence of VK on protests in Russia using the data on the large rallies against electoral fraud and on voting in the electoral cycle of 2011–2012. They find out that, paradoxically, VK penetration increases pro-government voting in the respective cities, but simultaneously has a positive effect on both the probability of protests and the number of protesters. There is not enough data to test possible alternative explanations of this effect, such as a polarizing influence of higher SNS penetration levels on the population. At the same time, the reliability of both voting data (given the unknown character of electoral fraud) and the protest data (given that they were taken from the media where the numbers reported by the protesters and by the police dramatically diverged) is highly questionable. Reuter and Szakonyi (2015), who use offline survey data only, obtain somewhat different results and find that the usage of international social networks Twitter and Facebook increased the awareness about electoral fraud, while the usage of domestic VK and Odnoklassniki did not.

Koltsova and Selivanova (2019) solve the problem of contextual enrichment of SNS data with the deep involvement of one of the researchers into the social movement they study. Similar to Goncharov and Nechay (2018), their goal is to evaluate the mobilization potential of VK communities. In particular, they study the effect of VK on the turnout of the movement activists at the poll stations in the role of independent observers on the voting day in 2014. As the movement created VK groups responsible for each of the 17 administrative districts in St. Petersburg, the researchers investigate them all and find that their size, activity and density are positively related to the overall turnout; however, offline observers are neither more active nor more connected members of online groups. The authors offer two alternative interpretations of this effect, based on their experience with the movement, still the problem of reliability of the offline turnout data is one of the unresolved issues.

Importantly, SNS data, though more “objectified” than self-reported or hand-coded data, are not always reliable either. Comparing Twitter discussions devoted to two resonant political murders in Russia and Argentina, Filer and Fredheim (2016) find out that a significant proportion of the Russian tweets, unlike Argentinean messages, are automatically generated (2016, 13). This has two implications: first, the value of Twitter as the data source in Russian political research is limited because of the network’s limited penetration. Second, the study of Filer and Fredheim leads us to a whole range of related research topics that include online state propaganda, fake news, bot-generated content, influence of all those factors on electoral outcomes and the problem of the use of personal SNS data for political purposes.

Of this stream of research, Russia-related research has a number of special features. First, some research on political propaganda is performed as bot detection (Stukal et al. 2017). However, not all bots are political (some are commercial), and not all propaganda is robotized (some is manual). Second, the phenomena that researchers try to trace—for example, trolls—are hard to define conceptually and even harder to find empirically. For instance, Badawy et al. (2018) perform an interesting descriptive research of a collection of tweet accounts identified as Russian “trolls.” Using one of the algorithms of automatic classification known as label propagation, they identify most trolls as conservative-leaning, while Botometer software ([botometer.iuni.iu.edu/#/](http://botometer.iuni.iu.edu/#/)) (another classification algorithm) allows them to determine that the majority of those who retweeted trolls were not bots. Those valuable findings are to a certain extent limited by the data used. While trolls are defined as “malicious accounts created for the purpose of manipulation” (Badawy et al. 2018, 258)—that is, intentionally deceptive—the authors use a list of Twitter accounts taken from the website of the US congress Committee for Intelligence via a publication at the [www.recode.net](http://www.recode.net) website. The list contains only Twitter IDs and usernames, and no other information is searchable. It is thus unclear whether the list creators followed the authors’ definition of trolls, and if so, how they learnt about users’ purposes. More broadly, finding empirical referents for concepts whose definitions are based on intention (e.g. deception) is a challenge, while taking out intention from the definition of political trolls deprives them from their core meaning and makes them lumped together with authors of unconventional, still legitimate opinions.

Third, this type of research is sometimes not entirely free from politicization. For instance, Sanovich (Sanovich 2017) refers to Barash and Kelly (2012) and Kelly et al. (2012) as the research that for the first time identified a “large-scale deployment of pro-government bots and trolls in Russia” in favor of President Medvedev. However, this is not exactly what the sources suggest. First of all, while the word “bot” is mentioned only once, “troll” is never mentioned in either of the sources. Second, Barash and Kelly (2012) show unusual distributions of activity around tweets related to Medvedev’s innovation policy program, while Kelly et al. (2012), using the same data, note that the whole “innovation cluster” of tweets disappears when they filter out “instrumental”

accounts—those that are likely to use search engine optimization (SEO) and automation (bots). The authors do not offer any interpretations, but this suggests that accounts that tweeted about innovation were likely to use commercial promotion methods, in particular automation. Since Kollanyi et al. (2016) show that automation was spread both among pro-Trump and, to a lesser extent, pro-Clinton Twitter accounts during the 2016 US electoral campaign, such accounts might be equally termed trolls. However, there is more sense in distinguishing between trolls and bots than placing them in one category. This does not mean that the Russian government has never used trolls but suggests a certain lack of accuracy when it comes to the Russian computational propaganda research.

To sum up, social media data may significantly enrich the repertoire of research in the field of political science in Russia by providing access to large volumes of data and thereby conducting large-scale research with the usage of automated methods of data analysis. In particular, the access to textual and network data makes it possible to grasp the substance of political discussions and track communication flows between different political parties. At the same time, the results of such research should be viewed through the lens of existing technical and methodological constraints. First, often accessible data allows producing only a limited range of conclusions, for example, descriptive ones. Different approaches to the conceptualization of key concepts which sometimes leads to inconsistent results are also of highly relevant issue. Finally, online data from social media are not free from specific limitations which may affect its reliability.

## 19.4 SOCIOLOGY

### 19.4.1 *Virtual Demography and Structure*

The relative openness of VK data has enabled a number of large-scale studies investigating VK population structure, composition and patterns of communication. By far the largest of them is the project “Virtual population of Russia” (Zamyatina and Yashunsky 2018). It is based on the analysis of approximately 200 million VK accounts and 3.5 billion friendship links, although only 88 million accounts have been found to claim their location in Russia. The most valuable outcome of this project is an interactive website [webcensus.ru](http://webcensus.ru) that contains various subsets of the initial sample at different levels of aggregation and visualizes the most important distributions in the form of charts and maps. The data include age, gender, education, friendship patterns, migration routes and others, along with their relations, for example, distribution of average friendship connectedness over the Russian regions. This data is an incredible resource for researchers seeking to assess the difference of their samples from the total VK population of Russia and thus to statistically test various hypotheses about specific features of the studied sub-populations. Russia is, to our knowledge, the only country for which such detailed virtual census exists. However, this data

has a serious limitation: it dates back to 2015 and, as such data collection is extremely expensive and is unlikely to be updated.

Some studies aim at restoring missing data in SNS accounts based on the available data from other accounts. Such data may include age, gender, geolocation and others. As this research mostly lies in the sphere of computer science, we omit it here, with an exception of two related papers based on the full VK data from the Russian city of Izhevsk. The first paper studies the impact of missing geolocation data on the features of the city friendship network (Kaveeva and Gurin 2018), while the second does the same in relation to fake accounts (Kaveeva et al. 2018). In the first paper, the authors train a classifier that restores users' city of residence based on the accounts in which this data is present. They find out that while the city friendship network grows substantially when the missing users are added, most of its important metrics, do not change, while modularity and the number of communities in the largest connected component experience a very modest growth. This suggests that using incomplete data based on the users who choose to report publicly is a valid research strategy in social network analysis. In the second paper the authors train a classifier to recognize fake accounts. After deleting them, friendship network experiences the reverse change, as compared to the first paper, that is modularity and the number of communities drop a little. One limitation of the second paper is the nature of its training set that is based on self-reported data from 32 users who were asked to assess their friends. The problem of fake account identification is close to troll and bot detection and has no easy solution as all these types of accounts mutate quickly in their attempts to imitate real users.

Other research devoted to SNS structural features attempts inferential design (Kisilevich et al. 2012; Rykov et al. 2018). Kisilevich et al. (2012) examine how age and gender are related to the amount of disclosed information, based on 16 million accounts from a Russian SNS My World (Moj Mir, [www.my.mail.ru](http://www.my.mail.ru)). This research allows not only to evaluate the completeness of self-reported user data but also to investigate their self-disclosure behavior. The authors report that self-disclosure dramatically drops with age, but find no substantial gender difference which, as they claim, differentiates the Russian SNS users from the Western users. However, the statistical procedure used to claim no gender difference is somewhat unclear, while the presented plots suggest that females may be less frequently sharing information about their political views but are more often sharing quite a number of other types of information.

#### 19.4.2 *Social Issues and Problems: Education, Ethnicity, Urbanity*

While virtual demography and related studies are interested in the online population per se, various research tackling more specific sociological topics uses SNS data to obtain results about offline reality, or about the role of SNS in the respective problem or issue, be it education, migration, ethnic relations, or urbanity.

For instance, Smirnov (2018) uses VK data on 4400 Russian students to predict their scores in Programme for International Student Assessment (PISA) test—an international test assessing learning outcomes in reading, mathematics and science among 15-year old students. He obtains the data on 73 thousand online communities which students belong to and reaches the correlation of about 0.5 between the predicted and the real PISA scores. The most interesting conclusion is that groups contributing most to high scores are related to arts and science, while those contributing to low scores are related to humor, sex and horoscopes. Although 0.5 is a high correlation in social science, it also means that VK group membership cannot be used as the only predictor of PISA scores. Since VK group membership cannot be used as a causal explanation of PISA performance either, the significance of such type of research should be treated with caution. In general, the predictive power of such models tends to drop the more the farther in time other studied datasets are from the original dataset.

Alexandrov et al. (2018) use VK self-reported geolocation data to study factors influencing outgoing educational migration. They examine a larger number of student migration destinations over a sample of 85 thousand VK users aggregated at the city level. They find that, quite predictably, Far East and Southern Siberia are gravitated to China, North-West—to the Nordic countries, and Muslim regions—to the Middle East. It is interesting that among significant predictors they find both offline factors, such as religious and geographic proximity, and online factors, such as the number of VK groups related to the country of destination in the donor city. However, of course, it is unknown how well self-reported data on users' secondary schools and universities reflect the overall educational migration flows. This poses a broader question of the extent to which online data represent offline reality.

Ethnicity has been another important topic in Russia as a multi-ethnic society, with studies asking how communication in social media either reflects ethnic tensions or influences ethnic relations. Bodrunova et al. (2017) study the posts of top LiveJournal bloggers to determine how different ethnic groups are treated. They first extract ethnicity-related topical clusters via topic modeling and then hand-code 30 most relevant texts in each of 33 ethnicity-related clusters. Among other things, they find that Central Asians are treated as relatively positive aliens, with North Caucasians being presented both as negatively assessed aliens and aggressors. But while North Caucasians are also sometimes victimized, Americans are always both negative and aggressive, which suggests that global political conflicts overshadow local inter-ethnic tensions. This research is one of the few addressing the problem of instability of topic modeling by running the algorithm several times and choosing only stable topics, which is virtually never done in empirical social research. However, it has problems with representativity both in terms of the size of the coded sample and in terms of the choice of LiveJournal popular bloggers as a source.

Urban studies is a subfield that is widely believed to benefit from social media data. However, it often results in the simple plotting of social media data

on geographical city maps that, unlike the mapping of political discussions or virtual demography, is often less useful. Human movement in urban spaces is much better detected with mobile phone data than with social media data, and the content or sentiment of SNS messages is not always related to the places where they have been created. Some studies still attempt to tie geomapping to practical purposes of urban planning. Thus Petrova et al. (2016) examine some hundred thousand posts and check-ins from different SNSs in the city of Samara in order to generate town planning recommendations.. They find that messages are concentrated in the city center and are differentiated by gender, type of place, and topic, while locations also differ by check-in intensity, predominant sentiment of messages and the prevailing type of visitors—either locals or tourists. The authors suggest to create more attractive places in the city periphery and also to unite the most visited and the most positively assessed places in the center by a single pedestrian pathway. This conclusion seems to be based on the visual examination of maps, as the paper describes neither analytic procedures nor any methods of posterior evaluation of the efficiency of the suggested town planning strategy (for more, see Chap. 32).

An interesting result about the nature of urban civic activity is presented in Voskresenskiy et al. (2016). The authors analyze 41 restricted-access and 132 open-access VK groups run by the neighbors sharing the same apartment blocks in St. Petersburg. Based on topic modeling, restricted-access groups are found to prefer such topics as mutual help, socialization and apartment repairs, while open access groups favor city-level initiatives, contentious initiatives, including court disputes with the city administration, and, paradoxically, the maintenance of their apartment blocks and yards. This research is a rare comparison of closed and open SNS groups in an urban context, although one should keep in mind that the majority of the former (91 of 132) had denied access to the researchers.

Compared to political research, sociological research of Russian SNSs is less numerous. While political science has one of its important objects of research, namely political discourse and discussion, readily available in the form of online content, sociological focus demands more links to offline reality, which makes the overall tasks more difficult. Additionally, sociological problems of Russia generally provoke less interest from the international research community than the country's political problems.

## 19.5 PSYCHOLOGY AND HEALTH STUDIES

### 19.5.1 *Health Studies*

Health studies, as a field of research lying at the intersection of sociology, policy studies, psychology and medicine, is very young in Russia, and the works in E-Health are few and mostly exploratory. A series of papers has been devoted to the VK groups of acquired immune deficiency syndrome (AIDS) denialists, people who deny the existence of AIDS or its relation to human

immunodeficiency viruses (HIV), and other AIDS-related groups. The first work by Meylakhs et al. (2014) uses netnography (an online variant of ethnography) to examine the largest community of AIDS-denialists in VK. The authors obtain a policy-relevant result that the motives of newcomers are often far from irrational and may result from negative experience with doctors or atypical medical history. In addition, persuasive strategies of the “old” community members are described which makes the authors set a new research task—to find a method that would discern the core of the community from its periphery. The significance of this task is based on the assumption that, while core members cannot be convinced to change their views, the periphery could be re-oriented. This task is addressed in Rykov et al. (2017) with the help of SNA and regression analysis. The authors indeed find a correlation between some network and activity measures, on the one hand, and the core-periphery status of a user as determined by hand coding of his/her messages, on the other. However, as in (Smirnov 2018) the set of the examined predictors is not sufficient to classify the users correctly, which is why hand coding seems to be still needed.

Meylakhs’ conclusions about the motivation of AIDS-denialism newcomers echo qualitative research on coping strategies of HIV-positive people (Dudina and Artamonova 2018). The authors exploit the anonymous character of the respective Russian-language forum obtaining confessions that, in the authors’ opinion, would have never been possible in face-to-face interviews. On the whole, this research describes well-known stages of coping with chronic illness and the related problems, such as shock, denial, acceptance, status disclosure and stigmatization.

An attempt to describe interests of drug users based on social network data is made in Yakushev and Mityagin (2014). The authors apply a keyword-based search while crawling Russian-language LiveJournal accounts in order to find users who write about drugs. They also exploit an LJ feature allowing users to include tags representing their interests and perform a statistical test to find out which interests are typical among those users who write about drugs, as compared to those who do not. The main problem with this approach, as the authors themselves indicate, is the lack of equivalence between those who write about drugs and those who actually use them.

Thus, studying of various online communities in Russian social media is one of the research avenues in health studies, which open up great prospects for in-depth study of the structure, communication network, and leadership phenomenon in different, including hidden and hard-to-reach, populations.

### 19.5.2 *Psychology*

As mentioned earlier, psychological research based on Russian-language social media is least focused on Russia, but more often attempts to establish online-offline connections by seeking to predict psychological traits or conditions with social media data. Thus, Semenov et al. (2015) try to predict depression



propensity with the data from VK accounts, including different network metrics, and reach area under ROC (receiver operating characteristic) curve (AUC) metric of 0.84 which is comparable to other research in the field. The main problem of this research, similar to Yakushev and Mityagin (2014), is that the training set of users with depression propensity is compiled of those who contributed to discussion threads on being suicidal in depression- and suicide-related VK communities. This problem may be resolved in two different ways: by either collecting ground truth on psychological conditions outside social media, as in Panicheva et al. (2016) or by using social media data not as an indicator of “true” psychological condition but as a source of users’ self-representations, as in Bogolyubova et al. (2018).

In the latter work, the authors compare Instagram images used by Russian-speaking and anglophone users to express psychological distress. The truthfulness of those expressions is thus left out; this lets the authors concentrate on the observable data and make interesting findings about significantly more frequent use of images containing text by anglophone users. The authors connect this finding to the lack of culture of verbal psychological self-expression in Russia. It should be noted that in this research the agreement between coders who manually assessed the images was not very high, which is a common problem for this type of studies based on human labeling.

Panicheva et al. (2016) develop a Facebook application to collect the ground-truth data about users’ psychological traits—in their case, the so-called dark triad. They manage to obtain both the completed questionnaires and the text data from almost 2000 Russian-speaking users which is a huge number for psychological research. Using text data to predict the dark triad, the researchers, however, refuse from constructing a single model: as they are interested in evaluating the effect of each linguistic feature, not in the accuracy of prediction; they acknowledge the problem of distortion of significance levels in the models with too many predictors and apply a special correction procedure. For some reason, this problem is seldom raised outside psychological community, although it is typical for other tasks using such high-dimensional data as texts, and attention to this problem is a special value of the research by Panicheva and colleagues. At the same time, in their work, as only a limited number of user messages are available, and the sample is not described, the biases that might be introduced by these factors may be in fact more significant than those that the authors are struggling against.

An interesting extension of this research is presented in Bogolyubova et al. (2018) where the authors relate users’ linguistic behavior, their psychological traits and the propensity to engage in harmful online behavior. They find out that one of the dark triad components—psychopathy—is the best predictor of such behavior. They also use a different strategy to deal with linguistic features by first representing words as word vectors (lists of words most closely associated with each given word) and then clustering them into 182 clusters. They use these clusters as harmful behavior predictors with the same procedure of significance correction as in their first paper. It should be said that, just like

topic modeling (for more, see Chaps. 23, 25 and 24), both word embeddings and clustering algorithm used are unstable, and when combined can produce an indefinite number of different solutions with the same data.

A task similar to Panicheva (2016) is addressed in Rubtsova et al. (Rubtsova et al. 2018): the authors seek to find associations between user account features in VK and the types of teenager personality accentuations as defined by Lichko (1983). A major limitation of this research is that while Lichko's classification contains 11 types of personality, the authors manage to survey only 88 teenagers. This again raises the problem of big data collapse into small data due to constrained access to one type of data needed. This problem is also present in the research by Belinskaya and Bronin (2015) which is a reduced replica of the famous study by Youyou et al. (2015). Both teams use quasi-experimental design to measure the accuracy of perception of the most important personality traits—the so-called Big Five (Piedmont 2014)—by FB or VK users, respectively. For this, they ask one group of subjects (the assessed) to fill in the Big Five questionnaire, while the other group of subjects (assessors) are asked to fill in the same questionnaire on behalf of the assessed subjects. The differences between the two studies are, however, more significant than their similarities. First, Kosinski's team tests the accuracy of those who know the assessed people well, while Belinskaya and Bronin focus on those who have only met their friends online. Second, Kosinski and colleagues, by developing and promoting a FB application, manage to collect 86,220 observations, while the Russian authors collect only 30 offline assessments from 15 assessors. This turns the problem of big data collapse into a problem of digital divide in science: while collecting big data online seems cheap at the first glance, this is not the case in practice. On the contrary, substantial financial resources and time are needed to conduct large-scale research with social media data.

## 19.6 CONCLUSION

In this chapter we reviewed both the works on the Russian-language social media and the Russia-related topics that can be studied with social media data in general. We have shown that Russian SNSs give very broad opportunities for research—broader than most international SNSs do. However, this potential stays somewhat underused due to a number of factors, including the lack of resources for researchers within Russia and the lack of interest to the opportunities given by the Russian SNSs from the international scholars. The sphere that generates the largest interest from the international researchers is Russian politics, and this is reflected in the dominant position of this topic among Russian SNS-based studies. Sociological research is somewhat fragmented, and psychology studies are least of all related to Russia, with some strong studies done by the Russian scholars not using Russian data at all (Buraya et al. 2018).

In our review we focused both on the opportunities and problems of social media research. Our goal was to go beyond the strengths and limitations of

concrete works and to highlight the common trends, especially the limitations of the field because they are seldom spoken about in research papers, since they tend to report success rather than failures.

The opportunities include the ability to obtain large observational data collected in a non-intrusive manner and the ability to scale the research that otherwise would be bound to very small laboratory experiments or qualitative field work. Additionally, the fact that the data of the Russian-language SNSs come mostly from the Post-Soviet space gives an opportunity to study various political, social and psychological phenomena outside the Western context where most social research data come from. Finally, social media are an important key to a society where other types of data are often less available than in more transparent countries.

The limitations are, however, also large. First, online digital traces, in order to be meaningful, often have to be combined with other types of data that are not so easy to collect and that become a bottleneck on the way to large samples. This is where we observe the effect of big data collapse. Second, SNS data have various problems of representativity in terms of their ability to represent both offline and online phenomena. Sampling network data and especially textual data is generally a poorly developed methodological area, while these types of data are the core of digital traces left by humans on social media.

Finally, methods of SNS data analysis are lagging behind the techniques available for data collection. The existing approaches are very complex, and they hide many caveats that social scientists are often unaware of. Instability of the majority of text-clustering techniques, absence of statistical inference methods for non-independent (networked) data, lack of approaches to work with power-law distributions so common for SNS data compromise the validity of many of the existing studies without social scientists being fully able to grasp the scale of the problem. Nevertheless, an open discussion of these methodological difficulties can enrich our understanding of the field of social media research and enhance its development.

**Acknowledgments** This work is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

## REFERENCES

- Alexandrov, Daniel, Viktor Karepin, Ilya Musabirov, and Daria Chuprina. 2018. Educational Migration from Russia to the Nordic Countries, China and the Middle East. Social Media Data. In *Companion of the The Web Conference 2018—WWW'18*, 49–50. Lyon, France: ACM Press. <https://doi.org/10.1145/3184558.3186923>.
- Badawy, Adam, Emilio Ferrara, and Kristina Lerman. 2018. Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258–265. IEEE.

- Barash, Vladimir, and John Kelly. 2012. Salience vs. Commitment: Dynamics of Political Hashtags in Russian Twitter. In *SSRN Scholarly Paper ID 2034506*. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2034506>.
- Belinskaya, E.P., and I.D. Bronin. 2015. Accuracy of Interpersonal Perception in Mediated Contacts in Social Media. *Social Psychology and Society* 6 (4): 91–108. <https://doi.org/10.17759/sps.2015060407>.
- Bodrunova, Svetlana S., Olessia Koltsova, Sergey Koltcov, and Sergey Nikolenko. 2017. Who's Bad? Attitudes Toward Resettlers from the Post-Soviet South Versus Other Nations in the Russian Blogosphere. *International Journal of Communication* 11: 3242–3264.
- Bodrunova, Svetlana S., Anna A. Litvinenko, and Ivan S. Blekanov. 2018. Please Follow Us: Media Roles in Twitter Discussions in the United States, Germany, France, and Russia. *Journalism Practice* 12 (2): 177–203. <https://doi.org/10.1080/17512786.2017.1394208>.
- Bogolyubova, Olga, Philipp Upravitelev, Anastasia Churilova, and Yanina Ledovaya. 2018. Expression of Psychological Distress on Instagram Using Hashtags in Russian and English: A Comparative Analysis. *Sage Open* 8 (4): 2158244018811409. <https://doi.org/10.1177/2158244018811409>.
- Brand Analytics. 2018. SNS in Russia (Report). *Brand Analytics 2018*. <https://brand-analytics.ru/blog/wp-content/uploads/2018/12/Sotsseti-Rossiya-osen-2018.pdf>.
- Bulovsky, Andrew. 2019. Authoritarian Communication on Social Media: The Relationship between Democracy and Leaders' Digital Communicative Practices. *International Communication Gazette* 81 (1): 20–45. <https://doi.org/10.1177/1748048518767798>.
- Buraya, Kseniya, Aleksandr Farseev, and Andrey Filchenkov. 2018. Multi-view Personality Profiling Based on Longitudinal Data. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, ed. Patrice Bellot, Chiraz Trabelsi, Josiane Mothe, Fionn Murtagh, Jian Yun Nie, Laure Soulier, Eric SanJuan, Linda Cappellato, and Nicola Ferro, vol. 11018, 15–27. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-98932-7\\_2](https://doi.org/10.1007/978-3-319-98932-7_2).
- Dudina, Victoria I., and Kristina N. Artamonova. 2018. Stigmatization of People Living with HIV/AIDS and the Problem of Disclosure: Analysis of the Online Forum Discussions. *Sociology* 11 (1): 66–78. <https://doi.org/10.21638/11701/spbu12.2018.106>.
- Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2018. Social Media and Protest Participation: Evidence from Russia. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2696236>.
- Filer, Tanya, and Rolf Fredheim. 2016. Sparking Debate? Political Deaths and Twitter Discourses in Argentina and Russia. *Information Communication & Society* 19 (11): 1539–1555. <https://doi.org/10.1080/1369118X.2016.1140805>.
- Forbes.ru. 2019. Vremâ Edinorogov. Kak v Runete sozdaetsâ biznes stoimost'û \$1 mlrd | Tehnologii. Forbes.Ru, February 21, 2019. <https://www.forbes.ru/tehnologii/372571-vremya-edinorogov-kak-v-runete-sozdaetsya-biznes-stoimostyu-1-mlrd>.
- Gel'man, V. 2014. The Rise and Decline of Electoral Authoritarianism in Russia. *Demokratizatsiya* 22 (4): 503.
- Goncharov, D.V., and V.V. Nechay. 2018. Anti-Corruption Protests 2017: Reflection in Twitter. *The Journal of Political Theory, Political Philosophy and Sociology of Politics Politeia* 88 (1): 65–81. <https://doi.org/10.30570/2078-5089-2018-88-1-65-81>.

- Kaveeva, Adelia, and Konstantin Gurin. 2018. 'VKontakte' Local Friendship Networks: Identifying the Missed Residence of Users in Profile Data. *The Monitoring of Public Opinion: Economic & Social Changes* 3: 78–90. <https://doi.org/10.14515/monitoring.2018.3.05>.
- Kaveeva, Adelia, Konstantin Gurin, and Valery Solovyev. 2018. How 'VKontakte' Fake Accounts Influence the Social Network of Users. *International Conference on Digital Transformation and Global Society*, 492–502. Springer.
- Kelly, John, Vladimir Barash, Karina Alexanyan, Bruce Etling, Robert Faris, Urs Gasser, and John G. Palfrey. 2012. Mapping Russian Twitter. In *SSRN Scholarly Paper ID 2028158*. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=2028158>.
- Kisilevich, Slava, Chee Siang Ang, and Mark Last. 2012. Large-Scale Analysis of Self-Disclosure Patterns among Online Social Networks Users: A Russian Context. *Knowledge and Information Systems* 32 (3): 609–628. <https://doi.org/10.1007/s10115-011-0443-z>.
- Kollanyi, B., P.N. Howard, and S.C. Woolley. 2016. Bots and Automation Over Twitter during the First US Presidential Debate. *Comprop Data Memo* 1: 1–4.
- Koltsova, Olessia, and Sergei Koltcov. 2013. Mapping the Public Agenda with Topic Modeling: The Case of the Russian LiveJournal. *Policy and Internet* 5 (2): 207–227. <https://doi.org/10.1002/1944-2866.POI331>.
- Koltsova, Olessia, and Galina Selivanova. 2019. Explaining Offline Participation in a Social Movement with Online Data: The Case of Observers for Fair Elections. *Mobilization: An International Quarterly* 24 (1): 77–94. <https://doi.org/10.17813/1086-671X-24-1-77>.
- Lichko, A.E. 1983. *Psihopatii i akcentuacii haraktera u podroستkov*. Medicina.
- Mediascope. 2018. WEB-Index: The Audience of Internet Projects. The Results of the Measurement: Desktop, November 2018, Russia 0+. [https://mediascope.net/data/?download=3179&date=2018%2011&set\\_filter=Y&FILTER\\_TYPE=internet](https://mediascope.net/data/?download=3179&date=2018%2011&set_filter=Y&FILTER_TYPE=internet).
- Meylakhs, Peter, Yuri Rykov, Olessia Koltsova, and Sergey Koltsov. 2014. An AIDS-Denialist Online Community on a Russian Social Networking Service: Patterns of Interactions with Newcomers and Rhetorical Strategies of Persuasion. *Journal of Medical Internet Research* 16 (11): e261. <https://doi.org/10.2196/jmir.3338>.
- Panicheva, Polina, Yanina Ledovaya, and Olga Bogolyubova. 2016. *Lexical, Morphological and Semantic Correlates of the Dark Triad Personality Traits in Russian Facebook Texts*. In 2016 IEEE artificial intelligence and natural language conference (AINL); 1–8. IEEE.
- Petrova, Marina, Aleksandra Nenko, and Kirill Sukharev. 2016. Urban Acupuncture 2.0: Urban Management Tool Inspired by Social Media. In *Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia*, 248–257. ACM.
- Piedmont, R.L. 2014. Five Factor Model of Personality. In *Encyclopedia of Quality of Life and Well-Being Research*, ed. A.C. Michalos. Dordrecht: Springer.
- Reuter, Ora John, and David Szakonyi. 2015. Online Social Media and Political Awareness in Authoritarian Regimes. *British Journal of Political Science* 45 (1): 29–51. <https://doi.org/10.1017/S0007123413000203>.
- Rubtsova, O., A.S. Panfilova, and V.K. Smirnova. 2018. Research on Relationship between Personality Traits and Online Behaviour in Adolescents (With VKontakte Social Media as an Example). *Psihologičeskaâ Nauka I Obrazovanie [Psychological Science and Education]* 23 (3): 54–66. <https://doi.org/10.17759/pse.2018230305>.

- Rykov, Yuri, Peter A. Meylakhs, and Yadviga E. Sinyavskaya. 2017. Network Structure of an AIDS-Denialist Online Community: Identifying Core Members and the Risk Group. *American Behavioral Scientist* 61 (7): 688–706. <https://doi.org/10.1177/0002764217717565>.
- Rykov, Yuri, Yadviga Sinyavskaya, and Olessia Koltsova. 2018. Accumulating Social Capital in an Online Urban Network: The Effects of User Behaviors. *SSRN Electronic Journal*. Higher School of Economics Research Paper No. WP BRP 83, 1–27. <https://doi.org/10.2139/ssrn.3301266>.
- Sanovich, Sergey. 2017. Computational Propaganda in Russia: The Origins of Digital Misinformation. *Working Paper, Computational Propaganda Research Project*. Oxford Internet Institute. <http://www.philosophyofinformation.net/wp-content/uploads/sites/89/2017/06/Comprop-Russia.pdf>.
- Semenov, Aleksandr, Alexey Natekin, Sergey Nikolenko, Philipp Upravitelev, Mikhail Trofimov, and Maxim Kharchenko. 2015. Discerning Depression Propensity Among Participants of Suicide and Depression-Related Groups of Vk.Com. In *Analysis of Images, Social Networks and Texts*, ed. Mikhail Yu, Natalia Konstantinova Khachay, Alexander Panchenko, Dmitry Ignatov, and Valeri G. Labunets, vol. 542, 24–35. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-26123-2\\_3](https://doi.org/10.1007/978-3-319-26123-2_3).
- Smirnov, Ivan. 2018. Predicting PISA Scores from Students' Digital Traces. *Twelfth International AAAI Conference on Web and Social Media*.
- Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2017. Detecting Bots on Russian Political Twitter. *Big Data* 5 (4): 310–324. <https://doi.org/10.1089/big.2017.0038>.
- Voskresenskiy, Vadim, Ilya Musabirov, and Daniel Alexandrov. 2016. Private and Public Online Groups in Apartment Buildings of St. Petersburg. In *Proceedings of the 8th ACM Conference on Web Science—WebSci'16*, 301–306. Hannover, Germany: ACM Press. <https://doi.org/10.1145/2908131.2908173>.
- Yakushev, Andrei, and Sergey Mityagin. 2014. Social Networks Mining for Analysis and Modeling Drugs Usage. *Procedia Computer Science* 29: 2462–2471. <https://doi.org/10.1016/j.procs.2014.05.230>.
- Youyou, Wu, Michal Kosinski, and David Stillwell. 2015. Computer-Based Personality Judgments are More Accurate than Those Made by Humans. *Proceedings of the National Academy of Sciences* 112 (4): 1036–1040.
- Zamyatina, Nadezhda, and Alexey Yashunsky. 2018. Virtual Geography of Virtual Population. *The Monitoring of Public Opinion: Economic&Social Changes* 1: 117–137. <https://doi.org/10.14515/monitoring.2018.1.07>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.







# Digitizing Archives in Russia: Epistemic Sovereignty and Its Challenges in the Digital Age

*Alexey Golubev*

## 20.1 INTRODUCTION

In March 2016, the Supreme Court of the Russian Federation examined a recent decision by the Ministry of Culture, the parent body of the Federal Archival Agency (Rosarhiv), to ban personal use of cameras, smartphones, and other technical devices for copying documents in Russian archives. During court hearings, a representative of the ministry argued that free and unlimited digitization without supervision by professional archivists would likely cause increased wear and damage to historical documents. She also argued that the current ban did not violate the right of free access to archival documents and hence to historical knowledge. This position received support from many professional archivists; one of their arguments was that, if unrestricted copying of archival materials was allowed, archives would be unable to guarantee the authenticity of copies, which, in turn, would lead to manipulations of historical evidence on a large scale. Nevertheless, the Supreme Court ruled that this decision was unlawful and it had to be repealed (Galanichev 2016, 299–300; Druzin 2018, 4–5).

---

I am grateful to David Brandenberger, Tatyana Doorn-Moisseenko, Joan Neuberger, Serguei Oushakine, the participants of the University of Houston Digital Research Commons lecture series, and the editors and reviewers of this volume for their insightful comments and suggestions that helped me improve this chapter.

---

A. Golubev (✉)  
University of Houston, Houston, TX, USA  
e-mail: [avgolubev@uh.edu](mailto:avgolubev@uh.edu)

Archive users did not have much time to enjoy free and unlimited copying. In April 2016, Rosarhiv was transferred from the Ministry of Culture to the Presidential Administration of Russia, signifying that its head now reported directly to Vladimir Putin. Using its new legal status, in September 2017, the agency introduced a new regulation that allowed the use of personal cameras for copying, but only by permission (that could take days or even weeks) and for a fee. This regulation was challenged in the Supreme Court as well, but this time the court ruled that Rosarhiv did not violate any law and the practice of charging archive users for making digital copies with their own devices was legal (Kurilova 2019). As a result, Rosarhiv regained full control over the digitization of archival materials, even when it is done by private users for a personal purpose.

This story highlights the extent to which Russian state agencies and officials are concerned with digital reproducibility of historical documents as a phenomenon that challenges their control over the production of historical knowledge. According to the code and practice of Russian law, access to historical documents is a civil right, and while this right is routinely violated by state institutions, in many cases it has been successfully enforced through legal battles. At the same time, digitization of historical documents turned out to be a more controversial issue. The relative ease of the digital access and reproduction implies that online archives of historical documents can be not only produced at a low cost but also be available to a much bigger audience than before (for an example of such archives, see Chap. 21). This offers new opportunities in terms of communication of knowledge, but it also means that the circulation and production of historical knowledge becomes less centralized as it moves away from the expert-controlled domains such as edited volumes published by academic presses into a more egalitarian information society that trespasses national borders and jurisdictions where digital archives can be curated by virtually anyone. Given how important the maintenance of a coherent national narrative is for many Russian top officials (Brandenberger 2015, 200–205; Shteynman 2016, 107–108), it is unsurprising that Rosarhiv is very cautious when it comes to releasing the documents from its collections on the World Wide Web.

Yet the politics of history represent only one aspect of the current situation with digital archives in Russia. The unwillingness of Russian archives to outsource digitization of their materials means that when they start their own digitization projects, as a rule, they end up with high-cost solutions. An August 2018 post published by the Archival Committee of St. Petersburg in its Facebook group claimed that it took the archives fifteen years to digitize the parish register books of St. Petersburg; the same post claimed that a complete digitization of all collections of the St. Petersburg archives would require at least 45 billion rubles (ca. \$650 million at the current exchange rate) worth of funding and presumably decades of work (Archival Committee of St. Petersburg 2018). As a result, managers of Russian archives have to choose strategically which documents and collections should be presented online. Since most of

the archives are chronically underfunded, they also actively solicit external funding both domestically and internationally in order to push forward their digitization efforts and get a better public outreach. The research and political agenda of funding agencies thus becomes part of the archival digitization process in Russia.

This chapter examines the production of digital archives in a broader context of the political economy of historical knowledge in Russia. The resolution of Rosarhiv to control which and how many documents its users copy digitally should be treated as a symptom signaling the epistemic anxieties of the state authorities. The archive, as the works of Michel Foucault, Jacques Derrida, Ann Stoler and many others have shown, is a key institution of the state power over history: it defines the dominant forms of knowledge, its limits and silences, establishes hierarchies of voices from the past, and produces experts with authority to interpret its documents. In other words, the archive is a powerful tool that transforms the totality of the historical experience preserved by a certain community or society into a structured hegemonic form that privileges some parts of this experience and silences other (Foucault 1977, 155–164; Foucault 2002, 146–148; Derrida 1996, 3–5, 10–13, 19–20; Stoler 2006, 44–51). However, modern information and communication technologies represent a formidable challenge to maintaining this epistemic sovereignty as they simplified to the extreme a precise reproduction of historical documents and production of digital archives, thus diluting the state's sovereign control over historical knowledge. The situation is further complicated by a relatively marginal place of Russian archives in the global economy of knowledge where the demand for the digitization of their materials often comes from outside of the Russian national borders and represents political and cultural interests of non-state agents. As a result, a study of digital archives in Russia provides a unique perspective on the ways in which the Russian state adapts to the digital age.

This chapter starts with a discussion of the early digitization efforts during the 1990s and early 2000s that were largely funded by international funding agencies, most prominently the Open Society Institute, whose activity is presently banned in Russia. It then examines the growing concern of the state authorities over the reproduction of historical documents online and their efforts to control the production of digital archives through specialized funding agencies (such as the Russian Foundation for Humanities that in 2016 merged with the Russian Science Foundation) and restrictive measures. In the concluding section of the chapter I discuss the current state of digital archives in Russia, which is not limited to the activities of federal and international agencies, but also involves a great deal of private initiative. Yet despite the multitude of actors and large-scale digitization efforts, I argue, Russian digital archives are curated to conform to the dominant paradigms of historical knowledge and have so far failed to induce an epistemological shift in the studies of Russian history.

## 20.2 ARCHIVAL REVOLUTION: INTERNATIONAL ACTORS AND EARLY DIGITIZATION EFFORTS

The early effort to digitize historical documents from the former Soviet archives was an integral part of the archival revolution that started at the turn of the 1990s following the perestroika and led to an unprecedented openness of Russian archival collections for both the scholarly community and general public (Raleigh 2002). In Russia, this effort was spearheaded by international organizations. Between 1992 and 1996, for example, the Hoover Institution on War, Revolution, and Peace with the UK-based company Chadwyck-Healey Ltd microfilmed approximately 14,500,000 pages of historical documents and inventory lists from the central Russian archives (Davies 1997, 101–102). In the mid-1990s, the Yale University Press launched the Annals of Communism book series with English translations of thousands of formerly classified documents of the Communist Party and the Soviet government. Another archival collection that attracted significant international interest in the early 1990s was the documents of the Third (Communist) International, or Comintern, stored in RGASPI (*Rossijskij gosudarstvennyj arhiv social'no-političeskoj istorii*, Russian State Archive of Social and Political History). Since the Comintern as an instrument of Soviet influence coordinated activities of Communist parties and affiliated groups in numerous countries between 1919 and 1943, its records amount to 20 million pages written in multiple languages. The enormous size of these archive collections and imperfect inventory lists made their analysis complicated, as even a preliminary research required long-term trips to Moscow.

In 1992, the Council of Europe, the International Council of Archives and Rosarhiv initiated a discussion of a large-scale project that involved leading scholars on the history of the Comintern. The discussion aimed to prepare a large-scale project to create a database describing the entire Comintern collection at RGASPI (over 200,000 documents) and to digitize the key documents from the collection (ca. one million pages out of the total number of 20 million). In June 1996, the International Council of Archives and Rosarhiv signed a framework agreement that established the International Committee for the Computerization of the Comintern Archives (INCOMKA) under the auspices of the Council of Europe. The funding for this large-scale digitization project came from numerous sponsors, including many European archives, the Library of Congress, and the Open Society Archives. By 2003, the project was completed including 1,059,354 digital images of archival documents (ca. 5% of the entire collection) and a searchable database of the entire collection with 239,602 entries (Doorn-Moisseenko 2005; Bachman 2005; Amiantov 2011).

Since the project was initiated at the dawn of the Internet era, the digital archive was initially hosted at the RGASPI without remote access, and its copies were distributed on CDs to the international project participants. Users had to be physically present at the RGASPI (a dedicated space was equipped with 17 workstations for this purpose) or in one of the partner institutions in order

to work with the archive. However, by the time of the project completion in 2003, this principle was already outdated, and the following year a joint venture company was formed by the Dutch academic publisher, IDC Publishers (purchased by Brill in 2006) and Russian corporation ELAR to provide paid online access to the dataset. This service was hosted at [www.comintern-online.com](http://www.comintern-online.com) and charged between 2000 and 7500 euro for an annual subscription depending on the buyer, with revenues shared between Rosarhiv, ELAR and IDC Publishers/Brill. The archive remained behind the paywall for the next ten years, until 2013, when it was published online as part of a new initiative *The Documents of the Soviet Era* of Rosarhiv that I will discuss later in this chapter.

The digital archive of the Comintern documents is a good example, illustrating the early stage of the digitization of Russian archives. The perestroika and the first post-Soviet decade in Russia were characterized by the state's temporary disinvestment in historical knowledge, and many national and international non-government agents sought to fill this gap proposing new epistemological models, producing new historical narratives, and publishing formerly classified documents, at first as books and then, as new information and communication technologies became increasingly widespread, as digital archives. The Open Society Institute (OSI)—a parent organization of the Open Society Archives which was one of the partner institutions of the INCOMKA project—was the most visible actor in this sphere. Between 1996 and 2000, it granted a total of \$1,500,000 United States Dollars (USD) for different projects in the archival sphere, including the first website of Rosarhiv (Okhotin 2001). Before the Russian authorities forced the OSI to quit its activities in Russia, it supported such projects as online databases of the archival collections of the Russian State Documentary Film and Photo Archive (Bajgarova et al. 2000), Gorbachev Foundation (Kolesov 2002), National Archive of the Republic of Kareliä (Kadymova and Kolesova 2003), and Perm' State Archive (Perm State Archive 2015). It was also one of the key sponsors of the society *Memorial*, a key non-government institution that publicizes knowledge about historical and contemporary state violence in Russia, that produced a database of the victims and later perpetrators of Soviet political repressions and several archives related to Soviet state violence, dissident movement, German forced labor, and oral history.

These projects had a clearly defined political agenda to educate audiences in Russia and abroad about the history of state violence, which was shared by many other international foundations that funded the digitization of archival documents in the post-Soviet countries.<sup>1</sup> The goal to ensure a broad public outreach for these digital archives meant that they were produced as free products. A similar political agenda drove German research foundations to actively support the production of digital archives related to Soviet citizens' experience of Nazi violence during the Second World War. One of the latest projects of this kind is a digital archive of oral history interviews of former Soviet prisoners-of-war and forced laborers *Ta storona* (Another Side) at [tastorona.su](http://tastorona.su). The

archive is based on a collection of interviews of the society *Memorial*, and was funded by the German Foundation Remembrance, Responsibility and Future and the Henrich Böll Foundation, as well as OSI. Currently it provides access to 167 interviews using specially designed software that optimized the presentation of interview transcripts, their commentary, audiovisual and geospatial data, and meta descriptions (Beilinson 2015).

### 20.3 RUSSIAN ARCHIVES AND COMMERCIAL CONTENT PROVIDERS

The early stage of digitization of archival collections in Russia was also characterized by another trend: the commercialization of scholarly content. Histories of state violence in Russia were a particularly attractive subject for commercial publishers, but the trend was much bigger and included many other fields. The opening of Russian archives coupled with a huge economic gap between Russia and the First World nations during the 1990s created a situation when Western publishers were able to sign lucrative digitization agreements with Russian archives and libraries that granted them exclusive rights to distribute the resulting digital collections and archives worldwide. As a rule, Western partners provided equipment and funding for digitization, and Russian partners received digital copies of their collections that they could provide to users within their computer networks (although sometimes they retained the domestic distribution rights) as well as a percentage of sales.

The Dutch company IDC Publishers mentioned above in the context of the international distribution of the Comintern Digital Archive was one of the earliest and most active players in the digitization market of Russian historical documents and publications. It started producing Russia-related collections of primary sources for commercial distribution on microfiche long before the collapse of the Union of Soviet Socialist Republics (USSR) using books and periodicals from the Slavonic Library of the National Library of Finland. As early as in 1987, it signed contracts with the Russian National Library getting a priority access to its collection of periodicals, rare books, and historical documents (Russian National Library 2004, 17). Over the next twenty years, the IDC Publishers signed similar contracts with the Library of the Russian Academy of Sciences, K.D. Ushinsky State Scientific Pedagogical Library, Russian State Archive of Literature and Art, Russian State Military Archive, Moscow State University Library, and Russian State Archive of Social and Political History. Its team used the materials from these institutions to produce digital archives of the Artek Pioneer Camp (1944–1967), early Soviet cinema (1923–1935), Russian military intelligence on Asia (1651–1917), Jewish Theater under Stalinism, as well as extensive digital collections of Russian and Soviet periodicals. The company was remarkable for its careful market analysis, which, prior to its merger with Brill, allowed it to become a large and successful commercial content provider in the field of Russian studies, including in digital

archives. However, following its merge with Brill in 2006, Russia- and Eastern Europe-related content lost its priority for the new company management. Access to the materials digitized earlier by the IDC Publishers is still sold through Brill either as an institutional subscription or as limited-term access for individual users (Doorn-Moisseeenko 2019).

A similar model was used by the Yale University Press (YUP) in a collaboration agreement with the Russian State Archive for Social and Political History to digitize Joseph Stalin's personal archive (*Fond 558*). The project was initiated in the late 2000s with the financial support of \$1,300,000 from the Mellon Foundation (Mellon Foundation 2007), and in 2011 YUP launched the first version of the Stalin Digital Archive at [www.stalindigitalarchive.com](http://www.stalindigitalarchive.com) that over time grew to include 400,000 pages of over 28,000 documents including Stalin's personal papers, his domestic and international correspondence, and books with his marginalia. Access to the digital archive was provided through institutional subscription only. Rosarhiv as the parent organization of the RGASPI, however, retained the domestic distribution rights for the digitized copy of Stalin's archive, and launched its free version for Russia-based users in 2013 as part of the portal *Documents of the Soviet Era* at [sovdoc.rusarchives.ru](http://sovdoc.rusarchives.ru) that I will discuss in the next section of the chapter.

Another example of a successful commercialization of digital archives of Russian historical documents and publications is represented by the East View Information Services, a Minnesota-based company founded in 1989 as the East View Publications, Inc. to distribute Soviet military journals for international audiences. Over the following three decades, it established itself as a major provider of digital content from Russia and other post-Soviet states as well as China, Afghanistan, Iran, and several African nations. Yet while for the latter regions its content is focused on exclusively contemporary affairs, its experience and expertise in the international distribution of Russian-language publications (one of its co-founders, Vladimir G. Frangulov, is a former research fellow at the Institute of World Economy and International Relations of the Russian Academy of Sciences) helped them diversify their business model by digitizing archives of Imperial-era, Soviet and post-Soviet periodicals and offering them as part of the subscription to their catalogue. The catalogue now includes digital archives of the key Soviet newspapers and magazines with a full-text search, which makes it particularly appealing for scholars (Lee and Frangulov 2014).

The fact that the archival revolution in Russia coincided with its transition to market economy and a subsequent crisis in the funding of state archives, on the one hand, and an astonishingly fast development of the new digitization technologies, on the other hand, opened a window of opportunity for Russian and Western non-governmental agents and for-profit companies to launch large-scale digital archive projects. In a way, they can be described as quasi-colonial projects: exploiting economic inequality between Russian and the First World countries, they treated the Russian historical experience represented in these newly built digital archives as a commodity to be sold to the First World



audiences (East View or Brill), or as a deviant experience, a lesson of what future generations had to avoid at all costs (the digital archives of the Comintern or Joseph Stalin). Yury Afanasyev, a professional historian who became an influential politician during the late-Soviet and early post-Soviet periods, characterized the unequal relationship between Russian archives and Western content providers in precisely these terms when he interpreted the aforementioned 1992 agreement between the Hoover Institution, Chadwyck-Healey Ltd, and Russian archives as a “typically colonial exercise” (Afanas’ev 1992). It is hardly surprising that, once the Russian state agencies were able to fund their own digitization projects, their political agenda became radically different.

#### 20.4 RUSSIAN FOUNDATION FOR HUMANITIES: PATCHING THE ARCHIVAL FABRIC

Throughout the 1990s and into the early 2000s, domestic funding of archival digitizing activities in Russia had significantly lagged behind the projects supported by Western funding. It was only in 1994 that the Russian government established a specialized funding agency for social sciences and humanities, the Russian Foundation for Humanities (RFH), which had supported projects in digital humanities since 1995 (Semenov 1997, 118–119). However, limited funds and high costs of digitization activities kept them relatively small-scale: in 2015, an average one-year grant in digital humanities was 500,000 rubles, or less than \$10,000 USD in that year’s exchange rates (Blinov 2012, 235). Nevertheless, the Russian Academy of Sciences and universities have actively used this funding source to digitize and make available their archival collections, and a large number of small digital archives have appeared since the late 1990s. A typical project was, for example, a digital archive of ethnographic field records, a personal archive of a famous cultural figure or scholar, a collection of folklore texts and performances, or a digital archive of historical periodical. The number of documents in these archives typically numbered in hundreds, sometimes thousands, although the Russian Foundation for Humanities also supported multi-year projects that resulted in bigger digital archives. A full list of digital archives produced, thanks to the funding of the RFH, currently numbers in over a hundred, so below I will only focus on a few cases to illustrate the goals, scope, and implementation of such projects.

One of the early digital archives produced with the funding from the RFH was an online collection of audio records and transcripts of folklore performances of the Karelian Research Center (KRC) of the Russian Academy of Sciences. With the earliest records dating back to the 1930s, the collection represents critical knowledge of traditional music and performances among the ethnic groups of Northwest Russia, including Russians, Karelians, Vepsians, Finns, Izhorians, and Sami. In 1999, the KRC received funding to produce an online catalogue and to digitize sample records from both the Open Society Institute and the Russian Foundation for Humanities. An early version of the

archive went online in early 2000 at [phonogr.krc.karelia.ru](http://phonogr.krc.karelia.ru); however, the number of actual records available online was just fourteen (Vdovicyn et al. 2000, 32–34). In 2008, the KRC received another grant from the RFH to make available transcripts of folklore texts recorded or written down by Soviet ethnographers, which resulted in the publication of over 400 original texts at [folk.krc.karelia.ru](http://folk.krc.karelia.ru). In 2012, yet another grant from the RFH allowed the center to add 70 audio records and their transcripts to the digital archive at [phonogr.krc.karelia.ru/folklor](http://phonogr.krc.karelia.ru/folklor). This amount represents a fraction of the records stored in the archive of the KRC, yet it provides valuable material for those scholars of the region and members of the general public who are interested in its traditional cultures. Similar projects have been implemented elsewhere in Russia, such as a digital archive of the folklore of the peoples of Siberia developed by the Siberian Branch of the Russian Academy of Sciences since 2014 [folk.philology.nsc.ru](http://folk.philology.nsc.ru), or a digital archive of ethnographic records from the Kaluga region of the Moscow Tchaikovsky Conservatory produced from a grant of the RFH in 2007 at [folk.rusign.com](http://folk.rusign.com).

The RFH also supported projects that sought to produce digital archives from fragmented collections, often in separate geographic locations, thus providing scholars and the general public with a single point of access to a certain periodical or a personal collection. The author of this chapter was a leading team member in one of such projects to create a digital archive of the surviving issues of the Russian imperial newspaper *Oloneckie gubernskie vedomosti* (OGV, or News of the Olonec Governorate). OGV was an official newspaper of the Olonec Governorate from 1838 to 1917; apart from official news and government decisions, it also published materials on history, economy, ethnography, and culture of Northwest Russia. By the mid-2000s, several libraries and archives in Petrozavodsk and St. Petersburg had partial collections of the newspaper, but in order to protect the fragile newspapers, most of them significantly restricted access to their collections (for example, the only categories of users who could work with the collection of the OGV issues from the library of the Karelian Research Center of the Russian Academy of Sciences were scholars with advanced degrees and graduate students). A digital archive of the newspaper simultaneously solved both problems by providing a consolidated collection of the surviving issues in free access. Given the complex logistics and a large scale of work, the RFH supported a three-year long project (2006–2008). It was implemented by a team from the Petrozavodsk State University, which digitized OGV issues from the National Archive of the Republic of Karelia, National Library of the Republic of Karelia, Academic Library of the Karelian Research Center, and Russian National Library in St. Petersburg. The result was a digital archive of initially 4670 issues of the newspaper (Golubev and Fotina 2007). After the completion of the project, it was maintained by the National Library of the Republic of Karelia that was able to digitize an additional 400 issues; as of now, the archive provides free access to 5064 issues at [ogv.karelia.ru](http://ogv.karelia.ru). Similar projects supported by the RFH include, among others, a digital archive of Ivan Bunin's personal papers at [bunin-rgali.ru](http://bunin-rgali.ru) that

consolidated the holdings of the Russian State Archive of Literature and Leeds Russian Archive Collections, and a digital archive of ethnographic records of field trips of the Moscow State University and St. Petersburg State University at [ethnoarchive.spbu.ru](http://ethnoarchive.spbu.ru). Apart from the Russian Foundation for Humanities and, since 2016, the Russian Science Foundation, similar small-scale digitization projects were supported by several federal and regional programs run by the Russian Ministry of Education and Russian Academy of Sciences.

## 20.5 THE RETURN OF THE STATE

This system of grants run by the RFH helped to bring online many small archives of Russian universities and institutes of the Academy of Sciences, but at the same time it had a priori too limited a scope to engage in a large-scale digitization of archival collections of the central and regional Russian archives. The latter work became possible due to a particular conjunction of state and business interests in Russia. In 2002, the Russian government launched a federal program called *Elektronnaâ Rossiâ* (Electronic Russia) with the goal to accelerate the use of new information and communication technologies in state administration. The program was designed for the period of 2002–2010, and while it failed to achieve the original goal to create a comprehensive online access to state services, one of its by-products was the appearance of a new segment in the Russian Information Technologies (IT) market, namely, a stable demand for commercial solutions in the sphere of e-governance. Among the companies that benefited from the new market conditions was Electronic Archive Corp. (ELAR). Established in 1992, it specialized in developing and providing of highly automated digitization technologies, and after 2002 it became the leading contractor of Russian government agencies, including Rosarhiv, in such areas as digitization of archives, development of digitization hardware and software, and digital archive management systems (Plotnikov 2014). Unlike Western commercial publishers and content providers such as Brill and EastView, ELAR uses a different business model: namely, state contracts to produce digital archives. In this model, Rosarhiv has a full authority over the selection of the content, and ELAR as the contractor does not acquire any distribution rights for the digitized content, which is published in open access. Since its profits come exclusively from the amount of completed work, its management is interested in lobbying more projects with the Russian government and its agencies, which has become an important driving force in the digitization of Russian archives.

The digital archives produced, at least partially, within this model include Rosarhiv's *Documents of the Soviet Era* and *The People's Memory* contracted by the Russian Ministry of Defense. *Documents of the Soviet Era* was launched in 2013 at [sovdoc.rusarchives.ru](http://sovdoc.rusarchives.ru) and from the very beginning comprised several earlier document collections, namely, the aforementioned digital archives of the Comintern and of Joseph Stalin. According to the agreement with the Yale University Press, the website restricts access to Stalin's archive from

non-Russian and non-Belorussian Internet protocol (IP) addresses. At the same time, several document collections were digitized exclusively for this project, including 30,000 electronic copies of documents of the *Politburo* for 1919–1932 (May 2013), 240,000 electronic copies of documents of the Soviet State Defense Committee (June 2015), 122,000 copies of documents about the Russian Revolution of 1917 (December 2017), and several smaller collections. In his interviews, the current head of Rosarhiv Andrej Artizov repeatedly emphasized that the purpose of the digital archive was to educate national audiences about the complexity of the Soviet period of Russian history (Artizov 2018). In doing so, the management of Rosarhiv—which in these contexts represents the historical profession as such—asserts its authority in the production of knowledge about Russian history, thanks to the seemingly comprehensive nature of their digital archive as well as vast possibilities to enlarge it by digitizing additional materials when necessary. Rosarhiv has used this positionality in order to advance the state’s agenda in such questions as the legitimacy of Russia’s annexation of Crimea, the treat of Ukrainian nationalism, and the complacency of Britain and France in the rise of Nazism in Europe, with the publication of digital copies of historical documents serving as a technique to challenge the widespread accusations of illegitimacy of Russia’s claims for Crimea, the Ukrainian claims that the Holodomor was an intentional act of genocide against the Ukrainians, and of the critical role of the Molotov-Ribbentrop Pact in provoking the outbreak of the Second World War (Artizov 2014, 7–10; Brandenberger 2015, 202–203).

The management of Rosarhiv places a special emphasis on its comprehensive approach to the digital publication of archival documents: instead of hand-picking sources to highlight certain aspects of the Soviet historical experience, Rosarhiv chose to make available entire document collections. In the logic of their creators, this approach should produce a more credible picture of the historical past than the partial digital archives of Vladimir Bukovskij ([bukovskiy-archives.net](http://bukovskiy-archives.net)), Aleksandr Yakovlev ([www.alexanderyakovlev.org](http://www.alexanderyakovlev.org)), and Dmitrij Volkogonov’s collection in the Library of Congress (National Security Archive 2017). Their authors, who at one or another point in the late 1980s and 1990s acquired access to classified collections of Russian archives, understandably focused on more sensational and controversial documents (Bukovskij 1996, 51–63). Yet, while providing authentic copies of important historical documents, these digital collections fail to present the totality of Soviet decision-making at the top level that the *Documents of the Soviet Era* can claim.

Needless to say, this claim of totality and objectivity disguises the epistemic and political foundations of the actual archival collections in the possession of Rosarhiv (Rosenberg 2001, 82–84). Moreover, while its management and staff are concerned with the questions of epistemological sovereignty, their understanding of how to “decolonize” Russian history is framed primarily in terms of data management with archives performing the function of a mediator, but also a censor, between historical documents and professional historians. This logic is based on the understanding of historical documents deposited in

archives as the most authentic and epistemologically reliable evidence about the Russian past, which is why, for example, Artizov argues that the production of digital archives such as the *Documents of the Soviet Era* is the best strategy to refute “false” interpretations of Russian history (Artizov 2014, 7). What is missing in this understanding is that the documents themselves do not provide new knowledge, but instead replicate the same interpretive categories that were laid in the foundation of original archival collections with their explicit and hidden hierarchies, silences, gaps, and exclusions.

The corporation ELAR played a key role in the development of another digital archive called *The People’s Memory* at [pamyat-naroda.ru](http://pamyat-naroda.ru) that combines archival data and copies of original documents related to Soviet servicemen during the Second World War. The archive is based on the collections stored in the Central Archive of the Russian Ministry of Defense, and currently includes digital copies of 425,000 original documents. Its structure is much more complex than that of the *Documents of the Soviet Era*: in addition to presenting electronic copies of historical documents, the developers of the archive extract biographical data from them and combine personal information from various sources into coherent biographies. The archive also provides access to ca. 100,000 digitized military maps to trace the movements of Soviet detachments and servicemen during the war, and a rich collection of original battle reports. Currently, the development team collaborates with German archives to add to *The People’s Memory* information about Soviet prisoners-of-war. The officially postulated goal of this archive is, like in the case of the *Documents of the Soviet Era*, to build up a critical mass of documentary evidence so that an unbiased and objective understanding of the Soviet Union’s participation in the Second World War would appear as a result of free access to these documents. At the same time, apart from epistemic concerns, this project also has an important humanitarian mission: to help the relatives of the Soviet servicemen to find relevant biographic information about their lives and deaths.

## 20.6 VERNACULAR ARCHIVES

In recent years, thanks to the ongoing digital revolution that makes digitization technologies increasingly cheap and easy to use, this discrepancy was addressed when a new phenomenon appeared that can be characterized as vernacular digital archives: namely, projects created, maintained, and supported by volunteers who are driven by a desire to preserve those aspects of the Russian historical experience that have been neglected due to a lack of funding or interest by the state archives. Two prominent projects include a digital archive of Soviet-era radiobroadcasts *Audiopedia* at [audiopedia.su](http://audiopedia.su) and a digital archive of diaries *Prozhito* at [prozhito.org](http://prozhito.org) (for more, see Chap. 21). *Audiopedia* grew from an earlier project *Staroe Radio* (Old Radio, [staroeradio.ru](http://staroeradio.ru)), when in 2007 Yury Metelkin, a former Soviet rock-musician, launched an online radio station that broadcast Soviet-era programs, including audiobooks, audio-plays, science broadcasts, and so on. Over the following years, the

archive of *Staroe Radio* grew through the efforts of volunteers who digitized their old records and donated them to Metelkin; in 2018, it acquired an entire archive of the Irkutsk radio station that includes federal and local records from the 1940s on, or ca. 80,000 phonograms. *Prozbito* is a later project that dates back to 2015 when its founder, Mikhail Melnichenko, decided to produce a historical corpus of texts that could be used to trace the use of certain concepts. The project employs a large group of volunteers to identify, scan, process, and upload diary entries (Nordvik 2018, 43–44); as of early 2019, it provides access to nearly 3000 diaries, mainly from the Soviet era.

Both *Audiopedia* and *Prozbito* are non-commercial and non-government projects, and as such they show a large potential of public and digital history in terms of preserving and communicating historical evidence. Yet both ultimately work within the same paradigm of history as a national project that the *Documents of the Soviet Era* and *The People's Memory* are part of. Metelkin explicated this logic in his explanation of why he decided to preserve, digitize, and make available old radio broadcasts, mentioning the “preservation of the [national] language, education, cultural traditions, and ultimately national self-identity”—or everything that is traditionally identified as functions of state power—as the driving factors of his project. The very fact that both projects inadvertently prioritize the experience of the Soviet educated class, that is, the people who are more likely to internalize and embody the national agenda, means that the digital technologies do not challenge the logic of the national archive but rather ensure a better and more effective communication of knowledge produced through it to national audiences.

## 20.7 CONCLUSION

The examples discussed above show that the Russian state has sought to use digital archives to firmly re-establish itself through its institutions (primarily Rosarhiv) as the main authority in the production of knowledge of Russian history, especially during the Soviet period, which remains extremely contested. Symptomatically, other post-Soviet states have used similar strategies: for example, Latvia whose government has persistently interpreted the period between 1940 and 1991 as an experience of double (Soviet and Nazi) occupation has recently made accessible documents of the Latvian office of the KGB (*Komitet gosudarstvennoj bezopasnosti*, Committee for State Security) at [kgb.arhivi.lv](http://kgb.arhivi.lv) with the names of its agents. The online publication of these documents by the State Archives of Latvia follows the same logic as the development of digital archives in Russia. The digitization of historical documents remains an expensive business that the state agencies and non-governmental organizations (NGO) use strategically to create an online presence of document collections that, from their perspective, benefit the common good, the understanding of which can vary dramatically from the national interests to global human rights to objective knowledge.



The archival revolution of the early 1990s and the fact that it was used by a number of international actors (Library of Congress, Hoover Institution, IDC Publishers, etc.) to reproduce valuable collections of historical sources put Russian government bodies such as Rosarhiv and the Ministry of Defense in a situation where, despite their full control of the most important archival collections from Russian and Soviet history, they had to take proactive steps to act as the authoritative sources of critical evidence about Russian history. They responded to this challenge with ambitious digitization programs that provided an unprecedented level of access to primary sources in Russian history. At the same time, the political and epistemic concerns that drove the production of these archives forced their patrons and developers to concentrate on a very narrow list of topics that is limited to political and military history. Grants for digital projects provided by the RFH addressed a broader number of themes in social, cultural, and intellectual history, but in a very fragmentary manner due to limited funding.

The digital reproduction of historical documents and their communication online thus performs largely the same functions as the traditional archive, namely, maintaining state sovereignty over history, reinforcing silences in dominant historical narratives, and endowing certain groups of experts with the authority to define the authenticity and validity of selected facts and sources. Even though such a phenomenon as the vernacular archive has become increasingly prominent in recent years, it has yet to challenge this situation, since the developers of these digital archives have so far followed the preexisting structures and hierarchies of knowledge that prioritize the historical experience of privileged political and social groups.

## NOTE

1. For example, during 2006–2007, the author was an investigator of the project *Missing in Karelia: Canadian Victims of Stalin's Purges* funded by the Canadian Social Sciences and Humanities Research Council (principle investigator Prof. Varpu Lindström of York University). The project resulted in a database of Finnish-Canadian and Finnish-American immigrants to the Soviet Union that compiled information from several thousand of archival documents from the National Archive of the Republic of Karelia (<http://missinginkarelia.ca/>, accessed December 13, 2013); after the domain name expired in December 2013, the digital archive was deposited by the National Archives of Finland and the National Archive of the Republic of Karelia (currently unavailable online).

## REFERENCES

- Afanasyev, Yury. 1992. Arhivnaâ 'Berëzka' [Archives as a Hard Currency Shop]. *Komsomol'skaâ pravda*, May 23.
- Amiantov, Ju.N. 2011. Dokumenty sovetskoi èpohi. Istoriâ i sovremennost'. K 90-letiu Rossijskogo gosudarstvennogo arhiva social'no-političeskoi istorii [Documents of



- the Soviet Era. History and the Present. To the 90-Year Anniversary of the Russian State Archive of Socio-Political History]. *Vestnik arhivista* 1: 192–198.
- Archival Committee of St. Petersburg. 2018. Mif #1: Nado ocifrovat' i vyložit' v Internet vse dokumenty [Myth No. 1: All Documents Must be Digitized and Published Online]. *Archives of St. Petersburg*, a Facebook Group. <https://www.facebook.com/spbarchives/posts/1126231160858846>.
- Artizov, A.N. 2014. Obščestvennâ missiâ rossijskih arhivov [The Social Mission of Russian Archives]. *Otečestvennye arhivy* 5: 3–11.
- . 2018. Â blagodaren arhivistam za ih samootveržennyj trud [I Am Grateful to Archivists for Their Selfless Work]. *Otečestvennye arhivy* 1: 3–6.
- Bachman, Ronald D. 2005. The Comintern Archives Database: Bringing the Archives to Scholars. *Slavic & East European Information Resources* 6 (2–3): 23–36.
- Bajgarova, N.S., Ju. A. Buhštab, and N.N. Evteeva. 2000. Sovremennâ tehnologiâ soderžatel'nogo poiska v ėlektronnyh kollekcijah izobraženij [Current Technologies of Content Search in Digital Collections of Images]. *Meždunarodnââ konferenciâ EVA 2000 Moskva: Ėlektronnye izobraženia i vizual'nye iskusstva* [International Conference EVA 2000. Moscow: Digital Images and the Arts]. October 30–November 3, 2000. <http://www.artinfo.ru/eva/EVA2000M/eva-papers/200008/Baigarova-R.htm>.
- Beilinson, Daniel. 2015. Publish Interactive Historical Documents with Archivist. *Medium*, September 14. [https://medium.com/@\\_daniel/publish-interactive-historical-documents-with-archivist-7019f6408ee6](https://medium.com/@_daniel/publish-interactive-historical-documents-with-archivist-7019f6408ee6).
- Blinov, A.N. 2012. Rossijskij gumanitarnyj naučnyj fond i sociogumanitarnye issledovaniâ v sovremennoj Rossii [Russian Foundation for Humanities and Social Sciences and Humanities in Russia]. *Vestnik MGIMO-Universiteta* 2 (23): 231–236.
- Brandenberger, David. 2015. Promotion of a Usable Past: Official Efforts to Rewrite Russo-Soviet History, 2000–2014. In *Remembrance, History, and Justice: Coming to Terms with Traumatic Pasts in Democratic Societies*, ed. Vladimir Tismaneanu et al., 191–212. Budapest: Central European University Press.
- Bukovskij, Vladimir. 1996. *Moskovskij process* [The Moscow Hearings]. Moscow: MIK.
- Davies, R.W. 1997. *Soviet History in the Yeltsin Era*. London: Palgrave Macmillan.
- Derrida, Jacques. 1996. *Archive Fever: A Freudian Impression*. Translated by Eric Prenowitz. Chicago: University of Chicago Press.
- Doorn-Moisseeenko, Tatiana. 2005. The Comintern Archives Online. In *Virtual Slavica: Digital Libraries. Digital Archives*, ed. Michael Neubert, 37–44. Binghamton, NY: Haworth Information Press.
- Doorn-Moisseeenko, Tatyana. 2019. Interview by Alexey Golubev. By Skype. Houston, TX, March 2.
- Druzin, M.V. 2018. Otečestvennye arhivy i pol'zovateli: problemy vzaimootnošenij v načale XXI veka [Russian Archives and Their Users: Communication Problems at the Beginning of the 21st Century]. *Istoricheskij kur'er* 2: 13. <http://istkurier.ru/data/2018/ISTKURIER-2018-2-13.pdf>.
- Foucault, Michel. 1977. Nietzsche, Genealogy, History. In *Language, Counter-Memory, Practice*, ed. Michel Foucault, 139–164. Ithaca, NY: Cornell University Press.
- . 2002. *Archaeology of Knowledge*. London: Routledge.
- Galanichev, A.V. 2016. K arhivam primenim zakon 'O zašite prav potrebitelej' [The Consumer Protection Law is Applicable to Archives]. *Istoričeskaâ ėkspertiza* 3: 295–301.

- Golubev, A.V., and Je.V. Fotina. 2007. Polnotekstovaa informacionno-poiskovaa sistema 'Oloneckie gubernskie vedomosti' [Full-Text Storage and Retrieval System 'News of the Olonets Governorate']. *Innovacii v usloviakh razvitiia informacionno-kommunikacionnyh tehnologij*, 354–355. Conference Proceedings, Sochi, October 1–10. Moscow: MIEM.
- Kadymova, O.P., and V.G. Kolesova. 2003. Evropejskie proekty IST v oblasti arhivnogo dela i rossijskie arhivy v seti Internet [European IT Projects in Digital Archives and Russian Archives on the Web]. *Cultivate-Russia*, 2, February. <https://www.itweek.ru/idea/article/detail.php?ID=60636>.
- Kolesov, Andrej. 2002. Avtomatizirovannaâ sistema Gorbačev-Fonda [Software of the Gorbachev Foundation]. *itWeek*, 323, May 2002. <https://www.itweek.ru/idea/article/detail.php?ID=60636>.
- Kurilova, Anastasija. 2019. Den'gi sdadut v arhiv: VS otklonil isk k Rosarhivu ob otmene platy za kopirovanie dokumentov [Cash to be Deposited in the Archive: The Supreme Court Declined a Case against Rosarhiv to Stop Charging for Document Copying]. *Kommersant*, November 22, 2018. <https://www.kommersant.ru/doc/3807340>.
- Lee, Kent D., and Vladimir G. Frangulov. 2014. East View Information Services Celebrates 25 Years in Business. *Slavic & East European Information Resources* 15 (4): 247–254.
- Mellon Foundation. 2007. Digital Stalin Archive [Rosarhiv]. *Grant Database*, <https://mellon.org/grants/grants-database/grants/yale-university/40700616/>.
- National Security Archive. 2017. Kolekcija Volkogonova. <https://nsarchive2.gwu.edu/rus/Volkogonov.html>.
- Nordvik, Vladimir. 2018. Èto 'prozhito' ne zrà! [This 'Prozhito' is Valuable!]. *Rodina* 8: 42–53.
- Okhotin, N.G. 2001. Institut 'Otkrytoe obšestvo (Fond Sorosa) v podderžku rossijskih arhivov (1996–2000) [Open Society Institute (Soros Foundation) and Its Support of Russian Archives]. *Vestnik arhivista* 2 (2001): 89–92.
- Perm State Archive. 2015. Istoriâ arhiva [History of the Archive]. <http://www.archive.perm.ru/about/history.php>.
- Plotnikov, Pavel. 2014. Sodejstvovat' informacionnomu progressu obšestva i gosudarstva [To Contribute to the IT Progress of the Society and State]. *Innovacii* 2 (184): 102–106.
- Raleigh, Donald J. 2002. Doing Soviet History: The Impact of the Archival Revolution. *The Russian Review* 61, no. 1 (January): 16–24.
- Rosenberg, William. 2001. Politics in the (Russian) Archives: The 'Objectivity Question,' Trust, and the Limitations of Law. *The American Archivist* 64 (1): 78–95.
- Russian National Library. 2004. *Fevral' 2004 g. Hronika* [February 2004: News Bulletin]. St.Petersburg: RNL. <http://nlr.ru/news/rnbinfo/2004/2-3.pdf>.
- Semenov, Evgeny V. 1997. Ponât' učénogo [To Understand a Scholar]. *Vestnik Rossijskoi akademii nauk* 67 (2): 118–124.
- Shteynman, Maria. 2016. Political Myth and Political Glory: Shaping Media Reality. *Russian Sociological Review* 15 (4): 96–113.
- Stoler, Ann. 2006. *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton: Princeton University Press.

Vdovicyn, V.T., V.P. Kuznecova, A.A. Bedorev, N.B. Lugovaja, S.M. Rusakov, and A.D. Sorokin. 2000. Sozdanie èlektronnoj versii arhiva fol'klornoj fonoteki IÁLI KarNC RAN [Development of the Digital Archive of the Folklore Phonogram Collection of the Institute of Language, Literature and History of the Karelian Research Center of the Russian Academy of Sciences]. *Vtoraá Vserossijskaá naučnaá konferenciá "Èlektronnye biblioteki: perspektivnye metody i tehnologii, èlektronnye kollekcii,"* 32–38. Conference Proceedings, Protvino, September 26–28.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Affordances of Digital Archives: The Case of the *Prozhito* Archive of Personal Diaries

*Ekaterina Kalinina*

## 21.1 INTRODUCTION

Despite being the guardians of public records Russian archives not always grant the right of public accessibility (TASS 2018; Rambler 2017; Komissiâ 2016; Slobodenyuk 2019). Not being able to get access to archival materials, professional historians and amateurs alike have to look for alternative ways of getting hold of historical sources, and in such a situation digital databases can become their salvation (Venyavkin 2017).

Today, one can get access to a wide range of historical documents online: project *Ustnaâ istoriâ* (Oral history) makes recorded talks with famous intellectuals accessible through a web platform (<http://oralhistory.ru>); project *Prozhito* digitizes and publishes personal diaries (<http://prozhito.org>); database *Otkrytyj spisok* (Open list) makes it possible to find information about persons who fell victim to the Soviet repression machine (<http://ru.openlist.wiki>); photo databases such as *Pastvu* (<http://pastvu.com>) and *Istoriâ Rossii v fotografiâh* (History of Russia in photographs, <http://russianphoto.ru>) grant access to a wide range of photo materials.

These and other similar initiatives carry a promise of wider access to historical sources. Scholars even stress that digital archives form “public alternatives to official constructions of the past and ways in which that past is to be studied” (Lapina-Kratasyuk and Rubleva 2018, 164; Garde-Hansen et al. 2009). Alexandra Herlitz and Jonathan Westin (2018, 451) believe that “[c]onflating

---

E. Kalinina (✉)  
Jönköping University, Jönköping, Sweden  
e-mail: [ekaterina.kalinina@ju.se](mailto:ekaterina.kalinina@ju.se)

different archives digitally makes the archival substance less vulnerable to any agenda of an archiving individual and may therefore lead to a greater objectivity and reliability of the accumulated material.” In Russia such digital platforms become essential players in political, social and cultural life by allowing people to learn about the past from other sources rather than those that are state approved. This means that digital archives have the potential to not only challenge established historical discourses but also question the leading role of the state in the production of historical narratives.

This idea about the democratizing potential of digital archives is rooted in the belief that digital technologies are the solution to many problems. Recent studies of digital media, however, show that scholars should be more careful when stressing the democratic potential of the web, as internet is not necessarily democracy’s magical solution (Fuchs 2017; Morozov 2011; Nisbet et al. 2012; Rød and Weidmann 2015; Stoycheff et al. 2016). When archival collections become digital, various legal constraints and the fragmentary nature of archival records might lead to the reproduction of already existing biases. So, while some scholars (Herlitz and Westin 2018) see digital archiving as a possibility to safeguard and disseminate data to wider publics, others (Aghostino 2016; Azoulay 2012) insist on a more critical approach to digital archives by arguing that the non-neutral and ideological nature of digital infrastructures predetermines what type of content becomes visible and how much visibility it gets. Hence, in order to study the potential of digital archives for the democratization of history, one should look at what users can and cannot do in order to create new narratives that can disrupt dominant discourses of power.

Therefore, the aim of this chapter is to study affordances of digital archives, that is, the properties that allow users to perform certain actions on the platforms, in order to explore their democratic potential. To be able to do that, affordance analysis, which allows one to investigate the technological and organizational structures of digital platforms, the amount and quality of data available, the social underpinnings entangled in technologies and the degree of user participation, is applied to the Russian digital archive of diaries *Prozbito*.

The research questions that guide this chapter are: what kind of data is (not) available in the archive and why? What information about affordances is available and what does it tell us about the composition, constraints, limitations and affordances of the archive? How much participation does the archive allow? What kind of participation does the archive support? These questions are addressed by focusing on *Prozbito* as an environment that allows certain actions and forms of participation, rather than on specific uses of *Prozbito*. This means that this study is not a reception study but rather a research of digital archives as media environments.

In order to address the above-mentioned scholarly interest, the chapter starts with a section outlining the theoretical framework and methodological guidelines, followed by the analysis of *Prozbito* and a brief conclusion summarizing major findings.

## 21.2 THEORETICAL TOOLS FOR UNPACKING ARCHIVES

Some scholars believe that archives play an important role in the formation of national consciousness and in the development of democratic liberal citizenship, because they preserve evidence which is paramount for the work of justice and crucial for keeping individuals, groups and institutions accountable (Joyce 1999). Jacques Derrida, for example, claims the centrality of the archive to the existence of democratic society by stating that “[e]ffective democratization can always be measured by this essential criterion: the participation in and access to the archive, its constitution, and its interpretation” (Derrida 1995, 4).

Scholars also point out that archives are hardly neutral collections of records but rather environments where both technical and organizational structures produce as much as record events (Derrida 1995, 17) and organize data in a way that might “lead later investigators in a particular direction” (Manoff 2004, 16). Gillian Rose (2012, 228) argues that archives “have effects on what is stored within them” and on “those who use them.” Jaimie Baron (2014, 109ff) calls this process *archive effect* and explains it as a human response toward archival material, which can be triggered by the engagement of a person with archival documents and even purposefully employed to activate the public.

As archives are environments with specific technological and organizational structures the users come in contact with, the concept of affordances could be applied in order to study archival properties and what users can and cannot do with them.

The term “affordance” was first coined by psychologist James J. Gibson in his seminal book *The Ecological Approach to Visual Perception* (1979) to describe the properties of the environment. Gibson explains affordances as interconnected action possibilities offered by the environment to the subject. He says: “an affordance is not bestowed upon an object by a need of an observer and his act of perceiving it. The object offers what it does because of what it is” (Gibson 1979, 139). In other words, an affordance is the possibility of an action available in the environment that is independent of the subject’s ability to perceive this possibility. At the same time an affordance exists relative to action capabilities of the subject. This means that an affordance of an environment might exist, but it only can be activated if the subject has the capacity for an action.

At the same time what a subject perceives about affordances depends much on the information he/she has about them. Depending on the presence or absence of affordances, and the presence and absence of information about them, one can divide affordances into the following types (Gaver 1991): *perceptible* affordances (both information and action exist), *false* affordances (information about the affordance might exist but the affordance itself is absent), *hidden* affordances (affordance exists but information is hidden), and *correct rejection* (absence of information and affordance) (Gaver 1991).

Gibson (1979) also suggests two types of affordances: *positive* and *negative*, where positive affordances stand for properties that allow certain actions and negative affordances for properties that do not allow certain actions. They can also be called *platform constraints*. In this text the term “platform constraints” rather than “negative affordances” will be used. Constraints will be divided into technological, legal and ethical constraints in order to cover the whole spectrum of systematic issues.

One has to keep in mind that environments can be characterized by multiple affordances that exist in relation to each other. The notion of *nesting* helps to conceptualize these relationships between different affordances of environments by suggesting some sort of hierarchy between them. Turner (2005) suggests dividing affordances into *simple* affordances, that is, usability of the environment or objects, and *complex* affordances, that is, properties of the environment that have an important cultural or historical significance when being used. Meanwhile, Wagman et al. (2016) suggest calling them *subordinate* and *superordinate* affordances, pointing to the connections between affordances of different levels: affordances at lower levels have means which allow higher level affordances to come into existence.

In order to trace the relationships between different levels of affordances, Gaver (1991, 82) suggests dividing affordances into *sequential* and *nested*, where the former emerge as a result of actions on perceptible affordances (Gaver 1991, 82) while the latter refer to affordances that serve as a context for other affordances. The best example for sequential affordances can be the functions that a user learns about when logged in as an editor or a page administrator. In such case, each affordance that emerges after logging in would be a sequential affordance. Nested affordances are actualized in temporal sequences as well, “yet time is not the sole basis for the nesting of affordances” as “nesting can also exist across levels that differ in order” (Wagman et al. 2016, 2).

Studying affordances is essential for understanding archival composition, which is in turn pivotal for comprehending why some documents are included while others are excluded or censored. Archival composition reflects biases of an archivist and/or donor, providing substantial information about societal, political and cultural contexts of the archives and in turn can shed light on degrees of participation in a given society.

Two key principles of archival practices—provenance and authenticity—help to unpack the composition of archives. In archival studies, while “provenance refers to the documentation of the origins and history of an archived item, authenticity denotes the preservation of the original object rather than the truth or accuracy of its content” (Kallinikos et al. 2013, 361). Both principles are important when it comes to the discussion of an archive’s possibility to evoke multi-vocal narratives. As historian Jane Stevenson (2013, 160; 170) puts it, archives hardly ever store enough material to fully reconstruct the whole life of a person; therefore, it is crucial to know the origins of the available documents as well as their history to be able to construct narratives that would give the fullest picture of the subject in question.



The principles of provenance and authenticity are challenged in the digital environment because it is difficult to trace the origins of the documents (Marton 2010). This difficulty arises when one gets into contact with digital copies, which usually have neither traces of materiality nor records about their inquisition. In other words, this difficulty is a result of the technological and organizational constraints (negative affordances) of the modern digital platforms that collect and make available historical documents.

### 21.3 METHODOLOGICAL FRAMEWORK AND STEPS TO UNPACK *PROZHITO* ARCHIVE

In order to learn what affordances an environment can offer, one has to engage with this environment. Activity theory, for example, postulates that environments can only be perceived through acting (Albrechtsen 2001). When acted upon, an environment reveals its hidden structures and its constraints as well as what are called *perceived* and *actual affordances*, where the first type stands for what a subject thinks he/she can do and the second for what a subject can actually do with/in the given environment (Norman 1988). Hence, in order to learn about affordances of digital archives, one has to try them out.

Taking off from the activity theory, Turner's simple and complex affordances (2005) and Gibson's (1979) positive affordances and constraints, the project investigates what actions are allowed and not allowed on *Prozhito*. First, the user interface was browsed to collect as much information as possible about the perceived and simple affordances of the platform. Second, the platform was tested in order to create a historical narrative. Preparing for the publication of a special issue of the journal *Baltic Worlds* dedicated to the Centenary of the Russian Revolution, the author of the chapter has used *Prozhito* to tell a story about the Russian Revolution from the perspective of its contemporaries. In order to do that diaries tagged with the year 1917 were selected for the analysis and publication (Kalinina and Kochergan 2018). To be able to speak about volunteer experiences, the author joined the community of one of the laboratories as well as followed some of the laboratories online. The choice of the type of participation was defined by the author's experiences and capabilities. Next, in order to learn more about the archival composition of *Prozhito*, the author interviewed the founders of the project Mikhail Melnichenko (2017) and Ilya Venyavkin (2017) as the information on the website was not sufficient despite a very encompassing description in the section "About the project" (*O proekte*—in Russian).

### 21.4 UNPACKING *PROZHITO*

*Prozhito* was founded in 2015 by a professional historian, Melnichenko, and his colleagues in order to collect and make available already published diaries and manuscripts. In 2019 *Prozhito* received the status of Research Institute of

Ego-documents at the European University in St. Petersburg, Russia, with a whole range of responsibilities, such as the collection and research of ego-documents and the organization of events and laboratories. In September 2019, *Prozhito* was accepted into the European Ego-Documents Archives and Collections Network (<http://prozhito.org>) and became recognized by other memory institutions as a legible member (Fig. 21.1).

At the moment of writing, around 1700 diaries are uploaded into the system, with 350 published for the first time. In total, the archive contains 460,000 daily entries. The diaries available in the archive fall into the following categories: (1) transcribed manuscripts found in the archives; (2) transcribed manuscripts donated by the authors or their relatives; (3) digitized published diaries and (4) published diaries available online (Interview with Melnichenko 2017).

As a historical source, diaries have their specificity. A diary is usually a notebook filled with handwritten notes arranged by date, which makes it an important historical source that allows one to date events. Entries reporting on everyday occurrences, reflections, emotional experiences and impressions are usually written down for the author's own use, and not with the intention of being published (Fig. 21.2).

Nevertheless, many authors are aware that their diaries can be read by others and even edit them to fit the public eye. While some intentionally write for profit or self-vindication with a possible reading public in mind, others develop secret coding systems in order to conceal information from the eyes of unwanted readers (Interview with Melnichenko 2017). Compared with memoirs diaries contain less “strategic lies”—intentional misrepresentations of events done with the aim of creating a more favorable self-representation (Interview with Melnichenko 2017). Still, using diaries as a historical source

The screenshot displays the Prozhito website interface. At the top right, there are social media icons for Facebook, YouTube, and VK, along with language options for Russian (РУС) and English (ENG), and a 'ВХОД' (Login) button. The main header features the Prozhito logo and navigation links for 'Новости' (News), 'Дневники' (Diaries), and 'О проекте' (About the project), plus a 'ХОТИТЕ ПОМОЧЬ?' (Do you want to help?) button. The main content area is titled 'Дневники и записи' (Diaries and notes) and includes a search filter section with 'ПАРАМЕТРЫ ПОИСКА' (Search parameters) and a date range of '01 ЯНВАРЯ 1937 - 31 ДЕКАБРЯ 1949'. Below this, a search result for the year 1937 is shown, indicating '10000 ЗАПИСЕЙ' (10,000 entries) and listing years (1853, 1859, 1937, 1938, 1939). A specific entry for January 1st, 1937, is displayed, featuring a portrait of Lazar Bronfman and the text: '1 января. Встречали Новый Год у Прокофьева. Было человек 20. Напились изрядно. Закончили в 8 утра. Перед отходом: — Георгий, так я в этом году. — Обязательно! Был Мехлис.' (January 1st. We celebrated New Year at Prokofiev's. There were about 20 people. We got quite drunk. Finished at 8 AM. Before going to bed: — Georgy, I'm in this year. — Definitely! There was Meksis.)

Fig. 21.1 *Prozhito*. User interface

**Абдрахманов Юсуп**  
**Абдрахманович**  
 28 декабря 1901 - 5 ноября 1938

Первый Председатель Совета Народных Комиссаров Киргизской АССР.  
 Репрессирован и расстрелян.

Статья на Википедии

Пол: Male  
 LANGUAGES: Russian

Статус: загружен

**Дневники & Записи**

Год	Записи (177)
1928	57
1930	27
1931	77
1932	16

Теги в записях

**Издание**

Издание: *Абдрахманов Ю. 1916. Дневники. Письма к Сталину / Вступит. ст. Дж. Дзугушалиев, И. Е. Семенов. Ф.: Киргизстан, 1991. 320 с.*

Подготовка текста к загрузке: Евгений Ржевский, Миша Мельниченко.

Текст взят из оригинальной публикации и подготовлен к загрузке на сайт волонтером. Участники «Прожито» не работали с рукописью. При необходимости цитирования — обратитесь к первопубликации или оригиналу дневника. О достоверности дневникового текста, соответствии оригинального текста опубликованному, ошибках и допущениях, возможных при загрузке дневника в «Прожито» см. «[О проекте](#)».

Fig. 21.2 *Prozhito*. Author page

requires critical reflection on the part of the reader. One has to understand that the spectrum of motivations of the author, who keeps a diary, can be much wider than a simple desire for impartial recording of events and therefore one needs to consult other sources to verify the provided information.

According to the information available on the website, “absolutely all diaries present interest to the project” ([prozhito.org](http://prozhito.org)), be it the diaries of famous persons or average citizens. Melnichenko says that diaries of average citizens are even more interesting as they rarely become available for the public eye but contain experiences and emotions anyone can relate to (Interview with Melnichenko 2017).

Organizationally *Prozhito* consists of a core team and a community of volunteers. The core team is built from people who are responsible for the overall concept of the project, the coordination of volunteers and laboratories, informational support, the development of specific projects, website development, support and editing ([prozhito.org](http://prozhito.org)).

As the creation of an archive is a very time- and resource-consuming endeavor, *Prozhito* (being a non-commercial project) actively engages volunteers who collect, index, upload and tag diaries in the system. The core team defines tasks for the volunteers, which are communicated in authors’ directories, where next to the name of the author’s diary there is a mark signaling what kind of work could be performed by volunteers: text search, proof reading, editing or indexing (Fig. 21.3).

This means that volunteers can decide themselves which diary they want to work with depending on their personal interests and capabilities. The list of tasks for volunteers can be found on the Help page (*Помощь*—in Russian). Reading this section reveals that while certain actions can be performed without any supervision (preparation of the text for being uploaded), some actions

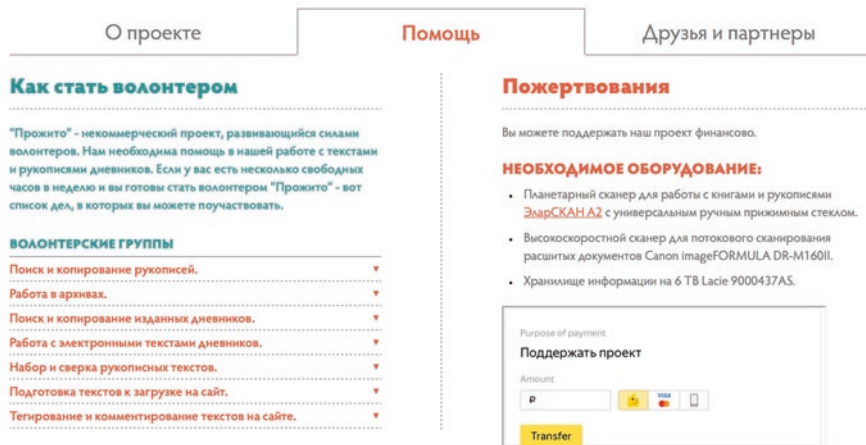


Fig. 21.3 *Prozhito*. The page *Pomosh'* (Help), where volunteers can learn how they can contribute to the project

demand additional coaching (such as proofreading and final editing) and require special access to the corpus of the texts (such as tagging).

Apart from collecting and making digital copies of diaries available online, *Prozhito* organizes special workshops called laboratories, where volunteers meet to transcribe and discuss manuscripts. These sessions usually last two or three hours, with curators providing historical context and some background information about the authors of the manuscripts. The goal of such labs is to work collaboratively on archival sources and by doing so sustain and educate the vibrant volunteer community. Usually such laboratories are organized at the GULAG (*Glavnoe upravlenie lagerej i mest zaklúcheniá*, Main Administration of Camps) museum in Moscow. Starting from autumn 2019 regular laboratories are also organized at European University in St. Petersburg. *Prozhito* also arranges workshops in other cities in Russia, but on an irregular basis.

By 2019 the project developed new technological and organizational solutions to improve the user experience. Nevertheless, certain constraints of technological and legal nature set boundaries on what becomes available in this archive. Therefore, in the following sections the technological and legal constraints are reviewed to understand what users can do on the platform.

## 21.5 THROUGH A GLASS DARKLY: FROM SIMPLE AFFORDANCE TO TECHNOLOGICAL CONSTRAINTS

*Prozhito* has a number of simple and perceived affordances that become evident when a user opens the archive's home page. *Prozhito's* interface suggests that a user, by clicking on action buttons, can search the data for (1) names of the author or individuals mentioned in the entries; (2) date of the entry; (3) key

words and tags. For example, if a user is interested in what happened on January 29, 1917, he/she can type the date into the search field to see all entries that exist in the database with this date tag. The system also suggests different filters, such as gender, age and geographical location. This means that all diaries are indexed and tagged to be searchable in the database. This feature allows users to work both with the diaries of specific authors and with the whole corpus of texts uploaded in the system. This also means that the platform has some hidden affordances that could be used for scientific research, but the information on them is not available because the developers did not want to scare “average users with much too many scientific tools” (Melnichenko 2017) (Fig. 21.4).

However, as the project is still in the making and not all texts are tagged and indexed, there is an uncertainty regarding the search output. This is a considerable constraint that might prevent users from relying on algorithmic search in the corpus and force them to do the search manually.

The fragmentary nature of the search output is not the only constraint of the archive. Altered and incomplete documents are another issue of the archive. Digitally stored information entails a set of specific problems: depending on

Избранные цитаты

**Павел Григорьевич Антокольский**

**5 апреля 1964** Суд был обставлен гнусно. Над дверьми суда висел плакат: здесь судят тунеядца. Письма Чуковского и Маршака были прочитаны вслух и не возымели никакого действия на судью и заседателей, которые вели себя вызывающе грубо, орали на подсудимого Бродского. Невозможно понять, кому и зачем требуется это издевательство над трудом писателя, т.е. над всей интеллигенцией. ...

ПЕРЕЙТИ

Fig. 21.4 *Prozhito*. A suggested page of a diary

the protocols and the types of data storage, not all types of documents or not all parts of documents can be stored or can be stored without distortion (for more on corpora, see Chap. 19). Hence, there is “potential for documents to be altered electronically in their content, provenance or through migration to new storage systems” (Glotzer 2013). In the case of *Prozbito*, the original documents when uploaded into the system lose any type of imagery, such as diagrams and tables, photographs and drawings, as the archive is primarily a textual corpus. As no images can be saved as part of the original document, the diaries lose a part of their identity, making it also difficult for a researcher to learn about the people who have written them. Meanwhile, illustrations may tell a lot about the time the diaries were written, as well as function as codes that may contain some information the authors wanted to conceal from potential readers. Even though the images can be annotated in the commentaries (the same goes for codes), working with the digital copy limits the actions of a researcher and therefore limits interpretation. Curators promise that in the nearest future they will add original images of the diaries for the users to be able to get a sense of what the manuscripts look like and therefore narrow down the gap between the original document, its digital copy and the reader (Interview with Melnichenko 2017).

There is, however, a solution to this problem, which is anchored in a sequential affordance of the platform. The curators of the project remind users that there is always a possibility to either ask for an original copy from the curators or the owners of the diary or, if the original is stored in a state archive, consult the original there. The information about the original manuscript is usually located in the annotation to the entry and contains either a bibliographical note (if the diary is published) or the address where the original can be found.

Another constraint users experience on *Prozbito* is the inability to trace how and by whom documents were written and revised. The information about authorization of some parts of the text is marked with <...>, but it does not allow tracing the history of the document, to see whether parts of a text were deleted or re-written, when exactly it happened and who did it. Such shortcomings undermine the validity of the sources and even historical narratives based upon them.

## 21.6 HIDDEN AFFORDANCES AS A RESULT OF LEGAL AND ETHICAL CONSTRAINTS

*Prozbito* informs users about various legal and ethical constraints that may restrict access to the texts and define how diaries are published. The first constraint is a direct consequence of intellectual property rights as written in the Civil Code of the Russian Federation, which protects authors' rights and the rights of owners and publishers. To abide by the law and make digital versions of already published diaries available the curators need to get permission from the publishers:



If we are lucky to find publishers, we ask for permission. If we do not have such possibility, we upload texts into the system, but keep them in such a “closed” regime that does not allow our users to read the texts, but only those entries that contain the names or the key words the users might be after. This means that these texts take part in the search, but the full access is not permitted. The search output is not yet available in the form of snippets, but in the form of the full diary entry. (Interview with Melnichenko 2017)

As it reads from the quote above, intellectual property right often determines the regime of visibility of uploaded diaries. Orphan texts, that is, texts that have no information about owners/authors, are only available in the citation regime allowed by the intellectual property legislation.

Melnichenko says that to enable the search function across as many documents as possible the curators strike agreements with publishers who allow digitization and indexing of their published books if *Prozhito* tells their readers where to buy these books. When it comes to unpublished manuscripts, the curators of the project try to get permissions from legal owners. If they cannot be found, diaries are uploaded to the database and remain there unless the owners show up and demand to take the diaries down. Such situations could be seen both as affordances and as constraints: the project curators dare to publish orphan texts and by doing so increase the number of texts participating in the search (an affordance); the possibility of removal, however, signals a constraint—a fact of censorship of an archival record.

Curators often collaborate with authors and owners to be able to publish diaries. Such practice often results in the editing of manuscripts. Melnichenko describes the process as follows:

We work in tight collaboration with the owners of manuscripts. After a text has been prepared for the upload it is given for a review to the owners, who have free hands to delete anything they consider sensitive. In other words, it means that we give the owners an opportunity to shorten the text. At the same time, we restrict users’ access to those texts we consider too personal. These texts are still indexed and if a user types a key word and this word exists in this diary, the user can see the entry with this key word, but cannot read the whole diary. (Interview with Melnichenko 2017)

Melnichenko says that all diaries published after 1942 could be heavily edited due to the above-mentioned regulations, while all texts dated after December 31, 1999, are to be made invisible for the readers in order to protect the privacy of the individuals mentioned in the texts (Interview with Melnichenko 2017).

Another constraint that defines which diaries see the light has both legal and ethical dimensions. Diaries often contain information about third persons and therefore fall under personal data protection regulations. Some texts might even contain insults and ungrounded accusations, which means that they fall under the defamation law. Melnichenko says that sometimes curators



themselves decide not to publish such texts due to personal reasons. He tells about a diary written by a woman full of frustration and hatred, which he felt reluctant to publish in order not to “let so much negative energy out into the world” (Interview with Melnichenko 2017).

The legal and ethical constraints mentioned above give the right to censor any information in the diary that can be deemed inappropriate or going against the current legislation of the Russian Federation. This means that the owners of the diaries and the curators of the project themselves can decide what information should be kept and what should be left out. These precautions taken by curators could also be seen as forced measures needed to safeguard the archive from the attacks of both Russian authorities and private individuals. As *Prozbito* makes information publicly available, it could be even considered a media source, which makes the platform vulnerable toward legislation that regulates mass media in Russia (The Law of the Russian Federation “On Mass Media” dated December 27, 1991, No. 2124-1). These regulations and laws guarding privacy of individuals and regulating sources of information can potentially hinder the construction of alternative historical narratives as information about perpetrators might never see the light, which might lead to the further silencing of the victims of the Soviet regime.

## 21.7 PARTICIPATION AS A COMPLEX AFFORDANCE

Digitization, that is, converting analogue documents into a digital format, is a time- and resource-consuming enterprise, which *Prozbito* resolves by using volunteer labor. Curators design tasks by taking into consideration people’s different skills and abilities and even allow volunteers to withdraw at any time. In the language of affordances it means that the platform allows for multiple levels of engagement and disengagement and has a number of sequential and nested affordances. These multiple affordances are related to each other in the following manner: the discussions of diaries during *Prozbito* labs are possible because the *Prozbito* volunteers found these diaries in the archives and copied them; indexing and tagging of diaries is possible because the *Prozbito* volunteers transcribed these diaries. In both cases volunteers enable multiple interconnected actions and are the key actors in the process of creation and functioning of the platform.

Some of the activities performed by volunteers are possible in the digital environment (indexing, tagging, transcribing), while some, such as work in the archives, can only be performed offline. In any case, any activities the *Prozbito* volunteers are engaged in assume direct contact with personal diaries and with their authors. It has its social and cultural benefits. First, by working with private documents, one gets an opportunity to learn more about individual experiences and reflections about historical events and as a result build some sort of solidarity with and develop compassion toward the authors of the diaries. Second, this revelatory power of personal documents, that is, archive power, has an “important social function”:

It gives a new sense of life to many people, especially old ones. Before *Prozhito*, they thought that their lives and their diaries mattered very little. After they found out about *Prozhito*, they became very busy. Now they run around and organize editorial meetings with their grandchildren and children. (Interview with Melnichenko 2017)

This family engagement in editorial work is a form of collective remembering—a representative of an older generation passes on memories to a younger generation. A family member's private story becomes the subject of a collective experience; it allows for understanding historical events from a personal perspective.

At the same time narratives that emerge from personal documents can challenge historical master narratives. For the lab session it means that a diary has to be carefully chosen for its capacity to make a certain statement. Meanwhile, the curator's work implies giving the volunteers different means both to recognize master narratives that have to be challenged and to create new narratives based on the material in the diaries that they work with during the lab.

During one of such labs, volunteers worked with a diary of Chernevsky Oleg (Chinar, born 1921, the diary volume started June 21, 1938, and finished September 10, 1938), archived by Russian historical and human rights organization Memorial. A crucial element of this diary is the presence of several layers of coding. One of the layers is the actual code, a system of signs that is used to encode a secret text that is not supposed to be understood by others than the owner of the diary himself. Another layer of coding is a lingo that is used during a specific period of time, in this case the time of the Great Terror. Decoding these layers reveals a matrix of visible and invisible aspects of Soviet life: what was allowed and what was not allowed to say out loud.

When working with diaries, volunteers get a chance to see snapshots of everyday life in the form of everyday descriptions in the diaries (Herlitz and Westin 2018, 453). By doing so they experience so-called *archival voyeurism*, the desire to see what they were not meant to see (Baron 2014). What is important is that during such labs people engage into what could be called a *collective archival voyeurism*, a process of collective reading of a diary that unites a group of volunteers together when they communicate with each other trying to decipher what is written in a manuscript. During such labs, volunteers engage in a process of meaning-making by sharing each other's fragmentary historical knowledge and attempting to guess the emotional state of the author of a diary during the period of writing. This collective experience has a bonding effect: by sitting together around a table and working on different pieces of one diary, people start feeling a connection to each other because they all secretly spy on somebody's personal life.

Hence, the role of these labs is to introduce the public into an archive through embodied (typing, transcribing) and sensual (empathizing or/and sympathizing) experiences (Chakrabarty 2013, 457). The diaries used in the labs are of fragmentary character (only some parts of the diaries are used for

work), which by nature can generate “a sense of the ‘presence’ of history” (Baron 2014, 12). The volunteers do not need to get the whole picture in its entirety presented to them. On the contrary, partial and incomplete information invites them to get engaged, to learn more. Being fragmentary, these diaries own a performative quality as they affect the volunteers and present a possibility for this audience to arise as active public through the interpretative act.

## 21.8 CONCLUSION

The aim of this chapter was to study *Prozbito*'s potential for the production of historical knowledge in Russia. After having analyzed *Prozbito*'s affordances several important aspects of the project have come forward.

First, the research has shown that affordance for democratic knowledge production is a nesting affordance, which is situated in several other simple and complex affordances, which in turn are sequential affordances. In practice it means that even though affordances exist independently of each other—for example, diaries are readable, indexable and searchable, they are also interdependent—diaries are readable because they are searchable and indexable. These simple affordances also allow for participatory practices, such as group discussions of Soviet history and collaborative work on the production of the *Prozbito* archive.

Second, sequences of these positive affordances emerge from a superordinate negative affordance—the impossibility to create the archive without the help of volunteers. As *Prozbito* does not have enough financial and technological resources, the work of collecting, preparing and editing texts is forced upon volunteers. This is a good example of how negative affordances of the environment, if handled creatively, condition positive changes and result in turn in the production of participatory infrastructures. Working on the creation of the archive gives people new meaning in life and ensures dialogue between different generations, family members and random people, who create their own networks by volunteering for *Prozbito*.

Third, archives like *Prozbito* in Russia indeed play an important political, social and cultural role by providing more democratic access to historical sources. By making alternative history narratives possible, archives mobilize communities for action, which results in independent learning and thinking.

By providing access to previously unavailable sources of information such archival initiatives also challenge state archives, which have always been central institutions for nation building and maintenance of political dominance through wielding power over the shape and direction of historical scholarship and collective memory.

However, being alternative public platforms for historical negotiations such archives become sites of conflict as they potentially might provide grounds for demanding justice. In the case of *Prozbito*, the curators put in place certain constraints to avoid potential conflicts of interests especially when it comes to

publication of diaries written during the last seventy-five years. In other words, ethical and legal constraints condition what data has to see the light and determine the composition of archive that also might have some consequences. As the archive deals with documents of personal origin, which might contain data about third parties, crucial information has to wait for some time before it can be published, hence minimizing chances for justice. Meanwhile, as information about the removal of sensitive information is not provided, it also makes it difficult to search for evidence as no traces of such evidence remain. Therefore, when celebrating the evident democratic potential of such initiatives, one must remember that it is still an archive and it is still subjected to certain politics of invisibility conditioned by various constraints.

## REFERENCES

- Aghostino, D. 2016. Big Data, Time and the Archive. *Symploke* 24 (1–2): 435–445.
- Albrechtsen, H., Andersen, Hans H.K., Bodker, S., and Pejtersen, Annelise M. 2001. Affordances in Activity Theory and Cognitive Systems Engineering. Risø National Laboratory. <http://www.risoe.dk/rispubl/SYS/syspdf/ris-r-1287.pdf>.
- Azoulay, A. 2012. Archive, Political Concepts: A Critical Lexicon. <http://www.politicalconcepts.org/archive-ariella-azoulay/>.
- Baron, J. 2014. *The Archive Effect: Found Footage and the Audiovisual Experience of History*. London: Routledge.
- Chakrabarty, D. 2013. Museums in Late Democracies. In *The Visual Culture Reader*, ed. Nicholas Mirzoeff, 3rd ed., 455–462. London: Routledge.
- Derrida, J. 1995. *Archive Fever: A Freudian Impression*. Trans. Eric Prenowitz. Chicago and London: University of Chicago Press.
- Fuchs, C. 2017. *Social Media: A Critical Introduction*. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: Sage.
- Garde-Hansen, J., A. Hoskins, and A. Reading. 2009. *Save as... Digital Memories*. New York: Springer.
- Gaver, W. 1991. Technology Affordances. In Proceedings of the ACM CHI 91 Human Factors in Computing Systems Conference, April 28–June 5, 1991, Scott P. Robertson, Gary M. Olson, and Judith S. Olson, eds., 79–84. New Orleans, Louisiana.
- Gibson, J. 1979. *The Ecological Approach to Visual Perception*. New Jersey, USA: Lawrence Erlbaum Associates.
- Glotzer, R. 2013. Archival Theory and Shaping of Educational History: Utilizing New Sources and Reinterpreting Traditional Ones. *American Educational History Journal*. <https://www.thefreelibrary.com/Archival+theory+and+the+shaping+of+educational+history%3A+utilizing+new+...-a0370031204>.
- Herlitz, J., and J. Westin. 2018. Assembling Arosenius – Staging a Digital Archive. *Museum Management and Curatorship* 33 (5): 447–466. <https://doi.org/10.1080/09647775.2018.1496847>.
- Joyce, P. 1999. The Politics of the Liberal Archive. *History of the Human Sciences* 12 (2): 35–49.
- Kak v raznyh stranah otkryvali dostup k arhivam KGB. 2018.

- Kakie dokumenty SSSR do sih por zasekrečeny. 2017.
- Kalinina, E., and A. Kochergan. 2018. The Russian Revolution in Personal Diaries. *Baltic Worlds* 11 (1): 42–50.
- Kallinikos, J., et al. 2013. The Ambivalent Ontology of Digital Artefacts. *MIS Quarterly* 27 (2): 357–370. <https://doi.org/10.25300/MISQ/2013/37.2.02>.
- Komissiâ 2016 = Komissiâ po gostajne otkazalas' rassekrečivat' arhiv NKVD. 2016. RBC, January 19, 2016. <https://www.rbc.ru/rbcfreeneews/569ea0d29a79477753125703>.
- Lapina-Kratasyuk, E., and M. Rubleva. 2018. Projects to Preserve Personal Memories: Digital Archives and Participatory Culture. *Steps* 4 (3–4): 147–163.
- Manoff, M. 2004. Theories of the Archive from Across the Disciplines. *Libraries and the Academy* 4 (1): 9–25. <https://doi.org/10.1353/pla.2004.0015>.
- Marton, A. 2010. The Transfigurability of Digital Objects. *Artnodes*, (10). <http://artnodes.uoc.edu/ojs/index.php/artnodes/article/view/n10-marton/n10-marton-eng>.
- Melnichenko, Mikhail. 2017. Interviewed by the author on March 21, 2017, in Moscow.
- Morozov, E. 2011. *The Net Delusion: The Dark Side of Internet Freedom*. New York: Penguin Books Ltd.
- Nisbet, E.C., E. Stoycheff, and K.E. Pearce. 2012. Internet Use and Democratic Demands: A Multinational, Multilevel Model of Internet Use and Citizen Attitudes About Democracy. *Journal of Communication* 62 (2): 249–265.
- Norman, D. 1988. *The Psychology of Everyday Things*. New York: Basic Books.
- Rambler 2017 = Kakie dokumenty SSSR do sih por zasekrečeny. 2017. Rambler.RU, August 29, 2017. <https://news.rambler.ru/other/37760141-kakie-dokumenty-sssr-do-sih-por-zasekrecheny/>.
- Rød, E.G., and N.B. Weidmann. 2015. Empowering Activists or Autocrats? The Internet in Authoritarian Regimes. *Journal of Peace Research* 52 (3): 338–351. <https://doi.org/10.1177/0022343314555782>.
- Rose, G. 2012. *Visual Methodologies: An Introduction to Researching with Visual Materials*. 3rd ed. London: Sage.
- Slobodenyuk, I. 2019. Dumaû, èto stanet povsemestnoj praktikoj. *Znak*, March 13, 2019. [https://www.znak.com/2019-03-13/gosarhiv\\_novosibirskoy\\_oblasti\\_zakryl\\_dostup\\_k\\_fondam\\_kgb](https://www.znak.com/2019-03-13/gosarhiv_novosibirskoy_oblasti_zakryl_dostup_k_fondam_kgb).
- Stevenson, J. 2013. 'I Can Never Find Anything Among the Piles of Old Paper and General Rubbish': Edward Burra and his Archive. In *All This Stuff! Archiving the Artist*, ed. Judy Vaknin, Karyn Suckey, and Victoria Lane, 157–170. Faringdon: Libri.
- Stoycheff, E., E.C. Nisbet, and D. Epstein. 2016. Differential Effects of Capital-Enhancing and Recreational Internet Use on Citizens' Demand for Democracy. *Communication Research* 10: 4345–4367. <https://doi.org/10.1177/0093650216644645>.
- TASS 2018 = Kak v raznyh stranah otkryvali dostup k arhivam KGB. 2018. TASS.RU, December 20, 2018. <https://tass.ru/info/5936863>.
- Turner, P. 2005. Affordance as Context. *Interacting with Computers* 17 (6): 787–800.
- Venyavkin, Ilya. 2017. Interviewed by the Author on September 10, 2017, online.
- Wagman, J.B., S.E. Caputo, and T.A. Stoffregen. 2016. Hierarchical Nesting of Affordances in a Tool Use Task. *Journal of Experimental Psychology: Human Perception and Performance* 42 (10): 1627–1642. <https://doi.org/10.1037/xhp0000251>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Open Government Data in Russia

*Olga Parkhimovich and Daria Gritsenko*

## 22.1 INTRODUCTION

Open data can be defined in different ways and based on different principles, but it generally entails that anyone can access, use, and share the data freely. For instance, according to the European Union, open government data describes “the information collected, produced or paid for by the public bodies (public sector information, PSI) and made freely available for re-use for any purpose” (European Data Portal [n.d.](#)). In the broader sense, open government data is not only datasets, but also open government initiatives, policies and strategies, data management and publication approach, and models for interaction with citizens, nongovernmental organizations (NGO) and business. The set of policies enabling open government data promotes transparency, accountability, and improved efficiency of public services. In this way, open data initiatives are closely aligned with the freedom of information (FOI) principle, which is considered a cornerstone of democratic governance (Ackerman and Sandoval-Ballesteros 2006). As a result, open data can allow citizens to develop socially significant services and applications, analyze government actions, and know how government spends their money. At the same time, open data poses important questions with regard to data collection, processing, maintenance, storage, and security.

---

O. Parkhimovich  
Saint Petersburg National Research University of Information Technologies,  
Mechanics and Optics (ITMO), St Petersburg, Russia

D. Gritsenko (✉)  
University of Helsinki, Helsinki, Finland  
e-mail: [daria.gritsenko@helsinki.fi](mailto:daria.gritsenko@helsinki.fi)



In Russia, the executive order to provide open government data was signed by the President Vladimir Putin in May 2012 (Decree No. 601 from May 7, 2012). In 2014, the Open Government Data Portal ([data.gov.ru](http://data.gov.ru)) was launched. According to this portal, open government data (referred to as “open data” on most occasions) is

information (including documented) created within the limits of its powers by government bodies, or received by the specified bodies and organizations, as well as by information and analytical organizations participating in the publication of its own open data in the territory of the Russian Federation, which is to be placed on the Internet in a format that ensures its automatic processing for the purpose of re-use without prior modification by a person (machine readable format), and can be used freely in any lawful purposes by any persons, regardless of the form of its placement (a simple collection of information, a database, etc.).

The initiative has been actively developed, and by the end of 2019, more than 22,500 datasets were published on it. The Open Data Portal has been tightly connected to another initiative—the Open Government—that was launched by the President Dmitry Medvedev to ensure transparency of the legislative and executive processes in Russia (for more, see Chap. 2). In May 2018, the Russian Federation Open Government initiative was abolished, and the functions of the minister of Open Government were not transferred to another portfolio, which significantly reduced the activity of government agencies in this sphere. Nevertheless, but the obligation to publish open data has not been canceled. Therefore, the study of the specifics of open data in Russia and their use in applications and services is still relevant.

This chapter proceeds as follows. First, it provides the general legal background on the freedom of information in Russia. Next, it presents the Open Government Data initiative in Russia. The following sections explore the Russian open data strategy from the policy and implementation perspectives. Next, the regional dimension of open data is discussed. The following section provides an overview on the forms of interaction between the state and the citizens based in the open data. Finally, the chapter gives examples of public, civil society, and business initiatives that were enabled by the Open Data initiative.

## 22.2 RUSSIAN FREEDOM OF INFORMATION ACT

Freedom of information (FOI) is usually considered as an extension of freedom of speech, one of the fundamental human rights recognized in the European Convention on Human Rights. FOI is the key legal concept that guarantees access by the general public to the government-held information. “From India to South Africa and Mexico to China, states of varying degrees of development, size, and political persuasion have embraced openness and FOI” (Hazell and Worthy 2010, 352). For a long time, many FOI laws—especially in non-democratic states—were criticized for lacking the implementation machinery,

so that the free access to information remained a right only on paper (Relly and Sabharwal 2009). Yet, the digital transformation has become a turning point at which FOI could be given a new substance through publishing government datasets online.

In Russia, there is a number of laws concerning the right of access to information. Article 29 of the Russian 1993 Constitution guarantees everyone the right to freely seek, receive, transmit, produce, and disseminate information by any lawful means. The list of information constituting a state secret is determined by a special federal law. Federal Law No. 149-FZ of July 27, 2006, “*Ob informacii, informacionnyh tehnologiáh i o zásite informacii*” (On Information, Information Technologies and Information Protection) is a key legal document in the field of freedom of information. For the first ten years of its existence, there were 25 editions of this law, demonstrating its highly sensitive and political nature. The mechanisms for implementing the Russian FOI law is regulated by a number of special laws: Federal Law of May 2, 2006, No. 59-FZ “*O porádke rassmotreniáh obrášenij graždan Rossijskoj Federacii*” (On the Procedure for Consideration of Appeals of Citizens of the Russian Federation), Law of the Russian Federation No. 2124-1 of December 27, 1991, “*O sredstvakh massovoj informacii*” (On the Mass Media), Federal Law No. 8-FZ of February 9, 2009, “*Ob obespečeníi dostupa k informacii o deátelnosti gosudarstvennyh organov i organov mestnogo samoupravleniáh*” (On providing access to information on the activities of state and local government agencies), Federal Law of December 22, 2008, Federal Law No. 262-FZ “*Ob obespečeníi dostupa k informacii o deátel’nosti sudov v Rossijskoj Federacii*” (On providing access to information on the activities of courts in the Russian Federation), and Federal Law of October 22, 2004, No. 125-FZ “*Ob arhivnom dele v Rossijskoj Federacii*” (On the archival business in the Russian Federation).

The Russian FOI law guarantees to its subjects of the right to information by determining the basis for the realization of the right to information, and establishing the principles, forms, and freedoms for obtaining information. The right to seek and receive information is provided for both individual citizens and organizations and can be exercised through an official request to the owner of the information to provide certain information. If a citizen requests information that affects his or her rights and freedoms, the government cannot refuse such a request. The law also grants a right to request, without justification, information on (1) normative legal acts on the rights and obligations of a person or organization, (2) information on the state of the environment, (3) information on the activities of government agencies and their use of budget, (4) information that accumulates in the open collections of libraries, museums and archives, and information systems, (5) information, access to which cannot be limited by law. If requested, information must be provided without conditions and limitations. Refusal to provide information can be appealed in a higher authority, the prosecutor’s office, or at the relevant court. The state can charge a fee for the provision of information only if this is expressly stated in the law. Information on the activities of the government bodies posted on the

Internet and information affecting the rights and duties of a person and in other legal cases are always supposed to be provided free of charge (Olenichev 2017).

A citizen, a journalist, a media outlet, an organization, or any other civil law entity may request information under the Russian FOI law. The request for information may be sent by regular or electronic mail or personally delivered to a government agency. Internet sites of federal executive bodies contain forms for filing electronic appeals of citizens. Some regional governments create a “single reception room” (a single form, for example, on the website of the regional government), and they themselves forward requests to the necessary regional executive bodies. In order to receive a reply, it is necessary that the request for information contains the name of the state body, the name and surname of the applicant, his or her postal or e-mail address, the date of appeal, and the signature (if the message was not sent by email). The appeal is considered for 30 days, and, in exceptional cases, the term can be extended for another 30 days. Consideration of the request ends with the direction of the response. This procedure applies to all subjects of civil law, with the exception of the media. The authorities must respond to requests from the media within seven days of receiving the request (or notify within three days that the information will be provided later, indicating the date and reason for the postponement of the deadline).

In accordance with the Russian FOI law, authorities are required to provide information not only on request, but also on a regular basis. In particular, they are obliged to publish information about their activities in the media, on the Internet, in the premises of government bodies, and in other places that they have specifically identified, to acquaint users with information on the activities of state bodies through library archives and funds, and to allow interested citizens to attend meetings of collegiate bodies.

Given the long-standing tradition of secrecy within most branches of government and state authorities in the Soviet Union that has been inherited by the Russian Federation, the 2010 FOI law has been a major legislative milestone. Yet, the discrepancy between the law and its implementation has been noticed. The Global Integrity Index 2010 revealed that while Russian citizens have a strong constitutional right to information (scoring 90 out of 100), the actual ability of citizens to utilize this right was very limited (scoring only 56/100) (Global Integrity. Global Integrity Report: Russian Federation–2010, <http://www.globalintegrity.org/report/Russian-Federation/2010/>). As a result, Freedom of Information Law is often left unused by members of the public as there is a lack of knowledge with regard to FOI and a lack of transparency culture (Henderson and Sayadyan 2011).

### 22.3 OPEN GOVERNMENT DATA INITIATIVE IN RUSSIA: POLICY, INSTITUTIONS, INFRASTRUCTURE

Open Government Data can be regarded as a special case of implementation of the freedom of information principles. From the legal point of view, the Open Data initiative in Russia is mainly regulated by the Decree of the President of the Russian Federation (No. 601 of May 7, 2012) “*Ob osnovnykh napravleniakh soveršenstvovaniâ sistemy gosudarstvennogo upravleniâ*” (On the main directions of improving the system of public administration) and the federal laws “*Ob informacii, informacionnykh tehnologiâh i o zâsite informacii*” (On Information, Information Technologies and Information Protection, No. 149-FZ, entered into force on July 27, 2006) and “*Ob obespečenii dostupa k informacii o deâtelnosti gosudarstvennykh organov i organov mestnogo samoupravleniâ*” (On providing access to information on the activities of state and local government agencies, No. 8-FZ, entered into force on February 9, 2009), with respective amendments.

The first systematic approach in the field of open data in Russia was developed in 2012–2014, when the *Concept of Open Data of the Russian Federation* was adopted and implemented (Concept 2014). This Concept laid down the institutional, legal, and technological foundations of the open data system as it exists today. The Concept outlined the movement toward Open Data as a four-fold process, consisting of the development of methodological and normative documentation, adoption of the main instruments of certification, registration, and publication of open data, adoption of the plans for the disclosure of state and municipal data, and, finally, the launch of the Open Data Portal of the Russian Federation ([data.gov.ru](http://data.gov.ru)). The main effort for the realization of the Open Data Concept were concentrated within the *Governmental Commission for Coordinating the Activities of an Open Government*—in short, the Open Government—an expert group working within the Russian government. The official documents about the open data activities on the national level can be found in the section about the project Open Data published on the Russian Open Government website (<http://opendata.open.gov.ru>). The website has not been updated since 2018, however, following the resignation of the Open Government Minister Abyzov.

By law, all the federal executive bodies, including ministries, federal services, and agencies, must publish open government data on their websites. Yet, in accordance with the Order of the Government of the Russian Federation issued in 2013 (No. 1187-r of July 10, 2013), not all information is a subject to mandatory disclosure. The information that is mandatory to be disclosed in the form of open data includes seven categories:

1. Names of territorial bodies and representative offices (representatives) of the federal executive authority abroad (if any).
2. Names of subordinate organizations (if any).

3. Plan for conducting inspections of legal entities and individual entrepreneurs for the next year.
4. The results of planned and unscheduled inspections conducted by the federal executive authority and its territorial bodies within their authority, as well as the results of inspections conducted in the federal executive authority, its territorial bodies and subordinate organizations.
5. Statistical information generated by the federal executive body in accordance with the federal statistical work plan, as well as statistical information on the results of planned and unscheduled inspections.
6. Information on vacant positions of the state civil service, available in the federal executive body and its territorial bodies.
7. Registers of licenses for specific types of activities licensed by federal executive bodies.

The publishing of other information in the form of open data is optional.

To facilitate the implementation of the Executive Order, in 2015, the *Governmental Commission for Coordinating the Activities of an Open Government* created the *Open Data Council* (<https://opendata.open.gov.ru/sovet/about/>), a consultative body consisting of representatives of federal authorities, business, and universities, headed by the minister of the Open Government. The Council has four main functions. First, it develops specific mechanisms for opening data and to help the government to solve all organizational, legal, and technical problems as efficiently as possible. Second, it was mandated to work with business and citizens, helping to measure the demand for open data, as well as to choose the priorities when disclosing government information. The third task of the Council is to collect and promote best practices, popularize the idea of open government data, and show specific opportunities for business development. Finally, it was asked to create an independent feedback mechanism to assess the overall economic and social impact from the disclosure of government databases. Meeting every two to three months, the main idea behind the *Open Data Council* was to discuss the questions related to different aspects of the open data, for example data about different topics or data from different federal government bodies. During the meetings, representatives of state bodies were invited to the Council to make presentations, exchange information, and help the Council to achieve its core tasks. As the Council was established with consultative functions, its recommendations have to be submitted to the *Governmental Commission for Coordinating the Activities of an Open Government*, which is responsible for coordinating different points of view and interests, as well as the consideration of expert opinions. Hence, the Governmental Commission, not the Council itself, had the power to issue final recommendations by governmental orders. In May 2018, after the reelection of Vladimir Putin to the presidential post and the formation of a new government, the *Open Data Council* was suspended. A new council or working group has not been created, but the need for a council or center of competence is being discussed by experts.

In course of its functioning, the *Open Data Council* developed recommendations on the development of the entire open data ecosystem. The action plan (“Road Map”) “Open Data of the Russian Federation” for 2015–2016 (Roadmap 2014) could be considered as the main outcome of its work. The Road Map presupposed a number of concrete action points. In 2015, all federal executive bodies were to create sections of open data on their Internet resources and disclose the so-called priority datasets, or socially significant datasets, grouped into 27 thematic areas, according to a certain schedule. The legislator was, at the same time, tasked with the development of the terms. Finally, the presumption of general availability of primary statistical data was introduced as an amendment to the Federal Law “*Ob oficial nom statističeskom učete i sisteme gosudarstvennoj statistiki v Rossijskoj Federaciji*” (On Official Statistical Accounting and the System of State Statistics in the Russian Federation, No. 282-FZ from November 29, 2007). While the roadmap has not been legally canceled, due to the abolition of the minister of open government and the lack of a responsible person in the federal government, the roadmap has completely disappeared from the public and internal agendas of the federal government and federal executive bodies. At the time of writing this chapter, the roadmap can be considered to be suspended.

Reports on the implementation of the “Road Map” were submitted by all federal executive bodies to the Ministry of Economic Development of Russia quarterly to monitor the quality and timeliness of the implementation of the plan. The reports were used to monitor the progress of Open Data policy implementation. Also, the federal executive bodies annually fill out a form of self-examination of the level of development of mechanisms and directions of openness, one of the tools of which are the open data (Self-examination form 2017), and were used to compile and the “open data rating.” According to the report produced by the *Open Data Council*, the openness self-perception among the federal executive organs has been growing, and new federal bodies are joining the Open Data movement (Expert Council Report 2016). The Ministry of Defense, Ministry of Energy, and Ministry of Finance occupied the top three positions in the perceived transparency in 2015.

In order to facilitate open data management, an Action Plan *Open Data of the Russian Federation* for 2016–2017 was developed, outlining the activities to be undertaken, expected results, schedule, and naming the responsible executors (Action Plan 2015). The Action Plan included actions to develop methodological support in the field of open data, the development of regulatory legal support, the development of an open data infrastructure, access to open data, the formation of an open data ecosystem, and the development of non-state institutions. There have been no follow-up action plans or other strategic documents published since then. In general, the 2015 Action Plan has been followed by the Russian Open Data Council. A significant part of the actions in this Action Plan consisted of discussions; therefore, the implementation of this plan did not lead to qualitative changes in the openness of key areas of data publication in the Russian Federation. For example, detailed data on quality of

life or a register of companies have not been disclosed. On the other hand, the data that are now most open in Russia (for example, data on public finances) were disclosed in parallel with the activities of the Open Data Council as part of the functions of the responsible public authorities.

From the infrastructural point of view, the main gateway to the open government data in Russia is the Russian Open Data Portal ([data.gov.ru](http://data.gov.ru)) launched in 2014 and maintained by the Russian Ministry of Economic Development. The portal contains datasets provided by the federal, regional, and local level government bodies, and some federal government websites are even configured to automatically upload data to the Russian Open Data Portal. The Portal is equipped through a search function that allows a user to do keyword searches. Each dataset is also assigned to a thematic category, such as “Government,” “Economics,” “Health,” “Transport,” “Tourism,” et cetera, and is promoted as the core of the open data ecosystem in Russia. Most datasets currently uploaded fall within the “Government” category (almost 15,000, or two-thirds of all uploaded datasets), while least data can be found under the categories “Cartography” (81), “Electronics” (29), and “Weather” (5) ([data.gov.ru](http://data.gov.ru), December 19, 2019). While some official agencies and authorities took proactive steps to disclose their information, others fulfill the requirements in a superficial way (Henderson and Sayadyan 2011). In a recent research, Repponen (2018) investigated open data availability of 75 Russian executive organs, including federal agencies and services, ministries and funds, revealing a tendency among the studied bodies to release datasets on contact information, thereby only fulfilling the minimum requirements of the 2012 executive order to provide open government data on the Internet. In 2019, Begtin et al. issued a special report under the auspices of the Russia Audit Chamber, suggesting a new instrument—an Openness rating—as a tool to monitor and provide specific recommendations to the federal authorities. The Openness rating measures three key dimensions—the openness of information, open data availability, and open dialogue. The first results demonstrate that the federal ministries show higher results on information and open data dimensions, while only about a third of them scored high on the open dialogue criteria. Similarly, federal agencies tend to score higher on information openness, while only 24% scored high on the open dialogue dimension.

Over the past year and a half, the Russian Open Data Portal has not been developed or supported, and funding has not been allocated for it. In the fall of 2019, the Ministry of Economic Development of Russia announced tenders and concluded contracts for technical support and refinement of the Portal. The contracts also included services such as webinars and hackathons and the development of recommendations. Yet, the cost, timing, and quality of work, the results of which could be observed at the moment of writing this chapter (December 2019), raise questions from the expert community.



## 22.4 OPEN DATA MANAGEMENT AND PUBLICATION APPROACH

To facilitate open data management and publication, the Russian government developed an Open Data Standard ([openstandard.ru](http://openstandard.ru)), including the Concept for the Openness of the Federal Executive Bodies (Concept 2014), the Methodological Recommendations for the Publication of open data by state bodies and local self-government bodies (Methodological Recommendations 2014), as well as technical requirements for the publication of open data. The Methodological Recommendations have become the main applied tool for the authorities, as this 100-page document contains specific guidelines for publishing open data for government bodies. They were developed to provide a relevant, structured, and targeted tool that helps ensure compliance with the legislation of the Russian Federation by explaining the law, suggesting best practices for compliance, providing examples of with applicable national and international technical standards. According to the Methodical Recommendations, open data is information placed on the Internet in the form of systematized data organized in a format that ensures its automatic processing without prior modification by a person for the purposes of repeated, free and free use. The Methodological Recommendations outline for the data owners (state and municipal employees) and their publishers (specialists of internal information technologies [IT] departments or companies involved on the basis of a contract) the requirements for the content of information resources, the technical requirements for formats for the presentation of open data, and the composition and principles of interaction of elements of the national open data infrastructure. The Recommendations are quite often criticized in the expert community for three main reasons: they are not well structured, their target audience is not clearly specified, and they do not answer some of the important questions that arise for data publishers. As a result, the experts often highlight the need to revise and update this document.

The Technical Requirements for the Publication of Open Data, an annex to the Methodological Recommendations, contain specifications on the requirements for the publication of the register of open data, open datasets, the passport of an open dataset (description of metadata), and the requirements for the structure of an open dataset in machine-readable form. Publication of the metadata, called a “dataset passport,” is required and is relevant for the end-user, as along with the identification number (id), name, description, owner, person in charge, link, and format, it includes important temporal information, such as the date of the first publication, the date and contents of the last modification, and links to previous versions of the dataset, as well as the version of methodical recommendations to which it adheres. When an authority creates an Internet page to provide access to its datasets, it should have a heading that clearly marks the content—“Open Data”—and the following elements: a register of open data, information on the total number of open datasets, in case there are more than 20 datasets—a search tool should be provided, and a tool

for requesting information in the form of open data. Importantly, each published dataset should have both a machine-readable and a human-readable representation, with requirements differentiating the data formats (such as csv, xml, json, html + rdfa). Currently, majority of the data is published in a simple comma-separated format (CSV), which on the one hand is a format familiar to lay users, but, on the other hand, is the format most prone to formatting mistakes which makes aggregating and cleaning of large datasets into a separate, laborious task.

The conditions for the use of open data—terms of use and/or license—periodicity of updates, and the public authority responsible for publication (publisher) should also be clearly identified. Often, in addition to the current version of the open dataset, user can also download archival versions. For datasets published on the Russian Open Data Portal ([data.gov.ru](http://data.gov.ru)), a change log is available. One important feature of open data is publishing under an open license. The license or terms of use of the dataset allow programmers to understand what actions they can do with published data: can third-party applications be created based on open data, can data be used for commercial services, et cetera.

The regulatory framework created in Russia provides public authorities with requirements and recommendations for the publication of open data. The open data infrastructure, which has been created mainly between 2012 and 2014, serves the purpose of providing access to government data to a variety of actors. Accounts Chamber of the Russian Federation, the parliamentary body of financial control in the Russia, still considers open data as one of the priority areas of work. Yet, as the federal government significantly rolled back the openness agenda, government data management in Russia has shifted from a priority to ensure openness to an internal inventory of data within authorities without additional publicity to this process.

## 22.5 REGIONAL OPEN DATA INITIATIVES

In 2013, the Russian government adopted a resolution “*Ob obespečenii dostupa k obšedostupnoj informacii o deâtel’nosti gosudarstvennyh organov i organov mestnogo samoupravleniâ v informacionno-telekommunikacionnoj seti ‘Internet’ v forme otkrytyh dannyh*” (On providing access to publicly available information on the activities of state bodies and local governments in the information and telecommunications network “Internet” in the form of open data, No. 583, July 10, 2013). This resolution, as already mentioned, contains rules pertaining to publicly available information placed by the federal and local government bodies on the Internet in the form of open data. Among other things, it introduces rules for the public authorities of the regions of the Russian Federation and local government bodies. The regional and local authorities are required to make publicly available information on the activities of the state authorities of the constituent entities of the Russian Federation and local government bodies established by these bodies or receive by them in the exercise of authority in the regions of the Russian Federation. The rules determining

the periodicity of publishing the open data on publicly available information on the activities of state bodies and local government bodies on the Internet, such as the timing of renewal, ensuring the timeliness of the implementation and protection of users' rights and legitimate interests, as well as other requirements for the placement of information in the form of open data, are also specified.

It is important to add that some local and regional data are collected on federal information systems. For example, each local and regional public body is required to add detailed data about their contracts on the official procurement website ([zakupki.gov.ru](http://zakupki.gov.ru)), while each budgetary autonomous institution should add data about planned and actual performance indicators, including balance sheets, on the official website for posting information about state and municipal (local) institutions ([bus.gov.ru](http://bus.gov.ru)).

Currently, the Russian Open Data Portal ([data.gov.ru](http://data.gov.ru)) has more than 9500 datasets published by the regional authorities and more than 3000 datasets provided by the local-level authorities, with more than 500 regional and about 400 local public bodies being registered on this website. These numbers suggest that the regional and local implementation of the Open Data strategy lags behind the federal implementation.

A more detailed picture on the open data publication and openness of information of the federal, regional, and local authorities in Russia can be inferred from the ratings that were prepared by the Russian nongovernment project center "Infometer" (<http://system.infometer.org>). A distinctive feature of these ratings is the availability of links to all sites being researched and references to each assessed parameter and to the relevant legislation.

The Infometer's rating "Regional Open Data 2016" estimates open data of all Russian subnational units. This instrument measures 84 parameters, such as the following:

- The availability of a separate page (section) "Open Data" or a separate site for the placement of open data or a section on the portal of the open data of the Russian Federation.
- There is no requirement for registration and authorization on the site for the use of open data.
- Information for developers, who made applications based on open data.
- Name of the person responsible for the content of the open datasets.
- Availability of datasets "Names of registry offices," "Names of executive authorities of the subjects of the Russian Federation," "Plan for conducting state ecological expertise," "Information on the results of state ecological expertise," "State forest registry (for forests located within the territory of the regions of Russia)," "Register of Licenses for Educational Activity," et cetera.

Seventeen out of 85 regions scored less than 30%, while 16 scored more than 70%, meaning that the majority of regions were rated average in the open

data performance. The City of Moscow, Bryansk, Tomsk, Tula, and Ulyanovsk regions, as well as Khanty-Mansi Autonomous Area have all scored 100%. The City of St. Petersburg scored 91.6%.

Another Infometer's rating "Local Open Data 2017" estimates open data in cities with population of more than 100,000 people using 60 parameters. The rating does not include Moscow, St. Petersburg, and Sevastopol, as these are subnational units in their own right, or "cities of the federal status" whose governments operate at the regional level. The 166 cities were included in the 2017 rating. Despite their obligation to provide open data, the index indicated that 68 cities did not publish anything, 10 cities published more than 20 datasets, 53 cities scored more than 50%, and only 2 cities published more than 100 datasets. The cities that scored more than 80% are Tula, Novomoskovsk, Domodedovo, Taganrog, Yekaterinburg, Nizhny Tagil, Obninsk, Nizhnevartovsk, Shakhty, and Bratsk.

## 22.6 CIVIL SOCIETY, BUSINESS, AND GOVERNMENT INTERACTIONS BASED ON THE OPEN DATA

The new quality of interactions between the state and its citizens is one of the central promises of open data. It is difficult, however, to provide a systematic assessment of the level of interaction between the civil society and the federal government on the topic of open data in Russia. The starting point for such assessment would be the analysis of open data requests. As mandated by law, each federal executive body has a form for the electronic appeal of citizens on its website; some also have a feedback form or additional email for open data requests or comments. Each appeal must necessarily be examined and answered within a month. Yet, the federal authorities do not publish detailed statistics about the requests they received and responded to, so it is not possible to single out requests or appeals for open data. In addition, federal executive bodies may underestimate the relevance of open data to citizens and programmers, as programmers often do not report the use of open data in their projects and do not send requests.

In 2014–2016, the Ministry of Finance of Russia organized meetings with developers on the topic of open data several times a year (Minfin 2016). It was a unique and effective mechanism that allowed software developers to hear presentations from the ministry and its contractors about the public data, as well as to ask their questions and get answers on the same day. Competitions, such as BudgetApps ([www.budgetapps.ru](http://www.budgetapps.ru)), the All-Russian competition "Open data of the Russian Federation." ([www.opendatacontest.ru](http://www.opendatacontest.ru)), and hackathons, for example Hackathon of the Accounts Chamber of the Russian Federation "Data Audit" (<http://data-audit.ru>), have provided another pathway for the programmers and the broader community to engage with the government around the use of open data.

The BudgetApps, an annual competition of projects based on open government financial data, was held by the Russian Ministry of Finance in 2015–2018. The popularity of the competition grew from year to year: whereas 45 projects were submitted in 2015, 155 projects were submitted in 2016 and 160 in 2017. The prize fund of the contest is about 500,000 rubles per year (around €7300 as of December 2019). The partners at the national level are the Federal Tax Service, the Federal Service for Regulation of the Alcohol Market, and the Federal Treasury. The NGO Infoculture acted as a contractor of the Ministry of Finance on the BudgetApps competition, providing the organizational work. In 2018, the format of the competition “BudgetApps” was changed to an independent search and selection of projects by an expert commission. It did not include hackathons, events for developers, neither activities on social networks, so the quality and quantity of projects decreased significantly.

One of the projects submitted for the 2015 BudgetApps competition was, for example, the Russian Schools project (<https://goodschools.ru>), a social service that accumulates in one place all the basic accessible information and knowledge about the activities of schools. The service is based on open data on state institutions, government contracts, exam results, and public reports of schools, providing an overview and rating of schools based on their funding, exam results, and personnel. It can be used by a variety of actors; for example, it can help parents to choose a school for their children, teachers—find out which schools are better paid, or provide public activists with information on how effectively taxpayer money is spent in the educational sphere. The Russian Schools project is still supported and developed.

The All-Russian competition “Open data of the Russian Federation” was planned as an annual competition of projects based on open data. Some federal ministers and federal agencies developed tasks for participants, including the Federal Treasury, Ministry of Culture, Ministry of Industry and Trade, Ministry of Transport, Ministry of Construction, Ministry of Labor, Ministry of Finance, Rosaccreditation, Roslesinfor, Rosnedr, Rospatent, Rosstandart, Rosstat, Rostrud, Rosturizm, Open Data Council, and the Federal Tax Service of Russia. It has been held by the Russian Open Government in collaboration with the Ministry of Economic Development of Russia and in partnership with the Analytical Center under the Government of the Russian Federation in 2015, 2016, and 2017, but not repeated in the subsequent years.

Russian Open Data Summit was first held by Russian Open Government in 2015. It was supposed to become an annual conference and a platform for communication of representatives of the state among themselves and with the developers. At the end of 2016, it did not take place and was moved to the beginning of 2017. In early 2017, it was postponed to the end of 2017 and was not carried out. Thus, the Open Data Summit was held only once in 2015 and now it is impossible to say whether it will be held in the future.

Interactions between the representatives of the open data community and the government can also occur at various conferences and in online communities, for example the online community for open data in the Telegram

messenger (<https://telegram.me/opendatarussiachat>) and on a platform called Slack (<https://opendatarussia.slack.com>). Federal executive bodies often issue press releases on their websites about the publication of new open datasets, which is a helpful way for the interested party to receive the latest updates. Another tool for interaction is public pages, maintained by federal executive bodies in social networks (VKontakte and Facebook). For example, the Russian Federal Antimonopoly Service and the Russian Audit Chamber not only maintain the pages, but also actively respond to user comments (although it is impossible to determine how much of this interaction revolves around open data).

Attention has also been paid to the capacity-building. In September 2016, the Russian Open Government has launched an online course “Open Data. Theory and Practice” (<https://open.gov.ru/events/5515416/>). The program is designed for both civil servants and IT service developers, as well as a wide range of other professionals, who want to learn how to work with open data. In order to gain access to video lectures, text, and test materials of the course, equivalent to 72 academic hours, it is necessary to register on a specially created website (the registration is free). The main requirement for attendees is knowledge of the basics of computer literacy. Based on the results of the training, certification is provided for two main profiles: “civil servant” and “IT specialist.” When it comes to the nonprofit sector, periodically different teams conduct webinars and hackathons with educational content on how to work with open data. For example, the NGO “Infoculture” in cooperation with the Open Government developed the “Open Data School” in 2013–2014 with offline lectures, seminars, and workshops.

## 22.7 OPEN DATA IMPACT IN RUSSIA

There are several cases of how open data can have an effect on increasing transparency and accountability, positive impact on the economy, and creation of new companies.

### 22.7.1 *Increasing Transparency and Accountability*

Open data could lead to improvements in government transparency and accountability in a number of ways: for example, supporting journalism and data journalism which uncovers wasteful spending, corruption, or other wrongdoing by government departments or officials; supporting the creation of applications which allow citizens to report on their experience of government services; supporting scrutiny of government decision-making; supporting greater citizen engagement in policy-making (Open Data Charter 2015).

In Russia, several examples of services based on open data related to state finance and public procurement are worth mentioning. “Government Spending” (<https://clearspending.ru>) is a nonstate project to increase public awareness of spending public funds. The automatic monitoring system allows

a user to study, understand, find violations, and reuse data on public spending, in particular, on grants and on state and municipal contracts. The aim of the project is to encourage the authorities to search for and implement ways to solve problems in the sphere of public spending and to eradicate abuses in the state procurement industry. The project has been featured in the news multiple times, and it regularly organizes webinars and open lectures for journalists to enhance their awareness of the service and the opportunities it provides. Opening of public procurement data also allowed *Transparency International Russia* to produce several influential research reports, including “How do the largest donors of political parties make money on government contracts?” (2017), “How do heads of state theaters pay themselves?” (2017), and “Siberian roads: how roads were repaired in six Siberian cities” (2017), all pointing out problematic patterns in using taxpayers’ money and the mechanism of state contracts. For instance, “Siberian roads” report was based on the open data on 575 contracts for road repairs in six Siberian cities—Barnaul, Irkutsk, Novosibirsk, Omsk, Tomsk, and Chita. The authors have identified schemes by which cartels and affiliated firms take more than 50% of all contracts. Another project, Open NGO (<https://openngo.ru>), aims at showing citizens how Russian nonprofit organizations are organized and funded from state sources by bringing together open data on subsidies from the federal budget, state contracts, grants of the presidential grants fund, and the register of nonprofit organizations of the Ministry of Justice.

### 22.7.2 *Economic Impact of Open Data*

Open data may impact on the economy, for example, through supporting existing businesses to lower their costs or become more efficient or through supporting better economic planning. Open government data can be used by entrepreneurs to make commercial or nonprofit services.

There are successful examples of companies earning money using government open financial data (Begtin 2016). According to the Open Data Impact Map, a project of the Center for Open Data Enterprise in partnership with the World Bank Group, there are 39 companies in Russia whose business is based on open data, while the Russian Open Data Portal enlists links to 255 applications based on open data, including both nonprofit and commercial uses. To name a widely known example, YandexTaxi, a taxi application run by the Russian tech giant Yandex, was launched after the registering of taxi licenses was made openly available. Another example includes technical solutions, which use and integrate the data of the Federal Tax Service of Russia (the register of legal entities), financial statements, and data of the Federal State Statistics Service Rosstat, such as *KonturFocus* and *Spark Interfax*. These applications allow users to perform due diligence checks. For example, the revenue of the company *KonturFocus* in 2016 amounted to 8.6 billion rubles.

Open Data can also be seen as beneficial in a wider economic context. The National Research University Higher School of Economics (HSE) estimated



**Table 22.1** Cumulative effect of using applications based on open data in Moscow's public transport system

<i>Factor</i>	<i>Annual saving (bln rub)</i>
Higher public transport coverage and more efficient use of equipment	3,668
Reducing travel time for passengers on public transport	8,890
Reduction of travel time by private transport	41,515
Decrease in waiting time at stops	11,967
Reducing gasoline consumption and revenues from its sale (negative economic effect)	7,287

Source: Author based on Artamonov et al. (2015)

the cumulative economic effect of using applications based on open data in public transport in Moscow (Artamonov et al. 2015). According to the study, the cumulative economic effect of using applications based on open public transport data in Moscow could amount to 58,753 billion rubles a year (Table 22.1).

## 22.8 CONCLUSION

The open government data has been developing in Russia since 2012. Within a short amount of time, the necessary regulatory framework was created, guidelines for the publication and management of open data were developed, an open data portal was launched, and an increasing number of government agencies, not only at the federal but also at the regional and local levels, were involved in the creation and publication of open data. Being a practical tool to the implementation of the Freedom of Information principles, open data in Russia has become a basis for a large number of public projects that provided tools for obtaining information from government agencies and interacting with them. Also, a community of data journalists has appeared.

Since 2018, a rollback in the area of open data has begun. Since May 2018, for the federal government the topic of open data has been replaced by an inventory of government data. The increasing internal and external economic challenges, domestic political changes, a decrease in Russia's interaction with international organizations that focus on an open data agenda, and a loss of Russia's interest in joining the Organization for Economic Co-operation and Development (OECD), all had a negative effect on the open data ecosystem development. And yet, open data movement in Russia continues both thanks to the regional and municipal authorities and especially to the community of developers and citizen activists. Also for some government agencies, the topic of open data remains not only relevant, but also a priority. The publication of open data by the Federal Tax Service of Russia, the launch of the project [spending.gov.ru](http://spending.gov.ru) by the Audit Chamber of the Russian Federation, the use of open data for interaction between the Ministry of Culture and its subordinate

organizations and cultural institutions can all serve as illustration of the continuity in open data development.

Currently, maintenance of the open data agenda is largely undertaken through the activities of the community and nonprofit organizations aimed at including open data in the federal government's data management agenda. The priorities for open data experts and NGOs are training public servants to work with open data, lobbying for the inclusion of open data topics in the created legal acts, and interacting with authorities to improve the quality of data and the convenience of its publication. Open Data Day (<http://opendataday.ru>) is a prime example of a community-driven annual event that brings together open data experts, activists, and developers in Moscow and other large Russian cities. State authorities often and actively participate in the Open Data Day as speakers in discussions and workshops, despite the decline in interest at the governmental level. As a result, despite the fragmentation of open data initiatives and the lack of a unified federal agenda, the open data movement in Russia remains in existence, and the ecosystem has fair chances to be further developed in the future, even if the speed and scope of the development are somewhat limited.

## REFERENCES

- Ackerman, J.M., and I.E. Sandoval-Ballesteros. 2006. The Global Explosion of Freedom of Information Laws. *Admin. L. Rev.* 58: 85.
- Action Plan. 2015. Action Plan "Open Data of the Russian Federation" 2016–2017. Accessed December 15, 2019. <http://opendata.open.gov.ru/upload/iblock/b3b/b3bd585774bf891c827ff3fc2f21d23a.doc>.
- Artamonov, R., S. Datiev, A. Zhulin, A. Kondrashov, N. Lavrentyev, E. Muleev, S. Plaksin, E. Styurin, and E. Yastrebova. 2015. *Assessment of the Socio-economic Effect of the Publication of Open Data on the Example of Moscow Public Transport Data*. Report. Moscow: HSE Publishing. Accessed December 15, 2019. [http://gos.hse.ru/downloads/2015/Socio\\_economic\\_effect\\_transport\\_Moscow\\_2015.pdf](http://gos.hse.ru/downloads/2015/Socio_economic_effect_transport_Moscow_2015.pdf).
- Begtin, I. 2016. Begtin: How to Earn on Open Data. Accessed December 15, 2019. <https://roem.ru/15-08-2016/230898/otkrytye-dannye-startap/>.
- Begtin, I., V. Burov, A. Sakoyan, O. Parkhimovich, M. Amara, M. Komin, and P. Demidov. 2019. *State Openness in Russia*. Expert Report. The Russian Federation Audit Chamber. Accessed December 15, 2019. <http://audit.gov.ru/about/document/Открытость-доклад.pdf>.
- Concept. 2014. *Concept for the Openness of the Federal Executive Bodies*. Accessed December 15, 2019. <https://rg.ru/2014/02/03/koncepciya-site-dok.html>.
- European Data Portal. n.d. What is Open Data? <https://www.europeandataportal.eu/clearing/en/module1/#/id/co-01>.
- Expert Council Report. 2016. On the Results of a Self-examination of the Level of Openness of Federal Executive Bodies and Updating of Departmental Plans for the Implementation of the Concept of Openness of Federal Executive Bodies. Open Government Expert Council. Accessed December 15, 2019. <http://open.gov.ru/upload/iblock/5c9/5c927e149c6cddc2580aeb6862ad4e81.pdf>.

- Hazell, R., and B. Worthy. 2010. Assessing the Performance of Freedom of Information. *Government Information Quarterly* 27 (4): 352–359.
- Henderson, J., and H. Sayadyan. 2011. Freedom of Information in the Russian Federation. *European Public Law* 17 (2): 293–311.
- Local Open Data. 2017. Local Open Data Rating 2017. *Infometr*. Accessed December 15, 2019. <http://system.infometer.org/ru/monitoring/420/rating/>.
- Methodological Recommendations. 2014. Methodological Recommendations for the Publication of Open Data by State Bodies and Local Self-government Bodies. Accessed December 15, 2019. <http://opendata.open.gov.ru/upload/iblock/7b8/7b8db735e85b68a1ca3f7e87bdd538b6.pdf>.
- Minfin. 2016. Meeting Plan for Meetings with the Users of Open Data in the Sphere of Finance and Budgeting for 2016. Ministry of Finance. Accessed December 15, 2019. <https://www.minfin.ru/ru/opendata/ecosystem/plans/>.
- Olenichev, M. 2017. The Right to Know. Report ‘Team 29’ on Access to Information in Russia. Accessed December 15, 2019. <https://team29.org/wp-content/uploads/2017/08/Access-to-information.pdf>.
- Open Data Charter. 2015. International Open Data Charter. Accessed December 15, 2019. <https://opendatacharter.net/>.
- Regional Open Data. 2016. Regional Open Data Rating 2016. *Infometr*. Accessed December 15, 2019. <http://system.infometer.org/ru/monitoring/371/rating/>.
- Relly, J.E., and M. Sabharwal. 2009. Perceptions of Transparency of Government Policymaking: A Cross-national Study. *Government Information Quarterly* 26 (1): 148–157.
- Repponen, I. 2018. Linked Ontology on Metadata on Russian Open Governmental Data. Accessed December 15, 2019. <https://blogs.helsinki.fi/digital-russia-studies/data/>.
- Roadmap. 2014. Road Map “Open Data of the Russian Federation.” Accessed December 15, 2019. <http://open.gov.ru/upload/iblock/2bd/2bd533adb07b9f91c53d286f3aca5d39.pdf>.
- Self-examination Form. 2017. Self-examination of the Level of Development of Mechanisms and Directions of Openness. Accessed December 15, 2019. <https://minvr.ru/upload/iblock/aab/Результат%20Самообследования%202016.pdf>.

#### LEGAL DOCUMENTS

- Decree of the President of the Russian Federation (No. 601 of May 7, 2012). *Ob osnovnykh napravleniakh soveršenstvovaniâ sistemy gosudarstvennogo upravleniâ* [On the Main Directions of Improving the System of Public Administration]. <https://rg.ru/2012/05/09/gosupravlenie-dok.html>.
- Federal Law of October 22, 2004, No. 125-FZ. *Ob arhivnom dele v Rossijskoj Federacii* [On the Archival Business in the Russian Federation]. <https://rg.ru/2004/10/27/arhiv-dok.html>.
- Federal Law of May 2, 2006, N 59-FZ. *O porâdke rassmotreniâ obrašeniij graždân Rossijskoj Federacii* [On the Procedure for Consideration of Appeals of Citizens of the Russian Federation]. <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102106413>.
- Federal Law No. 149-FZ (entered into force on July 27, 2006). *Ob informacii, informacionnykh tehnologiâh i o zaštite informacii* [On Information, Information

- Technologies and Information Protection]. <https://rg.ru/2006/07/29/informacia-dok.html>.
- Federal Law of November 29, 2007, No. 282-FZ. *Ob oficial'nom statističeskom učete i sisteme gosudarstvennoj statistiki v Rossijskoj Federacii* [On Official Statistical Accounting and State Statistics System in the Russian Federation]. [https://eng.gks.ru/law\\_on\\_stat](https://eng.gks.ru/law_on_stat).
- Federal Law of December 22, 2008, No. 262-FZ. *Ob obespečenii dostupa k informacii o deâtel'nosti sudov v Rossijskoj Federacii* [On providing access to information on the activities of courts in the Russian Federation]. <https://rg.ru/2008/12/26/sud-internet-dok.html>.
- Federal Law No. 8-FZ (entered into force on February 9, 2009). *Ob obespečenii dostupa k informacii o deâtel'nosti gosudarstvennyh organov i organov mestnogo samoupravleniâ* [On Providing Access to Information on the Activities of State and Local Government Agencies]. <https://rg.ru/2009/02/13/dostup-dok.html>.
- Federal Law No 583, July 7, 2013. *Ob obespečenii dostupa k obšedostupnoj informacii o deâtel'nosti gosudarstvennyh organov i organov mestnogo samoupravleniâ v informacionno-telekommunikacionnoj seti 'Internet' v forme otkrytyh dannyh* [On Providing Access to Publicly Available Information on the Activities of State Bodies and Local Governments in the Information and Telecommunications Network 'Internet' in the Form of Open Data]. <http://pravo.gov.ru/proxy/ips/?docbody=&nd=102166683>.
- Law of the Russian Federation No. 2124-1 of December 27, 1991, *O sredstvakh massovoj informacii* [On the Mass Media]. <https://rg.ru/1991/12/27/smi-zakon.html>.
- Order of the Government of the Russian Federation (No. 1187-r of July 10, 2013). <https://rg.ru/2013/07/29/info-site-dok.html>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Topic Modeling in Russia: Current Approaches and Issues in Methodology

*Svetlana S. Bodrunova*

## 23.1 INTRODUCTION

### 23.1.1 *Topic Modeling as a Scientific Method*

Topic modeling is a method of probabilistic clustering of textual documents mostly used for large text collections. It finds itself on the crossroads of probabilistic and predictive text classification, natural language processing methodologies, semantic analysis, and discourse studies. In this chapter, we look at how the teams involving Russian-speaking scholars have enhanced the topic modeling algorithms, tested their efficiency, and employed them for interpretation of real-world datasets, including those from today's social media—either in Russian only or for the Russian cases in comparison with those in other languages.

For many scholars, topic models are about latent semantic analysis, or LSA (Steyvers and Griffiths 2007), but, algorithmically, LSA appears to be only one option of topic modeling; a large variety of algorithmic approaches and extensions to them have been suggested within the last two decades (Blei and Lafferty 2009; Korshunov and Gomzin 2012). The main goal of using any topic modeling algorithm is to detect the so-called topics in a text collection. In communication terms, a topic is a theme around which the discussion is evolving; but, in topic modeling, topics express themselves via collections of words and/or documents that the modeling algorithm considers similar and/or related to each other.

---

S. S. Bodrunova (✉)  
St Petersburg State University, St Petersburg, Russia

The basis for the texts to be related to each other is word co-occurrence. The method implies that texts belonging to one topic may be described as those where particular words stay close to each other or, at all, can be found. This understanding leads to the probabilistic iterative process which sees the dataset as a “bag of words” where the word order and syntactic links between the words are ignored. It defines, by multiple iterations, which words most probably stand together in which documents. Computationally, a topic is a discrete (multinomial) probability distribution over terms in a given vocabulary (Mcauliffe and Blei 2008, 121). Thus, each document belongs to each topic with some probability (often negligibly small), but some texts belong to some topics with much higher probabilities, with an arbitrary threshold for where to cut the “long tail” of nonrelevant texts. The results of the modeling are represented in two matrices: the word-topic one (the probabilities of particular words to belong to a topic) and the topic-document one (the probabilities of a topic to be found in a particular document); but the end-users usually assess the top words (the words with the highest probability for the topics) and the most probable texts in the topics. For the end-user, a topic is a collection (cluster) of texts that belong with high enough probability to one theme slot and are expected to be linked by topicality of their content.

The quality of modeling—that is, how well the topics are separated from each other, how many texts they involve above the relevance threshold, and how interpretable they are—may be measured by the metrics of topic interpretability, coherence, robustness, et cetera. The baseline for topic quality assessment is human coders’ interpretation, but a lot of automated metrics of quality have been developed to make the topic quality assessment quicker and easier.

Of the Bayesian bag-of-words algorithms, the one based on Dirichlet distributions and called latent Dirichlet allocation (LDA) (Blei et al. 2003) is, undoubtedly, the most developed today. Along with it, for various types of data, several other algorithms have matured and also gained important extensions that allow the scholars to intervene and change the parameters of the algorithm. In terms of allowed intervention, topic modeling may be unsupervised, supervised (Mcauliffe and Blei 2008), semi-supervised (Bodrunova et al. 2013), or weakly supervised (Lin et al. 2011). Alternative promising approaches to topic detection mostly try to preserve the semantics that stem from word order and grammatical relations between words, like the approaches based on Markov chains (Gruber et al. 2007) or n-grams (Wang et al. 2007).

Topic modeling has advantages quite attractive for scholars, as well as shortcomings inherent for the method. Among the latter, there are the principal instability of clustering results (i.e. different runs resulting in a slightly different shape of topics) and an impossibility of a priori definition of the optimal number of topic slots for getting the most robust and interpretable topics. Due to this, multiple runs are practiced, with the varying number of slots for topics—usually, 50 to 400, depending on the nature of the dataset. Another inherent problem is dependence of the results upon the length of texts in the dataset: the longer the texts, the more material there is for an algorithm to analyze;

thus, the topics formed of shorter texts are more vulnerable to non-interpretability. Another set of complications lies in malformation of topics from the human viewpoint; for example, among “bad” topics, there are topics dominated by general words, mixed and “chained” topics, or those where one theme splits to several topics (Boyd-Graber et al. 2014, 235–37).

Technical issues about topic modeling are, first, its relatively low feasibility, as the data for topic modeling, especially the real-world datasets, demand several steps of preprocessing (including stemming, lemmatization, and cutting out stop-words) and then either human interpretation or automated quality assessment plus reading by coders; second, it is the dependence on available software and hardware, as collection and processing of large datasets demands a lot of resources.

But, despite the aforementioned discrepancies, topic modeling remains attractive to the scholars, as it has several key (even if arguable) advantages. The first one is that, in comparison with naïve keyword search, the topics unite the texts that might belong to a discussion subtheme but do not contain the keyword, thus enriching our understanding of how people discuss the theme and what it is linked to. The second advantage is that topic modeling may be easily combined with other methods and can serve as a processing tool for other computational goals, including dataset dimensionality reduction. Topic modeling has already proven to be efficient “for a wide range of research-oriented tasks, including multi-document summarization, word sense discrimination, sentiment analysis, machine translation, information retrieval, discourse analysis, and image labeling” (Boyd-Graber et al. 2014, 227).

The third advantage is that the method is believed to be language-independent (given that the language is not hieroglyphic): it means that the algorithms work with words as independent units of analysis, and this approach is suitable for any language. However, today, this assumption is questioned. Topic modeling *per primo* was created for analytical languages such as English, and synthetic languages including Russian, where a role of inflexions for transferring meanings is high, experience additional complications in word preprocessing. Thus, 12 possible case forms of a noun in singular/plural need to be distinguished from numerous forms of the same-root verb in singular/plural in three tenses; for modeling, both the noun and the verb need to “collapse” into singular-nominative (for nouns) or indefinite (for verbs) forms. Moreover, contextual linkages between words arranged, for example, with the help of diminutives, may be lost in stemming.

An overwhelming multitude of descriptions of topic modeling in general, with their advantages and shortcomings (Boyd-Graber et al. 2014; in Russian, Korshunov and Gomzin 2012), as well as particular algorithms, may be found elsewhere (for more detailed example of the procedures of topic modeling applied to a Russian language, see Chaps. 24 and 25). Here, we will focus on how the scholars who deal with the Russian-language datasets develop the topic modeling methods tackling the issues stated above, including



topic quality assessment, and interpret the public discussions in Russia with the help of topic models.

### 23.1.2 *Topic Modeling for the Russian Language*

To our best knowledge, there has so far been no extensive review of how topic modeling has developed for the Russian language. This gap exists despite the fact that Russian-oriented topic modeling studies appear to be one of the most developed beyond the English-language realm, outnumbering German, French, and Spanish in terms of methodological suggestions and cases of application. Also, topic modeling for Russian is considered the most developed among the highly inflected languages like Slavonic ones. Contributions by the scholars working with the Russian-language datasets have become internationally recognized.

To make our review more systematic, we will divide the works into groups. For Russian, topic modeling studies may be divided into *methodological* (that develop, compare, and extend models as well as evaluate their quality), *applied* (that apply topic modeling to extract the meanings from datasets), and *relational* (that relate topic modeling results to other features of the datasets or external factors). Of course, in the case of a rapidly developing method like topic modeling, nearly all the works that use it become methodological, as the method is used in a particular variation which needs to be chosen, grounded, and often reworked or extended. But still we see this distinction as fruitful to structure the results that have been achieved by the scholars. Also, a separate group of works focuses on topic quality assessment. We will also mention topic modeling for short texts like tweets, as, first, modeling for Twitter occupies a separate arena in international topic modeling studies and, second, it has also started to be developed in Russia (for more, see Chap. 30).

The chapter is, thus, organized as follows. In Sect. 23.2, we provide an overview of the methodological papers; here, we summarize the main directions of development of topic modeling for Russian and the main issues that the researchers work upon, including modeling for short texts. In Sect. 23.3, we review the works that deal with topic quality assessment. In Sect. 23.4, we focus on both Russian- and English-language papers about meaning extraction; here, we review the papers that link topic models to other text features, research methods, and contextual knowledge. In particular, we will look at how topic models are used in a wider context of aspect extraction and sentiment analysis. In concluding remarks, we indicate the potential research gaps and the prospects for future studies.

## 23.2 METHODOLOGICAL STUDIES OF TOPIC MODELING FOR THE RUSSIAN LANGUAGE

### 23.2.1 *Model-oriented Works: LDA and pLSA*

In the recent decades, there have been several groups within Russia who have been focusing on various topic modeling algorithms.

Thus, in a sequence of influential works, Koltsova and colleagues have been developing LDA (Koltcov et al. 2014) and a range of extensions and improvements to it. What this group has tried to tackle, with the help of Russian-language datasets from LiveJournal and VKontakte, were the dataset-level and the topic-level issues.

On the level of dataset, the group has dealt with instability of the results of modeling, nonexhaustive LDA results, the quality of sampling and optimization of the number of topics; we will now review the group's achievements in the stated order.

Thus, the topics that appear in two runs of the model are, logically, more stably present in the dataset than those that appear only once and may be occasional. Based on Kullback-Leibler divergence for topic models, the authors have introduced the normalized Kullback-Leibler topic similarity metric (NKLS) for multiple runs (Koltcov et al. 2014). They have used NKLS and also the Jaccard topic similarity metric (Bodrunova et al. 2017) to assess the stability of topics. They have also introduced several LDA extensions to make the results more stable: among others, one is granulated LDA (Koltcov et al. 2016a) similar to the idea of using  $n$ -grams (Batura and Strekalova 2018; Sedova and Mitrofanova 2017a), and another is LDA with local density regularization (Koltcov et al. 2016b).

Doing topic modeling in search for a particular result (say, the public opinion on a particular event or issue), a researcher cannot be sure that the topics (s)he finds in the modeling results represent the full picture of the public discourse. Thus, the group has introduced interval semi-supervised LDA (ISLDA) that links naïve keyword search with probabilistic clustering by attaching word labels to topic slots, thus making the algorithm “crystallize” the topics around keywords (Bodrunova et al. 2013). By attaching the same keyword to several topic slots, a researcher can exhaust the respective theme in the dataset, at the same time getting the topics “thin” enough to see multiple aspects of the discussion (Koltcov et al. 2017).

As to sampling, it is the core procedure of the method that defines in which order the words are sampled (metaphorically, “taken out of the bag of words”) to be probabilistically put together. Most researchers use Gibbs sampling for LDA (Blei et al. 2003), while expectation maximization (EM-algorithm; Mashechkin et al. 2013) and Expectation-Propagation algorithm can also be used (Minka and Lafferty 2002). After introducing the granulated LDA, Koltcov et al. (2016c) have also suggested an optimization for Gibbs sampling for granulated data.

And, last but not least, selecting the optimal number of clusters was tackled. The number of topics is crucial for the results, and, in unsupervised models, it is the only parameter set by the researcher. Usually, multiple runs with varying number of topics are necessary to choose the number closer to optimal, and automation of selection of the number of topics is a separate scientific task. Using the maximum entropy principle, Koltcov et al. (2018) have suggested applying Rényi and Tsallis entropies to find the optimum number of topics. Other groups of scholars have suggested using text representations by dense vectors and sentence embeddings for the same purpose (Krasnov and Sen 2019; Bodrunova et al. 2020).

Topic-level discrepancies of the method were less a focus of attention for this research group, but, in most of their works, they describe the coding experience and the problems of topic interpretability. Thus, they show that human interpretability is linked to the writing style of the authors of the texts in the dataset, as well as to the number of topics, and that the focus of the topic (“war” vs. “Israeli-Palestinian conflict”) matters much for qualitative studies (Koltsova and Koltcov 2013). For dealing with specifically Russian-related issues like the synthetic structure of the language, the group has successfully used pre-developed decisions on lemmatization and have involved contextual interpretations in their works described below, successfully linking the use of topic modeling to qualitative studies of social media and beyond (Koltcov et al. 2017). The group has developed its own software TopicMiner and has worked mostly with texts from the Russian LiveJournal, [VK.com](#), and other social media datasets.

Similarly, the works by Vorontsov and colleagues (e.g. Vorontsov et al. 2015a, 2015b; Vorontsov and Potapenko 2015) have been influential in exploring probabilistic LSA (pLSA) and its modifications based on non-Bayesian regularization. PLSA differs from LDA, as parameters of discrete distributions are estimated via likelihood maximization, with nonnegativity and normality constraints, while LDA uses Dirichlet distribution and additional parameters that help reduce overfitting (Potapenko and Vorontsov 2013, 784). In particular, Vorontsov and colleagues have shown that robust pLSA performs better than LDA for certain tasks; they have also suggested a generalized learning algorithm for probabilistic topic models (PTM), arguing that the currently used algorithms of topic modeling may all be viewed as specific cases of such an algorithm but with differing sets of algorithmic features like regularization, sampling, update frequency, sparsing, and robustness (Potapenko and Vorontsov 2013, 784).

Within this logic, and also advocating for avoidance of unnecessary probabilistic assumptions in natural language processing (Vorontsov and Potapenko 2015, 304), the group has developed ARTM—a non-Bayesian additive regularization of topic models. The authors have argued that, mathematically, “[l]earning a topic model from a document collection is an ill-posed problem of approximate stochastic matrix factorization” and that “[m]any requirements for a topic model can be more naturally formalized in terms of optimization criteria rather than prior distributions. Regularizers may have no probabilistic

interpretation at all” (Vorontsov and Potapenko 2015, 304). ARTM as a regularization framework that integrates many potential regularizers for topic modeling parameters, as the authors have shown. The authors’ claim of high efficiency of their approach, as well as of BigARTM, an open-source library for additive-regularized topic models (Vorontsov et al. 2015a, b), remains unchallenged (Kochedykov et al. 2017). Later, the group has developed TransARTM based on hyper-graph multimodal modeling for “transactional data” where transactions are interactions between network nodes, for example, users on social networks (Zharikov et al. 2018) and have suggested an ARTM improvement by relying on segmental structure of texts (Skachkov and Vorontsov 2018).

Also, this group of scholars has tested two algorithms for the Russian-language short texts, namely biterm topic modeling (BTM) and word network topic model (WNTM) (Kochedykov et al. 2017, 191). These algorithms were also tested against LDA for short texts including tweets (see below) and user queries (Völske et al. 2015).

Despite their varying algorithmic preferences, the research groups led by Koltsova and Vorontsov have collaborated on additive and regularized topic models (Apishev et al. 2016a, b). Also, Vorontsov and colleagues have published important methodological and review papers in Russian, including one on regularization, robustness, and sparsity of probabilistic topic models (Vorontsov and Potapenko 2012).

The similarity between these groups of scholars lies in their focus. First, they both develop the methodologies on the level of dataset, and the level of word in a text corpus mostly remains their secondary concern. This, it seems, stems from the fact that, second, they both treat Russian as “language as such”—just as English is used in topic modeling, often without discussing inherent linguistic or contextual limitations of analytical/inflective languages. This has its advantages, as the language is not treated as “local,” and thus the scholars avoid the “colonial” relations between more universal English and more localized other languages. Also, the authors’ contributions can be easily applied to other languages. But, at the same time, they, to some extent, overlook the word-level of topic modeling, being, of course, well aware of the achievements of Russian computer linguists in developing opinion mining for Russian.

### 23.2.2 *Computer-linguistic Approaches to Topic Modeling*

The latter efforts have, for decades, been concentrating in several groups vaguely linked to each other via the conference on computational linguistics and intellectual technologies called “Dialogue” dedicated to, inter alia, sentiment analysis and aspect detection (for details, see dialog-21.ru). For years, in the conference proceedings and individual papers, the notion of topicality and topic detection has been developing on the level of word semantics and lexical relations. Semantic proximity, ambiguity of meaning, inflections and their impact upon word semantics, sentiment, and other features of lexical units have

been the focus of attention of this sparse “school” or, rather, array of research groups.

Here, we find an understanding of goals of topic modeling that differs from that in the previously described studies. Topic modeling is seen here as a tool for resolution of grapheme-, word-, or fragment-level tasks, such as, for example, relevance detection for automatic text annotation (Mashechkin et al. 2013), automatic content filtration and genre detection (Voronov and Vorontsov 2015), or aspect-based (Rubtsova and Koshelnikov 2015) and non-aspect-based sentiment analysis (Koltsova et al. 2016a; Tutubalina and Nikolenko 2015). Such an approach shifts the very notion of what a topic is: thus, already as early as in 2000, Loukachevitch and Dobrov (2000) noted that topics may be viewed as semantically linked chains of words, thus stating the necessity for a topic to preserve both the grammatical and semantic relations between lexical units. Loukachevitch, Dobrov and their colleagues who, for over two decades, have been dealing with both hard and fuzzy classification methods for the Russian language have developed the notions of “thematic knots” and “thematic text representation” based not on co-occurrence but on semantic relatedness of words in documents (Loukachevitch and Dobrov 2009; for more, also see Chap. 18).

In accordance to this, within computational-linguistic approaches, topic modeling is often used for the tasks that deal with the level of a lexical unit, and not always with great success in comparison with other methods of computational linguistics. Thus, one recent work by Davydova (2019) unites LSA-base modeling with the use of contextual vectors for the task of disambiguation and differentiation of meaning. It successfully unites LSA with word-vector logic to detect thematic relevance of lexemes. In other works (see, e.g., Lopukhin and Lopukhina 2016; Lopukhin et al. 2017), though, it was argued that, for lexical disambiguation, word2vec approaches were more efficient than LDA and other topic modeling approaches based on bag-of-words logic, as topic modeling works on the level of document/dataset.

Thus, the two approaches to developing topic models—the method-oriented one and the computational-linguistic one—seem to be moving forward but without being interconnected, not integrating each other’s achievements into research practice, even despite co-publications and collaboration. There is an evident lack of works that would both develop the topic modeling algorithms *and* have in mind the peculiarities of the Russian language. Despite the evident necessity of integration of the two logics, it is rarely found also for other inflective languages; we see this logic explicitly employed by only one group working in Slovenian (see, e.g., Maučec et al. 2004, and later works). Beside this, several works by computer linguists have suggested decisions for the Russian language, including adding automated labeling to Russian-language topics (Mirzagitova and Mitrofanova 2016) and showing the possibility of domain term extraction by topic modeling (Bolshakova et al. 2013). Automatic topic labeling by a single word or phrase is expected to ease topic interpretation; working upon it continued in the recent years by

comparing quality of two labeling algorithms, namely the vector-based Explicit Semantic Analysis (ESA) and graph-based method, with the former one preferred by the authors (Kriukova et al. 2018).

### 23.2.3 *Topic Modeling for the Russian Twitter*

Unlike for longer texts, short-text modeling for Russian is also done within comparative international context. For instance, there are at least three methodological works that explore topic modeling for the Russian Twitter (Mimno et al. 2009; Sridhar 2015; Gutiérrez et al. 2016; for more, see Chap. 30) while developing multilingual modeling tools. The first two do not discuss individual results for any single language, and the third only observes one difference in description of sports between Russian- and English-language Twitter. Similarly, only a small handful of works applies topic modeling to Russian Twitter to detect substantial meanings or discussion features. Thus, one work (Chew and Turnley 2017) has shown the divergence between Russian- and English-language “master narratives” on Russian cyber-operations.

The works by Bodrunova and colleagues appear to be the only continuous effort (since 2013) to combine topic modeling for Twitter with various other instruments of automated text analysis, also in comparison with other languages (Bodrunova et al. 2019a, c). Thus, we have tested three topic modeling algorithms, namely unsupervised LDA, WNTM, and BTM (Blekanov et al. 2018), and have shown that BTM works best, as measured by normalized PMI and Umass (see below). We have also applied BTM to detect the dynamics of topicality in conflictual discussions (Smoliarova et al. 2018) and have demonstrated that the saliency of topics in time may help detect pivotal points in mediated discussions. Experiments with datasets on Twitter in three languages, including Russian (Smoliarova et al. 2018, Bodrunova et al. 2019a), show that sentiment of tweets is linked to topicality: thus, more interpretable topics are more sentiment-loaded, in particular negativity-loaded (Bodrunova et al. 2019a). Another study (Bodrunova et al. 2019b) has shown that topic interpretability may be linked to topic robustness and topic saliency.

## 23.3 QUALITY ASSESSMENT AND INTERPRETABILITY OF THE RUSSIAN-LANGUAGE TOPICS

All around the world, a vast array of works on topic modeling is dedicated to finding and testing the metrics of its quality. Arguably, these metrics may be divided into those assessing the overall quality of the modeling and those of the topic level. Here, we will review the contribution by the Russian scholars to topic modeling quality studies.

One of the first metrics that were used to assess the modeling itself was *perplexity*—a predictive metric of how well the current distribution matrices

predict the results for new samples. Perplexity has been assessed by Koltcov et al. (2014); they have shown that it is unclear how to use it for qualitative studies in topic modeling, due to inability to establish how dictionary-dependent perplexity is linked to human interpretability of topics. Instead, to measure the quality of modeling, the group has introduced word and document ratios that allow drastically cutting the dictionary of the dataset for computation and suggested a new metric for topic stability measurement. The idea of this metric is that “good” topics are both human-interpretable and stable in multiple runs. As we mentioned above, Koltsova and colleagues have introduced normalized Kullback-Leibler divergence-based metric of topic similarity (NKLS) that allows for detecting similar and stable topics.

They have also improved another traditional metric such as term frequency–inverse document frequency (*tf-idf*). Tf-idf calculates values for each word in a document through an inverse proportion: frequency of the word in a particular document against the percentage of documents the word appears in—which gives a hint on how relevant a given word is in a given document. Tf-idf values allow for calculating the tf-idf coherence metric, to see whether the topics are composed of the words highly relevant for them (Koltcov et al. 2017).

*Coherence* as a measure of topic quality is one of the basic metrics suggested in early years of topic modeling, but later, other automated metrics were introduced. An extensive study of nine automated metrics juxtaposed to the human-coding baseline was performed by Nikolenko (2016). The author has looked at several classes of metrics, including coherence, pairwise pointwise mutual information (PMI), and metrics elaborated by the author based on distributed word representations where each word is represented as a vector in a semantic space (word2vec approach). The author shows that normalized PMI (NPMI) suggested in the paper outperforms PMI as well as other conventional metrics like tf-idf, but vector-based metrics work even better than NPMI. But the question remains whether both NPMI and word2vec metrics work well for short texts, as there is evidence that NPMI marks the topics as good while they remain low-interpretable for human coders (Bodrunova et al. 2019b). For automated topic assessment versus human interpretability, an important attempt to introduce a quality metric has recently been made. Mavrin et al. (2018) have introduced a new interpretability score for top words, based both on assessing the word probability against an external dataset of frequently used words and on pairing the words and assessing the pairs’ coherence. In parallel, Alekseev et al. (2018) have suggested intra-text coherence as a metric to improve interpretability, fairly arguing that topic coherence and interpretability cannot stand for each other, due to a very small percentage of text volume covered by the topic’s top words. Another work has discussed metrics based both on linguistic and probabilistic similarity for hierarchical topic modeling, a special sort of topic modeling (Belyy et al. 2018).

But none of these works has primarily focused on the causes in human (non-)interpretability of the topics, mostly seeing human coding as a baseline—perhaps because, for longer texts, when interpretability was at stake, the



models performed well enough. Thus, Koltsova and Koltcov (2013) have shown that, for long texts like LiveJournal posts, circa two-thirds of the topics are interpretable after LDA has been applied. They have also identified three types of uninterpretable topics: “language” (other than Russian), “style” (writing styles, including offensive language), and “noise” (uninterpretable texts/combinations of texts) (Koltsova and Koltcov 2013, 218). In our pilot studies, though, we have seen that topics for Twitter are less interpretable, with only up to 40–45% identified as such in all the three languages (Bodrunova et al. 2019a, b); thus, it is not the nature of Russian alone that seems to be causing lower topic interpretability in the case of Russian Twitter. Also, we examined the features of top words and found that their negative sentiment could actually raise topic interpretability (Bodrunova et al. 2019a).

### 23.4 USE OF TOPIC MODELING FOR CONTENT INTERPRETATION

In this part of our chapter, we provide a short overview of how the topic models have been applied to social and language studies. A detailed review, though, would demand a separate chapter, as many findings by scholars working with the Russian data are illuminating enough; here, we will only indicate the examples of content-exploring research aiming to demonstrate the variety of possible applications of topic modeling for today’s social science. Also, many works have already been discussed above, and, here, we will only mark the major findings.

The works exploring content may be divided into “purely applicational” and “relational.” The former apply the methods to generate findings relevant for social science; the latter relate such findings to other phenomena or research methods. Also, content-exploring research has scrutinized both social media and text collections beyond them.

In social media studies, topic modeling was first employed to map the agenda of the Russian LiveJournal (Koltsova and Koltcov 2013), finding that the topical structure of posts of the top 2000 Russian LiveJournal authors was quite stable across time and, thus, challenging the notion of dissipative social media agendas. Later, this structural finding was amplified by analyzing the structure of co-commenting communities in LiveJournal (Koltsova et al. 2016b) which showed that the role of individual authors and active commentators was higher than that of topics for the stability of commenting structure.

The two major themes explored via topic modeling have been politics and ethnicity. Thus, Koltsova and Shcherbak (2015) have shown how the bias in political LiveJournal posts correlated with the ratings of the leading parties and presidential candidates in the 2011–2012 election campaigns in Russia. This chapter is an example of combining topic modeling as a dataset reduction instrument with manual coding and descriptive statistics performed for the reduced dataset. Also, Smoliarova and colleagues (2018) have shown that

assessment of topic saliency (i.e. which topics stick out and when) may help detect pivotal moments in development of conflictual political discussions online.

Other important works add to media effects theory, including agenda setting and media framing. They, *inter alia*, demonstrated how agendas on the Ukrainian conflict were gradually diverging on the Russian and Ukrainian TV, thus coming from different framing to building differing agendas (Koltsova and Pashakhin 2017), and that the agendas in news and user comments on Russian regional news portals diverge (Koltsova and Nagornyy 2019). Later, a full-cycle methodology was suggested for co-analysis of news topicality and user feedback (Koltsov et al. 2018). Another group of scholars has also applied LDA to analysis of newspaper coverage on climate change in 2000–2014 (Boussalis et al. 2016) identifying national-level and newspaper-level factors influencing the volume and framing of the coverage.

In a separate line of research, the scholars have explored ethnic content of the Russian social media (Apishev et al. 2016b; Nagornyy 2018a), including detection of most hated ethnicities (Bodrunova et al. 2017), as well as user ethnicity and gender versus attitudes toward ethnic groups (Nagornyy 2018b). Here, topic modeling has produced results unavailable by means of surveys or field research. It has been shown that Americans (outside Russia) and Caucasian nations (inside Russia) provoke the most negative discussion; also, a clear division of attitudes in the Ukrainians-related topics had shown up in LiveJournal much before the Ukrainian conflict started.

Last but not least, beyond the social networking realm, LDA has been applied to Russian and English prose with the aim of facilitating translation of fiction (Sedova and Mitrofanova 2017b) and to a corpus of musicological texts, with the purpose of automated defining syntagmatic and paradigmatic relations between terms (Mitrofanova 2015). In the former work, the authors have added bigrams to the LDA algorithm to detect the differences in various translations of novels. The paper shows high differences in topical structure between English and Russian versions of novels but shows that this diversity may be used for lexical and topical comparison of prose translations. The latter paper is of descriptive nature and was conducted to show that automated text clustering provides the results that are in line with expert knowledge on musicology.

## 23.5 CONCLUSION

Among highly inflected languages, Russian is today the most researched upon in terms of topics models and their applications. The scholars working with Russian-language data have successfully employed the existing methods and have suggested both their universally applicable modifications and new quality metrics. Significant results going much beyond the modeling methodology have been achieved in analysis of social structures of online communication, agenda setting and framing, ethnic studies, and political factors of user discussions.

At the same time, we have identified a gap between method-oriented works that develop topic modeling for Russian as “language as such” and the math-linguistic approach that is Russian-oriented but often sees topic modeling as a secondary, not very useful tool for aspect extraction. Also, there is already a slight “method fatigue” among the researchers who have, to a large extent, reached the limits of the method and are willing to combine it with other methods for resolving tasks in social science. Topic modeling suits well for mapping subthemes inside a stable corpus of documents or understanding the configuration of a particular subtheme beyond the naïve search; it fits a bit less for regular monitoring or precise classification of highly noisy data from social media. There is also lack of studies of human interpretability of Russian-language topics and the factors behind it. In future, we need more discussion on how the properties of Russian influence the modeling results, how text semantics may be used to enhance topic extraction, and whether topic modeling may be used to monitor the dynamics of the discussions. Also, within practically all Slavic languages, no attempts have so far been made to use topic detection in image studies; all these fields are open for rigorous research.

**Acknowledgments** This chapter is supported by presidential grants of the Russian Federation for young Doctors of science, grant MD-6259.2018.6.

## REFERENCES

- Alekseev, Vasily A., Vladimir G. Bulatov, and Konstantin V. Vorontsov. 2018. Intra-Text Coherence as a Measure of Topic Models’ Interpretability. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2018*, 1–13.
- Apishev, Murat, Sergei Koltcov, Olessia Koltsova, Sergei Nikolenko, and Konstantin Vorontsov. 2016a. Additive Regularization for Topic Modeling in Sociological Studies of User-Generated Texts. In *Proceedings of Mexican International Conference on Artificial Intelligence (MICAI)*, 169–184. Cham: Springer.
- . 2016b. Mining Ethnic Content Online with Additively Regularized Topic Models. *Computacion y Sistemas* 20 (3): 387–403.
- Batura, Tatyana, and Svetlana Strelkova. 2018. Podhod k postroeniû rasširenykh tematičeskikh modelej tekstov na russkom âzyke [An Approach to Constructing Extended Topic Models in the Russian Language]. *Bulletin of Novosibirsk State University, Information Technologies Series* 16 (2): 5–18.
- Belyy, Anton, Maria Seleznova, Aleksei Sholokhov, and Konstantin Vorontsov. 2018. Quality Evaluation and Improvement for Hierarchical Topic Modeling. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2018*, 110–123.
- Blei, David M., and John D. Lafferty. 2009. Topic Models. In *Text Mining: Classification, Clustering, and Applications*, ed. Ashok Srivastava and Mehran Sahami, 101–124. Chapman and Hall/CRC.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3: 993–1022.

- Blekanov, Ivan, Nikita Tarasov, and Alexey Maksimov. 2018. Topic Modeling of Conflict Ad Hoc Discussions in Social Networks. In *Proceedings of the 3rd International Conference on Applications in Information Technology*, 122–126. ACM.
- Bodrunova, Svetlana, Sergei Koltsov, Olessia Koltsova, Sergei Nikolenko, and Anastasia Shimorina. 2013. Interval Semi-Supervised LDA: Classifying Needles in a Haystack. In *Proceedings of the Mexican International Conference on Artificial Intelligence*, 265–274. Berlin – Heidelberg: Springer.
- Bodrunova, Svetlana S., Olessia Koltsova, Sergei Koltcov, and Sergei Nikolenko. 2017. Who's Bad? Attitudes Toward Resettlers from the Post-Soviet South Versus Other Nations in the Russian Blogosphere. *International Journal of Communication* 11: 3242–3264.
- Bodrunova, Svetlana S., Ivan Blekanov, and Mikhail Kukarkin. 2019a. Topics in the Russian Twitter and Relations Between Their Interpretability and Sentiment. In *Proceedings of the IEEE International Workshop on Sentiment Analysis and Mining of Social Networks (SAMSNS)*, 549–554. IEEE.
- . 2019b. Topic Modelling for Twitter Discussions: Model Selection and Quality Assessment. *Proceedings of the 6th SWS International Scientific Conference on Social Sciences* 6 (5): 207–214. Sofia: STEF92 Technology.
- Bodrunova, Svetlana S., Ivan Blekanov, Anna Smoliarova, and Anna Litvinenko. 2019c. Beyond Left and Right: Real-World Political Polarization in Twitter Discussions on Inter-Ethnic Conflicts. *Media and Communication* 7 (3): 119–132.
- Bodrunova, S. S., Orekhov, A. V., Blekanov, I. S., Lyudkevich, N. S., & Tarasov, N. A. (2020). Topic Detection Based on Sentence Embeddings and Agglomerative Clustering with Markov Moment. *Future Internet* 12 (9): 144–160.
- Bolshakova, Elena, Natalia Loukachevitch, and Michael Nokel. 2013. Topic Models Can Improve Domain Term Extraction. In *Proceedings of European Conference on Information Retrieval*, 684–687. Berlin – Heidelberg: Springer.
- Boussalis, Constantine, Travis G. Coan, and Marianna Poberezhskaya. 2016. Measuring and Modeling Russian Newspaper Coverage of Climate Change. *Global Environmental Change* 41: 99–110.
- Boyd-Graber, Jordan, David Mimno, and David Newman. 2014. Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements. In *Handbook of Mixed Membership Models and Their Applications*, ed. Edoardo M. Airoldi et al., 225–255. Chapman and Hall.
- Chew, Peter A., and Jessica G. Turnley. 2017. Understanding Russian Information Operations Using Unsupervised Multilingual Topic Modeling. In *Proceedings of International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 102–107. Cham: Springer.
- Davydova, Yulia. 2019. Defining Thematic Relevance of Messages in the Task of Online Social Networks Monitoring in Providing Information-Psychological Security. *International Journal of Open Information Technologies* 7 (4): 11–18.
- Gruber, Amit, Yair Weiss, and Michal Rosen-Zvi. 2007. Hidden Topic Markov Models. *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (PMLR)* 2: 163–170.
- Gutiérrez, Elkin D., Ekaterina Shutova, Patricia Lichtenstein, Gerard de Melo, and Luca Gilardi. 2016. Detecting Cross-Cultural Differences Using a Multilingual Topic Model. *Transactions of Association for Computer Linguistics* 4: 47–60.

- Kochedykov, Denis, Murat Apishev, Lev Golitsyn, and Konstantin Vorontsov. 2017. Fast and Modular Regularized Topic Modelling. In *Proceedings of 21st Conference of Open Innovations Association (FRUCT)*, 182–193. IEEE.
- Koltcov, Sergei. 2018. Application of Rényi and Tsallis Entropies to Topic Modeling Optimization. *Physica A: Statistical Mechanics and Its Applications* 512: 1192–1204.
- Koltcov, Sergei, Olessia Koltsova, and Sergei Nikolenko. 2014. Latent Dirichlet Allocation: Stability and Applications to Studies of User-Generated Content. In *Proceedings of ACM Conference on Web Science*, 161–165. ACM.
- Koltcov, Sergei, Sergei Nikolenko, Olessia Koltsova, and Svetlana Bodrunova. 2016a. Stable Topic Modeling for Web Science: Granulated LDA. In *Proceedings of 8th ACM Conference on Web Science*, 342–343. ACM.
- Koltcov, Sergei, Sergei Nikolenko, Olessia Koltsova, Vladimir Filippov, and Svetlana Bodrunova. 2016b. Stable Topic Modeling with Local Density Regularization. In *Proceedings of International Conference on Internet Science (INSCI)*, 176–188. Cham: Springer.
- Koltcov, Sergei N., Sergei I. Nikolenko, and Elena Y. Koltsova. 2016c. Gibbs Sampler Optimization for Analysis of a Granulated Medium. *Technical Physics Letters* 42 (8): 837–839.
- Koltcov, Sergei N., Sergei I. Nikolenko, and Olessia Koltsova. 2017. Topic Modelling for Qualitative Studies. *Journal of Information Science* 43 (1): 88–102.
- Koltsov, Sergei, Sergei Pashakhin, and Sofia Dokuka. 2018. A Full-Cycle Methodology for News Topic Modeling and User Feedback Research. In *Proceedings of the International Conference on Social Informatics (SocInfo)*, 308–321. Cham: Springer.
- Koltsova, Olessia, and Sergei Koltcov. 2013. Mapping the Public Agenda with Topic Modeling: The Case of the Russian Livejournal. *Policy and Internet* 5 (2): 207–227.
- Koltsova, Olessia, and Oleg Nagornyy. 2019. Redefining Media Agendas: Topic Problematisation in Online Reader Comments. *Media and Communication* 7 (3): 145–156.
- Koltsova, Olessia, and Sergei Pashakhin. 2017. Agenda Divergence in a Developing Conflict: Quantitative Evidence from Ukrainian and Russian TV Newsfeeds. *Media, War and Conflict*. <https://doi.org/10.1177/1750635219829876>.
- Koltsova, Olessia, and Andrey Shcherbak. 2015. ‘LiveJournal Libra!’: The Political Blogosphere and Voting Preferences in Russia in 2011–2012. *New Media & Society* 17 (10): 1715–1732.
- Koltsova, Olessia, Svetlana Alexeeva, and Sergei Kolcov. 2016a. An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2016*, 277–287.
- Koltsova, Olessia, Sergei Koltcov, and Sergei Nikolenko. 2016b. Communities of Co-Commenting in the Russian LiveJournal and Their Topical Coherence. *Internet Research* 26 (3): 710–732.
- Korshunov, Anton, and Andrey Gomzin. 2012. Tematičeskoe modelirovanie tekstov na russkom âzyke [Topic Modeling of Texts in the Russian Language]. *Proceedings of the Institute of Systemic Programming of the Russian Academy of Science* 23: 215–243.
- Krasnov, Fedor, and Anastasiia Sen. 2019. The Number of Topics Optimization: Clustering Approach. *Machine Learning and Knowledge Extraction* 1 (1): 416–426.
- Kriukova, Anna, Aliia Erofeeva, Olga Mitrofanova, and Kirill Sukharev. 2018. Explicit Semantic Analysis as a Means for Topic Labelling. In *Proceedings of the 7th International Conference on Artificial Intelligence and Natural Language Processing (AINL)*, 110–118. Cham: Springer.

- Lin, Chenghua, Yulan He, Richard Everson, and Stefan Ruder. 2011. Weakly Supervised Joint Sentiment-Topic Detection from Text. *IEEE Transactions on Knowledge and Data Engineering* 24 (6): 1134–1145.
- Lopukhin, Konstantin, and Anastasia Lopukhina. 2016. Word Sense Disambiguation for Russian Verbs Using Semantic Vectors and Dictionary Entries. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2016*, 393–405.
- Lopukhin, Konstantin, Boris Iomdin, and Anastasia Lopukhina. 2017. Word Sense Induction for Russian: Deep Study and Comparison with Dictionaries. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2016*, 121–134.
- Loukachevitch, Natalia V., and Boris V. Dobrov. 2000. Issledovanie tematičeskoj struktury teksta na osnove bol'shogo lingvističeskogo resursa [Studying Topical Structure of Text with the Help of a Large Linguistic Dataset]. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2016*, 252–258.
- . 2009. Avtomatičeskoe annotirovanie novostnyh klasterov na osnove tematičeskogo predstavleniâ [Automated Annotation of News Clusters Based on Topical Representation]. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2009*, 299–305.
- Mashechkin, Igor, Mikhail Petrovsky, and Dmitry Tsarev. 2013. Metod vyčisleniâ relevantnosti fragmentov teksta na osnove tematičeskikh modelej v zadače avtomatičeskogo annotirovaniâ [Methods of Relevance Calculation of Textual Fragments for Automated Annotation Based on Topic Models]. *Vyčislitel'nye metody i programirovanie [Computational Methods and Programming]* 14 (1): 91–102.
- Maučec, Miriam S., Zdravko Kačič, and Bogomir Horvat. 2004. Modelling Highly Inflected Languages. *Information Sciences* 166 (1–4): 249–269.
- Mavrin, Andrey, Andrey Filchenkov, and Sergei Koltcov. 2018. Four Keys to Topic Interpretability in Topic Modeling. In *Proceedings of AINL Conference*, 117–129. Cham: Springer.
- Mcauliffe, Jon D., and David M. Blei. 2008. Supervised Topic Models. *Advances in Neural Information Processing Systems* 20: 121–128. Neural Information Processing Systems Foundation.
- Mimno, David, Hanna M. Wallach, Jason Naradowsky, David A. Smith, and Andrew McCallum. 2009. Polylingual Topic Models. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* 2: 880–889. ACL.
- Minka, Thomas, and John Lafferty. 2002. Expectation-Propagation for the Generative Aspect Model. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, 352–359. San Francisco: Morgan Kaufmann Publishers.
- Mirzagitova, Aliya, and Olga Mitrofanova. 2016. Automatic Assignment of Labels in Topic Modelling for Russian Corpora. In *Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics (ExLing)*, 115–118. ExLing.
- Mitrofanova, Olga. 2015. Probabilistic Topic Modeling of the Russian Text Corpus on Musicology. In *Proceedings of the International Workshop on Language, Music, and Computing*, 69–76. Cham: Springer.
- Nagorny, Oleg. 2018a. Topics of Ethnic Discussions in Russian Social Media. In *Proceedings of the International Conference on Digital Transformation and Global Society*, 83–94. Cham: Springer.
- . 2018b. User Ethnicity and Gender as Predictors of Attitudes to Ethnic Groups in Social Media Texts. In *Proceedings of International Conference on Internet Science (INSCI)*, 33–41. Cham: Springer.



- Nikolenko, Sergei. 2016. Topic Quality Metrics Based on Distributed Word Representations. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Developments in Information Retrieval*, 1029–1032. ACM.
- Potapenko, Anna, and Konstantin Vorontsov. 2013. Robust PLSA Performs Better than LDA. In *Proceedings of the European Conference on Information Retrieval*, 784–787. Berlin – Heidelberg: Springer.
- Rubtsova, Yuliya, and Sergey Koshelnikov. 2015. Aspect Extraction from Reviews Using Conditional Random Fields. In *International Conference on Knowledge Engineering and the Semantic Web*, 158–167. Cham: Springer.
- Sedova, Anastasia, and Olga Mitrofanova. 2017a. Tematičeskoe modelirovanie russkoâzyčnyh tekstov s oporoj na lemmy i leksičeskije konstrukcii [Topic Modeling of Russian Texts Based on Lemmata and Lexical Constructions]. *Komp'ûternaâ lingvistika i vÿčislitel'nye ontologii [Computer Linguistics and Computational Ontologies]* 1: 132–144.
- . 2017b. Topic Modelling in Parallel and Comparable Fiction Texts (the Case Study of English and Russian Prose). In *Proceedings of the International Conference on Internet and Modern Society (IMS)*, 175–180. ACM.
- Skachkov, Nikolay, and Konstantin Vorontsov. 2018. Improving Topic Models with Segmental Structure of Texts. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2018*, 652–661.
- Smoliarova, Anna S., Svetlana S. Bodrunova, Alexander V. Yakunin, Ivan Blekanov, and Alexey Maksimov. 2018. Detecting Pivotal Points in Social Conflicts via Topic Modeling of Twitter Content. In *Proceedings of the International Conference on Internet Science*, 61–71. Cham: Springer.
- Sridhar, Vivek K.R. 2015. Unsupervised Topic Modeling for Short Texts Using Distributed Representations of Words. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 192–200. Association for Computational Linguistics.
- Steyvers, Mark, and Tom Griffiths. 2007. Probabilistic Topic Models. In *Handbook of Latent Semantic Analysis*, ed. Thomas K. Landauer et al., 427–448. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Tutubalina, Elena, and Sergei Nikolenko. 2015. Inferring Sentiment-Based Priors in Topic Models. In *Proceedings of the Mexican International Conference on Artificial Intelligence (MICAI)*, 92–104. Cham: Springer.
- Völske, Michael, Pavel Braslavski, Matthias Hagen, Galina Lezina, and Benno Stein. 2015. What Users Ask a Search Engine: Analyzing One Billion Russian Question Queries. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, 1571–1580. ACM.
- Voronov, Sergei O., and Konstantin V. Vorontsov. 2015. Avtomatičeskaâ fil'traciâ russkoâzyčnogo naučnogo kontenta metodami mašinnogo obučeniiâ i tematičeskogo modelirovaniâ [Automated Filtration of Russian-Language Academic Content by Machine Learning and Topic Modeling]. *Computational Linguistics and Intellectual Technologies: Materials of DIALOGUE 2016*. [dialog-21.ru/media/1143/voronovsvorontsovkv.pdf](http://dialog-21.ru/media/1143/voronovsvorontsovkv.pdf).
- Vorontsov, Konstantin, and Anna Potapenko. 2012. Regulârizaciâ, roblastnost' i razrežennost' veroâtnostnyh tematičeskikh modelej [Regularization, Robustness, and Sparsity of Probabilistic Topic Models]. *Komp'ûternye issledovaniâ i modelirovanie [Computer Research and Modeling]* 4 (4): 693–706.



- . 2015. Additive Regularization of Topic Models. *Machine Learning* 101 (1–3): 303–323.
- Vorontsov, Konstantin, Oleksandr Frei, Murat Apishev, Peter Romov, Marina Suvorova, and Anastasia Yanina. 2015a. Non-Bayesian Additive Regularization for Multimodal Topic Modeling of Large Collections. In *Proceedings of the 2015 Workshop on Topic Models: Post-Processing and Applications*, 29–37. ACM.
- Vorontsov, Konstantin, Oleksandr Frei, Murat Apishev, Peter Romov, and Marina Dudarenko. 2015b. BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections. In *Proceedings of International Conference on Analysis of Images, Social Networks and Texts*, 370–381. Cham: Springer.
- Wang, Xuerui, Andrew McCallum, and Xing Wei. 2007. Topical N-Grams: Phrase and Topic Discovery, with an Application to Information Retrieval. In *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM)*, 697–702. IEEE.
- Zharikov, Ilya, Murat Apishev, and Konstantin Vorontsov. 2018. Gipergrafovye mnogomodal'nye veroâtnostnyye tematicheskie modeli tranzakcionnyh dannyh [Hypergraph Multimodal Probabilistic Topic Models of Transactional Data]. In *Proceedings of the Conference “Intellektualizaciâ obrabotki informacii”*, 148–149.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Topic Modeling Russian History

*Mila Oiva*

## 24.1 INTRODUCTION

The power of history rests on its capability to interpret and contextualize past phenomena to explain continuities, differences, and specificities. To a large extent, this process depends on the work and abilities of the historian. With the introduction of computational analysis methods, historians can now use auxiliary means to enhance this process. Topic modeling is a highly useful computational analysis method used to understand and identify the content and patterns of text corpora. This method also allows for additional close reading. Based on complex calculations of word co-occurrences, the method summarizes the text by identifying the topics (i.e. the constellations of words that tend to come up in a discussion) (Mohr and Bogdanov 2013, 547) that can be used to analyze, categorize, and compare text corpora. It can be used to studying the content of a large number of texts as well as a “microscope” that extracts patterns that are otherwise difficult to detect in a small number of texts.

One can view topic modeling as a machine that condenses the studied text into topics that consist of a collection of words connected in a statistically coherent way. For example, if researchers wanted to extract ten topics from past issues of *Pravda* newspaper over the course of a century, they would most likely get one topic consisting of the terms “the Party,” “meeting,” and “resolution”; another topic containing terms such as “team,” “gymnastics,” and “skiing”; and a third topic with terms such as “plan,” “development,” and “production.” If the researcher then extracted more detailed topics from the texts, they would come across terms related to topics such as “Stakhanovite,” de-Stalinization,” and “Perestroika.”

---

M. Oiva (✉)  
University of Turku, Turku, Finland  
e-mail: [mila.oiva@utu.fi](mailto:mila.oiva@utu.fi)

The “machine” of topic modeling is based on a statistical calculation that reveals patterns of co-occurring words in a text. Using probability distribution, topic modeling categorizes individual words in the studied text and analyzes which words are statistically used most often in connection with each other, and these words then form a topic (Brett 2012; Mohr and Bogdanov 2013, 546; Nelson n.d.). In this way, topic modeling allows for the consistent detection and examination of patterns in a large text without the need for sampling.

“Topic modeling” is the term commonly used in text mining to signify a large group of computational algorithms that aim to detect patterns in an unstructured collection of documents. Topic modeling algorithms often use unsupervised machine learning, which allows researchers to identify patterns in the text without prescribing in advance what should be looked for. However, this does not apply to all forms of topic modeling (Isoaho et al. 2019; Boumans and Trilling 2016; Goldstone and Underwood 2014). There are several statistical methods used to calculate the number of topics in a text, with the most frequently used method being the latent Dirichlet allocation (LDA). In addition, there are other types, such as structural topic modeling, dynamic topic modeling, and sub-corpus topic modeling, that produce more nuanced results (Hakkarainen and Iftikhar in press; Isoaho et al. 2019; Roberts et al. n.d.; Blei and Lafferty 2006; Tangherlini and Leonard 2013). This chapter highlights the possibilities of MALLETT, a basic LDA topic modeling tool. MALLETT is a Java-based, open source, and free text analysis program developed by Andrew McCallum (2002).

Topic modeling provides exciting ways in which to analyze the content of larger corpora that encapsulate the content and “see” the text computationally. Therefore, it is not surprising that it has become one of the most popular methods of text analysis in the humanities and social sciences. It has been successfully used to reveal the themes studied texts consist of, the temporal variations of said themes, and the differences between texts (see Gritsenko 2016; Goldstone and Underwood 2014; Tangherlini and Leonard 2013). Numerous scholars in the humanities and social sciences now consider topic modeling a ubiquitous “digital humanities method that solves all the problems” without problematizing or examining which research purposes it can be used for.

The aim of this chapter is to show how topic modeling can be applied to research in Russian and East European studies, with an emphasis on historical research and the choices a researcher will face when using topic modeling. First, the chapter charts the steps that need to be taken when preparing a data set for topic modeling and describes how different choices can affect the results of the analysis. Second, the chapter discusses how the results of topic modeling can be interpreted. Third, the chapter explores the uses of topic modeling in Russian history sources, as well as the associated challenges and opportunities in this context.

## 24.2 PREPARING A TEXT FOR TOPIC MODELING

The results of topic modeling respond to the following questions: What kinds of topics are present in the text? How prevalent are they? Where do these topics appear? The algorithm produces two types of outputs—namely, word–topic and topic–document proportions of the text. It thus creates groups of words that form a topic and identifies how frequently each topic appears in the text. Unlike a human reader, however, the topic modeling program does not understand the text: It only calculates the statistical co-occurrence of words and produces results based on this calculation that offer a statistical perspective of the text. Topic modeling often provides predictable results that are in accordance with the impressions of human readers, but it can also produce nonsensical results or reveal unexpected aspects of the text. It is thus up to the human to decide, based on their knowledge of the data and method, which results should be relied on and which should be discarded.

When beginning topic modeling, and similar to most natural language processing (NLP) analyses, the researcher needs to arrange the data to correspond with the research question, name the documents systematically, preprocess the text, prepare an adequate stop-word list, and determine the specificity of the results by selecting the number of topics. This chapter explains the steps that need to be taken when using Mallet, but it is important to note that different topic modeling algorithms require different approaches regarding the arrangement or naming of data. The choices made at this stage affect the results of the analysis and comprise a crucial element of the process. This stage is also the most time-consuming.

### 24.2.1 *Arranging the Data*

Arranging the data to correspond with the research question is the first step in preparing the text for topic modeling. The arrangement of the data, whether in one large document or in several smaller documents, and according to certain categories, determines what the topic modeling analysis will reveal. Combining all the studied texts into one large document provides a general view of the data set, whereas separating the text according to preset categories allows for the detection of their differences and similarities. For example, if a researcher is simply interested in what kinds of topics exist in studied texts, the texts can be merged into a large text document. If, in contrast, the researcher wishes to study how the topics of a newspaper have evolved over the years, it is useful to arrange the texts so that all the issues of one year or one month are in one document, another year or month in the second document, and so on. This arrangement would then provide data on topic changes on an annual or monthly basis. In a project that studied the reception in Soviet media of French singer and actor Yves Montand when on a tour of the Soviet Union in December 1956, we arranged the data so that each individual article was downloaded as a separate document and saved under a name that indicated the publication date

and newspaper it was published in (Johnson et al. 2019). This then allowed us to detect how the depiction of Montand varied between publications and over time.

### 24.2.2 *Systematic Naming*

The second step, naming the documents in a systematic, expressive, and concise manner, assists the later stages of interpretation of the output. For example, using a document name such as “1953\_F\_Pravda.txt” for an article written by a female journalist and “1953\_M\_Pravda.txt” for an article written by a male journalist condenses the essential information of the document in a comprehensive way and does not confuse the computer program. Document names that are too long or that have spaces between words often cause problems when running the program.

### 24.2.3 *Preprocessing the Text*

Once the data have been arranged and named accordingly, the third step is to preprocess the text. Preprocessing is not mandatory, but it does make the final results clearer due to the simplistic assumptions inherent to the topic modeling algorithm. This step simplifies and standardizes the text for the computational analysis so that certain elements of the text can be revealed. Preprocessing involves various stages, including lemmatization, stemming, the removal of punctuation, and the conversion of uppercase letters into lowercase letters. Lemmatization refers to the process that converts the words in the text into their basic form (e.g. “studying” becomes “to study”). Stemming changes words into their root forms (e.g. “studying” becomes “stud”) (Arnold and Tilton 2015). In the context of Russian-language texts, these processes are highly useful, as Russian is a highly inflected language, and the same words can appear in different forms in a text. Although a human reader recognizes different forms of the same word, the computer program sees them as different words. Because the program does not recognize that the words “studying” and “studied” are different forms of the verb “to study,” the results of the analysis do not attribute the correct weight to the words, thus meaning that the results are distorted. Lemmatization produces more nuanced results than stemming, as the text is simplified to a greater extent (Sharoff et al. 2012; Jabeen 2018). In topic modeling, which explores the statistical relations between words, identifying the correct dictionary-based basic form of the word using lemmatization is often useful. However, simplifying the text does improve the final results if the research looks to explore how different word cases or tenses appear in the text. In these cases, the scholar should not stem or lemmatize the text, as it will result in the decreased quality of the final results.

The Mallet program removes punctuation and converts all the letters into lowercase but does not lemmatize or stem the text. Thus, researchers who wish to lemmatize their Russian-language texts can use programs such as MyStem,

TreeTagger, Language Analysis Service LAS, Python, or R programming language packages for natural language processing (see for example MyStem [n.d.](#); TreeTagger [n.d.](#) or Language Analysis Service LAS [n.d.](#)).

#### 24.2.4 *Composing the Stop-Word List*

The fourth stage of preparing for topic modeling comprises the composition of a stop-word list. In this stage, researchers often remove the most frequent nonmeaning-making words from the text, including “the,” “and,” “or,” and “but.” These are referred to as stop words. While they appear in the texts frequently, they are irrelevant when analyzing the content of texts using statistical means. The Mallet program does not contain a ready stop-word list in Russian, meaning that users need to download their own stop-word lists. Luckily, there are stop-word-lists available online that can be easily applied to Mallet. When downloading a stop-word list in Russian, it should be saved in an eight-bit Unicode Transformation Format (UTF-8) to ensure that the Cyrillic appears correctly.

Although ready-made stop-word lists are available, for serious analysis, it is important to customize the stop-word list for the purposes of the study. Ready-made stop-word lists can contain words that are important in the text, but the specifics of the study might also require the removal of words that appear too frequently in the text. For example, a digitized collection of newspaper articles can contain repeated names of the days of the week (indicating the day of publication of the issue) or the authors of the articles. The researcher might want to add the names of the week and the journalists’ names to the stop-word list to avoid these words being overemphasized and distorting the analysis of the articles. However, having these words in the stop-word list removes them completely from the texts, and this affects the topic modeling results.

#### 24.2.5 *Selecting the Number of Topics*

The fifth stage in the topic modeling process comprises the selection of the number of topics, the  $k$ -value. The difficulty in determining the “correct”  $k$ -value is considered one the greatest weaknesses of the method. There is no one way to determine the correct number of topics, and although there are computational means to determine the optimal number (Isoaho et al. [2019](#); Oiva et al. [2019](#)), the researcher ultimately chooses which  $k$ -value to use. The researcher determines the  $k$ -value depending on how detailed an outcome the study requires. The optimal number of topics depends on the size of the data, the nature of the research question, and the content of the text.

For example, in studies exploring large long-term data sets, scholars have worked with one hundred topics (see Underwood [2012](#)), while in the Yves Montand project, we found ten topics to be meaningful due to the small volume of the data set (Johnson et al. [2019](#)). The data analyzed in the Montand project was preselected and contained only texts that discussed Montand’s tour

during a short period of time. This meant that the expected variation of the topics was small. If the content of the data is not a preselected sample but covers a wide variety of different texts, the number of expected topics will obviously be higher. Similarly, if the researcher's aim is to understand the general variation of the topics in the text, a smaller  $k$ -value may be useful; if the aim of the research is to extract nuances from the text, the  $k$ -value should be higher.

When determining the number of topics, the usability of the results guides the selection of reasonable  $k$ -value, as a large number of topics does not summarize the text in a way that is understandable for humans. For example, many scholars find three hundred topics too large a number to be analyzed. Tangherlini and Leonard argue that the risk of using too many topics is lower than the risk of using too few (2013, 732). Often, if the  $k$ -value is set to be larger than the “actual” number of topics in the text, the researcher can easily understand which topics belong to the same family of topics. For example, in a project that studied the contexts in which Finland was discussed in the Russian news and Russia discussed in the Finnish news, the vast coverage of sports news was divided into different types of sports news segments that were easy for researchers to identify (Gritsenko et al. 2018). As Tangherlini and Leonard state, perhaps the best—and most informal—advice is that given by Doyle and Elkan, according to whom a useful way in which to determine the number of topics is to look at whether the proposed topics are plausible (Tangherlini and Leonard 2013, 731).

While there is no singular and clear-cut way to determine the correct  $k$ -value, running the topic modeling algorithm with different numbers of topics is not difficult. This allows the researcher to determine the best  $k$ -value through exploration. Thus, a good way to explore the optimal number of topics is to run topic modeling with different  $k$ -values and decide, based on the results, which one to focus on. When analyzing the results, it is useful to consider the results of other  $k$ -values and discuss the reasons for selecting the studied number of topics for the study.

For example, when analyzing the articles for our Montand project, we eventually extracted ten topics after testing different  $k$ -values. We ran the algorithm with five, ten, and twenty topics. The analysis with five topics produced overly general topics that did not appear to provide any additional insights. Twenty topics provided more detailed results, but these were so detailed that the topics did not accurately summarize the texts. Ten topics provided sufficiently detailed results that also successfully summarized the studied articles. It is important to note that the topics we received in the smaller number of topics seemed to contain the topics that we received with the larger number of topics. This finding confirmed that the results were not random—rather, they followed certain logic—and that we just needed to select the resolution we wanted to operate with.

Although the launching of a topic modeling project has been depicted as a straightforward process, in reality the process often develops by repeated testing and alterations, as shown in the previous paragraphs. This comprises a



normal way of developing a research project, and to ensure that the process does not become confusing, it is advisable to keep track of the steps taken as well as the reasons behind them. Checking the steps made throughout the process at the end will help to explain and justify the choices made in the research.

After arranging and naming the data, preprocessing the text, customizing the stop words, and selecting the number of topics to be studied, the researcher can then run the data through the topic modeling algorithm. The *Programming Historian* journal offers lessons on how to conduct topic modeling with the Mallet program (Graham et al. 2012).

### 24.3 INTERPRETING THE RESULTS OF TOPIC MODELING

After the topic modeling algorithm has been run, the program produces lists of words that together form a topic. The percentage coverage of each topic in the analyzed documents is also produced. Although the researcher has been working with the text for a while at this stage, the results of the topic modeling only *launch* the actual analytical process.

Since the results of topic modeling can easily be misleading, it is important to validate the choices made and assess the output. At this stage, the researcher should evaluate how the preprocessing choices and modeling parameters affect the results, how well the topics model the phenomenon under investigation, and how interpretable and plausible the outcomes are (Isoaho et al. 2019). The overall assessment of the quality and robustness of the topic model results is crucial, as it forms the basis for the whole analysis. Several scholars have suggested metrics and solutions for computational quality assessment, both concerning the overall and topic-level quality (see Chap. 23; Chuang et al. 2015; Mimno et al. 2011).

After assessing the results, the analysis of the results can begin by naming the topics, as the algorithm only produces groups of words and does not evidence what kind of meaning the words that appear together make. However, it is not always necessary to name the topics. For example, if the researcher is interested in studying the appearance of certain keywords or in creating a relevant sample out of a vast corpus for close reading, it is reasonable to keep the “names” as Topic 1, Topic 2, et cetera. When naming the topics, at first glance, the lists of words may seem nonsensical. However, after some close reading, the common themes become clear. Word lists can be large and probabilistic, and the same word can belong to several topics to a varying extent. Luckily, the sequence of words is meaningful, as it shows the proportion of the words in the topic. The words that are more central to the topic come first in the list, thus helping the researcher to identify what is crucial to the topic. Researchers sometimes name the topics according to the first word of the word list, especially when they are operating with large numbers of topics. When operating with a smaller number of topics, it is useful to ascribe them more precise titles.

To concretize the naming of topics, below are two example topics from the Montand project. Using TreeTagger, the words have been lemmatized into their basic form so that different declinations of one word appear as the same word.

**Topic 1:** montan iv moskva sin'ore simona sssr franciâ pariž pevec vestibûl' dekabr' svâzi večer press gazeta zal dežurstvo šurupova

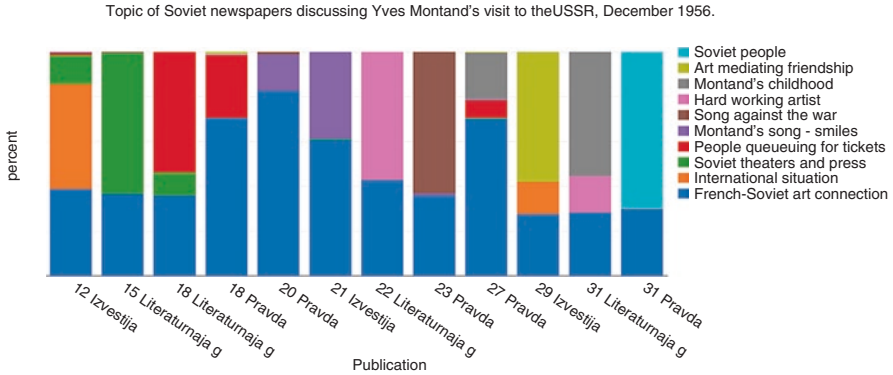
**Topic 2:** montan iv pet' lûdej iskusstvo slov pariž pesnâ lûbov' pevec lûbit' čajkovskogo serdce vystuplenie koncert

Topic 1 contains Montand's name and his spouse, Simone Signoret, and words including "the USSR," "France," "singer," "lobby," "December," "evening," "press," "newspaper," "hall," "shift," and the surname "Šurupova." This combination of words comprises a topic that discusses Yves Montand's arrival in the Union of Soviet Socialist Republics (USSR) from France in December, Soviet newspapers writing about him, and Soviet people eagerly waiting to buy tickets to his concerts. The words "lobby," "hall," "shift," and "Šurupova" refer to a specific article that discussed the queues of Soviet people waiting to buy tickets to Montand's concerts. Topic 1 could be titled the "Reception of Montand."

Topic 2 contains, in addition to Montand's name, words including "to sing," "people," "art," "word," "Paris," "singer," "love," "heart," "performance," "Tchaikovsky" (a famous concert hall in Moscow), and "a concert." This topic describes the songs Montand sang in his concerts, their lyrics, and the positive emotions the articles reported them evoking in the Soviet audiences. This topic could be titled "Montand's emotional songs." As these examples demonstrate, the naming of the topics depends, in addition to the words emerging in the group of words that represent the topic, on the researcher's interpretation and knowledge of the context.

In addition to the word lists, the output shows the percentage coverage of the topics in the studied texts. In terms of the interpretation of topic modeling output, this part is important. A good method of grasping the variation of the topics is to visualize them to allow for an understanding of the proportions of the topics that the program suggests.

The ability of topic modeling to provide new insights into the text becomes especially visible when analyzing the scope and content of the topics. In the Montand project, this was exemplified in an interesting way: We had already read the articles before conducting the topic modeling, meaning that we had a suitable understanding of the topics that we expected to emerge. However, despite our established knowledge, based on a close reading of the articles, the results of the topic modeling provided a new kind of angle. The results highlighted that the topic we had labeled "French-Soviet art connection" consistently prevailed in the articles (see Fig. 24.1). While this result made sense, it is likely that with simple human reading we would not have identified the topic as being so prevalent. It was clear that all the articles discussed the



**Fig. 24.1** The ten topics covered in newspaper articles that discussed Yves Montand's 1956 tour of the Soviet Union

French–Soviet artistic exchange, exemplified by Yves Montand's tour of the Soviet Union. However, the more abstract level of understanding, whereby the transnational interaction addressed the cultural, diplomatic, and political spheres more generally, only became evident when the program had displayed the results.

Topic modeling is considered a particularly useful method for analyzing large data sets. The shortcut it provides in understanding large amounts of texts is so efficient that it is impossible for a human reader to analyze it within a reasonable amount of time. However, the Montand project demonstrates that it is also possible to use topic modeling as a type of “microscope” that provides a statistical overview of the studied text. When topic modeling a smaller data set, however, one needs to remember that the statistical reliability of the results decreases with a smaller amount of data. Nevertheless, the algorithmic reading of texts provided by this quantitative approach complements qualitative research by adding another analysis layer to human reading.

In another example, it is demonstrated how topic modeling can allow for the detection of patterns that are not self-evident to a human reader. This example comprises a project in which I analyzed the annual reports of the Polish Chamber of Foreign Trade between 1950 and 1980. Again, in this project, I had read through the documents before beginning the topic modeling analysis. Based on my reading, I assumed that one topic would dominate all the studied texts throughout the years. However, the result of the topic modeling showed that said topic was dominant in just one document (it was also present in the other documents to a lesser extent). It appears that when reading through the documents, I had read the text in which the topic was dominant at an early stage in my close reading, and when I continued to read, I paid special attention to it. In this way, the topic became important in my mind. This exemplifies how a human reader reads in different ways: While they pay attention to one element, they may omit other issues that seem irrelevant but are not in

actuality. A human reads texts by immediately interpreting the meaning and paying attention to issues of interest. Thus, although one cannot say that computer reading leads to truer results (because the results of computer-assisted analyses can sometimes be nonsensical), this example shows how topic modeling provides other useful and statistically based interpretations of the text.

Topic modeling provides new insights into the studied text because its results are based on the systematic categorization of text without understanding its meaning. In contrast, a human reader interprets the text immediately and pays attention to the significant similarities or differences regarding their understanding of the topic. However, the results of a computational analysis are also dependent on human interpretations in terms of preprocessing, arranging, customizing the stop words, and selecting the number of topics. As well as this, the algorithms behind the program are all based on human construction and selection, and this affects the outcomes of the analysis. The strength of computer reading lies in its inability to understand and its extraordinary capacity to calculate the text in a systematic way. The combination of the algorithms' analysis and human's interpretative skills leads to new findings. Thus, the use of the interpretative power of a human reader and the systematic reading of a computer renders topic modeling a powerful tool.

Alongside small data sets, topic modeling is highly useful for determining general patterns in larger text corpora. The results of topic modeling of large text corpora can help identify interesting sub-corpora, guide further analysis, and even give rise to new research questions (Nelson n.d.). For example, conducting a topic model of all the issues of *Życie Gospodarcze*, a Polish economics newspaper, between 1950 and 1980, led to the creation of new research questions, as it revealed radical temporal alterations of the most frequent topics (see Fig. 24.2).

Upon closer inspection, the results showed that the topics of production planning and the need to increase production prevailed in the newspaper during the entire period. But from 1953 to 1963, the *tone* of the discussion was different from the tone used in 1964 onward. The change in tone was so

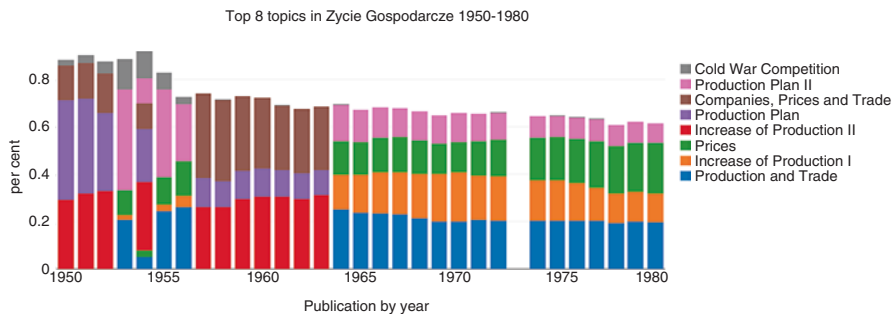


Fig. 24.2 The top eight topics in the Polish *Życie Gospodarcze* newspaper, 1950–1980

prevailing that the topic modeling program identified the production planning discussion before and after 1964 as separate topics.

Sometimes the program classifies a topic—that appears to a human reader as one topic—as two separate topics. This occurs if the researcher sets the number of topics relatively high. These results are often important indicators of radical shifts in the way in which issues are discussed. Conducting topic modeling with more topics can thus lead to the identification of more sensitive topic alterations. This topic modeling result is significant and calls for further exploration. The result also evidences the power of topic modeling in leading to new findings, as these results are extremely difficult to arrive at without the help of statistical computing. A human reader, reading through thirty years of newspaper articles, would most likely have sensed the change of tone but would have been unable to show it in a systematic way.

According to Guldi (2018), 90 per cent of topic modeling results reveals information that we already know. In a sense, this makes researchers confident that the method works and that the countless hours of work done by the preceding generations of scholars have not been done in vain. The remaining 10 per cent reveals insights that have not been identified by preceding research. In order to identify which results belong to the 90 per cent category and which to the 10 per cent category, one needs to understand the context and preceding research. The shift of topics in *Życie Gospodarcze* is unsurprising for a scholar researching Polish economic history, but the comprehensive nature of the change in tone is something that nobody has been able to demonstrate in this way before.

When analyzing the results, one should remember that topic modeling is a probabilistic method, therefore meaning that it provides *probabilities* rather than fixed end-results. The program calculates the probability of the topics several times when processing the text, and the results it provides comprise the average of these calculations. This is visible in practice, as the results are not always absolutely fixed, and topic modeling the same data set several times provides results that are not exactly the same. Thus, the results of topic modeling can give an approximate sense of the topics but not the exact truth. It is useful to view topic models as lenses that allow researchers to view a textual corpus in a different light and scale, where well-informed hermeneutic work is also needed in order to interpret the meaning of the results (Mohr and Bogdanov 2013, 560). For a historian, this variation is usually not problematic because we are accustomed to using interpretative data. When reading an eyewitness description of an event, for example, we don't expect the document to reveal on a word-by-word basis what was said. Rather, we expect an interpretation of the tone of the discussions in the event. Similarly, topic modeling provides the approximate form of the studied text.

In addition to paying attention to the most prevailing topics, it is also worth considering the topics that do not appear in the results to get an overall understanding of the phenomenon. If a topic is prevailing, it means that it has been discussed extensively, but if a topic does not appear in the results, it does not

mean that the topic is not important. Often issues that are taken for granted are not discussed, but the issues that arouse controversies are discussed. When interpreting the results of topic modeling, one should reflect on the limitations embedded within the chosen data set. For example, in the annual reports of the Polish Chamber of Foreign Trade, the issue of press advertising was discussed extensively in the mid-1950s when the chamber promoted the increased use of advertising in the Polish foreign trade. The use of print advertisements increased, but as it then became a normal aspect of foreign trade activities, it did not need to be discussed anymore. In addition, as the reports were written for the Polish Ministry for Foreign Trade, among others, the topics tackled in the reports were issues that the chamber wanted the ministry to be aware of, while the topics it did not want the ministry to be aware of were most likely not discussed.

Thus, one cannot emphasize enough the importance of understanding the context in topic modeling outcomes. To avoid being blindly guided by the topic modeling outcomes, one needs to understand the nature of the source, how the topic modeling processes influence the results, and the relationship between the emerging topics and the issues studied. As Mohr and Bogdanov incisively state, one might think that running any text through a topic modeling program like MALLETT would produce brilliant research. However, it is still the quality of knowledge about the case and the clarity of thinking about the phenomena that determine the utility and richness of the analysis (2013, 559). When using topic modeling in Russian and East European studies, one needs to understand the context of the research.

#### 24.4 TOPIC MODELING: RUSSIAN AND EAST EUROPEAN STUDIES

Although the method of topic modeling itself is universal, its application to studies of Russian and East European history comes with certain requirements and challenges. The languages of the region and their “special” characters (from the perspective of English) are obvious specificities that need to be considered. The program used in the examples above recognizes numerous scripts, including Cyrillic and the Polish alphabet, once the text is in UTF-8.

The greatest vulnerability that prevents the use of topic modeling in studies of Russian history is the dispersed nature of digitized sources and the lack of systematic computer-readable text collections that contain adequate metadata. Russian state archives, museums, libraries, scholarly projects, private companies, and initiatives of nongovernmental organizations, together with private individuals, are digitizing historical sources in increasing amounts. However, the problem lies in the fact that text collections seldom form systematic series that cover long periods that would be needed for systematic computational analysis. Furthermore, too often, “digitization” means uploading a scanned image of a text to the internet, with no computer-readable text or possibility to

download the text to one's own computer for further research. Fortunately, Russian digital history scholars have taken the initiative to collect the available sources into link collections that facilitate finding the available sources (see Perm University Digital Humanities Center's project [n.d.](#); Perm Province Newspapers 1914–1922; Historical Materials and Oral History [n.d.](#)) (for more, see Chap. 20; Kizhner et al. 2019).

Currently, the digitized text collections that can be used for the research of Russian history form a kaleidoscopic landscape. For example, the Russian National Library has not produced systematic digital collections with computer-readable text or metadata, comparable to the digital newspaper collections in many other countries. As a general rule, the memory organizations' digitization efforts produce openly accessible samples of scanned images on nationally interesting topics (see, for example *Artistic Legacies of Anna Akhmatova*). Digitization initiatives are important openings for the popularization of history, but they do not serve the purposes of the big data approach to digital text analysis due to their focus on random samples and general lack of access to computer-readable texts. Furthermore, private companies have systematically digitized Soviet-era newspapers, and their collections form long-time series. However, access to their sources lies behind a paywall, and it is currently impossible to have complete data sets downloaded onto your own computer, which makes the use of big data approaches impossible. Lastly, private initiators and nongovernmental organizations also produce digital collections (e.g. *Prozbito* [n.d.](#)). These collections comprise an important addition to the digitization efforts made by other parties, but they are often only small data sets and are not easily downloadable.

The lack of voluminous collections of digitized historical texts with adequate metadata means that the Big Data approach to Russian historical studies has to be reconsidered. If one wants to apply this approach to studies of Russian history, it is necessary to shift the focus from volume to the other two Vs of Big Data—namely, variety and velocity (Schöch 2013). Instead of seeking a uniform analysis of one vast collection, it is necessary to develop intelligent ways of exploring vast collections of heterogeneous data and linking the results of smaller data sets together to form a meaningful whole. In this way, digital text analyses of Russian historical studies could contribute to the overall development of Big Data studies.

## 24.5 CONCLUSION

As this chapter has demonstrated, using topic modeling in studies of Russian and East European history is useful and can provide new ways to understand the past. For this, understanding the context, the specifics of the data, how the algorithm works, and the stage that the research literature is currently at is extremely important. Without these basic components, the researcher is unable to explain in an adequate manner the results of topic modeling and its wider meaning. The low number of usable digitized collections restrains



opportunities to topic model large text corpora in studies of Russian history. Luckily, as this chapter has demonstrated, topic modeling can also be useful in analyzing smaller data sets. However, if we are willing to understand current complex problems that are studied with the help of big data, we should make an effort to understand the historical roots of these developments. For that, we should develop ways to combine the kaleidoscopic multitude of digitized historical sources into long time series of data that would correlate with the big data produced today.

## REFERENCES

- Arnold, Taylor, and Lauren Tilton. 2015. *Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text*. New York: Springer.
- Blei, David M., and John D. Lafferty. 2006. Dynamic Topic Models. Proceedings of the 23rd International Conference on Machine Learning, Pittsburg, PA. [https://mimno.infosci.cornell.edu/info6150/readings/dynamic\\_topic\\_models.pdf](https://mimno.infosci.cornell.edu/info6150/readings/dynamic_topic_models.pdf).
- Boumans, Jelle W., and Damian Trilling. 2016. Taking Stock of the Toolkit. *Digital Journalism* 4 (1): 8–23. <https://doi.org/10.1080/21670811.2015.1096598>.
- Brett, Megan R. 2012. Topic Modeling: A Basic Introduction. *Journal of Digital Humanities* 2, 1. <http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/>.
- Chuang, Jason, Margaret E. Roberts, Brandon M. Steward, Rebecca Weiss, Dustin Tingley, Justin Grimmer, and Jeffrey Heer. 2015. TopicCheck: Interactive Alignment for Assessing Topic Model Stability. In *Human Language Technologies*, 175–184. Colorado, USA: Denver.
- Goldstone, Andrew, and Ted Underwood. 2014. The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us. *New Literary History* 45 (3): 359–384. <https://doi.org/10.1353/nlh.2014.0025>.
- Golubev, Alexey. in press. Digitizing Archives: Epistemic Sovereignty and its Challenges in the Digital Age. In *Handbook of Digital Russia Studies*, Daria Gritsenko and Marielle Wijermars, eds.
- Graham, Shawn, Scott Weingart, and Ian Milligan. 2012. Getting Started with Topic Modeling and MALLET. <https://programminghistorian.org/en/lessons/topic-modeling-and-mallet>. Accessed 25 Nov 2019.
- Gritsenko, Daria. 2016. Vodka on Ice? Unveiling Russian Media Perceptions of the Arctic. *Energy Research & Social Science* 16: 2–12.
- Gritsenko, Daria, Andrey Indukaev, Viktorija Tkachenko, Ilona Repponen, Alexey Igonen, Mila Oiva, Miika Lampi, and Arturs Polis. 2018. Russia–Finland. Reflection on Neighbours Next Door. Digital Humanities Hackathon, May 25–June 1, 2018. Helsinki.
- Guldi, Jo. 2018. Topic Modeling. Lecture in Digital History Research Methods, University of Oulu, Jyväskylä, Finland, January 29–February 9, 2018. <https://digi-histfinlandroadmapblog.wordpress.com/topic-modeling/>.
- Hakkarainen, Heidi, and Zuhair Iftikhar. in press. The Many Themes of Humanism: Topic Modelling Humanism Discourse in Early Nineteenth-Century German-Language Press. In *Digital Readings of History: History Research in the Digital Era*, ed. Mats Fridlund, Mila Oiva, and Petri Paju. Helsinki: Helsinki University Press.

- Historical Materials. n.d. Accessed November 25, 2019. <http://istmat.info/>.
- Isoaho, Karoliina, Daria Gritsenko, and Eetu Mäkelä. 2019. Topic Modeling and Text Analysis for Qualitative Policy Research. *Policy Studies Journal* 25 (June): 1–25. <https://doi.org/10.1111/psj.12343>.
- Jabeen, Hafsa. 2018, October 23. Stemming and Lemmatization in Python. Data Camp tutorial. <https://www.datacamp.com/community/tutorials/stemming-lemmatization-python>.
- Johnson, Bruce, Mila Oiva, and Hannu Salmi. 2019. Yves Montand in the USSR. Mixed Messages of Post-Stalinist/Western Cultural Encounters. In *Entangled East and West: Cultural Diplomacy and Artistic Interaction during the Cold War*, ed. Simo Mikkonen, Jari Parkkinen, and Giles Scott-Smith, 241–261. Berlin: De Gruyter Oldenbourg.
- Language Analysis Service LAS. n.d.. <https://github.com/jiemakel/las/>. Accessed 25 Nov 2019.
- McCallum, Andrew Kachites. 2002. MALLET: A Machine Learning for Language Toolkit. <http://mallet.cs.umass.edu>.
- Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing Semantic Coherence in Topic Models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272. Scotland, UK: Edinburgh. <https://dl.acm.org/doi/abs/10.5555/2145432.2145462>.
- Mohr, John W., and Petko Bogdanov. 2013. Introduction—Topic Models: What They Are and Why They Matter. *Poetics: Topic Models and the Cultural Sciences* 41 (6): 545–569. <https://doi.org/10.1016/j.poetic.2013.10.001>.
- MyStem. n.d.. <https://tech.yandex.ru/mystem/>. Accessed 25 Nov 2019.
- Nelson, Robert K. n.d. Mining the Dispatch. <http://dsl.richmond.edu/dispatch/>. Accessed 24 Oct 2018.
- Oiva, Mila, Asko Nivala, Hannu Salmi, Otto Latva, Marja Jalava, Jana Keck, Laura Martínez Domínguez, and James Parker. 2019. Spreading News in 1904. The Media Coverage of Nikolay Bobrikov’s Shooting. *Media History* 25 (3): 1–17. <https://doi.org/10.1080/13688804.2019.1652090>.
- Perm Province Newspapers. 1914–1922. <http://permnewspapers.ru/>. Accessed 25 Nov 2019.
- Perm University Digital Humanities Center’s project. n.d. Istoriko-orientirovannyye informacionnyye sistemy. <http://digitalhistory.ru/>. Accessed 25 Nov 2019.
- Prozhito. n.d.. <http://prozhito.org/>. Accessed 31 Oct 2018.
- Roberts, Molly, Brandon Steward, and Dusting Tingley. n.d. STM: An R package for Structural Topic Model. <https://www.structuraltopicmodel.com/>. Accessed 31 Oct 2018.
- Schöch, Christof. 2013. Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities* (November). <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>.
- Sharoff, Serge, Tomaz Erjavec, Anna Feldman, Mikhail Kopotev, Dagmar Divjak, Joakim Nivre, and Marilena di Bari. 2012. Russian Statistical Taggers and Parsers. <http://corpus.leeds.ac.uk/mocky/>.
- Tangherlini, Timothy R., and Peter Leonard. 2013. Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research. *Poetics* 41 (2013): 725–749.

TreeTagger. n.d.. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>. Accessed 25 Nov 2019.

Underwood, Ted. 2012. What Can Topic Models of PMLA Teach Us About the History of Literary Scholarship? The Stone and the Shell: Using Large Digital Libraries to Advance Literary History (blog). <https://tedunderwood.com/category/methodology/topic-modeling/bayesian-topic-modeling/>. Accessed 14 Dec 2012.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Studying Ideational Change in Russian Politics with Topic Models and Word Embeddings

*Andrey Indukaev*

## 25.1 INTRODUCTION

Ideas are both a promising and challenging object for political and social science, especially in the case of Russian studies. The challenges and promises are of methodological and theoretical order. At the theoretical level, the study of the Russian political system quite often discards ideas because of the prevalent rent-seeking behavior of political and economic actors that make interests, not ideas, reign (Gel'man 2016b). However, ideas are of importance for politics and policy processes in any context (Carstensen and Schmidt 2016). Indeed, recent research shows that many aspects of Russian politics cannot be understood without taking into account the ideational dimension, making it a promising direction in the field of Russian studies (Wengle 2015; Dabrowska and Zweynert 2015). Pursuing this direction is challenging. First, ideas are hard to grasp and cannot be studied without a thorough and context-aware examination of meaning expression. Quite often, that implies using methods relying on a “close reading” of texts and requires that the texts where ideas are expressed are available. Within the context of Russian electoral authoritarianism, public expression of ideas in political arenas, through media and other channels, faces constraints that should be taken into consideration. Since the parliament is not a place of political debate, one does not have the data that could serve as a reference for capturing the legislature’s ideological landscape, thus making it difficult to study the ideas in the Russian politics (Lowe et al. 2011; Slapin and

---

A. Indukaev (✉)  
University of Helsinki, Helsinki, Finland  
e-mail: [andrey.indukaev@helsinki.fi](mailto:andrey.indukaev@helsinki.fi)

Proksch 2008). The political context and historical factors are also a source of the “public aphasia”—the lack of the register for the public discussion of the issues of common interest where ideas and ideological positions are expressed (Vakhtin and Firsov 2016).

The proliferation of digital communication, leaving a multitude of textual traces, implies that the study of ideas can significantly widen its scope. It is particularly relevant for Russia, where the Internet became a primary medium for oppositional politics and emerged as an arena for public debate, reducing the “public aphasia” and making the public expression less constrained (for more, see Chap. 2). In addition, a greater volume of textual data and new computational methods of textual analysis are becoming available. They promise a possibility to study the ideational dimension of politics without relying on big corpora of texts frequently and explicitly invoking political ideas, such as parliamentary debates or party manifestos. Indeed, ideational content can be captured even when it is sparsely distributed across a large volume of texts. Digital data and computational methods of text analysis give the opportunity to complete or scale up insights of qualitative analysis based on scarce sources of ideationally dense political texts.

Word embeddings (WE) and topic models (TM) refer to two groups of techniques of text processing and mining that are often used by researchers in social science and humanities (SSH) to study the ideational dimension of politics. This volume provides a discussion of both of them in detail (for more, see Chaps. 26, 23, and 24). An important feature of TM and WE, when used in SSH, is that there are no guidelines on how to apply them to research problems. Instead of treating them as “ready-to-use” methods, a sensible use of TM and WE in SSH implies nowadays developing a research design that takes into account the specificities of the research question and the data at hand. Thus, I find it important to complement contributions to this volume focusing on the overview of the methods with an illustration of their application. To do that, I will use WE and TM to study how ideas, influential in Russian politics, change over time. Particular attention will be accorded to explaining why and how each method was used given the peculiarities of the research question, of the Russian context, and of the data available. To allow the larger audience to apply the chapter’s ideas in their own research, I will give preference to solutions that can be easily implemented in the R programming language ([www.r-project.org](http://www.r-project.org)) and indicate specific R packages used in each case.

The empirical focus of the chapter is on the ideas of innovation, technology, and economic development that played a key role in the modernization agenda of Dmitry Medvedev, on their evolution when the agenda was abandoned, and when some elements of it resurfaced in Russian politics thanks to Putin’s fourth term’s emphasis on digitalization as a key priority. More specifically, I explore the evolving relationship that the innovation, technology, and economic development maintain, in public discourse, with the political liberalization idea. Through its focus on digitalization, this chapter is connected to the first section

of the handbook, which studies digitalization as a sociopolitical phenomenon, and to the chapter of Anna Lowry on the digital economy.

The chapter will be organized as follows. In Sect. 25.2, I present the key ideational dimensions of Medvedev's modernization agenda and their evolution on the basis of a qualitative study. Then I formulate the research questions: (1) Did the ideas associated with modernization and its demise manifest themselves in the Russian media? and (2) How did the concept of digitalization embed itself in the existing set of ideas on technology and politics? Section 25.3 focuses on the overview of the methodology. In Sect. 25.3.2, I discuss topic modeling, and in Sect. 25.3.3, I discuss word embeddings and how it can be used to detect complex semantic relationships between words, revealing social and political representations. Finally, in Sect. 25.4, I apply TM and WE to the Russian media data by using specific approaches given the type of data at hand. I show that the modernization agenda influenced public discourse in Russia by promoting the idea that innovation, technology, and economic development are associated with political and social change, that this idea disappeared from the public discourse, and that the rise of the digitalization agenda did not bring it back.

## 25.2 IDEAS OF MODERNIZATION

As suggested above, qualitative analysis is essential for studying the ideational dimension of politics. Thus, any study of ideas using quantitative techniques should be accompanied by qualitative analysis or should build on such analysis done previously. In this chapter, I will apply TM and WE to study a case that I extensively studied in my doctoral dissertation, using a variety of qualitative techniques (Indukaev 2018). My focus will be on the ideas on the political role of innovation, technology, and economic development. These ideas have played an important role in Russia recently because of the political agenda of modernization that Dmitry Medvedev embraced during his presidency in 2008–2012. They were subject to a major transformation after the modernization program was abandoned. The transformation was a nontrivial one, making it an interesting object for the study. In the following section, I will describe the political context of the case study, outline the key features of the ideational change I am focusing on, and state the research questions and hypothesis.

### 25.2.1 *Politics of Innovation, Technology, and Economic Development During Medvedev's Presidency and After*

Medvedev's political platform positioned him as a more liberal and reform-minded president, without directly opposing him to Putin. Medvedev's political manifesto "Go, Russia!" presented economic and technological modernization of the country as a top priority, but also promised political liberalization. The latter promise relied, in part, on the planned change of the country's political system—including giving more power to the parliament and

making elections more inclusive. However, at the discursive level, the political change was subordinated to the imperative of economic modernization since, in Medvedev's reading of history, "democracy occurred on a mass scale, not earlier ... than when the level of the technological development of the Western civilization made it possible to gain universal access to basic amenities: education, healthcare and information" (Medvedev 2009, n.p.). Technological and economic modernization is presented as a precondition to political modernization: "the technological development is a societal and political task of top priority because the scientific and technological progress is inextricably linked with the progress of political systems" (Ibid, n.p.). In Medvedev's political program, the idea of technological and economic development connoted the idea of political liberalization and social change, while the concept of modernization englobed both ideas.

The key projects associated with Medvedev's modernization agenda aiming at technological and economic development were associated with the ambition of political liberalization. For example, the organizational design of the Skolkovo Innovation Center was influenced by the idea that the state should leave more space for the bottom-up initiative, making its mission focused less on concrete projects but more on the development of an "ecosystem" providing opportunities for unfettered innovative activities (Indukaev 2018). Rusnano, an institution aiming at nanotechnology development in Russia and created under Putin's patronage before Medvedev's election, associated itself with the political ambition of the modernization even more explicitly. Anatoly Chubais, the head of Rusnano, published on the organization's official website a short polemic text intended to defend the idea that modernizing the economy within the nondemocratic context is worth doing. One of his main arguments was that developing an innovative economy will bring to life a class of "scientific and technological intelligentsia," and that "true democracy will appear in the country only when there is a social class that really needs it" (Chubais 2009). Thus, Rusnano's investment in high-tech companies was framed as serving the cause of democratic transition.

Medvedev did not run for a second term and was not able to advance his political program. The ambitions of the political liberalization and social transformation that Medvedev's project included were discarded and have never completely regained their political standing. The situation is different with the ambitions of technological and economic modernization. They lost their priority status after Medvedev's departure. During Putin's third term, the projects inherited from the modernization era were not at the forefront of the political leadership's agenda, sidelined by the conflicts in the international arena and the conservative turn in the country-level politics. Many experts believed that the policy projects associated with modernization would be stopped, in particular Skolkovo (see, for example, Gel'man 2015). However, the project survived, and its budget was not significantly cut. Rusnano and other projects that aligned with the modernization agenda also remained active. Moreover, these organizations managed to align themselves with the import substitutions



agenda, *importozamêsenie* (for more, see Chap. 17), which was central to the field of the technological and economic development (Indukaev 2018). More importantly still, the Medvedev-era economic and technological development policy instruments regained political importance when Putin presented his fourth mandate as being centered around the ambitions of radical “digital transformation” (Rus. *cifrovizaciâ*) and of the “breakthrough” (Rus. *proryv*), the accelerated economic and technological development. Skolkovo, for example, immediately associated itself with Putin’s agenda. Promoting technological and economic development, even though framed more in the digitalization than in the modernization terms, revived some elements of Medvedev’s project.

### 25.2.2 *Research Questions*

The story I outlined above implies that in 2008–2012 innovation and technological and economic development were associated in Russian politics with the promise of political liberalization. This association was coined in the concept of modernization, which also served as a keyword (for more, see Chap. 17) of Medvedev’s political program. When Putin replaced Medvedev as the head of state, the ideational configuration of modernization was discarded; innovation, technology, and economic development were not associated with the promise of liberalization any more.

In my previous research (Indukaev 2018), I detected the described change by qualitative analysis of political speeches and manifestos, policy documents, and institutional arrangements. Thus, the observed change concerns ideas expressed by top-level politicians and reflected in policy decisions. The first question I want to address in this chapter is whether the described ideational configuration and its change was reflected in the way innovation, technology, and economic development were discussed in the media.

The second research objective of this chapter is to extend the scope of my analysis to a new element, which started playing an important role in the political discourse on technology, innovation, and economic development, namely the idea of digitalization. At the top-level of the official discourse, I have not found any indices that Putin’s promise of digitally enabled development was associated with the promise of political liberalization. Instead, digitalization is framed as prioritizing merely the quality of the citizens’ life, and, not less important, the country’s standing in the international arena. In contrast, digital technology was associated with liberalization during Medvedev’s time, who suggested, “The growth of modern information technologies, something we will do our best to facilitate, gives us unprecedented opportunities for the realization of fundamental political freedoms, such as freedom of speech and assembly” (Medvedev 2009, n. p.). Moreover, the development of digital tools promising political empowerment and democratization was actively supported

by the state after 2012 and at the level of local politics (Chap. 3). One may suggest that digitalization is associated with political liberalization in public discourse, despite this association not being explicitly expressed by the political leadership. The second question of this chapter is whether this suggestion is valid.

## 25.3 DATA AND METHODS

To extend the analysis based on writings and speeches by political leaders and policy documents to the ideas expressed by a wider audience, one needs appropriate data, such as Russian media data used in this chapter. Despite the limited freedom of speech, Russian media are not mere translators of the political leadership's perspective and can be used to assess how ideas spread within the Russian public. To analyze these data, I will use two families of computational textual analysis methods, topic modeling (TM) and word embeddings (WE). Apart from methodological reasons, described below, the choice is determined by the fact that topic modeling is among the most widely used among these techniques (Isoaho et al. 2019), and word embeddings could be expected to take the lead in the coming years.

### 25.3.1 Data

Integrum is the largest database of Russian media. It is a commercial product primarily aimed at business clients but is also used by researchers in their studies of Russian language and society (Chap. 17). In this chapter, I use this database. The research strategy is to assemble the corpus focused on technological and economic issues to detect how the political issues appear there. Thus, the query did not include the word *modernization*, since it has explicit political connotations. Instead, the query was made of terms related primarily to technological and economic development, innovation, and digitalization, but not to political change. The query looked as follows:

“иннов\* OR роснан\* OR сколков\* OR венчур\* OR нано\* OR цифровиз\* OR Электронная Россия)/W2 OR (Цифровое Развитие)/W2”,

where the symbols in Cyrillic represent stemmas of Russian words, “OR” is an operator, and “W2” is a context that is considered in the query.

When forming the query, I used wildcards to include all possible morphological forms of a word corresponding to a concept of interest (for the description of the search options, see Chap. 17). The promotion of innovative activities

was an important part of the modernization program, so any form and cognate word for *innovaciá* (innovation) could be used in relation to this program. I use the stem with a wildcard “иннов\*” (innov\*) to capture all these forms. The stem “венчур\*” (venčur\*) refers to venture capital, a specific form of investments in early-stage innovative firms, which was an important reference for the state’s effort to promote innovation. Skolkovo was a flagship project of modernization, so I used “сколков\*” (skolkov\*) to get it mentioned. This part of the query returned a limited number of irrelevant documents because of the Russian word *skolok* (plural *skolki*) meaning, “pricked pattern,” which should not influence the analysis because of the word’s rarity. However, when building a query, a user should be aware that Russian words may have more frequent homonyms, which makes searching tricky.

Nanotechnology promotion and the designated organization Rusnano were major projects of technological development and were also associated with modernization. Again, the “нано\*” (nano\*) part of the query returned a lot of irrelevant documents because of many words beginning with “нанос” (“nanos”), in particular the verb *nanosit’*, the meaning of which is “to inflict.” That included, for example, a significant amount of crime news. The corresponding documents were removed during the corpus preparation. The query “цифровиз\*” (“*cifroviz\**”) aims at various forms of the word *cifrovizaciá* (digitalization), a distinctive term that Putin introduced into political language as a Russian equivalent of digitalization. The query also included the names of two major policy instruments in the field of digitalization, *Elektronnaá Rossiá* (Electronic Russia) and *Cifrovoe razvitie* (Digital Development) programs.

The list of media included 12 sources from the category “Central press,” 2 from “Central news agencies,” 39 from “Central internet media,” 13 from “Central TV and radio,” and 20 regional newspapers, news agencies, and internet media as well as the websites of the Russian government and the presidency. The list composition aimed at a coverage of a variety of sources, including pro-government and more oppositional ones, and also media specialized in technology or economics, regional media, in particular from the regions actively engaged in development projects, such as Tomsk, Novosibirsk (Indukaev 2019), and Tatarstan. The time period covered spans from October 1, 2007, until January 1, 2019, starting about half a year before Medvedev’s inauguration. The query produced 320,000 documents, among which a random selection of 160,000 was used to work with, because of computational limitations of the used setup. The corpus was preprocessed: all characters were transformed to lowercase, and punctuation and number were stripped. Using the collocation functionality of text2vec R package ([CRAN.R-project.org/package=text2vec](https://CRAN.R-project.org/package=text2vec)), the most common multi-word expressions, such as *tehnologičeskoe razvitie* (technological development) were transformed into tokens, such as *tehnologičeskoe\_razvitie*. The stopwords were removed using “stopwords-iso” list from R stopwords package ([CRAN.R-project.org/package=stopwords](https://CRAN.R-project.org/package=stopwords)). The resulting corpus had 45,295,399 tokens.

### 25.3.2 *Topic Modeling*

The fact that modernization became a slogan for President Medvedev’s term sparked active research in the field of Russian studies focused on a vast range of subjects connected to the topic of modernization (Gel’man 2016a; Mustajoki and Lehtisaari 2017). The only use of methods of quantitative textual analysis that I am aware of is the study of “the attitudes of the people towards modernization” in Russia. It was done through exploring the media publication available in the Integrum database (Chap. 17; Laine and Mustajoki 2017). The authors showed how economic, educational, and political preconditions of modernization were debated. To do so, the authors focused on the uses of the word *modernizaciâ* (modernization) that explicitly refer to country modernization. Applying an iterative search procedure to a dataset containing about 10,000 occurrences of the word, the authors extracted 100 passages where the necessary preconditions to the modernization of the country were discussed.

In this chapter, I analyze the political concept of modernization in the context of a larger set of ideas on the political role of innovation, technology, and economic development. This leads me to use the corpus that covers quite a large spectrum of discussions of innovation, technology, and the corresponding government’s activities and to look there for evidence regarding the research questions focused on political ideas. To do that, I will approach the corpus in a way that gives the opportunity to explore the totality of its ideational content but also to focus on particular ideas and concepts. Topic modeling is a great method to start this exploration.

To put it simply, topic modeling is based on the assumption that the documents in a given corpus are generated as a mixture of a determined number of topics—technically, bags of words grouped together based on their tendency to co-occur in the corpus (for more, see Chap. 24). Many variations and extensions of the method are available (for more, see Chap. 23); however, the basic intuition stays the same. Initially, topic modeling was developed as a tool for the retrieval of information that can summarize the thematic content of a large collection of documents. Yet, the key issue that researchers in social sciences and humanities encounter when using topic modeling is that there is no universal rule for interpreting the output of the topic model—the “topic” that emerges as output—as well as no universal way to integrate TM into the research design and to adapt it to a specific research question (Isoaho et al. 2019). In what follows, I discuss how to use the method to answer research questions related to the study of ideas.

In many studies using topic modeling, the thematic content of a corpus is predetermined, and the method is used instead to detect various ideological perspectives on a given topic. In these approaches, a topic or a set of topics, detected by the model, are interpreted as being associated with a specific perspective on the thematic content. Typically, scholars analyze these perspectives using the concept of “issue dimension” (Nowlin 2016) or, more commonly,

that of a frame (see e.g. DiMaggio et al. 2013; Fligstein et al. 2017; Ylä-Anttila et al. 2018). However, quite often, using an issue-specific corpus does not guarantee that the topic model will output topics corresponding to the frames. In the known examples of the use of topic modeling in frame analysis by Fligstein et al. (2017) and DiMaggio et al. (2013), both interpret some sets of topics among the output of the model as corresponding to the frames, while not attributing other topics to any frame. Indeed, the topic model outputs, even within an issue-specific corpus, cannot be automatically seen as frames in most cases (see Isoaho et al. 2019). The association between a topic produced by a topic model and a frame, an “issue dimension,” or any other comparable analytical category is a matter of interpretation which does not rely exclusively on the topic model output but invokes other quantitative or qualitative methods and a theoretical perspective on the issue.

Another use of TMs for studying ideas is to treat TM output more as topics in the literal sense—basically, a coherent theme appearing in the corpus—and not to interpret them as ideational perspectives. When other methods are used to reveal these perspectives, topic modeling can be used to offset the influence of thematic content of analyzed texts on the ideational perspective (Jelveh et al. 2018; Lauderdale and Clark 2014). Other approaches suggest modification of the Topic Modeling algorithm in a way that assumes that word choice in texts is determined both by the ideological perspective and by the topic in the mainstream understanding of a term—the theme of a text (Magnusson et al. 2018; Ahmed and Xing 2010).

In the next section, I apply TM to summarize the thematic content of the corpus. Then, I focus on the topics that are of interest for the study of ideas on innovation, technology, and economic development. I will not use the family of approaches described in the previous paragraph. However, the insight that there are a variety of possible relationships between a topic detected by TM and an ideological perspective—from equivalence to independence—will be key to understanding the limitations of TM-based analysis of ideas. To overcome these limitations, I will use another family of techniques based on word embeddings.

### 25.3.3 *Word Embeddings: Semantic Change and Interpretable Dimensions*

Word embeddings is a family of techniques that represent words as numerical vectors in a way that the relative positions of vectors in the embedding space reflect the relations of semantic proximity of corresponding words (for more, see Chap. 26). To put it simply, the semantic proximity of two words corresponds to the geometric proximity of two vectors that represent the words. The term “word embeddings” is often used to refer both to the vectors representing the words and to the techniques used to obtain the vectors.

The capacity to represent semantic proximity as a geometric one opens avenues for many advanced approaches for studying the ideational dimension in

large corpora. One of these approaches comes from the studies of how the meanings of words change in time: diachronic lexical semantics change. Distributional semantics is one of the advanced computational approaches to semantic change in linguistics. Since the introduction of neural word embeddings, the methods of distributional semantics have manifested significant progress (for a comprehensive review, see Tahmasebi et al. 2018; Kutuzov et al. 2018; and also Tang 2018).

Distributional semantics using WE analyzes semantic shifts following the logic that a relative position of words vectors in multidimensional embedding space is a reflection of the meaning shift. The techniques used may vary. In most cases, researchers use diachronic corpora: for example, a bigger corpus sliced into a set of subcorpora corresponding to consecutive time periods. Then, word embeddings are created for each subcorpus. The vectors representing a word of interest and corresponding to different time periods in the subcorpora may have a different position relative to other words' vectors. That could imply a change of semantics of the word of interest. One of the most used techniques is to focus on the change of semantic “neighbors” of a word—the words whose vectors are the closest to the vector representing the word of interest. For example, a word is expected to have changed its meaning if there was a significant change of the list of top ten words most semantically similar to it. For example, in the word embedding space based on the corpus of English texts dating to the 1850s, the word “broadcast” had words like “seed,” “sow,” and “scatter” as its nearest neighbors, but in the embeddings based on a 1990s corpus, it neighbored “bbc,” “radio,” and “television.” That suggests that the old meaning “throwing seeds” was replaced by the new one, “disseminating information” (Hamilton et al. 2018, 2).

The methods of distributional semantics can be used to analyze ideational change, even though they were not designed for this purpose. Within the study of semantics, the change of word meaning can be explained by “sociocultural” causes (Kutuzov et al. 2018, sec. 2), which opens an avenue for research that interprets semantic change not as a language’s internal affair, but as an indicator of an ideological transformation in the society. Also, the methods of distributional semantics can be used to analyze synchronic variation instead of diachronic change. For example, Azarbondyad et al. (2017) used word embeddings-based metrics of semantic similarity to contrast the viewpoints of Labor and Conservative parties on democracy.

The malleability of words and concepts, the fact that their meaning can vary in time and across different social and political contexts of use, is an essential feature of political language. In the case of the studies of Russian politics, this malleability is of great importance. The change of political language is not primarily associated with public debates on political arenas but is related to opaque political processes that are not always intelligible. Moreover, compared to democratic systems, abrupt political change, and, correspondingly, changes in political discourse are not a feature of Russian politics. At the surface, the political system manifests continuity, and its political discourse is subject to a

gradual change. That makes this change less obvious. Using the methods of distributional semantics, I will show how the concept of modernization, central to Medvedev's political program, gradually changed its meaning while staying an important element of the political discourse on technology, innovation, and economic development.

When analyzing the ideational change through looking at how concepts central to the political discourse change their meaning, a question arises of how to include new concepts in the analysis. Indeed, the ideational configuration can evolve not only through semantic drifts of its key elements. One of the possible paths to ideational change is the rise of new ideas and new concepts. In my inquiry on political ideas on innovation, technology, and economic development, one can easily detect such a new element—the concept of digitalization. In this case, analyzing how new politically important concepts are different compared to old ones becomes an important problem, and word embeddings provide an opportunity to do it.

An important feature of word embedding is that any two words can be characterized not only by the distance between them—that means the *length* of the difference vector obtained by subtracting the vector representing word A from the vector representing word B. In addition to it, the *direction* of the difference is informative, as it can reveal fine-grained aspects of the semantic relationship between two words. For example, the vectors for the words “queen” and “king” can have a relatively small distance and be neighbors in the embedding space built on a sufficient volume of data—a trivial result, since both words designate a monarch. However, if one looks at the direction of the difference between two-word vectors in an embedding space, one can make an interesting observation. The difference between the vector “king” and the vector “queen” will be almost the same as the difference between the vector “man” and the vector “woman” (Mikolov et al. 2013). Thus, one can conclude that it is possible to determine in the embedding space a vector whose direction summarizes the semantic difference between male and female, or in other words, a “gender” dimension. This logic can be extended to other forms of semantic relationship, for example those opposing “rich” and “poor,” or the “affluence” dimension. This approach is thoroughly presented by Kozłowski et al. (2019) in a recent article. The authors calculated word embeddings on Google Ngram's corpus with the help of standard techniques but used the resulting vectors in a way that made it possible for them to extract what they call “cultural dimensions.” The technique assembles antonym pairs for a dimension, such as “poor”-“rich” for the “affluence” dimension, and then calculates the difference vector for each pair and the average difference vector. Thus, any word in the corpus can be located as being more or less related to the affluence. Authors show, for example, how certain activities are located on an “affluence” dimension, tennis, for example, being more related to affluence than boxing. The method was proved to capture cultural representations existing in society and revealed through other means, such as surveys or experimental studies. Comparable approaches are being actively developed, such as one proposed by



Bodell et al. (2019), who modified a word embeddings algorithm in a way that the resulting embedding space dimensions are interpretable.

The approaches like the one developed by Kozłowski et al. (2019) convincingly show that one can construct, in an embedding space, the vectors that capture the semantic relationships corresponding to cultural representations within a society. Such approaches do not have to focus exclusively on culture but can be applied to the study of political ideas and representations. For example, Rheault and Cochrane (2019) use a modified version of the word2vec model, which, based on a parliamentary debate corpus, creates an embedding space that, after applying the dimensionality reduction, produces a vector that represents the opposition between the right and the left ideological perspectives.

One of the questions of this chapter is how a new conceptual element—namely, digitalization—fits into the existing ideational landscape. To answer it, I will rely on the approaches described above by constructing vectors that correspond to key dimensions of this ideational landscape.

## 25.4 RESULTS

In this chapter, I analyze the Russian media to see how its language reflected the events in which the association between political liberalization and innovation, technology, and economic development was brought into political discourse by Medvedev, but vanished after his departure. Also, I am going to look in the media, for evidence that the digitalization agenda revived in the public discourse the political liberalization promise of the modernization agenda.

First, it is important to get an idea of the corpus thematic composition, to understand whether it can be used to answer the research questions. The keywords used in the query, in particular “innov\*,” match with words that have a multiplicity of meanings. For example, the word “innovacija” (an innovation) is often used to refer to new features of products. As a consequence, the corpus is composed of many documents unrelated to the research question. In general, one does not know precisely what is being discussed in the corpus. In this situation, topic modeling is an appropriate method to start with, as it can reveal the composition of the corpus.

The corpus was analyzed using the text2vec library for R (Selivanov and Wang 2018). This library has an advantage of being developed with computational efficiency in mind. Topic modeling is implemented there using WarpLDA algorithm, which is significantly more efficient than other algorithms for Latent Dirichlet Allocation (Chen et al. 2016). The disadvantage of this implementation is that it does not take into account topic correlation and does not allow the inclusion of covariates, such as date, in the topic modeling process, which is possible to do with the much slower Structural Topic Models (STM) version of TM (Roberts et al. 2019). However, as this chapter uses a large corpus necessary to calculate word embeddings, a more efficient but less sophisticated algorithm was preferred.

Given the size of the corpus, the number of topics to be calculated had to be set high. I first ran a model with 50 topics, a part of which were interpreted as relevant to the chapter's research questions. Then, to check the stability of these relevant topics, I ran models with 45 and 55 topics, respectively, and saw the same topics reappear. This technique was used to validate the model, showing that the results are robust enough to resist minor changes in the model parameter—number of topics.

Analyzing TM output showed that the corpus includes many documents that discuss themes irrelevant to the research questions. For example, many topics focus on specific products and services, such as mobile phones or cloud services; others correspond to themes dominating the Russian media space, such as Ukrainian politics or war in Syria. As mentioned, many topics were interpreted as being relevant, such as the one focused on nanotechnology. However, one topic stands up as being central to my analysis.

The topic that is the most prevalent in the corpus is the one that is clearly associated with the modernization agenda. I analyzed the 50 most representative words for this topic (using the `tex2vec` function `get_top_words`, setting `lambda` to 0.3). The list includes various forms of the words and expressions *gosudarstvo* (state), *èkonomika* (economy), *modernizaciã* (modernization), *èloveèeskij\_kapital* (human capital), *srednij\_klass* (middle class), *strana* (country), *reforma* (reform), *otstavanie* (retardation), *proizvoditel'nost'\_truda* (labor productivity), *konkurencii* (genitive for competition), *preobrazovanij* (genitive for transformations), *peremeny* (changes), *strukturnyh\_reform* (genitive for structural reforms), *obšestva* (genitive for society), and *razvityh\_stran* (genitive for developed countries). These words are characteristic for the topic and suggest that it is associated with Medvedev's idea of modernization. First, it appeared in a corpus that was built without using *modernizaciã* (modernization) in the query but focusing on documents mentioning innovation and policy tools in the fields of innovation, technology and economic development. It suggests that the debate of modernization is associated with the debate on innovation and technological and economic development, as it was in Medvedev's program. Second, the topic combines words referring to economic development with words referring to social and political change and reforms. Third, terms like “developed countries” and “retardation” suggest the importance of the rhetoric where the country's modernization is seen as “catching up” with the most developed countries. Last, the words referring to the state are frequent in the topic, suggesting that the modernization is considered at the state level. All these dimensions of modernization are present in Medvedev's manifesto “Go, Russia” (Medvedev 2009). A close reading of the top ten documents where the topic is the most prevalent confirms my interpretation. All the documents debate the ideas that are present in Medvedev's program.

As I described in Sect. 25.3.2., TM is often used to analyze political ideas by associating a topic with a certain ideational perspective. The “Modernization” topic that I described can be associated with a specific perspective on the relationship between political and social change and economic development,

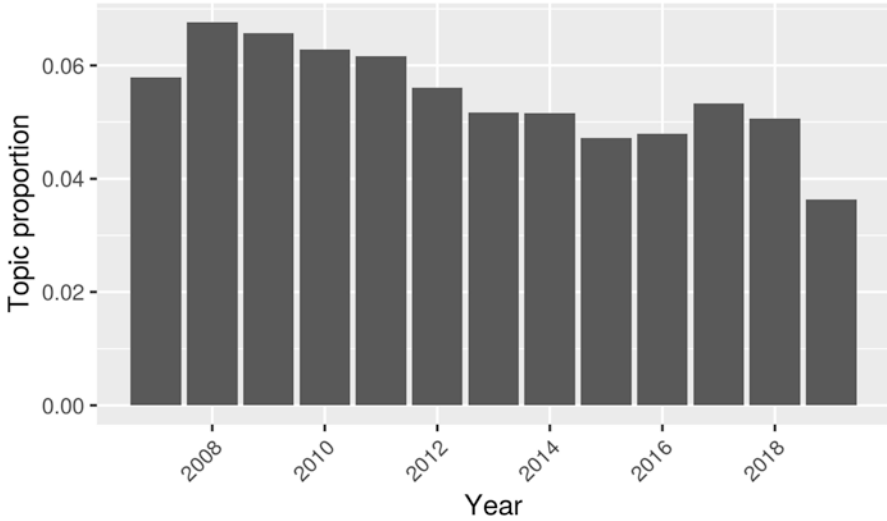


Fig. 25.1 “Modernization” topic prevalence dynamic

technology, and innovation. If one accepts the idea that this topic is an indicator of a certain ideological position, one can attempt to assess the ideational change looking at topic dynamics. The topic prevalence in time corresponds, in part, with what could be expected based on the case description in Sect. 25.2. The topic peaked in 2008 and declined gradually after that (Fig. 25.1). However, after reaching a minimum in 2015, the topic started rising again, with a second peak in 2017, the year Putin started promoting his agenda of digitalization. That could suggest that the revival of technological development as a central element of the political leadership agenda revived the connotation between technological and sociopolitical change. However, using the methods of distributional semantics described in Sect. 25.3.3, I will show that this interpretation does not hold.

Modernization, despite its clear association with Medvedev’s political program, is a concept that has a rich and a malleable meaning, and actors can use it in ways that can highlight various dimensions of the meaning and even attempt to redefine it. I will show that there was a change of meaning which erased the Medvedev era’s ideological association between modernization and political reform, as suggested by the qualitative analysis in Sect. 25.2.

To analyze meaning change, I used a technique based on word embeddings. To calculate word embeddings, I used an implementation of the GloVe algorithm (Pennington et al. 2014) provided by text2vec package in R.

The data used is the same as described in the corresponding section, but with one major adjustment, which is due to my choice of research design appropriate for detecting the change in word meaning and use. Dubossarsky et al. (2019) recently demonstrated that the Temporal Referencing technique

has significant advantages over other approaches of detecting genuine semantic change. The idea of the method is, first, to focus on a limited set of words whose change is going to be studied. Then, the corpus is not sliced into sub-corpora corresponding to different time intervals; instead, the word embeddings are calculated for the entire corpus. However, the corpus is modified: the words of interest are replaced by “time-specific tokens.” It simply means that, for example, if one wants to study how the meaning of the word “modernization” changes from year to year, one replaces the word “modernization” in the documents dated by 2007 by “modernization\_2007” and does the same for every other year. The rest of the words, whose semantic change is not analyzed, stay intact. In this chapter, I use the described method to trace the semantic change of two words: *modernizaciâ* (modernization in singular) and *innovacii* (innovations in plural). To keep the research design simple, I worked with two periods January 1, 2007–January 1, 2012 (label “\_before”), and January 2, 2012–January 1, 2019 (label “\_after”). This change was made because by the end of 2011, it was clear that Medvedev would not keep the presidency, and the promise of political reform was not to be fulfilled.

To detect the change in the meaning of *modernizaciâ* “modernization,” I compare the list of semantic neighbors of *modernizaciâ* before January 1, 2012, to semantic neighbors of *modernizaciâ* after 2012. In addition, I analyze how the list of neighbors changed: which words became less semantically close to the word of interest and which words got closer. When one looks at the “neighbors” of *modernizaciâ*, one sees quite a radical change. In Table 25.1 are provided the top 30 words closest to *modernizaciâ* before and after January 1, 2012.<sup>1</sup> Modernization after 2012 does not have a semantic proximity to democracy (*demokratiâ*), fight with corruption (*bor'ba\_s\_korruptsiej*), reforms (*reformy*), and politics (*politika*), being associated mostly with terms related to technological advances, efficiency (*povyšenie\_èffektivnosti*), and retooling (*tehnìčeskoe\_perevooruženie*). The meaning of the concept changed, and the specific association between the modernization and the promise of political liberalization evaporated after Medvedev’s departure. This result refutes the idea based on the topic modeling analysis of the “modernization” topic, which suggested that Medvedev’s modernization discourse resurrected around 2017: instead, the very concept of modernization changed its meaning.

The fact that the concept of modernization lost its association with political change does not completely rule out that the other key concepts referring to technological and economic development do not manifest it. Digitalization became the most important concept in the technological and economic development projects by the government after 2016. Revealing the association of this concept with the idea of political liberalization in public discourse is a good way to assess the scope of the ideational change that happened after 2012. Would it be possible that the digitalization project took the role of a technology development project bearing also a promise of a political change? To explore it, I used the approach following the insight that vectors in the embedding space can capture “dimensions of cultural meaning” and ideological

**Table 25.1** Top-30 semantic neighbors of the word *modernizaciâ* (modernization) before and after January 1, 2012. Neighbors appearing only in one list are highlighted

<i>Before January 1, 2012</i>	<i>After January 1, 2012</i>
nevozmožna	rekonstrukciâ
modernizaciï	modernizaciï
modernizaciû	masštabnaâ
reforma	optimizaciâ
innovacionnoe_razvitie	zamena
bor'ba_korruptcij	innovacionnoe_razvitie
koreннаâ	proizvodstvennyh_mošnostej
nužna	provedena
politika	tehničeskoe_perevooruženie
nikakaâ	perestrojka
neobhodima	povyšenie_èffektivnosti
demokratiâ	modernizaciï
strukturnaâ	vozmožna
perestrojka	kompleksnaâ
revolûciâ	èkspluataciâ
inovacii_before	obnovlenie
vozmožna	dal'nejšaâ
innovacionnaâ_èkonomika	koreннаâ
èkonomičeskaâ	infrastruktury
strategiâ	nevozmožna
političeskaâ	realizaciâ
ser'eznaâ	kardinal'naâ
reformy	privatizaciâ
demokratizaciâ	reforma
nastoâšaâ	strukturnaâ
dolžna	bama
razvitie	tehnologičeskaâ
diversifikaciâ	transformaciâ

dimensions (Kozłowski et al. 2019, 905). To operationalize this insight, I followed the approach by van Lange and Futselaar (2019), which is less robust than the one proposed by Kozłowski et al. (2019), but requires less data and less preparation. The authors suggest that to create a vector that captures the distance of any word to a given perspective, it is enough to detect the words that indicate the perspective, then to create an aggregate vector that is an average of vectors of each of these words. The proximity of any word to a given perspective is then measured as a cosine distance between the words' vector and the aggregated vector representing the perspective. Van Lange and Futselaar's strategy to construct the aggregated vector is to focus on a concept that epitomizes an ideological perspective and then to find all the words that refer to a concept and do not have multiple meanings.

My analysis was limited by constructing two vectors, corresponding to two ideational perspectives on technology: innovation and economic development. The first perspective—"Political liberalization"—frames these phenomena as associated with social and political change. The second one, however, is focused on development, efficiency, competitiveness—the issues that economy and

technology face and that are seen as apolitical. I labeled this perspective “Economy and technology.” To construct the vectors for each perspective, I first compiled, based on qualitative knowledge of the case, the list of words that are markers of a position. Next, among these words, I selected those that are frequent in the corpus (more than 150 occurrences) because the embedding vector’s quality is sensitive to word frequency. Next, I looked at the top neighbors of each word and selected, based on a qualitative analysis, those that are good markers of the ideational perspective, again selecting only frequent words. Finally, I excluded words with multiple meanings. As a result, the “Political liberalization” vector consists of *reformy*, *demokratiâ*, *demokratii*, *liberal’noj*, *liberalizaciâ*, *liberalizacii*, *svobody*, *prav\_svobod*,<sup>2</sup> *strukturnye\_reformy*. The “Economy and technology” vector consists of *diversifikaciâ*, *diversifikacii*, *diversificirovat’*, *importozamêsenie*, *konkurentosposobnost’*, *povyšenie\_konkurentosposobnosti*, *èkonomičeskoe\_razvitie*.

I calculated the distance to the two aggregated vectors for the vectors representing words that are central to the research question, including the two vectors for *modernizaciâ* before and after January 1, 2012, and the same for *innovacii*. Fig. 25.2 shows that *modernizaciâ* before January 1, 2012, was closer to the “Political liberalization,” but during the period after January 1,

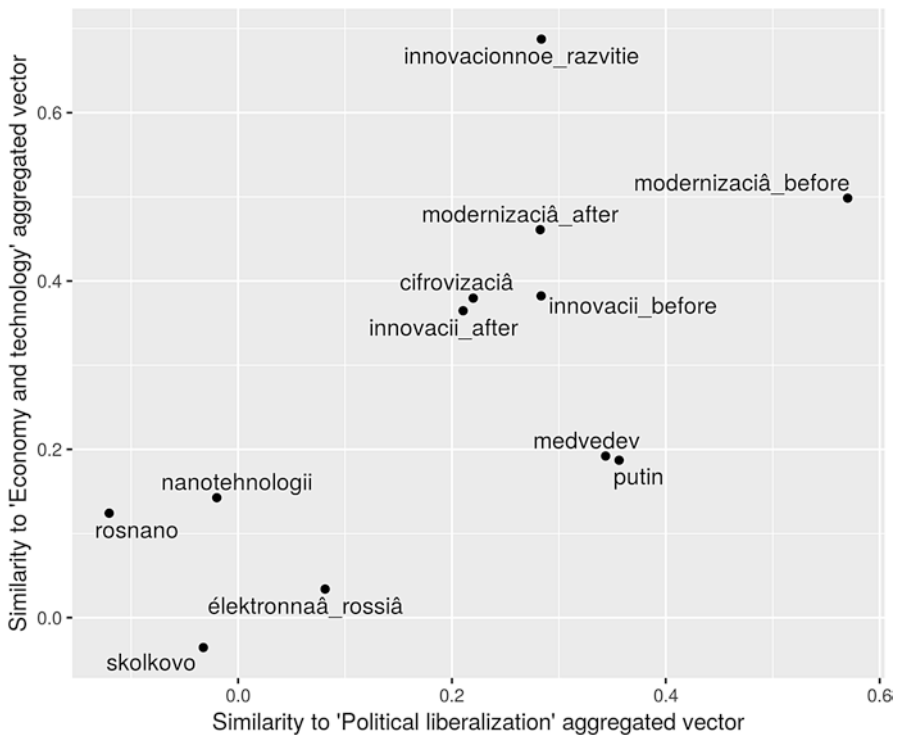


Fig. 25.2 Projection of keywords on a two-dimensional space

2012, it joined other terms, such as *innovacii*, becoming less associated with the idea of political change. The graph also gives an answer to our second research question. The position of *cifrovizaciâ* (digitalization) vis-à-vis the “Economy and technology” and “Political liberalization” vectors is almost identical to that of *modernizaciâ* after January 1, 2012, and that of *innovacii*. That suggests that the digitalization program, in contrast to Medvedev’s modernization, is not associated with the question of political liberalization at the level of public discourse.

## 25.5 CONCLUSION

In this chapter, I used computational methods of textual analysis to study a recent case of ideational change in Russian politics. Based on my prior qualitative analysis of policy documents and political communication of political leadership, I outlined the main contours of this change. When Dmitry Medvedev was president, innovation and technological and economic development were associated with the promise of political liberalisation and played key role in the modernization agenda endorsed by the new president. The modernization agenda was abandoned after the end of Medvedev’s mandate, but technology and innovation regained political importance when Putin chose digitalization as a priority project. However, the political liberalization was not associated with this new promise of technological and economic development.

I showed that the described story of ideational change could be observed at the level of Russian media discussing innovation, high technology, and policy projects of technological development. A good illustration of this case is the semantic change of the concept of modernization. This key concept of Medvedev’s agenda had a close connotation to political and social change but became an apolitical term referring to mere economic and technological development. Moreover, the analysis corroborated the hypothesis that the digitalization, as a new political concept, does not have connotations to political and social change.

From the methodological point of view, this chapter serves as an example of the application of two popular methods of text mining—word embeddings and topic models—to the study of ideational change. An interesting result is that the analysis revealed how topic modeling can provide misleading results and how the methods of distributional semantics and multidimensional ideological mapping can help to avoid an erroneous interpretation. More precisely, believing that a topic can indicate the presence of a certain ideological position across time may lead to errors. As I showed, the topic centered on modernization, while being coherent and well present in the corpus, cannot be seen as a proxy for Medvedev era’s ideational perspective on the political role of technology and economic development. This insight seems to be of great relevance for Russian politics, where ideational change takes place with not much public debate and can be overlooked by a researcher.



This chapter also provides a successful exploration of the possibilities that word embeddings provide for the study of the ideational dimension of politics. The capacity of WE to capture complex semantic relationships gives an opportunity to construct multidimensional “ideational spaces” in which words can be located according to their proximity to two or more ideational perspectives. I believe that this promising and actively developing branch of text analysis will be of great use for Russian studies, given the lack of simple ways to identify ideational oppositions that structure Russian public life.

## NOTES

1. The authentic context is *bor'ba s korruptsiej*, but the stopword “s” (“with”) is deleted.
2. The authentic context is *prav i svobod*, but the stopword “i” (“and”) is deleted.

## REFERENCES

- Ahmed, Amr, and Eric P. Xing. 2010. Staying Informed: Supervised and Semi-Supervised Multi-View Topical Analysis of Ideological Perspective. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1140–1150. EMNLP '10. Stroudsburg, PA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1870658.1870769>.
- Azarbonyad, Hosein, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps. 2017. Words Are Malleable: Computing Semantic Shifts in Political and Media Discourse. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1509–1518. CIKM '17. New York, NY: ACM. <https://doi.org/10.1145/3132847.3132878>.
- Bodell, Miriam Hurtado, Martin Arvidsson, and Måns Magnusson. 2019. Interpretable Word Embeddings via Informative Priors. *ArXiv:1909.01459 [Cs, Stat]*, September 10.
- Carstensen, Martin B., and Vivien A. Schmidt. 2016. Power Through, Over and in Ideas: Conceptualizing Ideational Power in Discursive Institutionalism. *Journal of European Public Policy* 23 (3): 318–337. <https://doi.org/10.1080/13501763.2015.1115534>.
- Chen, Jianfei, Kaiwei Li, Jun Zhu, and Wenguang Chen. 2016. WarpLDA: A Cache Efficient O(1) Algorithm for Latent Dirichlet Allocation. *ArXiv:1510.08628 [Cs, Stat]*, March. <http://arxiv.org/abs/1510.08628>.
- Chubais, Anatolij. 2009. O Modernizacii i Liberalizacii [On Modernization and Liberalization]. 28 December. <http://www.rusnano.com/about/press-centre/first-person/76380>.
- Dabrowska, Ewa, and Joachim Zweynert. 2015. Economic Ideas and Institutional Change: The Case of the Russian Stabilisation Fund. *New Political Economy* 20 (4): 518–544. <https://doi.org/10.1080/13563467.2014.923828>.
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding. *Poetics, Topic Models and*

- the Cultural Sciences* 41 (6): 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>.
- Dubossarsky, Haim, Simon Hengchen, Nina Tahmasebi, and Dominik Schlechtweg. 2019. Time-Out: Temporal Referencing for Robust Modeling of Lexical Semantic Change. *ArXiv:1906.01688 [Cs]*, June. <http://arxiv.org/abs/1906.01688>.
- Fligstein, Neil, Jonah Stuart Brundage, and Michael Schultz. 2017. Seeing Like the Fed: Culture, Cognition, and Framing in the Failure to Anticipate the Financial Crisis of 2008. *American Sociological Review* 82 (5): 879–909. <https://doi.org/10.1177/0003122417728240>.
- Gel'man, Vladimir. 2015. *Modernizacija, instituty i «poročnyj krug» postsovetskogo neopatrimonializma [Modernization, Institutions and the “Vicious Circle” of Post-Soviet Neo-Patrimonialism]*. Saint Petersburg: European University at St.Petersburg Press. [https://tuhat.helsinki.fi/portal/files/46596892/M\\_41\\_15\\_final.pdf](https://tuhat.helsinki.fi/portal/files/46596892/M_41_15_final.pdf).
- . 2016a. *Authoritarian Modernization in Russia: Ideas, Institutions, and Policies*. London: Routledge. <http://ebookcentral.proquest.com/lib/helsinki-ebooks/detail.action?docID=4644407>.
- . 2016b. The Vicious Circle of Post-Soviet Neopatrimonialism in Russia. *Post-Soviet Affairs* 32 (5): 455–473. <https://doi.org/10.1080/1060586X.2015.1071014>.
- Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2018. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *ArXiv:1605.09096 [Cs]*, October. <http://arxiv.org/abs/1605.09096>.
- Indukaev, Andrey. 2018. *Revaloriser La Recherche En Russie Moderne (2000–2018). De La Politique De L'innovation à L'entrepreneuriat Scientifique?* Thesis, Paris Saclay. <http://www.theses.fr/2018SACLN047>.
- . 2019. Siberian Software Developers. In *From Russia with Code: Programming Migrations in Post-Soviet Times*, ed. Mario Biagioli and Vincent Lepinay, 195–212. Durham, London: Duke University Press.
- Isoaho, Karoliina, Daria Gritsenko, and Eetu Mäkelä. 2019. Topic Modeling and Text Analysis for Qualitative Policy Research. *Policy Studies Journal*. <https://doi.org/10.1111/psj.12343>.
- Jelveh, Zubin, Bruce Kogut, and Suresh Naidu. 2018. *Political Language in Economics*. SSRN Scholarly Paper ID 2535453, Social Science Research Network, Rochester, NY. <https://papers.ssrn.com/abstract=2535453>.
- Kozłowski, Austin C., Matt Taddy, and James A. Evans. 2019. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review* 84 (5): 905–949. <https://doi.org/10.1177/0003122419877135>.
- Kutuzov, Andrey, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic Word Embeddings and Semantic Shifts: A Survey. *ArXiv:1806.03537 [Cs]*, June. <http://arxiv.org/abs/1806.03537>.
- Laine, Veera, and Arto S. Mustajoki. 2017. Preconditions of Russian Modernisation: A Media Analysis. In *Philosophical and Cultural Interpretations of Russian Modernisation*, ed. Arto S. Mustajoki and Katja Lehtisaari, 175–190. London; New York: Routledge, Taylor & Francis Group.
- Lauderdale, Benjamin E., and Tom S. Clark. 2014. Scaling Politically Meaningful Dimensions Using Texts and Votes. *American Journal of Political Science* 58 (3): 754–771. <https://doi.org/10.1111/ajps.12085>.

- Lowe, Will, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly* 36 (1): 123–155. <https://doi.org/10.1111/j.1939-9162.2010.00006.x>.
- Magnusson, Måns, Richard Öhrvall, Katarina Barrling, and David Mimno. 2018. Voices from the Far Right: A Text Analysis of Swedish Parliamentary Debates. *SocArXiv*, April. <https://doi.org/10.31235/osf.io/jdsqc>.
- Medvedev, Dmitri. 2009. «Rossiâ, Vperëd!» [“Go, Russia!”]. *Gazeta.Ru*, September 10. [http://www.gazeta.ru/comments/2009/09/10\\_a\\_3258568.shtml](http://www.gazeta.ru/comments/2009/09/10_a_3258568.shtml).
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *ArXiv:1301.3781 [Cs]*, January. <http://arxiv.org/abs/1301.3781>.
- Mustajoki, Arto S., and Katja Lehtisaari, eds. 2017. *Philosophical and Cultural Interpretations of Russian Modernisation*. London; New York: Routledge, Taylor & Francis Group.
- Nowlin, Matthew C. 2016. Modeling Issue Definitions Using Quantitative Text Analysis. *Policy Studies Journal* 44 (3): 309–331. <https://doi.org/10.1111/psj.12110>.
- Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543, Association for Computational Linguistics, Doha, Qatar. <https://doi.org/10.3115/v1/D14-1162>.
- Rheault, Ludovic, and Christopher Cochrane. 2019. Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. *Political Analysis*: 1–22. <https://doi.org/10.1017/pan.2019.26>.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. Stm: Package for Structural Topic Models. *Journal of Statistical Software* 91 (2). <https://doi.org/10.18637/jss.v091.i02>.
- Selivanov, Dmitriy, and Qing Wang. 2018. *Text2vec: Modern Text Mining Framework for R* (version 0.5.1). <https://CRAN.R-project.org/package=text2vec>.
- Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science* 52 (3): 705–722. <https://doi.org/10.1111/j.1540-5907.2008.00338.x>.
- Tahmasebi, Nina, Lars Borin, and Adam Jatowt. 2018. Survey of Computational Approaches to Diachronic Conceptual Change. *ArXiv:1811.06278 [Cs]*, November. <http://arxiv.org/abs/1811.06278>.
- Tang, Xuri. 2018. A State-of-the-Art of Semantic Change Computation. *Natural Language Engineering* 24 (5): 649–676. <https://doi.org/10.1017/S1351324918000220>.
- Vakhtin, Nikolai, and Boris Firsov, eds. 2016. *Public Debate in Russia: Matters of (Dis) Order. Russian Language and Society Series*. Edinburgh: Edinburgh University Press.
- van Lange, Milan, and Ralf Futselaar. 2019. Debating Evil: Using Word Embeddings to Analyse Parliamentary Debates on War Criminals in the Netherlands. *Contributions to Contemporary History* 59 (1). <http://ojs.inz.si/pnz/article/view/322>.
- Wengle, Susanne A. 2015. *Post-Soviet Power: State-Led Development and Russia's Marketization*. New York: Cambridge University Press.
- Ylä-Anttila, Tuukka, Veikko Eranti, and Anna Kukkonen. 2018. Topic Modeling as a Method for Frame Analysis: Data Mining the Climate Change Debate in India and the USA. *SocArXiv*, March. <https://doi.org/10.31235/osf.io/dgc38>.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Deep Learning for the Russian Language

*Ekaterina Artemova*

## 26.1 INTRODUCTION

Deep learning has conquered the natural language processing (NLP) research area in the mid-2010s. Most research publications were focused on English and showed a significant improvement of results on major datasets. However, languages other than English were out of the scope of early deep learning research. Russian-oriented research first appeared on Russian local venues, such as Dialogue, Artificial Intelligence and Natural Language (AINL) and Analysis of Images, Social networks, and Texts (AIST). Early papers addressed such tasks as text classification and part of speech tagging. As of the late 2010s, a new trend for multilingual model development was established, which resulted in quite a few models for Russian, released by non-Russian universities and technology companies, such as Google or Facebook.

The deep learning breakthrough is grounded on the efficient use of large amounts of data, without any handcrafted features. While traditional statistical-based machine learning algorithms require a lot of manual annotation of textual data, the deep learning methods discover hidden patterns in the data without human help. Before the deep learning era, an NLP practitioner had to manually set hundreds of features: starting from such surface features as “is a word capitalized,” or “is there a comma before the word,” up to complex features that try to encode semantics. This resulted, among other things, in creating linguistic corpora, such as Russian National Corpus (<http://www.ruscorpora.ru/>) and OpenCorpora (<http://opencorpora.org>) (for more, see Chap. 17).

---

The original version of this chapter was revised. A correction to this chapter can be found at [https://doi.org/10.1007/978-3-030-42855-6\\_33](https://doi.org/10.1007/978-3-030-42855-6_33)

---

E. Artemova (✉)  
Higher School of Economics (HSE University), Moscow, Russia  
e-mail: [echernyak@hse.ru](mailto:echernyak@hse.ru)

The advantages of the deep learning approach to text processing are two-fold: first, it produces efficient word and sentence representations, sometimes addressed as word and sentence embeddings, which are capable of modeling lexical and grammatical meaning; second, due to multiple nonlinear transformations applied to word and sentence representations inside the deep model, language patterns are learned from actual observations, rather than from human annotations.

Although deep learning treats data differently from traditional machine learning, **training a model** is core to both approaches. The “black box” is a common metaphor to describe what a model is. We can treat any traditional machine learning or deep learning as a black box, which inputs some observations and outputs target labels. For example, for the task of sentiment analysis, the inputs are the users review and the outputs are either “positive” or “negative” labels (for more on sentiment analysis, see Chap. 28). Inside the black box are mathematical functions and objects that have many settings. The model is developed in two stages. During the first stage, which is addressed as the training stage, the model is **trained** to make correct predictions. The model is presented both with the inputs and correct labels and the settings of the model are adjusted so that the model is capable to produce correct answers. The correct labels help to rule the behavior of the model: if the predictions of the model are correct, it is encouraged to behave the same way, otherwise it is punished for incorrect predictions. It is common to say that the model is **supervised** while receiving feedback from correct labels. During the second stage, **prediction** or **inference** stage, the model is only used for prediction and the settings of the model are unchanged.

The procedure of training a model can be compared to the learning-by-doing, educational approach. The model is not presented with any theoretical statements, but rather is trained to perform in an expected way. While traditional machine learning exploits a variety of different models, deep learning apparatus is based on a single notion of artificial neural network, which is loosely inspired by the human brain. The usage of neural networks allows to develop more versatile models, as different types of neural networks are used as building blocks for specific tasks. This makes the models more reusable and easier to adjust to new tasks. Together, the ability to generalize well along with versatility turns deep learning into a powerful framework that is appealing for use in NLP, as it allows to attain a very high performance across many different NLP tasks.

This chapter provides an overview of deep learning applied to Russian NLP. The remainder is organized as follows: Sect. 26.2 introduces the main deep learning architectures, that is, neural network building blocks. Section 26.3 presents a few NLP tasks and Russian-language examples along with the lists of available datasets and models. Section 26.4 concludes.

## 26.2 DEEP LEARNING ARCHITECTURE OVERVIEW

The process of designing a neural network is similar to cooking a layered cake. An NLP practitioner first thinks of a preliminary sketch of the model and understands what the input to the model is, and what the model should output. Next, the layers are added one by one to the model. The lowest layer is responsible for reading the textual input and creating an efficient representation of the input. The upper layers are aimed at solving the task under consideration and preparing the desired output. The middle, or the hidden, layers do most of the work: hidden language patterns are discovered here by applying numerous nonlinear transformations.

Neural network architectures are constructed from various types of building blocks or layers. A crucial component of neural networks is the embedding layer. It maps words to vectors in a low dimensional space. These vectors, referred to as word embeddings, can be manipulated as any mathematical object: not only is it possible to calculate a similarity between them, but also to sum them up or to subtract them. The closer the words are by lexical meaning, the closer the corresponding word embeddings should be. The construction of word vectors can be treated either as a standalone task (see Sect. 26.3.1 of this chapter) or as a part of the whole neural network training. Word embeddings can be seen as a broad understanding of the grammar and semantics. When pretrained on a large general corpus, such as Wikipedia, word embeddings reveal the understanding of general language that can be adopted for a more specific domain. Word embeddings are shallow representations that only incorporate previous training in the input layer of the network. The upper layer of the network still needs to be trained from scratch.

Two major neural network architectures are Feed Forward Networks (FFNs) and recurrent neural networks (RNNs). The main difference between these architectures is in the way these architectures input the textual data.

FFNs treat the input text as a so-called “bag of words,” disregarding grammar and word order and taking only word frequency into account. For example, the sentence “the cat sat on the mat” would be turned into the following tuple: ([the, cat, sat, mat, on], [2, 1, 1, 1, 1]). Although FFNs are capable of combining the words in a meaningful way, it is still a significant disadvantage for languages with free word order, where the word order heavily affects the meaning of the sentence.

The design of RNNs overcomes the disadvantages of FFNs by introducing a built-in memory mechanism that summarizes the input text. RNNs can be seen as a tool which reads the input text sequentially in a word-by-word fashion. As the memory is updated after reading a new word, RNNs are endowed with memorizing the word order and the understanding of the current word context. RNNs are usually treated as the analytical module of the whole network and are rarely used as a standalone component. The power of RNNs is in their ability to produce context-aware word representations, which help, for example, to disambiguate word senses. RNNs often work in tandem with



FFNs, so that the output of the RNN is fed into a FFN, intended for final prediction.

The duality of feedforward and recurrent neural networks is caused by the difference of two widely used models for text representation. In contrast to the bag-of-words model exploited by FFNs, the recurrency targets at language modeling, which is central to the majority of NLP tasks. A language model has a double purpose: first, it assigns a probability to a sequence of words. Second, it predicts the next word based on a number of previously used words. The probability of a sentence, estimated by a language model, is closely related to the quality and correctness of the sentence. Language models help to evaluate the quality of machine translation or any other natural language generation task. By predicting the next word, the language model creates context-dependent word and sentence representations.

Although one of the early works by Bengio et al. (2003) shows that FFN can be treated as a language model, RNN outperforms by far FFNs for the task of language modeling. Finally, technical limitations of vanilla RNNs are resolved by gated architectures, such as long short-term memory (LSTM) and gated recurrent unit (GRU) networks. Both LSTM and GRU are very efficient as language models and are de-facto baseline NLP architectures.

The building blocks of neural network architectures are not limited to feedforward and recurrent layers. Convolutional neural networks (CNNs) are an extension of the FFN architecture. CNNs excel in discovering local patterns. They can be seen as a magnifier, which moves over a word sequence and identifies important features. CNNs are often utilized on the lowest network layers to process not words, but rather characters, to discover long orthographic and derivational patterns. Many applications in Russian, a morphologically rich language, benefit from the ability of CNNs to capture derivational word suffixes and endings. It helps to handle rare words, such as family names, terminology, toponyms, and slang, as well as to take surface features into account (Fig. 26.1).

When compared to feedforward and convolutional neural networks, recurrent neural networks are much slower to train, since they pose long-term dependencies and it is hard to parallelize recurrent computations. The recently introduced transformer layer combines the best of two approaches. It consists of multiple feedforward layers and a powerful attention mechanism that is analogous to human attention in the same way the artificial neural networks model biological neural networks. The attention mechanism directs focus to a certain part of the task while maintaining a background understanding of the whole task. It models word-by-word interactions on each feedforward layer, so that different types of dependencies are considered. The self-attention mechanism is used both to produce context-aware word embeddings, and also measures how strong the dependencies are between the words.

At the core of the recent paradigm shift in NLP, are pretrained language models that are built with rare exceptions with transformer blocks. Not only word embeddings, but the whole neural network is now pretrained as a language model. It becomes possible, since the language modeling objective, next

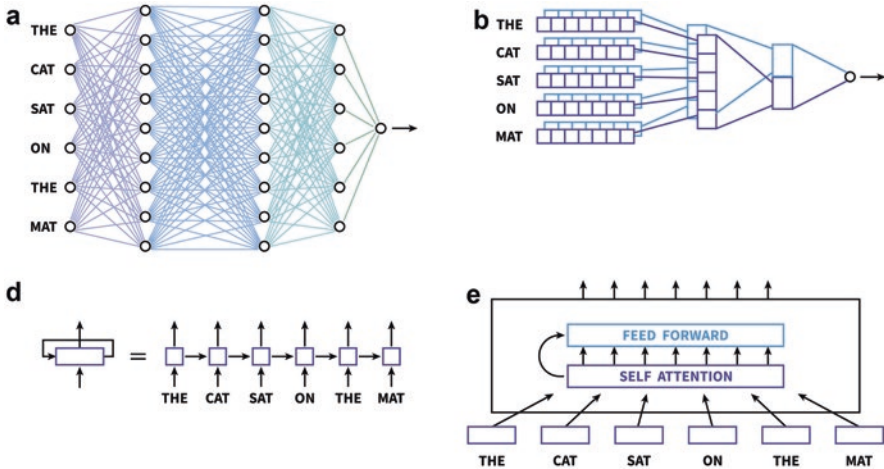


Fig. 26.1 Neural network layers. (a) feed forward layer, (b) convolutional building layer, (c) recurrent layer, (d) transformer layer

word prediction, does not require any human annotation. The training data comes for free and the amount of training data available in almost every language are potentially unlimited. Transformer-derived language models seem to capture many facets of language relevant for other NLP tasks. When pretrained on large and diverse corpora, they can be fine-tuned for downstream tasks and surpass previous results in almost every application (for more on corpus linguistics, see Chap. 17).

Despite having excellent results for NLP tasks, neural networks have some disadvantages. First of all, they are frequently treated as black boxes as they lack interpretability. There have been several attempts to find a plausible explanation of how exactly neural networks operate. One of the hypotheses states that the neural network follows the common linguistic pipeline of staged processing of the language. It has been shown that if the neural network is deep enough, lower layers may become morphology aware, middle layers model syntactic dependencies, while the upper layers discover complex semantic patterns. Secondly, deep learning technologies require a lot of data and computational sources. Modern computations, which may take about a month of training, are worth thousands of dollars. Thirdly, ethical concerns arise when training a model on textual data collected from the Web. A model can become unfair when trained on all misconceptions, offensive and biased judgments, fake news and false facts, published on the Web (for more on Runet, see Chap. 16). Finally, the fluency of text generation models may lead to potentially harmful usage. New breed of text generation models impresses with their ability to generate coherent text from minimal prompts. When provided with a headline, such a model will compose a news story; when provided with a movie title, it will compose a movie plot. Text generation models can often

give the appearance of common sense and intelligence, so that it may become quite challenging to recognize, whether a text was composed by a human or by a machine. This frustrates research progress in language generation development, as, it sees, text generators may be misused to generate fake news or propaganda or to increase the amount of spam on the Web. It is of crucial importance, the release of a powerful text generator is accompanied with a tool, which is capable of recognizing machine generated text and can be used to tackle online disinformation.

## 26.3 NLP TASKS

### 26.3.1 *Word Embeddings: How Do Computers Understand Lexical Meaning*

Word embedding stands for a group of methods which are used to map words from a large vocabulary, to vectors. These vectors should consist of real numbers, have few zeros and be of relatively small dimensionality: it is common to construct 300-dimensional word embeddings. These vectors are treated as mathematical objects: not only similarity (or distance) between them can be computed, but also they can be added together or subtracted. At the core of numerous methods for word embedding construction is the distributional hypothesis: words that occur in the same contexts tend to have similar meanings (Harris 1954). Word embedding models are trained on large text corpora. They aim at finding words that share contexts and represent them with such vectors that would be close, according to a mathematical similarity measure. For example, the embeddings of such words as *kofe* (“coffee”) and *čaj* (“tee”) should have a high similarity degree, since they are used in a similar way, along with the words *pít’* (“to drink”), *čaška* (“cup”), *nalít’* (“to pour”), et cetera. What is more, advanced word embedding models allow to conduct arithmetical operations: *kofe* (“coffee”) to *utro* (“morning”) = “*čaj* (“tee”) to *věčer* (“evening”); *Moskva* (“Moscow”) to *Rossiá* (“Russia”) = *Berlin* (“Berlin”) to *Germaniá* (“Germany”). Of course, these associations are corpus-specific and may not be present in other models. The examples are provided by RusVectors (<https://rusvectors.org>), a free online service which provides, and which computes semantic relations between words in Russian and provides pretrained distributional semantic models (word embeddings), including contextualized ones.

Word embeddings may serve as input to a neural network model, which further will be trained for any downstream task, and may be used as a stand-alone model for studies of language usage. Word embeddings help to detect semantic shift, caused by either diachronic (Kutuzov et al. 2018) or social changes (Solovyev et al. 2015). Bilingual word embeddings help to develop dictionaries and find similar concepts in different languages (Gordeev et al. 2018).

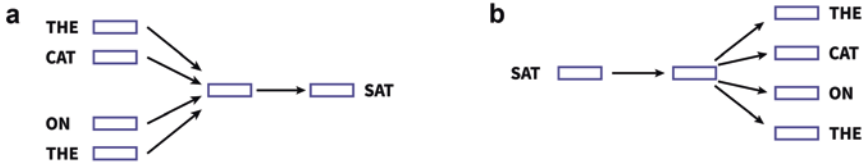


Fig. 26.2 Word2vec configurations. (a) continuous bag of words, (b) skip-gram

The most popular word embedding model is word2vec (Mikolov et al. 2013) and its extension fasttext (Joulin et al. 2017). Word2vec exploits neural networks to compute word embeddings. It has two configurations: in a continuous bag of words, CBOW, it predicts a word based on surrounding words (two to the left and two to the right). In skip-gram, SGNS, it predicts surrounding words based on the given central word (Fig. 26.2).

SGNS is a de-facto state of the art model for word embeddings and is almost a default choice for many NLP applications for the English language. However, for the Russian language SGNS might not be the best choice. When trained on raw texts, SGNS does not take into account the derivational forms of the words. As a result, for the word *kot* (a cat) there might be up to ten possible vectors for each possible derivational form. This would make a similarity measure almost invalid, since the closest words to the vector *kot* (a cat) would be the vectors of the derivational forms *kotu* (to the cat), *kote* (about the cat), et cetera. To overcome this issue, a preliminary normalization is required to replace each word with its base form. Normalization methods, however, may either have limited vocabulary and introduce some mistakes while processing out of vocabulary words or require word embeddings. This vicious circle is broken by the fasttext model that does not modify word2vec mathematics but treats the words differently. Instead of computing a single vector for a given word, it computes multiple vectors for all character n-grams (sequences of two to five characters) and then combines them to get the final vector.

Fasttext allows to capture such properties of rich morphology in Russian as derivational patterns in suffixes and endings. It is strongly recommended to use fasttext for the Russian language as the word embedding model. See Table 26.1 for available pretrained word embedding models and Table 26.2 for word embedding training tools.

Word embedding models often fail when faced with such complex language phenomena as antonyms or homonyms. Although word embeddings are exceptionally powerful for finding words that share a similar meaning, they often mistake for words that have opposite meanings, such as *proigrat'* (“to lose”) or *vyigrat'* (“to win”), as they occur in similar contexts. Word embedding models suffer from polysemy and homonymy. Such words as *luk* (“onion” or “bow” or “a look”) and *zamok* (“castle” or “lock”) get a single vector, despite having multiple sense. A few models, such as AdaGram (Bartunov et al. 2016) and SenseGram (Pevlevina et al. 2016), try to overcome this issue by

**Table 26.1** Word embeddings for Russian

<a href="http://rusvectors.org">http://rusvectors.org</a>	Multiple Russian-only word and sentence embeddings (Kutuzov and Kuzmenko, 2017)
<a href="http://docs.deeppavlov.ai/en/master/features/pretrained_vectors.html">http://docs.deeppavlov.ai/en/master/features/pretrained_vectors.html</a>	Multiple Russian-only word and sentence embeddings
<a href="https://fasttext.cc/docs/en/crawl-vectors.html">https://fasttext.cc/docs/en/crawl-vectors.html</a>	Google fasttext embeddings trained with limited preprocessing for 157 languages
<a href="http://vectors.nlp.eu/repository/">http://vectors.nlp.eu/repository/</a>	Thirteen Russian word embedding models trained with clearly stated hyperparameters, on clearly described and linguistically preprocessed corpora
<a href="https://github.com/bheinzerling/bpemb">https://github.com/bheinzerling/bpemb</a>	BPE embeddings for 275 languages

**Table 26.2** Tools to train word embeddings

<i>Library</i>	<i>Language</i>	<i>URL</i>
Gensim	Python	<a href="https://radimrehurek.com/gensim/">https://radimrehurek.com/gensim/</a>
AllenNLP	Python	<a href="https://github.com/allenai/allennlp">https://github.com/allenai/allennlp</a>
flair	Python	<a href="https://github.com/zalando-research/flair">https://github.com/zalando-research/flair</a>
fasttext	C++/terminal interface	<a href="https://fasttext.cc">https://fasttext.cc</a>
Deeplearning4j	Java/Scala	<a href="http://deeplearning4j.org">http://deeplearning4j.org</a>

simultaneous word sense disambiguation, and word embedding training. However, current pretrained language models are a much more efficient solution to this issue, as they search for context-dependent word embeddings.

As of the mid 2010s, using pretrained word embeddings as an input to any machine learning or deep learning has become a must. The word embeddings can be fine-tuned while training the model for a downstream task or remain constant. Fine-tuning of word embeddings may help to resolve some issues related to antonyms or homonyms. When fine-tuned for sentiment classification (for more on Sentiment analysis, see Chap. 28), embeddings for words *horošij* (“good”) and *plohoj* (“bad”), which may be initially close, will be pushed apart from each other.

Last but not least, an alternative approach to word tokenization, called byte pair encoding (BPE; Heinzlerling and Strube 2018), suggests not to use whole words as text units, but rather split the words into subwords, based on frequent n-grams. BPE tokens resemble to a certain degree, morphemes, and seem quite promising for Russian.

To conclude this section, we will list a few pretrained word embedding models for Russian in Table 26.1.

All these models are available for downloads as single files. The models are trained on large freely available corpora, such as Wikipedia, Taiga,<sup>1</sup> and

Araneum.<sup>2</sup> The vocabulary of the models ranges from 100K to 700K unique tokens and the model size ranges from 200MB to 3GB.

RusVectores additionally provides web interface for exploration of word embedding models, along with visualization and semantic calculator.

Table 26.2 lists tools freely available to train embedding models from scratch. Gensim is one of the most popular Python libraries for building word embedding models and topic models, though Gensim does not provide deep learning functionality. In contrast to Gensim, AllenNLP and flair provide reference implementations for deep learning models for NLP, including word2vec and fasttext. These libraries provide tools for processing textual data and share similar functionality, though target different audience. AllenNLP is more advanced and flair is designed as a very simple framework. Both AllenNLP and flair have Python interfaces. Fasttext is available as a console application of the same name. Deeplearning4j is a general deep learning framework that provides scripts for training deep learning models.

### 26.3.2 *Text Classification*

The task of text classification is to assign categories to texts. This is a common supervised task: given labeled data (i.e. texts, annotated with class labels), a model should be first trained, and then applied to unlabeled test data.

Text classification is one of the most demanded industrial NLP tasks. Sentiment analysis and information filtering are the most common applications of text classification algorithms. Sentiment analysis is widely used for marketing research. Companies use sentiment classification for product analytics, brand monitoring, customer support, and market research. One of the main information filtering techniques is spam filtering, which exploit classification algorithms to distinguish between spam and ham incoming emails. In general, email categorization is a powerful idea which facilitates the work of an office employee. Other information filtering applications may include identification of trolls, obscene content detection, ad blocking and privacy protection. What is more, hotlines use text classification for language identification.

Virtual personal assistants, such as Apple Siri or Amazon Alexa, are becoming an internal part of our daily lives. They use the whole range of NLP methods, including text classification. Each user utterance is classified according to its intent, according to the desired action of the user (i.e. whether the user meant to launch an application, make a call, write a note, etc.).

The classification of Russian texts is almost no different from English text classification and follows a standard pipeline:

1. Word embeddings are used as an input to the model
2. Multiple hidden CNN- or RNN-derived layers are used for input processing
3. A feed forward layer is used for final prediction.

The labels in the training set, that is, the correct answers, are used for supervision. When presented with correct answers, the model is able to adjust its own parameters so that its predictions become correct.

The quality of the classification task is evaluated according to the ratio of correct predictions and the ratio of erroneous predictions.

There are a few recent Russian-language datasets for text classification:

- A large-scale dataset for sentiment analysis, which consists of texts from social media (Rogers et al. 2018).
- A dataset for sentiment analysis of product reviews on e-commerce sites (Smetanin and Komarov 2019).
- A collected dataset for humor recognition in short stories (Baranova-Bolotova et al. 2019).
- RusIdiolect<sup>3</sup> is a dataset for experimental studies of the idiolect of a native Russian speaker, such as deception detection (Litvinova et al. 2017).
- RusProfiling is a popular dataset for author profiling, including gender identification. Current state of the art results are achieved by Sboev et al. (2018).

These datasets are available to download from the Web. In contrast to major English datasets gathered in Natural Language Toolkit<sup>4</sup> (NLTK), there is no unified application programming interface (API) to access Russian datasets.

Finally, the major component of fasttext (Joulin et al. 2017) functionality is a simple yet strong classification algorithm. It is very fast and easy to use and is strongly recommended as a strong baseline.

Finally, there are a few applications of word embeddings outside linguistic field. For example, (Panicheva and Litvinova 2019) report on using word embeddings to measure speech coherence of patients, affected by schizophrenia. “Semantic coherence” is defined as mean pairwise similarity between words in a sample text, written by a patient. Word embeddings allow to measure semantic coherence, as they provide a simple approach to measure word similarity. The schizophrenia status of a patient along with text samples is provided in RusIdiolect corpus. The findings of Panicheva and Litvinova show that semantic coherence features allow to distinguish between healthy patients and patients, who suffer from schizophrenia. This is comparable to results reported for similar task in English. This research project aims at studying various phenomena present in the schizophrenia and by no means calls to replace traditional medical diagnostics.

### 26.3.3 *Sequence Labeling*

The task of sequence labeling is to assign categories to single words. Common examples of a sequence labeling tasks are part-of-speech (POS) tagging or named entity recognition (NER). POS tagging is the task of labeling a word with a corresponding POS tag. NER seeks to identify such named entities as



**Table 26.3** Two examples of sequence labeling tasks

	<i>Boris (Boris)</i>	<i>Pasternak (Pasternak)</i>	<i>rodilsá (was born)</i>	<i>v (in)</i>	<i>Moskve (Moscow)</i>
POS tags	PROPN	PROP	VERB	PREP	PROPN
NE tags	Person	Person	O	O	Location

POS tagging (first line), named entity recognition (second line). Each word is assigned with two tags: a POS tag and a named entity tag. If the word is not a named entity, the tag “O” is used

persons, locations, organizations, et cetera, and assign them with a corresponding tag. See Table 26.3 for examples of POS tagging and NER.

Sequence labeling applications range from linguistics tasks, such as POS tagging, which can be treated as a preliminary step for further analysis, up to more complex tasks, such as coreference and gapping resolution. NER, as a sequence labeling task, can be treated as a preliminary step for machine translation. Named entities should be identified and treated differently from regular words for proper translation. When used in Legal Tech or medical applications, NER helps to discover important features, such as legal condition or diseases, used further for decision-making. In Russian realities, Legal Tech applications are very much in demand. This motivates several research groups to develop NER methods for specific domains.

Sequence labeling helps virtual assistants to understand user needs better. While text classification helps to detect user intent, sequence labeling methods are able to fill in slots, that is, to discover specific details, such as what exactly application should be launched or which contact should be addressed. Gapping and coreference are crucial for handling messaging history. Gapping resolutions helps to find omitted predicates in consequent turns, while coreference resolutions helps to connect nouns and names with corresponding pronouns.

RNN and its variations are widely used for sequence labeling tasks due to its ability to process a sequence word by word. We can think of RNN as an attentive reader that reads each word carefully, thinks over the context of the word, and then makes a decision as to what tag to assign. It is worth noting that bidirectional variations of RNN, capable of both left-to-right and right-to-left reading, are suited to model languages with free word order as they maintain both left and right contexts.

The pipeline of the sequence labeling task does not differ significantly from the text classification pipeline:

1. Word embeddings are used as an input to the model. Word embeddings may be extended with convolved character representations, which would take care of derivational patterns.
2. Multiple hidden RNN-derived layers are used for processing input and for producing context-aware word representations.
3. Each word representation is fed into a feed forward layer for final prediction and each word is assigned with a label. Alternatively, another model,

conditional random field (CRF), may be used on top of the recurrent layer to reweight its prediction.

The main difference between text classification and sequence label affects the final layer. When used for text classification, the final layer is applied only once to get one class label. However, for sequence labeling it is applied to each individual word representation from previous layer.

In contrast to text classification task, sequence labeling seems to be more complicated from a linguistic point of view. Tasks, modeled as sequence labeling, are more advanced and range from POS tagging to coreference and gapping resolution.

There are several Russian datasets for the sequence labeling task:

- Universal dependencies<sup>5</sup> project presents four Russian corpora annotated with POS tags
- Persons-1000 (Gareev et al. 2013) and FactRuEval (Starostin et al. 2016) are large-scale datasets for named entity recognition
- AGGR-2019 (Smurov et al. 2019) is a corpus for gapping resolution
- RuCor and AnCor (Toldova et al. 2014) are corpora used for coreference and anaphora resolution
- SberQUAD, a dataset for question answering, treats answer generation as a retrieval of a relevant fragment of text.

### 26.3.4 *Transfer Learning in NLP*

Since 2017, NLP field has witnessed the emergence of transfer learning methods and algorithms. Transfer learning stands for the process of training a model on a large-scale dataset to conduct a simple task, such as language modeling. Next, this pretrained model is trained for the second time for more complicated tasks. The transfer learning process is comparable to the way a child is educated. Children acquire the language from their environment, and only in the school they are taught to complete grammar tasks. The same way models gain language understanding while being pretrained and then are supervised for specific tasks.

Transfer learning led to a paradigm shift in NLP. Instead of using every time pretrained word embeddings and training the whole model from scratch, now a pretrained model is fine-tuned for downstream tasks. This requires much less annotated data and leads to superior results simultaneously. Word embeddings were an imperfect way to store language representation, which suffered from language ambiguity. Pretrained models are less prone to polysemy and antonymy and are able to handle multilinguality at the same time.

Despite the fact that transfer learning paradigms leads to superior results in comparison to previous approaches, so far it has not enabled any exceptionally new applications.

Inside transfer learning models are transformer layers (see Fig. 26.1d) that are more advanced from a technical point of view when compared to other layers. The architecture of transfer learning models is sophisticated, enumerates millions of parameters, and take weeks to be pretrained.

Not only transfer learning models established new state of the art for several existing NLP tasks, they also appear to be efficient in new generations of tasks. For example, there is evidence that the tasks that require commonsense understanding can be conducted using transfer learning techniques. This is supported by the idea that excessive pretraining results in a subtle understanding of language patterns.

When pretrained on the corpus of multiple languages or on parallel corpus, transfer learning models become aware of several languages at the same time and can be shared across several languages for the same downstream task. For example, Piskorski et al. (2019) show how NER in four Slavic languages can be approached by a multilingual model.

Even though pretraining of a large model is expensive and time-consuming, new models appear almost every month as of late 2019. Among others, ELMo (Peters et al. 2018) and BERT by Google (Devlin et al. 2019) are the most popular models. BERT's successors, ALBERT, RoBERTa, XLNet, and T5, released by Facebook, Microsoft, and other technology companies, are larger and outperform BERT by far. At the same time, they are heavily criticized for being unaffordable for smaller institutions. Indeed, few universities in Russia have enough resources to train transfer learning models. Table 26.4 lists transfer learning models available for the Russian language. RusVectors poses both word and sentence embeddings model. RusVectors provides not only word embeddings, but also a pretrained ELMo model, which can be treated as sentence embedding models.

Transfer learning models can be exploited as a standalone sentence embedding tool. Sentence embeddings are massively used in those applications, which require modeling of sentence similarity. Consider, for example, the task of finding an answer to a frequently asked question (FAQ). Imagine that the answers to some FAQs are already known, and a user asks a new question. The most similar question to the new one can be found by using an embedding-based similarity measure. With a high chance, the answer to the retrieved question should fit the new question, too.

**Table 26.4** Transfer learning models for Russian

<a href="http://docs.deepavlov.ai/en/master/features/models/bert.html">http://docs.deepavlov.ai/en/master/features/models/bert.html</a>	BERT in DeepPavlov (Kuratov and Arkhipov 2019)
<a href="http://rusvectors.org">http://rusvectors.org</a>	Multiple Russian-only word and sentence embeddings (Kutuzov and Kuzmenko 2017)
<a href="https://github.com/vlarine/ruberta">https://github.com/vlarine/ruberta</a>	Russian RoBERTa, RuBERTa

Transfer learning models are excellent not only in solving complex downstream tasks, but also in text generation. Some researchers are afraid of the fluency of these models and raise ethical questions of the harmless strategy of releasing the models. When misused, the transfer learning models can generate fake news and offensive utterances and be disturbing. However, these concerns are vivid for English-spoken communities and do not reach Russia so far.

## 26.4 CONCLUSION

This chapter discusses the applications of deep learning methods to Natural Language Processing tasks and is particularly oriented at the Russian language. We traced the development of deep learning methods for NLP from early stages of using feed forward networks to recent developments in transfer learning. Two basic text representation models, namely bag-of-words and language models, were presented and related to the duality between convolutional and recurrent neural networks. We have recognized a recent paradigm shift, caused by new advances in architecture design and development of transformer layers. Several analogies between human intelligence and neural networks were drawn. Neural networks aim at resembling human by using artificial neurons and attention mechanisms, and acquiring language from textual data.

With no doubt, deep learning is a leading paradigm in modern language technology. Unfortunately, the Russian language resources do not provide enough resources to exploit deep learning scope fully. The Russian research community is facing a need for both keeping track of worldwide challenges and, if necessary, reapply the methods initially developed for the English language to the Russian language.

The latter requires an increase not only of computational powers, which is rather a financial matter but also of the amounts of annotated data. Recent government decisions and AI-centered strategies seem to provide financial support to the research community that may help to narrow the gap between English and Russian language resources.

Not only is Russian different from English from a linguistic point of view, but also different language technology applications are demanded in Russia and English-spoken countries. The major NLP applications in Russia are related to marketing, e-Government transformation, and call center automation. Whole domains such as Legal Tech, medical NLP, educational NLP still stay out of business focus and are subject to further development. Minority languages currently are not supported by major language technology applications with Yandex.Search (the web search engine and the core product of Yandex) being the only exception.

At the same time, while language technologies become more and more sophisticated, the entry threshold to the NLP field is lowered. Recent advances in programming tools and programming languages made it possible to develop high-level languages, which can be easily comprehended by users with little or

no previous programming experience. Successful implementations of many deep learning architectures have substantially facilitated the development of practical applications. The complexity of deep learning models comes along with the flexibility of fine-tuning and reuse across practical applications. The nearest future will likely witness the transformation of learnable approaches to daily routines.

## NOTES

1. [https://tatianashavrina.github.io/taiga\\_site/](https://tatianashavrina.github.io/taiga_site/).
2. [http://ucts.uniba.sk/aranea\\_about/](http://ucts.uniba.sk/aranea_about/).
3. <https://rusidiolect.rusprofilinglab.ru>, yet not published.
4. <https://www.nlTK.org>.
5. <https://universaldependencies.org>.

## REFERENCES

- Baranova-Bolotova, V., V. Blinov, and P. Braslavski. 2019. Lightning Talk-Humor Recognition in Russian Language. In *Companion Proceedings of the 2019 World Wide Web Conference*, 1268–1269. ACM.
- Bartunov, S., D. Kondrashkin, A. Osokin, and D. Vetrov. 2016. Breaking Sticks and Ambiguities with Adaptive Skip-Gram. *Artificial Intelligence and Statistics*: 130–138.
- Bengio, Y., R. Ducharme, P. Vincent, and C. Jauvin. 2003. A Neural Probabilistic Language Model. *Journal of Machine Learning Research* 3 (Feb): 1137–1155.
- Devlin, J., M.W. Chang, K. Lee, and K. Toutanova. 2019. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1. (Long and Short Papers), 4171–4186.
- Gareev, R., M. Tkachenko, V. Solovyev, A. Simanovsky, and V. Ivanov. 2013. Introducing Baselines for Russian Named Entity Recognition. In *International Conference on Intelligent Text Processing and Computational Linguistics*, 329–342. Berlin, Heidelberg: Springer.
- Gordeev, Denis, Alexey Rey, and Dmitry Shagarov. 2018. Unsupervised Cross-Lingual Matching of Product Classifications. In *Proceedings of the 23rd Conference of Open Innovations Association FRUCT*, 62. FRUCT Oy.
- Harris, Z.S. 1954. Distributional Structure. *Word* 10 (2–3): 146–162.
- Heinzerling, B., and M. Strube. 2018. BPEmb: Tokenization-Free Pre-Trained Subword Embeddings in 275 Languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. LREC-2018.
- Joulin, A., E. Grave, P. Bojanowski, and T. Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics 2, Short Papers*, 427–431.
- Kurатов, Y., and M. Arkhipov. 2019. Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language. *arXiv preprint arXiv:1905.07213*.

- Kutuzov, A., and E. Kuzmenko. 2017. WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In *Analysis of Images, Social Networks and Texts, AIST 2016. Communications in Computer and Information Science*, ed. D. Ignatov et al., 661. Cham: Springer.
- Kutuzov, A., L. Øvrelid, T. Szymanski, and E. Veldal. 2018. Diachronic Word Embeddings and Semantic Shifts: A Survey. In *Proceedings of the 27th International Conference on Computational Linguistics*, 1384–1397.
- Litvinova, Olga, Pavel Seredin, Tatiana Litvinova, and John Lyell. 2017. Deception Detection in Russian Texts. In *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, 43–52.
- Mikolov, T., I. Sutskever, K. Chen, G.S. Corrado, and J. Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Advances in Neural Information Processing Systems*, 3111–3119.
- Panicheva, Polina, and Tatiana Litvinova. 2019. Semantic Coherence in Schizophrenia in Russian Written Texts. In *Proceedings of the 25rd Conference of Open Innovations Association FRUCT*, 240. FRUCT.
- Pelevina, M., N. Arefiev, C. Biemann, and A. Panchenko. 2016. Making Sense of Word Embeddings. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, 174–183.
- Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *Proceedings of NAACL-HLT*, 2227–2237.
- Piskorski, J., L. Laskova, M. Marcińczuk, L. Pivovarova, P. Přibáň, J. Steinberger, and R. Yangarber. 2019. The Second Cross-Lingual Challenge on Recognition, Normalization, Classification, and Linking of Named Entities across Slavic Languages. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, 63–74.
- Rogers, Anna, Alexey Romanov, Anna Rumshisky, Svitlana Volkova, Mikhail Gronas, and Alex Gribov. 2018. Rusentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian. In *Proceedings of the 27th International Conference on Computational Linguistics*, 755–763.
- Sboev, Alexander, Ivan Moloshnikov, Dmitry Gudovskikh, Anton Selivanov, Roman Rybka, and Tatiana Litvinova. 2018. Deep Learning Neural Nets Versus Traditional Machine Learning in Gender Identification of Authors of RusProfiling Texts. *Procedia Computer Science* 123 (2018): 424–431.
- Smetanin, S., and M. Komarov. 2019. Sentiment Analysis of Product Reviews in Russian Using Convolutional Neural Networks. In *2019 IEEE 21st Conference on Business Informatics (CBI)*, vol. 1, 482–486. IEEE.
- Smurov, I.M., M. Ponomareva, T.O. Shavrina, and K. Drozanova. 2019. Agrr-2019: Automatic Gapping Resolution for Russian. *Computational Linguistics and Intellectual Technologies*: 561–575.
- Solovyev, Valery D., Vladimir V. Bochkarev, and A.D. Kaveeva. 2015. Variations of Social Psychology of Russian Society in Last 100 Years. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, 519–523. IEEE.
- Starostin, A.S., V.V. Bocharov, S.V. Alexeeva, A. Bodrova, A.S. Chuchunkov, S.S. Dzhumayev, and M.A. Nikolaeva. 2016. FactRuEval 2016: Evaluation of Named

Entity Recognition and Fact Extraction Systems for Russian. In *Computational Linguistics and Intellectual Technologies. Proceedings of the Annual International Conference Dialogue (2016)*, vol. 15, 702–720.

Toldova, S.J., A. Roytberg, A.A. Ladygina, M.D. Vasilyeva, I.L. Azerkovich, M. Kurzukov, and Y. Grishina. 2014. RU-EVAL-2014: Evaluating Anaphora and Coreference Resolution for Russian. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue”*, 681–694.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.







# Shifting the Norm: The Case of Academic Plagiarism Detection

*Mikhail Kopotev, Andrey Rostovtsev, and Mikhail Sokolov*

## 27.1 INTRODUCTION

Plagiarism currently tends to be viewed as a problem connected primarily with students, albeit more prominent authors such as William Shakespeare and George Friedrich Handel were accused of it long ago. Plagiarism continues to be widespread in educational institutions, predominantly due to single-click technology, but another contributing factor that helps make it common practice is the tolerance of plagiarism on the part of educators and academia in general. In 2004, for instance, it was estimated that 10 percent of student projects in the United States and Australia involved plagiarism (Oakes 2014, 60). By contrast, in Russia, 36 percent of respondents admitted to having regularly copied the texts of others (Kicherova et al. 2013, 2); as many as 36.7 percent of undergraduate students in 8 Russian universities took personal credit for material they had, in fact, downloaded from the Internet (Maloshonok 2016).

---

M. Kopotev (✉)  
Higher School of Economics (HSE University), Saint Petersburg, Russia  
e-mail: [mkopotev@hse.ru](mailto:mkopotev@hse.ru)

A. Rostovtsev  
Institute for Information Transmission Problems RAS, Moscow, Russia

M. Sokolov  
European University at Saint Petersburg, Saint Petersburg, Russia  
e-mail: [msokolov@eu.spb.ru](mailto:msokolov@eu.spb.ru)

The problem of plagiarism is certainly not limited to undergraduate students. For example, two cases of plagiarism were documented in PhD dissertations published in Germany in 2011. These cases, which were analyzed in detail by the *GuttenPlag* community, led to the monograph titled *False Feathers: A Perspective on Academic Plagiarism* (Weber-Wulff 2014). However, plagiarism is arguably exceedingly prevalent and more deeply rooted in Russia than in Europe (see Golunov 2014; Denisova-Schmidt 2016). One reason for this may be that the symbolic value of scholarly achievements in Russia has been widely appropriated by politicians, civil servants, businesspersons, and administrators from educational and medical fields. These professionals have been awarded degrees by lenient defense panels for dissertations that have been entirely copy-pasted from other sources. This would be even more prevalent among those in power if strong opposition had not been voiced by the academic community. This led to the establishment of “Dissernet,” a network that purports to expose large-scale plagiarism in Russian scientific publications. Our focus in this chapter is on Russian doctoral and post-doctoral dissertations,<sup>1</sup> which constitute merely the tip of an academic iceberg that includes articles, monographs, coursebooks, and other scholarly works. In fact, the post-Soviet publishing market is flooded with texts of questionable originality.

The current availability of material and ease of use raises more general questions. For example, what is the *textual authenticity* and what are the *norms* of textual authenticity for scholars at a time when everything is “a copy of a copy of a copy” (Palahniuk 1996)? Western academic culture presupposes that the origin of the words and ideas in a scholarly text, from the first word to the last, are from the author or authors accredited in connection with the title, with the exception, of course, of properly attributed quotations from other scholarly works, or paraphrases of them. Even within these norms, however, exactly what is meant by “original from the first word to the last” is somewhat ambiguous (Korbut 2013).

One of the principal subjects in sociology since the time of Durkheim is social norms; in other words, the rules of conduct that are considered proper, right, and socially desirable. In recent decades, digitalization has made it possible to analyze compliance with various norms using digital traces of naturally occurring behaviors rather than self-reporting, official statistics, or other less reliable resources. Areas of conduct analyzed in this manner vary—from using dirty words (McEnery 2004) to observing meritocratic principles in the selection of professors (Clauzet et al. 2015). Due to the increasing digitalization of Russian society together with emergent methods of analysis, it is now possible to study the level of support for a particular norm, specifically one that requires authenticity in academic writing, and to analyze conditions under which this norm is likely to be transgressed.

The *norm of textual authenticity* requires that any academic text be fully original in compliance with the highest academic standards, which permit quotation and paraphrasing with correct and appropriate attribution to the source. This could be deconstructed into two different norms requiring (1) that the text be written in full by its presumed author and (2) that the text is written for one and only one purpose or publication outlet. The latter, which forbids any

recycling of an academic text, is more restrictive than the former in that it bans all forms of reuse including that of one's own texts. The focus of this chapter is on the first, less restrictive norm of texts that are written entirely by an author. The assumption is that dissertation authors can reproduce sections of their dissertation in articles and that this is universally regarded as a permissible and even a desirable practice.

The norm of textual authenticity requires identification of what constitutes a form of expression, such as widely used terms or stock phrases, and what is the true content of the academic text. Some forms of expressions or presentation style may or may not qualify as unauthorized borrowing. These include the use of certain truisms and clichés such as “to the best of our knowledge,” design layouts, and fonts. To apply the norm of textual authenticity thus requires constant discrimination between what is the “mere form” of an academic message and what is “the message itself,” with the form being considered part of academic convention and without authorship. The digitalization of scholarly production together with software development facilitate the study of particular variations in the norms of textual authenticity.

We begin the analysis for this chapter by describing the challenge that academic plagiarism poses for digital humanities in an era when sophisticated tools make it possible to detect inappropriate academic activity, and we focus specifically on Russian dissertations. Second, we examine the changing norms of academic integrity in terms of the sociology of science. Thus, in Sect. 27.2, we describe the various types of plagiarism and the computational tools that have been created to detect fraudulent texts. Section 27.3 comprises a review of available digitized resources, including dissertations, articles, and abstracts published by the Russian academic press. In Sect. 27.4, we provide an overall picture of the Dissernet findings when these tools were applied to large-scale (greater than 50%) plagiarism in dissertations that have been defended in Russia. Section 27.5 presents a case study of small-scale plagiarism based on the same academic genre. This study analyzes and traces the shifting authenticity norms in Russia since post-Soviet times. Finally, Sect. 27.6 concludes the chapter.

## 27.2 TYPES OF PLAGIARISM AND TOOLS ENABLING ITS DETECTION<sup>2</sup>

The *Modern Language Association (MLA) Style Manual and Guide to Scholarly Publishing* defines plagiarism as follows:

Forms of plagiarism include the failure to give appropriate acknowledgment when repeating another's wording or particularly apt phrase, paraphrasing another's argument, and presenting another's line of thinking. (Modern Language Association 2008, 166)

Two types of plagiarism are commonly distinguished in the scholarly literature, which Bela Gipp refers to as *copy&paste* versus *shake&past* (Gipp 2014, 12; see also Potthast et al. 2010). The former refers to copying someone’s text unchanged without proper acknowledgment, whereas the latter implies minor modifications, such as varying the word order or using synonyms—again without acknowledging the source. Several services are currently available that can detect plagiarism in Russian-language texts (see Nikitov et al. 2012). Below we describe several of the most advanced technologies applicable to textual plagiarism. We do not address evidence of fraudulent publication such as image and diagram falsification, carbon-copied lists of references, or data manipulation (for example, wild data or loose correlation).<sup>3</sup>

*Copy-and-paste*, or *cut-and-paste* refers to “involving or relating to the cutting and pasting of printed material, or (Computing) the ‘cut’ and ‘paste’ functions on a computer” (OED, c.v. *cut-and-paste*). Technically, the basic commands available on any computer can create the simplest form of plagiarism, and hence the most alluring, is when a source is used but not cited properly. This is easy to identify, even when the text under suspicion has been—intentionally or otherwise—modified or corrupted. Detection is based on identifying similar chains of symbols and their possible modifications. Some of these modifications reflect deliberate distortions by the borrower-creator, such as Cyrillic letters replaced with identical Latin ones, whereas others derive from optical character recognition (OCR) (see Table 27.1).

The plagiarism in each of these cases can be detected by conducting a basic similarity test or by using a more sophisticated technique such as the *Levenshtein distance*, which is the number of required symbol substitutions for one word to be changed into another (Levenshtein 1966). This approach is exemplified by a tool called *Disserrorubka* (literally “the Thesis-grinder”) and was developed by the Dissernet community. Another service that is available online, albeit a commercial one, is antiplagiat.ru, which is specifically designed to detect plagiarism in Russian texts. The available techniques and services allow copy-and-paste plagiarism to be effectively detected by taking into account specific issues related to the Cyrillic alphabet, such as the Cyrillic “P” replaced with Latin “P,” and the confused recognition of “Ф” as “%.”

**Table 27.1** A source text (left) and the copy-pasted text after OCR (right)<sup>a</sup>

<p><b>Специфика воинской деятельности в сочетании</b> с высочайшим напряжением всех духовных и физических сил, с возможностью и <b>необходимостью самопожертвования</b> во имя Родины, определяют значимость духовного фактора для армии.</p>	<p><b>С п е ц и ф и к а в о и н с к о й д е я т е л ь н о с т и в с о ч е т а н и и</b> с высочайшим напряжением всех духовных и физических сил, с возможностью и <b>не обходимостью само пожертвования</b> бо имя Qодины, определяют значимость духовного %актора для армии.</p>
---	---

The distortions are in bold.

<sup>a</sup>The examples are fictional and were constructed by the authors: any correspondence to actual texts is accidental.

In the case of *paraphrasing*, different linguistic techniques are used to rework the source texts, including word removal, word replacement, synonym substitution, word-order modification, grammatical changes, and patchwriting (for example, by combining fragments from several texts) (Oakes 2014, 60). The nature of these changes depends on whether the paraphrase had been generated by means of manual text editing or automatically (Gupta et al. 2011, 1), as shown in Table 27.2.

Dictionary-based methods are used to detect this type of plagiarism, requiring a lexicon that contains all possible changes, substitutions, and transformations. All modifications are weighted, with the slighter ones prioritized, and those that are more substantial being downgraded. For instance, word-order modification and word replacement are both automatically detectable, but the former is weighted more heavily than the latter because it preserves more of the original source. An application of this approach to the Russian, Ukrainian, and English languages, developed by K. Kuznetsov and M. Kopotev, can be found online at <http://dissercomp.ru>. Thus far, the service is able to detect paraphrased plagiarism in Russian, Ukrainian, and English texts.

Another case of paraphrasing is *interlingual plagiarism*, when a text is “paraphrased” in a sense from one source language to another. This process may involve manual or automatic translation. When automatic translation is involved, the output of the machine translator usually undergoes post-editing, along with obfuscation, which makes a comparison of the sources with the plagiarized text substantially more difficult while at the same time displaying evidence of translation (Table 27.3).

Detecting this type of plagiarism poses a challenge and tests the very limits of the methods available to scholars in digital humanities. Those engaged in this endeavor have turned to distributional neural net modeling, and specifically to distributional semantics.

The initial idea behind this approach reflects the understanding of meaning through context, as proposed by J. R. Firth: “You shall know a word by the company it keeps” (Firth, J. R. 1957, 11). The main objective in distributional semantics is to analyze the co-occurrence of linguistic entities (usually words)

**Table 27.2** A source text (left) and the paraphrased text (right)

<p>Некоторая часть <b>начальников</b> и <b>преподавательского состава</b>, обладая <b>неплохими</b> теоретическими знаниями, сами имеют <b>слабые</b> практические навыки, поэтому они не могут <b>правильно</b> учить <b>курсантов</b>.</p> <p>English translation</p> <p>Some of the <b>heads</b> and <b>faculty</b>, possessing <b>goodish</b> theoretical knowledge, have themselves <b>weak</b> practical skills, so they can not <b>properly</b> teach the <b>cadets</b>.</p>	<p>Некоторая часть <b>командиров</b> и <b>учителей</b>, обладая <b>хорошими</b> теоретическими знаниями, сами имеют <b>плохие</b> практические навыки, поэтому они не могут <b>хорошо</b> учить <b>студентов</b>.</p> <p>Some of the <b>commanders</b> and <b>teachers</b>, possessing <b>good</b> theoretical knowledge, have themselves <b>poor</b> practical skills, so they cannot teach <b>students well</b>.</p>
---	--

The paraphrasing is indicated in bold.

**Table 27.3** A source text (left) and the translated text (right)

In a crisis, the whole educational system was reformed: the structure of educational institutions changed; the throughput of schools increased.	В условиях кризиса была проведена реформа всей системы образования: изменилась структура учебных заведений, увеличилась пропускная способность училищ.
---	--

and to summarize this distribution statistically on multidimensional “semantic spaces.” For example, the English noun *plagiarism* regularly collocates with the same words as the nouns *falsification*, *obscenity*, and *misbehavior*:

...accused of plagiarism/falsification/obscenity/misbehavior in...

Among the many applications for this paradigm, one that is based on the word2vec modeling was specifically developed to expose translated plagiarism. The authors call their method “semantic fingerprinting” (see Kutuzov et al. 2016); the service is also available online: [www.dissnet.org/dissesearch](http://www.dissnet.org/dissesearch).

### 27.3 AVAILABLE ELECTRONIC RESOURCES

A well-functioning computational tool does recognize plagiarism effectively. If they are to achieve results, experts also need access to the relevant textual data. Numerous (preferably all) academic texts are required in order to compare the plagiarized text with potential sources by applying an algorithm that can make searches. The full range of texts, both online and offline, would be available in a perfect world, but real life poses additional challenges. An accepted presupposition here is that both the copycat who scans for a suitable source to rewrite, and the unmasker who is intent on revealing the copycatting are most likely to be relying on the same resources, in other words (publicly), available digitized texts.

How many scientific text documents in Russian have been digitized and made available to the public? In answer to this question, we consider different categories of academic texts. The first category includes doctoral and post-doctoral dissertations which are referred to as *autoreferats*, a formal abstract of the dissertation. An *autoreferat* is a summary of the main results reported in a work that the author compiles and it usually consists of 20–30 pages abstracted from the full text. These abstracts also contain basic information on the formal public defense such as the date and place of the event, the name of the academic supervisor, the official opponents, and so on. The degree candidate in Russia is required to deposit both the dissertation and the abstract in the main libraries of the Russian Federation. The RSL (*Rossijskaâ gosudarstvennaâ biblioteka*, Russian State Library) in Moscow has been a major repository for these texts from 1944 onwards. In 2003, the RSL management decided to ensure broad public availability and preservation of dissertations electronically. Thus far, this has led to the creation of the most comprehensive electronic collection of abstracts (*autoreferats*) of domestic doctoral and post-doctoral dissertations

in the world. To date, the collection incorporates more than 919,000 full texts. The dissertations defended in 1994 and thereafter were digitized rather systematically, whereas the collection of abstracts (*autoreferats*) covers the time period from 2007 up to the present. Most, but not all, dissertations and abstracts from previous years have also been digitized.

All of the aforementioned documents are available in the Digital Dissertation Library at <http://diss.rsl.ru> upon registration. Registered visitors receive free and unrestricted, open access to the abstract collection. Access to the copyright-protected part of the Digital Dissertation Library is provided at the RSL in Moscow or in its virtual reading rooms, of which there are more than 600 in Russia and worldwide. Most of the reading rooms located abroad are accessible through local university libraries. Readers who are registered individually are also offered the opportunity to access the full texts remotely. However, they are limited to viewing at most five dissertations per day, and no more than fifteen per month. Beginning in 2014, prior to their public defenses, all post-graduate students have been required to publish their dissertations and their abstracts online and in open-access forums. As a result, the number of available dissertations is increasing annually by approximately 30,000 texts. The RSL with its Digital Dissertation Library nevertheless remains the only central collection of these documents in Russia.

All types of scientific publications apart from dissertations are accessible in many electronic libraries, both in Russia and beyond. Russia's most comprehensive and ambitious repository is the Russian Scientific Electronic Library, available at [elibrary.ru](http://elibrary.ru), which also offers many other categories of scientific publications. Another category comprises books and book chapters, of which more than 122,000 full texts are available in the Electronic Library, and more than 55,000 of them are open access. Collected papers constitute a further category of digitized documents available at the same website. There are also more than 127,000 volumes and papers available, and approximately 87,000 of them are open access. Conference and similar short papers are assigned a separate category among the digitized documents: there are more than 982,000 of them with 779,000 being open access. The last of these groups consists of academic articles or publications in scholarly periodicals, and this group naturally represents the largest category of digitized scientific documents with approximately 4.5 million papers written in Russian available at [elibrary.ru](http://elibrary.ru), and of these, about 3.3 million are open access.

The impressive collections of academic texts described above have become available, thanks to public funding. They are key sources of successful scientific work in Russia and/or of data in Russian for projects ranging from conducting basic bibliographic searches to discovering trends in Russian science. These data provide the groundwork for the detection of plagiarism in academic texts. Plagiarism detection rests on two crucial conditions: effective algorithms and the availability of source texts to which a suspicious text is compared in order to find similarities. The available data in Russian meets both conditions that allow the effective detection of plagiarism and deal with this social



phenomenon in depth. In the next two sections, we explore two case studies that utilize available resources. The first case concerns large-scale plagiarism that involves the copying of more than half of the source text, which provokes general observations of fake academic activity in Russia. By contrast, the second case focuses on small-scale plagiarism and discusses cross-cultural variation in interpretations of authenticity norms.

#### 27.4 THE BEST PRACTICES OF DISSERNET IN THE DETECTION OF LARGE-SCALE PLAGIARISM

The volunteer network known as Dissernet was established in 2013 to counter fraud and dishonesty in academia, specifically in fabricated dissertations and in the conferring of false university degrees. According to its manifesto, Dissernet is “a networking community of experts, researchers and reporters seeking to unmask swindlers, forgers and liars,” whose members “oppose abusive practices, machinations and falsifications in the fields of scientific research and education, in particular in the process of defending theses and awarding academic degrees in Russia” (English translation from <https://en.wikipedia.org/wiki/Dissernet>).

It is now possible to detect plagiarism in thousands of dissertations, primarily through the application of in-house tools, introduced in Sect. 27.2, to the data described in Sect. 27.3 of this chapter. The abstract, or *autoreferat*, serves as a prerequisite for identifying suspected cases of plagiarism in that it is available online and is thus indexed by search engines such as Google and Yandex. This works even when the dissertation itself is not indexed, based on the assumption that when a dissertation contains a large amount of plagiarized text, its *autoreferat* will retain fragments of the plagiarized sources. Dissernet software is able to pick up the abstracts one by one by utilizing search-engine indices to search for textual coincidences within the entire, publicly available mass of Russian digitized texts, including articles, monographs, and dissertations as well as their abstracts. This is essentially how the technological part of the process works, and hundreds of thousands of texts are automatically checked in this manner. Dissernet is principally aimed at detecting large-scale plagiarism, which is determined to be the illegal use of equal to or greater than 50 percent of a text. In an extreme but real-life example, a source text was utilized in full, with the automatic replacement of “dark chocolate” with “local beef,” and “confectionery” with “meat and dairy.” As at beginning of 2020, Dissernet had identified almost 9,000 plagiarized dissertations, both doctoral and post-doctoral, that had been defended in the previous two decades.

At the next level of its investigation, Dissernet exposes established practices that are corrupt, such as when an *omertà*-like community repeatedly produces fraudulent dissertations. Dissernet findings clearly indicate that as soon as rampant plagiarism is detected in one dissertation, plagiarism is likely to be discovered in other dissertations defended before the same defense panel or under

the same supervision. Many of those who produce these dissertations work in a “conveyor-belt” mode by using exceedingly limited sets of scientific texts as sources. The graph below (Fig. 27.1) demonstrates the density of such practice that one dissertation-defense panel established at MGPU (*Moskovskij pedagogičeskij gosudarstvennyj universitet*, Moscow Pedagogical State University). This panel approved more than 90 “doctored” dissertations from 2001 to 2012, with the same actors playing interchangeable roles first as *kandidat nauk* (doctoral degree candidate) or *doctor nauk* (post-doctoral degree candidate) and later as *naučnyj konsul'tant* (supervisors) or official opponents (see Fig. 27.1).

First and foremost, Dissnet activity targets plagiarism among top-ranked Russian politicians and administrators, both in academia and beyond. Thus, the results cannot be interpreted as representing the whole landscape across all disciplines over the entire country. However, the number of dissertations tested (more than 20,000) allows us to draw a number of preliminary conclusions. First, the number of heavily plagiarized dissertations varies significantly depending on the academic field. Most of the identified fake dissertations (44%) were in the field of economics. Other academic fields deeply infected by fraud include pedagogy (16%) and law (12%), followed by the medical sciences, political science, engineering, and the social sciences. However, this type of fraud is less common in the natural sciences. It is important to mention that this

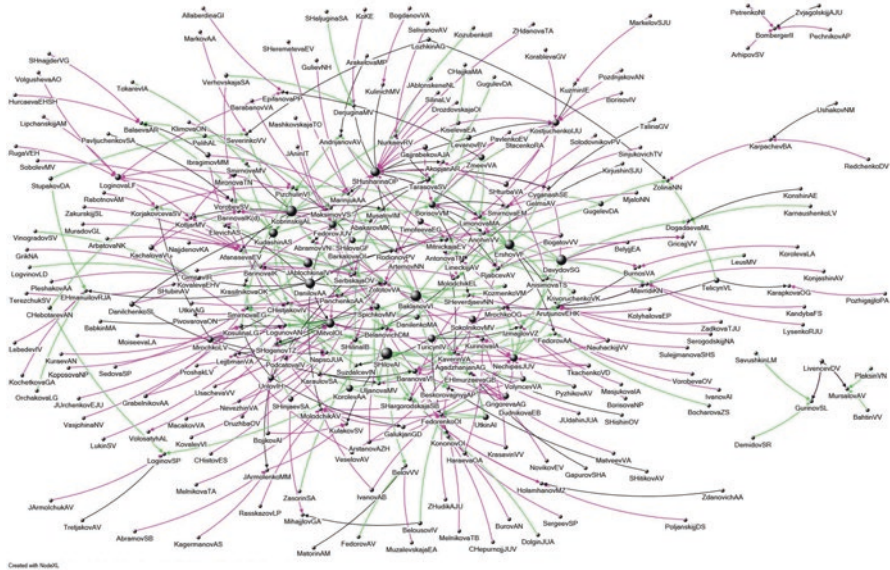


Fig. 27.1 A network in the MGPU producing large-scale plagiarism (A. Abalkina, Dissnet.org). The full interactive graph is available at: [https://www.dissnet.org/publications/mpgu\\_graf.htm](https://www.dissnet.org/publications/mpgu_graf.htm)

distribution is symptomatic because it represents the main bottlenecks in modern Russia: economics, law, and education.

Second, universities have been predominantly responsible for faking academic production, whereas the research institutions of the Russian Academy of Sciences, the RAS, have produced relatively small numbers of detected plagiarism cases. The two most prominent universities in terms of producing faked material during the last fifteen years are Moscow State Pedagogical University and the Russian Presidential Academy of National Economy and Public Administration. Yet other “leading contenders” include the Russian State University for Humanities and the Russian State Social University, as well as the country’s leading seat of learning, Moscow State University. By way of contrast, the RAS, which comprises hundreds of research institutions across the country, was ranked 23rd on the plagiarism list—the frauds being exclusively represented by its Caucasus-based branch.

Finally, the majority (approx. 50%) of those holding questionable academic degrees are working as administrative staff in universities. Not coincidentally, large-scale plagiarism was detected in 66 dissertations (21.22%) defended by rectors (311 of those awarded during the last fifteen years in Russia were checked). Politicians and businessmen fell behind in this regard with only about fifteen percent of their numbers engaging in plagiarism.

Large-scale plagiarism in Russia is, by its very nature, a special case when the numbers are compared to those recently disclosed in Western Europe (for example, see Weber-Wulff 2014). Whereas a Western plagiarist endeavors to present a text that has been copied from others as original research, the high-profile swindler in Russia may well not have even seen the plagiarized text prior to the public defense, having received it ready for publication from ghost-writers. When this occurs, wholesale plagiarism is not disguised; instead, the “dissertation” is composed with a crazy quilt of texts with fully automated replacements.

This pervasive academic corruption inevitably raises various questions. For example, does the widespread occurrence of badly adapted texts indicate a local trend that exclusively features pseudo-academics who attempt to enhance their value among their own kind? Or does it foretell greater changes in acceptable norms that academic communities have faced thus far? We address these questions in our second case study, presented in Sect. 27.5 below.

## 27.5 SMALL-SCALE PLAGIARISM AND SHIFTING NORMS OF TEXTUAL AUTHENTICITY

While the detection of small-scale plagiarism also involves the same tools and collections as those described above, it is more dependent on manual processing in that a small piece of text may be a legitimate quotation or a paraphrase with a valid reference. This challenge calls for deeper conceptual reasoning on the shifting norms of textual authenticity.

As is the case with many other norms, *justifications* for the norm of textual authenticity are subject to deeper disagreement than the norm itself. Those who attempt to provide grounds for accepting this norm tend to present one of two arguments. The first is that either copy-pasting from the texts of other persons is defined as an infringement of these authors' intellectual property and thus as a type of theft, or they regard copy-pasting as a fraudulent way of obtaining intellectual distinction that is not actually deserved, and thus akin to cheating on an exam. The latter interpretation is based on the assumption that an individual with a university-level degree is able, single-handedly, to produce a text that meets certain stringent requirements. Nonetheless, both justifications can be disputed in specific cases. In contrast to more obvious cases of theft, dissertation plagiarism does not necessarily damage the rightful owner of the property, who probably loses little in terms of professional recognition given that dissertations are rarely read. Moreover, as a reason for condemning plagiarism, it becomes irrelevant if an author of the borrowed source raises no objections. The Dissnet studies nevertheless revealed that a person's supervisor and/or opponents are the most likely sources of unauthorized large-scale borrowing (see Sect. 27.4 for details). In all probability, in such cases, the text is borrowed with the author's full consent, thus in the true sense of the word, no theft occurs of intellectual property. As for the second justification, although the copy-pasting of an entire text by another person is obviously incompatible with originality, borrowing some parts of it (such as the literary review or descriptions of procedures) is apparently possible without compromising the originality of the research results. One could therefore argue that the authentic reproduction of the whole text is much less serious than producing substantive original results, particularly in light of the aforementioned disagreements regarding the meaning of originality and authenticity. Despite a certain shakiness concerning the grounds on which it rests, the norm requiring full textual authenticity evolved in Western publishing, and it was officially supported by the VAK (*Vysšaâ attestacionnaâ komissiâ*, All-Russian Attestation Committee)—a state agency based in Moscow that verifies both doctoral and post-doctoral degrees.<sup>4</sup>

Researchers who use the software and data described above could determine how closely the norm of textual authenticity was adhered to by large numbers of academics and identify the deviants who did not follow it. Two hypotheses could be posited here. The first is the “weakness hypothesis” that deviation from the norm of textual authenticity is associated with academic weakness. In other words, this concerns those authors who are unable to produce texts of an acceptable quality and therefore accept the risk associated with plagiarizing. In a slightly different form, this hypothesis predicts that when academics decide whether or not to plagiarize, they self-sort themselves into two groups. There are those for whom the costs of writing an authentic text are greater than the costs of being revealed as a plagiarizer, multiplied by the estimated probability of such a revelation. The second group consists of those for whom the opposite is true (Spence 1973, 2002). The “convention hypothesis,” on the other hand,

holds that some academics disregard the norm because they disagree with its justifications, and may not be fully aware that others support it.

Several predictions follow from the “weakness hypothesis” as to where plagiarism is to be found. In the case of Russia, one would expect plagiarism to occur primarily in disciplines that were the least developed during the Soviet period, but which expanded after the collapse of the Soviet Union, that is, the social sciences. Second, one might expect less borrowing in institutions in which the prime research forces are concentrated, namely, the Academy of Sciences and the top universities. Third, individuals who conduct highly esteemed research are presumably less likely to borrow than those whose results are less prominent.

The “convention hypothesis” does not generate predictions, but it does explain why expectations based on the “weakness hypothesis” may be falsified. If no correlation occurs between borrowing and intellectual weakness, then the social sciences may not differ from the natural sciences, and the best institutions and scholars may not differ from their weaker counterparts. In this case, the principal variable deciding who plagiarizes and who does not is the degree of contact with Western academia and its standardized norms. Indeed, institutions which conduct the highest quality research are also likely to be more globalized. However, this correlation is probably weak, given that there are a few intervening variables.

To determine which hypothesis has more support, we analyzed 2,468 post-doctoral dissertations (*Doktor nauk*, see note 1 above), which were randomly selected from the pool of all dissertations defended in Russia in the years 2006–2015.<sup>5</sup> We utilized the antiplagiat.ru online service, which allowed us to assess the selected texts against many sources, including the Digital Dissertation Library of the RSL.

Figure 27.2 presents the overall distribution of plagiarism that occurred across disciplines. The figure in the graph is a boxplot. It divides the amount of borrowed materials found in each discipline into four quartiles, from the highest to the lowest, and indicates where the boundaries of each of them are situated. The band inside the box corresponds to the median, crosses (X) stand for averages, and points outside of the upper “whisker” are outliers with an extraordinarily high amount of borrowing for a given discipline. Three aspects of plagiarism immediately become apparent. First, inappropriate borrowing is almost universally present.<sup>6</sup> Second, the disciplines differ dramatically in what an “extraordinary” amount of borrowing means to them, such that the exceptional case of borrowing around 30 percent of a text in philology would be close to the average in agriculture. Third, cases of large-scale plagiarism similar to those discovered by Dissernet are rather rare. Thus, from the sample of 2,468 post-doctoral dissertations, we determined that 44 contained borrowing that exceeded 60 percent (1.7%). We checked these 44 manually, and three cases were false positives. Overall, large-scale plagiarism exceeding 50 percent was found in 149 of the 2,468 dissertations (6%). Thus, we further focus on relatively small-scale plagiarism.

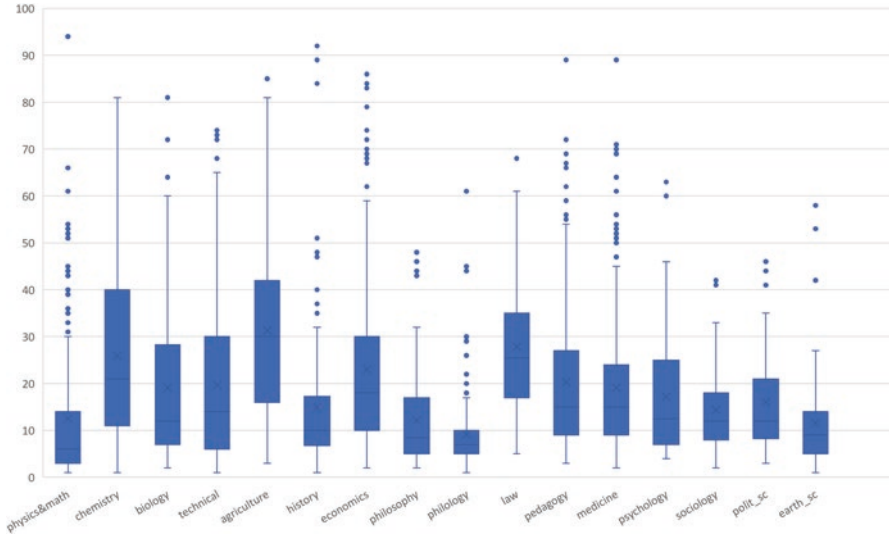


Fig. 27.2 The overall distribution of small-scale plagiarism across disciplines

In contrast to what is posited in the weakness hypothesis, no straightforward connections were discovered between the character of a discipline (humanities, social sciences, or natural sciences; predominantly theoretical or predominantly applied), the degree of its expansion in post-Soviet times, and the degree of plagiarism. Thus, of the three disciplines with the highest levels of unauthorized borrowing, agriculture (natural sciences, predominantly applied) has witnessed moderate expansion, chemistry (natural sciences, including both theoretical and applied subfields) is shrinking, and law (social sciences, both theoretical and applied) is expanding enormously. In the case of specific disciplines, apparently traditions play a key role, which sometimes differ in otherwise closely related subjects such as chemistry and biology, or economics and sociology. In general, it seems that neither the lower-level development of scholarship in a given field in Russia nor its recent expansion played a prominent role in tolerating unauthorized borrowings. The weakness argument does not appear to be valid for the moderate infringement of the norm for textual authenticity. It is interesting, however, that the logic does seem to be applicable in another sense: among the relatively sizable disciplines, philology (in Russia, this includes both literary studies and linguistics) displayed the least amount of borrowing.

We discovered some limited support for the hypothesis positing that aversion to plagiarism will be strongest among institutions of the Russian Academy of Sciences and universities that participate in Project 5-100, as their objective is to have at least five Russian universities among the top one hundred in the world university rankings. However, Russia's leading institutions are the most highly internationalized. For example, the top Russian universities are



evaluated according to the number of foreign students and faculty they employ. The leading institutions also serve as the gateways through which international norms find their way into Russia. In this sense, the “convention hypothesis” that resulted from the adoption of international norms may explain the aversion of these institutions to plagiarism (Table 27.4).

Finally, we examined individual publication profiles in the Russian Index for Scientific Citing. We selected 10 percent of the representatives of each discipline with the highest and the lowest number of borrowing and compared their publication profiles. The formulation of the sample thus eliminated the influence of differences in profile. Table 27.5 presents the results. Although some statistically significant differences emerged in the amount of plagiarism among researchers who publish widely in international publications compared to those who publish exclusively in second-rate domestic editions, such differences are relatively minor in absolute terms. Again, one could infer that according to the “convention hypothesis,” scholars with the most impressive international publication records are also those with the highest exposure to the norms of international publication.

Overall, our findings cast considerable doubt on the validity of the “weakness hypothesis.” It appears that the norm of textual authenticity is not widely accepted in Russia. Although borrowing larger amounts of a text (as in exceeding 50%) is rather rare, recycling the minor parts of other people’s texts is almost a universal practice (probably 75% of dissertations include at least a few slightly re-written paragraphs from the works of others, without attribution).

Some Russian scholars justified borrowing by describing a dissertation and its public defense as a “mere formality” and decrying “senseless conventions.” Others questioned the possibility of dividing collaborative work into personalized scientific contributions. There are no reasons to believe that the tendency to provide such explanations in any way correlates with the authors’ intellectual competency. Regretfully, widespread tolerance toward borrowing in Russia greatly impedes the addressing of more notorious types of plagiarism because it renders the difference between borrowing some technical paragraphs and borrowing the whole text a matter of degree rather than a matter of principle.

**Table 27.4** Percentage of borrowing in dissertations defended at various Russian institutions

<i>Organization</i>	<i>Median percent plagiarized (%)</i>
Top universities <sup>a</sup>	10.1
Russian Academy of Sciences	10.1
Other	15.9
<i>ALL</i>	<i>14.4</i>

<sup>a</sup>This includes 21 participants in Project 5-100, as well as Moscow and Saint-Petersburg state universities, in effect, 23 institutions in all.



**Table 27.5** Differences in publication and citation performance among authors demonstrating the highest and the lowest amount of borrowing

Variable	Averages		Average treatment effect ( $\Delta$ )
	Plagiarism top 10%	Plagiarism bottom 10%	
Publications RISC <sup>a</sup> core, %	19.96	24.99	-5.03*
Citing from RISC core, %	17.79	24.29	-6.50**
Impact factor, published	0.37	0.45	-0.07*
Impact factor, cited	0.42	0.52	-0.10*
Articles in foreign publications, %	3.74	6.84	-3.10***
Citing from foreign publications, %	6.94	10.4	-3.45**

\*Statistically significant values; their numbers reflect the degree of confidence from less (\*) to most (\*\*\*) significant

<sup>a</sup>The Russian Index for Scientific Citing, RISC, includes a “core” of editions receiving the highest evaluations in a survey of Russian academics. It is also partially integrated with the Scopus and Web of Science databases, which enable the tracing of publications and citations from non-Russian-language editions.

## 27.6 CONCLUSION

The aim of this chapter was to describe the tools and resources that are available to detect plagiarism, as well as to establish how academic plagiarism in Russia, detected by automatic means, can be interpreted from different perspectives. The most visible manifestation of this, and the one that is most hotly debated in the media, is the spread of large-scale plagiarism in dissertations by those in power, who believe that possessing an academic degree will advance their careers. Less commonly discussed, but no less interesting, is the range of interpretations of the authenticity norm that underlies the notion of plagiarism. Tolerance toward utilizing someone else’s text, which is evident in Russia, may be the sign of an impending global shift in academia, because it perfectly matches the *Zeitgeist* of digital post-modernity, or as Roland Barthes once observed: *La mort de l’auteur*, “the death of the author” (Barthes 1968).

## NOTES

1. Russia has two higher academic degrees: *kandidat nauk* (Candidate of Science, roughly equal to a Ph.D.) and *doktor nauk* (Doctor of Science, roughly equal to *doctor habilitatus* in some European countries). Henceforth in this chapter, we distinguish doctoral (=PhD) and post-doctoral (=habilitation) dissertations respectively.
2. This section is adapted from an article by M. Kopotev et al.; see (Smirnov et al. 2017).
3. There is currently no software that detects mathematical formulas. The possibilities for detecting “borrowed” graphs and figures are also rather limited, although the situation is rapidly changing (see, for example, Acuna et al. 2018; see also the survey by Eisa et al. 2015).

4. It is important to note that the verification does not extend to research papers.
5. The study reported in this section was conducted between March and November of 2018, at the Centre for Institutional Analysis of Science and Education in collaboration with the Centre for the Sociology of Education of the Russian Presidential Academy. The authors would like to thank Katerina Guba, Alexandra Makeeva, Nadezhda Sokolova, and Anzhelika Tsivinskaya for their help in this project.
6. Antiplagiat produces some false positives. For example, it sometimes counts lists of referenced literature as borrowing, or it may not recognize alternative spellings of an author's name. We checked more than 800 dissertations manually and for most disciplines found medians that were approximately five percent lower than in the case of automatic search. However, the manual check was rather conservative and probably underestimated the scale of the borrowing. The actual statistics are therefore somewhere in between these estimates. There were no significant differences between the relative propensity of disciplines to borrow as estimated by automatic and manual procedures, with one notable exception: automatic checks probably overestimate the borrowing in dissertations on law, most likely due to the highly formulaic forms of speech in this genre.

## REFERENCES

- Acuna, D. E., P. S. Brookes, and K. P. Kording. 2018. Bioscience-scale Automated Detection of Figure Element Reuse. Preprint at bioRxiv. Accessed February 1, 2019. <https://doi.org/10.1101/269415>.
- Barthes, R. 1968. La mort de l'auteur. *Manteia* 5: 12–17.
- Clauset, Aaron, Samuel Arbesman, and Daniel B. Larremore. 2015. Systematic Inequality and Hierarchy in Faculty Hiring Networks. *Science Advances* 1 (1): e1400005.
- Denisova-Schmidt, E. 2016. Corruption in Russian Higher Education. *Russian Analytical Digest* 191: 5–9.
- Eisa, Taiseer, Naomie Salim, and Salha Alzahrani. 2015. Existing Plagiarism Detection Techniques: A Systematic Mapping of the Scholarly Literature. *Online Information Review* 39: 383–400. <https://doi.org/10.1108/OIR-12-2014-0315>.
- Firth, J.R. 1957. A Synopsis of Linguistic Theory, 1930–1955. In *Studies in linguistic analysis. A Special Volume of the Philological Society*, 1–32. Oxford: Oxford University Press.
- Gipp, B. 2014. *Citation-based Plagiarism Detection. Detecting Disguised and Cross-language Plagiarism Using Citation Pattern Analysis*. Springer Fachmedien Wiesbaden. Accessed February 1, 2019. <http://link.springer.com/book/10.1007%2F978-3-658-06394-8>.
- Golunov, S. 2014. *The Elephant in the Room: Corruption and Cheating in Russian Universities*. Columbia University Press.
- Gupta, P., K. Singhal, P. Majumder, and P. Rosso. 2011. Detection of Paraphrastic Cases of Mono-lingual and Cross-lingual Plagiarism. In *Proceedings of ICON-2011*, Macmillan Publishers, India. Accessed February 1, 2019. [http://users.dsic.upv.es/~prossor/resources/GuptaEtAl\\_ICON11.pdf](http://users.dsic.upv.es/~prossor/resources/GuptaEtAl_ICON11.pdf).
- Kicherova, M., D. Kyrov, P. Smykova, et al. 2013. Plagiat v studenčeskikh rabotah: analiz sušnosti problemy [Plagiarism in Students' Papers: Toward the Roots of the

- Problem]. *Online Journal Naukovedenie (IGUPIT)*, 4. Accessed February 1, 2019. <http://naukovedenie.ru/PDF/83pvn413.pdf>.
- Korbut, A. 2013. Plagiat i konstitutivnyj porjadok dissertacionnogo teksta [Plagiarism and the Constitutive Order of a Dissertation Text]. *Sociologičeskoe obozrenie* 12 (2): 145–171.
- Kutuzov, Andrey, Mikhail Kopotev, Tatyana Sviridenko, and Lyubov Ivanova. 2016. Clustering Comparable Corpora of Russian and Ukrainian Academic Texts: Word Embeddings and Semantic Fingerprints. Accessed February 1, 2019. <https://arxiv.org/pdf/1604.05372.pdf>.
- Levenshtein, Vladimir I. 1966. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics* 10 (8): 707–710.
- Maloshonok, N. 2016. Kak vospriat'ie akademičeskoj čestnosti srede universiteta vzaimosvazano so studenčeskoj vovlečennost'ju: vozmožnosti konceptualizacii i èmpiričeskogo izučeniâ [How Perception of Academic Honesty at the University Is Linked with Student Engagement: Conceptualization and Empirical Research Opportunities]. *Voprosy obrazovaniâ* 1: 35–60.
- McEnery, Tony. 2004. *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Routledge.
- Modern Language Association. 2008. *MLA Style Manual and Guide to Scholarly Publishing*. New York: Modern Language Association of America.
- Nikitov, A., et al. 2012. Plagiat v rabotah studentov i aspirantov: problema i metody protivodejstviâ [Plagiarism in Under- and Postgraduate Students' Papers: Problem and Actions Against ]. *Universitetskoe upravlenie: praktika i analiz* 5: 61–68.
- Oakes, M. 2014. *Literary Detective Work on the Computer*. Amsterdam: John Benjamins Publishing Company.
- OED Online. June 2020. Oxford University Press. Accessed July 01, 2020. <https://www.oed.com/view/Entry/272993?rskey=Ma8aD8&result=2&isAdvanced=false>.
- Palahniuk, Chuck. 1996. *Fight Club*. New York: W. W. Norton & Company.
- Potthast, M., B. Stein, A. Barrón-Cedeño, and P. Rosso. 2010. An Evaluation Framework for Plagiarism Detection. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters: 997–1005*. Accessed February 1, 2019. [http://www.uni-weimar.de/medien/webis/publications/papers/stein\\_2010p.pdf](http://www.uni-weimar.de/medien/webis/publications/papers/stein_2010p.pdf).
- Smirnov, I., R. Kuznetsova, M. Kopotev, A. Khazov, O. Lyashevskaya, L. Ivanova, and A. Kutuzov. 2017. Evaluation Tracks on Plagiarism Detection Algorithms for the Russian Language. In *Computational Linguistics and Intellectual Technologies. Proceedings of the International Conference Dialogue-2017*, 285–297. Moscow: RSUH.
- Spence, M. 1973. Job Market Signaling. *The Quarterly Journal of Economics* 87 (3): 355–374.
- . 2002. Signaling in Retrospect and the Informational Structure of Markets. *American Economic Review* 92 (3): 434–459.
- Weber-Wulff, D. 2014. *False Feathers. A Perspective on Academic Plagiarism*. Berlin: Springer.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Automatic Sentiment Analysis of Texts: The Case of Russian

*Natalia Loukachevitch*

## 28.1 INTRODUCTION

Automatic sentiment analysis of texts, that is, the identification of the author's opinion about the subject discussed in the text, has been one of the most significant tasks in natural language processing in the past two decades. The interest in sentiment analysis is connected to the large volume of electronic texts available on social networks and online recommendation services that contain an abundance of individuals' opinions on various issues: from products and services to the current political and economic situation (for more on corpora, see Chap. 17).

A large number of scholarly works is devoted to sentiment analysis of user reviews stored in recommendation services (Pang et al. 2002; Pang and Lee 2008; Liu 2012). Another important area of sentiment analysis is the so-called reputation monitoring that tracks positive and negative feedback about a company and its products (Amigo et al. 2012). Sentiment analysis of financial reports and financial news is used to determine trends in the stock and currency markets (Nassirtoussi et al. 2015). The sentiment of mentioning terms in scientific articles is used to predict the most important concepts and scientific trends (McKeown et al. 2016). Sentiment information extracted from texts can be used to determine the personal characteristics of the author (Volkova et al. 2015).

The role of automatic sentiment analysis of social network messages for political and social research is growing. Such studies include the identification

---

N. Loukachevitch (✉)  
Lomonosov Moscow State University, Moscow, Russia

of political preferences (Volkova et al. 2014), the prediction of election results (Vepsäläinen et al. 2017; Vilares et al. 2015), and the identification of attitudes toward various political decisions. Also, automatic sentiment analysis can be used to recognize hate speech and calls for violence or fake news (Volkova and Bell 2016).

The first approaches to sentiment analysis aimed to determine the overall sentiment of the document or its fragment (Pang et al. 2002). This level of analysis assumes that a document expresses a unanimous opinion about a single entity, such as in a review of a product. Since the document can express multiple attitudes in relation to the different entities it contains, at the next stage scholars studied the tasks of sentiment analysis aimed toward specified entities mentioned in the text (Amigo et al. 2012; Jiang et al. 2011; Loukachevitch et al. 2015; Loukachevitch and Rubtsova 2016). Finally, an even more detailed level of sentiment analysis is the analysis of opinions on specific properties or parts (the so-called aspects) of the entity (Liu and Zhang 2012; Pontiki et al. 2016; Popescu and Etzioni 2007).

Liu and Zhang (2012, 4) define opinion as a five tuple  $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ , where  $e_i$  is the name of an entity to which the opinion relates,  $a_{ij}$  is an aspect (part or characteristic) of  $e_i$ ,  $s_{ijkl}$  is the sentiment regarding the entity and its aspect,  $h_k$  is the author of the opinion (opinion holder), and  $t_l$  is the time when the opinion is expressed by  $h_k$ . The sentiment  $s_{ijkl}$  may be positive, negative, or neutral, or may be expressed with varying degrees of intensity that is measured, for example on a scale of 1–5.

In this chapter, we first describe the problems that can be encountered in automatic sentiment analysis. Then, we briefly consider the main methods to conduct sentiment analysis and approaches to creating sentiment vocabularies. Finally, Russian-specific components of automatic sentiment analysis are described, including publicly available vocabularies and sentiment-related shared tasks.

## 28.2 PROBLEMS IN SENTIMENT ANALYSIS

If we ask native speakers what the most significant problems in sentiment analysis would be, the respondents often name irony and sarcasm. Certainly, the difficulties with these language phenomena really exist but problems of automatic sentiment analysis are much more diverse. In what follows, six additional challenges of sentiment analysis are presented.

### 28.2.1 *Multiple Opinions in a Single Text*

Approaches to extracting the main components of opinion largely depend on the genre of the analyzed text. One of the most studied genres of text in the task of sentiment analysis are user reviews on products or services. Such texts usually consider a single entity (but, perhaps in its different aspects), and the

opinion is expressed by one author, namely the reviewer (Pang et al. 2002; Pang and Lee 2008; Liu 2012).

Another popular type of texts for sentiment extraction is Twitter messages (Pak and Paroubek 2010; Rosenthal et al. 2017; Loukachevitch and Rubtsova 2016). Tweets (Twitter posts) were limited to 140 symbols before 2017, when they were extended to 280 characters. Such short texts often require precise sentiment analysis but most of them mention the only opinion target and opinion holder (for more on Twitter analysis, see Chap. 30). The following tweet shows an example of a negative attitude toward Russian phone company Megaphone, presented in sarcastic form, which requires the use of sophisticated methods to reveal the correct attitude:

*Megafon, spasibo tebe za zablockirovannye uvedomleniâ ot Rajffajzena* [Megaphone, thank you for the blocked notifications from Raiffeisen]

It can seem that in longer texts the author's opinion can be repeated several times in different ways, which would facilitate the analysis. However, long texts may include various entities and related sentiments (Choi et al. 2016; Loukachevitch and Rusnachenko 2018) and they may mention opinions of different persons. If the task is to find an attitude toward the entities mentioned, then the problem of determining the scope of the sentiments arises. For example, sentiment extraction is often carried out in relation to an entity mentioned in the same sentence. However, the author can refer to an entity using the means of reference, for example, pronouns. In addition, if the entire text is devoted to the discussion of one entity, then it can be explicitly mentioned far from the sentiment location (Ben-Ami et al. 2014).

In such document genres as news texts, or especially analytical texts, many opinions from different sources can be simultaneously mentioned. These texts contain opinions conveyed by different subjects, including the author(s)' attitudes, the positions of cited sources, and the relations of the mentioned entities to each other. Analytical texts usually contain a lot of named entities, and only a few of them are subjects or objects of a sentiment attitude (Loukachevitch and Rusnachenko 2018). It is clear that in texts with multiple subjects and/or objects of opinion, the complexity of high-quality automatic analysis of sentiment increases manifold.

### 28.2.2 *Implicit vs. Explicit Sentiment*

It is usually assumed that sentiment is expressed using specialized sentiment words (such as *good*, *bad*, *awful*), which is an explicit way of conveying attitudes. However, sentiment can be expressed also implicitly with the so-called sentiment facts (Liu 2012; Loukachevitch and Levchik 2016; Tutubalina 2015) or words with connotations (Feng et al. 2013).

According to the definition provided by Liu (2012, 26), an implicit opinion is an objective statement, from which the sentiment follows, that is, an implicit



opinion that conveys a desirable or undesirable fact. In preparation of datasets for testing sentiment analysis systems, such sentiment facts can be specifically annotated (Loukachevitch et al. 2015; Nozza et al. 2017). For example, Russian restaurant reviews may include such sentences as: “*Dolgo ždali*” (Waited for a long time) or “*Našli muhu v supe*” (Found a fly in the soup), which, on the one hand, describe what happened (report real facts), but on the other hand convey sentiment.

Connotation is a feeling or idea that is suggested by a particular word, although it need not be a part of the word’s meaning. Connotations often convey positive or negative sentiment (Feng et al. 2013). The appearance of words with positive or negative connotations in a text correlates with the corresponding sentiment expressed in the text. For example, in movie reviews, names of famous actors usually have positive connotations. In restaurant reviews, the noun *muha* (fly) is associated with a negative sentiment in different contexts, for example:

*No sil’no dulo ot okna, pri ètom letala nazojlivaâ muha i ne hvatalo oficiantov* [But there was a strong draft from the window, while an annoying fly was flying around and there were not enough waiters].

*Prišli v kafe na Ozernoj, oficiantku ele doždalis’ ležala muha mertvaâ na stole* [Went to the cafe on Lake street, barely waited for the waitress, there was a dead fly on the table].

An interesting example of a word with specific connotations in Russian restaurant reviews is the word *majonez* (mayonnaise). Many sources indicate *majonez* as a key component of Soviet and Russian cuisine (Shearlaw 2014; Whalley 2018). However, when mentioned in contemporary Russian restaurant reviews, this word usually conveys negative sentiment, for example:

*Absolútno vse salaty sodëržat majonez, pričem ego vezde mnogo* [Absolutely all salads contain mayonnaise, and lots of it in everything].

*Edinstvennye teplye rolly byli táželovaty vvidu naličiâ v nih majoneza* [The only hot rolls were heavy due to the presence of mayonnaise].

In news and analytical texts, we can find a lot of words with international negative connotations such as *war*, *unemployment*, *segregation*, or *traffic jam*. Positive connotations are often associated with achievements of a nation. For example, in Russia positive connotations are associated with cosmos-related concepts such as *sputnik*, *Yuri Gagarin*, or *MKS* (International Space Station).

Gradual adjectives (such as *long – short*, *large – small*, etc.) can often convey sentiment facts but their sentiment orientation is very dependent of the context (Cambria et al. 2010). For example, the word *long* can be both negative and positive in the digital camera domain: if it has a long battery life, it means the battery is good; if you need to adjust the focus for a long time, then the opinion about the camera is negative.

Because of the existence of implicit sentiments and connotations, it is impossible to create general sentiment lexicons, which can be equally useful across many domains. It is, therefore, necessary to develop specialized sentiment lexicons using domain-specific text collections or update existing general lexicons to adapt them to a specific domain. (Hamilton et al. 2016; Severyn and Moschitti 2015; Chetviorkin and Loukachevitch 2012).

### 28.2.3 *Ambiguity of Sentiment Words*

Difficulties with the interpretation of explicit sentiment vocabulary may also arise. Sentiment words can be ambiguous: in one sense, they can be neutral, while in other senses they are negative or positive (Akkaya et al. 2009; Baccianella et al. 2010). For example, the Russian word *presnyj* (fresh) bears a positive connotation in the phrase *presnaá voda* (freshwater), while in other senses of the word *presnyj* (tasteless for food and uninteresting as in movie reviews) this word is negative.

A word can change or lose its polarity depending on the subject area or the current context. For example, the Russian sentiment words *verolomstvo* (treachery) and *predatel'stvo* (betrayal) cannot be considered as conveying an opinion in movie reviews, because they are usually mentioned in a movie synopsis to retell the plot of a movie. The word *smešnoj* (funny), most likely, is negative in the sphere of politics, yet indicates a positive orientation when it is used in reviews of comedies. When characterizing other movie genres, this word can be both positive and negative.

### 28.2.4 *Sentiment Modifiers*

The appearance of sentiment words in the text may be accompanied by sentiment modifiers that enhance (for example, *much*, *more*), reduce (*too*, *less*) or inverse prior word sentiment (negation: *no*, *not*). Thus, when analyzing the sentiment, such modifiers should be taken into account, and it is necessary to have some numerical model that modifies the original polarities of the word (Taboada et al. 2011; Wilson et al. 2005; Wiegand et al. 2010). One of the common models of accounting for polarity modifiers ascribes some coefficients to them, which are considered as factors modifying the initial polarity of the words to which these modifiers relate.

Another important issue is determining the scope of the polarity modifier in a particular sentence (Taboada et al. 2011). Most approaches suppose that polarity modifiers, such as negation, modify sentiment of neighbor words, but long-distance influence is also possible. For example, in the sentence “*Á ne dumaú, čto èto zasluživayet upominaniá*” (I do not think it is worth mentioning), the negation changes the sentiment orientation of the word *zasluživayet* (worth) from positive to negative.

If negation stands before several sentiment-bearing words, it is important to calculate the overall sentiment of the whole group and then to apply negation

to it. In the following sentence, we see the phrase “*ne boitsá raskola*” (is not afraid of a split), where negation stands before two words with negative sentiment. To obtain the positive mood of the sentence, it is necessary to determine the sentiment of the phrase as negative and then to apply negation to it:

*Sekretar’ prezidiuma gensoвета “Edinoj Rossii,” zampredsedatelá Gosdumy Sergej Neverov v subbotu zaávil, čto partiá ne boitsá raskola v svázi s poávleniem v nej raznyh ideologičeskikh platform* [Secretary of the Presidium of the General Council of “United Russia,” Deputy Chairman of the State Duma Sergei Neverov, on Saturday stated that the party **is not afraid of a split** in connection to the appearance of various ideological platforms within it].

Polarity modifiers can also form groups such as double negation. We can see such double negation *ne bez* in well-known Russian proverb “*V sem’ye ne bez uroda,*” which translates into English without any negations: “Every family has its black sheep.” In this example, we see negative sentiment as if negation coefficients were multiplied.

### 28.2.5 *Factors of Irreal Context*

When analyzing the sentiment, it is important to consider how a proposition conveying sentiment corresponds to reality (Saurí and Pustejovsky 2012; Taboada et al. 2011; Wilson et al. 2005). For example, in the sentence “*My nadeáliš’, čto nam ponravitsá kino*” (We hoped that we would like the movie), one can see the positive word *ponravitsá* (like), but it says nothing about whether the author really liked the movie.

In linguistics, this is covered by the concept of irreals or irreal mood, which is a group of grammatical means that is used to denote that what is said in a sentence does not refer to what really happens (Taboada et al. 2011). In every language, there are some factors showing that the proposition is not factual (the so-called irrealis markers). In Russian modal verbs, private-state verbs, such as *nadeát’sá* (to hope), *ožidat’* (to expect), *dumat’* (to think), can be used as such markers.

According to Kuznetsova et al. (2013, 72), for the Russian language, such function words as *esli, by, li, esli by* also often mark the irrealis mood. When selecting parameters on the training set, Kuznetsova et al. (2013, 72) indicated that the prior sentiment scores of sentiment words found in the sentences with irrealis markers should be decreased (but not nullified).

### 28.2.6 *Comparisons*

Comparisons complicate the process of determining sentiment, because additional entities are mentioned in the text, and some sentiments can refer to them. It is often supposed that comparisons are conveyed with the so-called comparative constructions such as *lučše čem* (better than) or *dorože čem* (more

expensive than). In most cases, comparisons may be introduced without any specialized constructions. Additional entities mentioned for comparison are sometimes very difficult to detect, and it can also be a complex task to single out the attitudes related to them. For example, in the following extract from a restaurant review, the comparison is marked with word *drugoy* (another), and positive words *naslaždalis'* (enjoyed) and *volšebnym* (wonderful) characterize a restaurant distinct from the restaurant under review (example from Loukachevitch et al. 2015, 8):

*My rešili ne brat' zdes' desert i kofe, a pošli v drugoj restoran, gde naslaždalis' volšebnym zaveršeniem našego večera* (We decided not to have dessert and coffee there, but instead went to another restaurant where we **enjoyed a wonderful** end to our evening).

### 28.2.7 Irony and Sarcasm

The processing of irony and sarcasm is a serious problem for sentiment analysis systems, since the sentiment of an ironic (sarcastic) utterance differs from its literal sentiment (Wilson and Sperber 2007). In Benamara et al. (2017, 37), a generalized understanding of irony is proposed as “an incongruity between the literal meaning of an utterance and its intended meaning.” Most often, a positive-looking statement (containing more positive sentiment words or an equal number of positive and negative words) hides a negative opinion, for example, “*Sberbank—naibolee krupnaâ set' nerabotaûsîh bankomatov v Rossii*” (Sberbank is the largest network of nonoperating ATMs in Russia). Sarcasm is regarded as a sharper, more aggressive, possibly degrading form of irony (Benamara et al. 2017).

The annotation of textual data for the study of irony and sarcasm is a complex task. Interesting data for analyzing these phenomena are Twitter messages that the user can mark with special hashtags: *#irony*, *#sarcasm* and some others (Reyes et al. 2013; Sulis et al. 2016). However, recent studies of irony in Twitter show that ironic tweets marked with hashtags and annotated by experts have different characteristics (Kunneman et al. 2015). In addition, in the Russian segment of Twitter, users do not use similar Russian hashtags in the same way as American or European audience (Zefirova and Loukachevitch 2019, 48). The “*ironiâ*” (irony) hashtag is mostly used as a description for images or jokes alongside with such hashtags as “*#šutka*” (joke), “*#smel'*” (laugh) and does not seem to express the desired content.

## 28.3 METHODS AND RESOURCES USED IN SENTIMENT ANALYSIS

Automatic analysis of sentiment can utilize two main types of approaches (Liu 2012; Pang and Lee 2008): knowledge-based methods using sentiment lexicons and rules (Taboada et al. 2011; Kuznetsova et al. 2013) and approaches

based on machine learning (Liu 2012; Pang and Lee 2008). Knowledge-based methods require the creation of a specialized sentiment lexicon for a specific domain. Linguistic rules are necessary to sum up sentiment scores of several sentiment words and for accounting for the word context (sentiment modifiers, unreal context, etc.).

Supervised machine learning requires preliminary annotation of a training collection. Depending on the task, different classification algorithms, features of the text representation, and feature weights can be chosen (Pang et al. 2002; Pang and Lee 2008; Liu 2012). Currently, the best results in machine-learning sentiment analysis are achieved by deep learning with neural networks (Rosenthal et al. 2017; Cliché 2017; Arkhipenko et al. 2016), which substituted a previous leader: Support vector machine (SVM) classifier (Pang et al. 2002; Pang and Lee 2008; Chetviorkin and Loukachevitch 2013).

At present, there exist also approaches that integrate available sentiment vocabularies (both manually created and automatically generated) into machine learning methods, transforming them into specialized features (Rosenthal et al. 2017; Mohammad et al. 2013; Loukachevitch and Levchik 2016). The use of preliminary created lexicons helps to overcome data sparsity of training collections (Loukachevitch and Rubtsova 2016). Below, we consider some approaches to creating sentiment lexicons and publicly available Russian lexicons.

Most sentiment vocabularies look like lists of words and expressions with scores of their sentiment (Wilson et al. 2005). Some vocabularies also provide additional characteristics of the word sentiment called “strength.” Sentiment scores can also be assigned to specific senses of ambiguous words (Baccianella et al. 2010; Loukachevitch and Levchik 2016).

For many languages, general sentiment vocabularies have been published. Despite the fact that in each particular domain, specialized vocabularies are needed, general lexicons are also useful since they can serve as source material, which can then be adapted to a domain. Domain-specific sentiment vocabularies are usually generated with automatic or semiautomatic methods using domain-specific text collections (Hamilton et al. 2016; Severyn and Moschitti 2015; Chetviorkin and Loukachevitch 2012).

For Russian, Chetviorkin and Loukachevitch (2012) have described an automatically generated Russian sentiment lexicon in the domain of products and services called ProductSentiRus (ProductSentiRus 2012). The ProductSentiRus lexicon is obtained by applying a supervised machine-learning model to several domain review collections. It is presented as a list of 5000 words ordered by the decreased probability of their sentiment orientation without any positive or negative labels. For example, the most probable sentiment words in ProductSentiRus are as follows: *bespodobnyj* (peerless), *nevñatnyj* (slurred), *obaldennyj* (awesome), *otvratnyj* (disgusting), et cetera.

The general Russian lexicon of sentiment words and expressions, RuSentiLex (RuSentiLex 2017), was created in a semiautomatic way (Loukachevitch and Levchik 2016). The entries of the RuSentiLex lexicon are classified according to four sentiment categories (positive, negative, neutral, or positive/negative)

and three sources of sentiment (opinion, emotion, or fact). The words in the lexicon that have different sentiment scores in different senses are linked to the appropriate concepts of the thesaurus of the Russian-language RuThes (Loukachevitch and Dobrov 2014; RuThes 2016), which can help disambiguate sentiment ambiguity in specific domains or contexts. The lexicon was gathered from several sources: opinionated words from general Russian thesaurus RuThes, slang and curse words extracted from Twitter, and objective words with positive or negative connotations from a news collection (Loukachevitch and Levchik 2016; for more on RuThes, see Chap. 18). For example, the description of word *presnyj* in RuSentiLex is as follows (labels in quotes correspond to the names of RuThes concepts):

- *presnyj*, Adj, *presnyj*, negative, emotion, “NEVKUSNYJ” [*tasteless*];
- *presnyj*, Adj, *presnyj*, negative, opinion, “NEINTERESNYJ” [*insipid*];
- *presnyj*, Adj, *presnyj*, positive, fact, “PRESNAĀ VODA” [*fresh water*]

The Russian sentiment lexicon LINIS Crowd was created via crowdsourcing (Koltsova et al. 2016; LINIS Crowd SENT 2016). The lexicon is aimed at detecting sentiment in user-generated content (blogs, social media) related to social and political issues. Each word was assessed by at least three volunteers in the context of different texts. The words were scored from  $-2$  (negative) to  $+2$  (positive). For example, the word *anarhizm* (anarchism) obtained three 0 (neutral) scores and three  $-1$  (weakly negative) scores in the considered contexts.

Several international lexicons were automatically constructed for Russian. The Chen-Skiena’s lexicon (2876 words) (Chen and Skiena 2014; Chen-Skiena’s Lexicon 2014) was generated for 136 languages via graph propagation from seed words. However, from the human point of view, the words included in this automatically generated lexicon seem extremely strange. For example, positive words in the Chen-Skiena’s Lexicon include such words as *tipa* (type of), *post* (post), *sootvetstvenno* (correspondingly), *sovsem* (at all), et cetera.

Mohammad and Turney (2013) generated the Russian variant of the EmoLex lexicon (EmoLex 2017) with automatic translation from the English lexicon obtained by crowdsourcing (4412 Russian words).

Kotelnikov et al. (2018) studied available Russian sentiment lexicons and found that all the lexicons have relatively small intersection with each other. Besides, the translated lexicons (EmoLex and Chen-Skiena’s lexicon) have a smaller intersection with other lexicons than on average (10.0%), and at the same time are relatively similar to each other (18.2%). Kotelnikov et al. (2018) also compared available Russian lexicons as features in machine-learning text categorization. Users’ reviews from five domains (books, movies, banks, hotels, and kitchens) were used as text collections for the experiments. The study found that the best results of classification using a single lexicon in all domains

were obtained with ProductSentiRus (Chetviorkin and Loukachevitch 2012). The union of all lexicons gives slightly better results.

As was mentioned, useful sentiment lexicons should be fine-tuned or constructed specially for the domain under analysis. Therefore, to apply sentiment lexicons in a specific domain, it is recommended to gather all available lexicons and to collect a domain-specific text collection as large as possible. Having such data, it is possible to filter out sentiment words and constructions relevant in the domain.

## 28.4 RUSSIAN SENTIMENT-RELATED SHARED TASKS

For the evaluation of Russian sentiment analysis systems, several shared tasks have been organized. In 2011–2013, two evaluations of document-level sentiment approaches were carried out. Two types of text collections were used for the evaluation: users' reviews in three domains (movies, books, and digital cameras) and news quotations (Chetviorkin and Loukachevitch 2013).

For training in the review track, users' reviews were collected from recommendation services (Imhonet and Yandex.market). The reviews had users' scores on a ten-point scale for the Imhonet reviews (movies and books) and on a five-point scale for the Yandex reviews (digital cameras). The participants could choose any of tracks classifying reviews into two, three, or five classes. The reviews for the test collections were extracted from social network messages. The sentiment annotation was created manually by human experts. The participants utilized various machine-learning and knowledge-based approaches, but the best methods in all review-related tasks were SVM-based classifiers.

In the quotation track, direct or indirect speech fragments extracted from news reports had to be classified into three classes (positive, negative, or neutral). About 5000 fragments each were prepared for the training and test collection. Both collections were annotated manually; therefore, the size of the training collection was much smaller than for the review task. In this quotation task, the knowledge-based approaches showed the best results (Chetviorkin and Loukachevitch 2013).

The second series of Russian sentiment analysis evaluations (SentiRuEval 2014–2016) was devoted to the entity-oriented and aspect-based tasks of sentiment analysis. Namely, the tasks included aspect-based analysis of reviews in two domains (car and restaurant reviews). Using the prepared collections, Russian training and test datasets were further utilized in the international SemEval aspect-based sentiment evaluation in 2016 (Pontiki et al. 2016; ABSA SemEval-2006 2016).

The entity-oriented task was based on Twitter messages. The participants were asked to classify messages into three classes from the point of view of reputation monitoring (positive, negative, or neutral) in two separate domains: banks and mobile operators. For example, positive tweets could contain a



positive opinion or positive fact about the company. The training and test collections were prepared via crowdsourcing (SentiRuEval-2016 data 2016).

The approaches of the participants for the Twitter sentiment analysis differed significantly in 2015 and 2016 (Loukachevitch and Rubtsova 2016). In 2015, the basic approach was the SVM classifier trained on only the training collection without any additional data (unlabeled text collections or sentiment lexicons). Due to this, the participating systems could make mistakes in the classification of the test tweets if a tweet contained sentiment words absent in the training dataset (Loukachevitch and Rubtsova 2016).

In 2016, the best approach was based on neural networks, which used word embeddings (vector representations of words) calculated on a large collection of user comments (Arkhipenko et al. 2016). Such representations allowed the winner to overcome the differences in the training and test collections because words that have semantic similarity also have similar vector representations. The next most successful approaches in terms of the quality of results combined machine learning and the existing Russian lexicons (Loukachevitch and Rubtsova 2016).

## 28.5 CONCLUSION

Automatic sentiment analysis of texts is among the popular applications in natural language processing of texts. In this chapter, we described the problems that can be encountered in automatic sentiment analysis. Then, we briefly considered the main methods for sentiment analysis and approaches to creating sentiment vocabularies. Finally, Russian-specific components of automatic sentiment analysis—publicly available vocabularies and sentiment-related shared tasks—were presented.

The current state of affairs in sentiment analysis (including in its application to the Russian language) can be characterized as follows: approaches to sentiment analysis of some text genres, such as user reviews or short posts on social networking sites, are well studied, but there are a lot of complicated phenomena in sentiment analysis that require further research, especially in the processing of full-text news and analytical articles.

From the practical point of view, there are at least four Russian sentiment vocabularies currently available on the Internet. To apply sentiment lexicons in a specific domain, it is recommended to gather all available lexicons and to collect a domain-specific text collection as large as possible. Having such data, it is possible to filter out sentiment words and constructions that are relevant in the domain.

## REFERENCES

- ABSA SemEval-2016. 2016. Data for Aspect-Based Sentiment Analysis, SemEval-2016. <http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools>.

- Akkaya, Cem, Janyce Wiebe, and Rada Mihalcea. 2009. Subjectivity Word Sense Disambiguation. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, vol. 1, 190–199. Association for Computational Linguistics.
- Amigo, Enrique, Adolfo Corujo, Julio Gonzalo, Edgar Meij, and Maarten Rijke. 2012. Overview of RepLab. 2012: Evaluating Online Reputation Management Systems. *CLEF-2012 Working Notes*. <http://ceur-ws.org/Vol-1178/CLEF2012wn-RepLab-AmigoEt2012.pdf>.
- Arhipenko, Konstantin, Ilya Kozlov, Yuriy Trofimovich, Kirill Skorniakov, Andrey Gomzin, and Denis Turdakov. 2016. Comparison of Neural Network Architectures for Sentiment Analysis of Russian Tweets. *Proceedings of International Conference on computational linguistics and intellectual technologies Dialog-2016*, 50–58.
- Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. *Proceedings of Language Resources and Evaluation Conference LREC-2010*, vol. 10, 2200–2204.
- Benamara, Farah, Maite Taboada, and Yannick Mathieu. 2017. Evaluative Language Beyond Bags of Words: Linguistic Insights and Computational Applications. *Computational Linguistics* 43: 201–264.
- Ben-Ami, Zvi, Ronen Feldman, and Binyamin Rosenfeld. 2014. Entities' Sentiment Relevance. *Proceedings of Association for Computational Linguistics Conference ACL-2014*, 87–92.
- Cambria, Eric, Amir Hussain, Catherine Havasi, and Chris Eckl. 2010. Sentic Computing: Exploitation of Common Sense for the Development of Emotion-Sensitive Systems. *Development of Multimodal Interfaces: Active Listening and Synchrony* 5967: 148–156. Berlin and Heidelberg: Springer, LNCS.
- Chen, Yanqing, and Steven Skiena. 2014. Building Sentiment Lexicons for All Major Languages. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics ACL-2014*, vol. 2, 383–389.
- Chen-Skienna's Lexicon. 2014. Multilingual Sentiment Lexicons, Including Russian. <https://sites.google.com/site/datascienceslab/projects/multilingualsentiment>.
- Chetviorkin, Iliia, and Natalia Loukachevitch. 2012. Extraction of Russian Sentiment Lexicon for Product Meta-Domain. *Proceedings of COLING-2012*, 593–610.
- . 2013. Evaluating Sentiment Analysis Systems in Russian. *Proceedings of the 4th Biennial International Workshop on Balto-Slavic natural Language Processing*, 12–17.
- Choi, Eunsol, Hannah Rashkin, Luke Zettlemoyer, and Yejin Choi. 2016. Document-level Sentiment Inference with Social, Faction, and Discourse Context. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL-2016*, 333–343.
- Cliché, Mathieu. 2017. BB twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs. *Proceedings of the 11th International Workshop on Semantic Evaluation SemEval 17*, 572–579.
- EmoLex. 2017. NRC Word-Emotion Association Lexicon, Version 2017. <http://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>.
- Feng, Song, Jun Seok Kang, Polina Kuznetsova, and Yejin Choi. 2013. Connotation Lexicon: A Dash of Sentiment Beneath the Surface Meaning. *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics, ACL-2013*, 1774–1784.
- Hamilton, William, Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 595–604.

- Jiang, Long, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target Dependent Twitter Sentiment Classification. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics ACL-2011*, 151–160.
- Koltsova, Olesya, Svetlana Alexeeva, and Sergey Kolcov. 2016. An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. *Proceedings of Computational Linguistics and Intellectual Technologies Conference Dialogue-2016*, 277–287.
- Kotelnikov, Evgeny, Tatiana Peskischeva, Anastasia Kotelnikova, and Elena Razova. 2018. A Comparative Study of Publicly Available Russian Sentiment Lexicons. *Conference on Artificial Intelligence and Natural Language. AINL 2018. Communications in Computer and Information Science 930*, 139–151. Cham: Springer.
- Kunneman, Florian, Christine Liebrecht, Margot van Mulken, and Antal van den Bosch. 2015. Signaling Sarcasm: From Hyperbole to Hashtag. *Information Processing and Management 51*: 500–509.
- Kuznetsova, Ekaterina, Natalia Loukachevitch, and Ilya Chetviorkin. 2013. Testing Rules for a Sentiment Analysis System. *Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog-2013*, vol. 2, 71–81.
- LINIS crowd SENT. 2016. Russian Sentiment Lexicon, Version of 2016. <http://linis-crowd.org/>.
- Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Liu, Bing, and Lei Zhang. 2012. A Survey of Opinion Mining and Sentiment Analysis. In *Mining Text Data*, 415–463. Springer.
- Loukachevitch, Natalia, and Boris Dobrov. 2014. RuThes Linguistic Ontology vs. Russian Wordnets. *Proceedings of the Seventh Global Wordnet Conference GWC-2014*, 154–162.
- Loukachevitch, Natalia, and Anatoly Levchik. 2016. Creating a General Russian Sentiment Lexicon. *Proceedings of Language Resources and Evaluation Conference LREC-2016*, 1171–1176.
- Loukachevitch, Natalia, and Yuliya Rubtsova. 2016. SentiRuEval-2016: Overcoming Time Gap and Data Sparsity in Tweet Sentiment Analysis. *Proceedings of the Annual International Conference Dialogue-2016*, 416–427.
- Loukachevitch, Natalia, and Nicolay Rusnachenko. 2018. Extracting Sentiment Attitudes from Analytical Texts. *Proceedings of Computational Linguistics and Intellectual Technologies, Papers from the Annual Conference Dialog-2018*, 459–468.
- Loukachevitch, Natalia, Pavel Blinov, Evgeny Kotelnikov, Yuliya Rubtsova, Vladimir Ivanov, and Elena Tutubalina. 2015. SentiRuEval: Testing Object-Oriented Sentiment Analysis Systems in Russian. *Proceedings of International Conference of Computational Linguistics and Intellectual Technologies Dialog-2015*, vol. 2, 2–13.
- McKeown, Kathy, Hal Daume, Snigdha Chaturvedi, John Paparrizos, Kapil Thadani, Pablo Barrio, and Luis Gravano. 2016. Predicting the Impact of Scientific Concepts Using Full Text Features. *Journal of the Association for Information Science and Technology 67* (11): 2684–2696.
- Mohammad, Saif, and Peter D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence 29* (3): 436–465.
- Mohammad, Saif, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NrcCanada: Building the State-of-the-Art in Sentiment Analysis of Tweets. *Proceedings of Second Joint Conference on Lexical and Computational Semantics (\*SEM)*, vol. 2, 321–327.

- Nassirtoussi, Arman K., Saeed Aghabozorgi, Teh YingWah, and David Ngo. 2015. Text Mining of News-Headlines for FOREX Market Prediction: A Multi-layer Dimension Reduction Algorithm with Semantics and Sentiment. *Expert Systems with Applications* 42 (1): 306–324.
- Nozza, Debora, Elisabetta Fersini, and Enza Messina. 2017. A Multi-view Sentiment Corpus. *Proceedings of EACL-2017*, 273–280.
- Pak, Alexander, and Patrick Paroubek. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *Proceedings of Language Resources and Evaluation Conference LREC-2010*, 1320–1326.
- Pang, Bo, and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135.
- Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, 79–86.
- Pontiki, Maria, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Núria Bel, Salud María Jiménez-Zafra and Gülşen Eryiğit. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. *Proceedings of the 10th International workshop on Semantic Evaluation, Semeval-2016*, 19–30.
- Popescu, Ana-Maria, and Orena Etzioni. 2007. Extracting Product Features and Opinions from Reviews. In *Natural Language Processing and Text Mining*, 9–28. London: Springer.
- ProductSentiRus. 2012. Russian Sentiment Lexicon for Product and Services. <http://www.labinform.ru/pub/productsentirus/productsentirus.txt>.
- Reyes, Antonio, Paolo Rosso, and Tony Veale. 2013. A Multidimensional Approach for Detecting Irony in Twitter. *Language Resources and Evaluation* 47: 1–30.
- Rosenthal, Sara, Noara Farra, and Preslav Nakov. 2017. SemEval-2017 Task 4: Sentiment Analysis in Twitter. *Proceedings of the 11th International Workshop on Semantic Evaluation SemEval-2017*, 502–518.
- RuSentiLex. 2017. Russian Sentiment Lexicon, Version of 2017. [http://www.labinform.ru/pub/rusentilex/rusentilex\\_2017.txt](http://www.labinform.ru/pub/rusentilex/rusentilex_2017.txt).
- RuThes. 2016. Thesaurus of Russian Language, Version of 2016. [http://www.labinform.ru/pub/ruthes/index\\_eng.htm](http://www.labinform.ru/pub/ruthes/index_eng.htm).
- Saurí, Roser, and James Pustejovsky. 2012. Are you sure that this happened? Assessing the Factuality Degree of Events in Text. *Computational Linguistics* 38 (2): 261–299.
- SentiRuEval-2016 data. 2016. Training and Test Collections for Tweet Classification in Russian. <https://goo.gl/GhX3vU>.
- Severyn, Aliaksei, and Alessandro Moschitti. 2015. On the Automatic Learning of Sentiment Lexicons. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics*, 1397–1402.
- Shearlaw, Maeve. 2014. Understanding Russia’s Obsession with Mayonnaise. *The Guardian*. <https://www.theguardian.com/world/2014/nov/21/-sp-understanding-russias-obsession-with-mayonnaise>.
- Sulis, Emilio, Delia Farías, Paolo Rosso, Viviana Patti, and Giancarlo Ruffo. 2016. Figurative Messages and Affect in Twitter: Differences between #irony, #sarcasm and #not. *Knowledge-Based Systems* 108: 132–143.

- Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based Methods for Sentiment Analysis. *Computational Linguistics* 37 (2): 267–307.
- Tutubalina, Elena. 2015. Target-based Topic Model for Problem Phrase Extraction. In *European Conference on Information Retrieval*, 271–277. Cham: Springer.
- Vepsäläinen, Tapio, Hongxiu Li, and Reima Suomi. 2017. Facebook likes and Public Opinion: Predicting the 2015 Finnish Parliamentary Elections. *Government Information Quarterly* 34 (3): 524–532.
- Vilares, David, Thelwall Mike, and Miguel Alonso. 2015. The Megaphone of the People? Spanish SentiStrength for Real-Time Analysis of Political Tweets. *Journal of Information Science* 41 (6): 799–813.
- Volkova, Svitlana, and Eric Bell. 2016. Account Deletion Prediction on RuNet: A Case Study of Suspicious Twitter Accounts Active During the Russian-Ukrainian Crisis. *Proceedings of NAACL-HLT*, 1–6.
- Volkova, Svitlana, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring User Political Preferences from Streaming Communications. *Proceedings of ACL-2014*, vol. 1, 186–196.
- Volkova, Svitlana, Yoram Bachrach, Michael Armstrong, and Vijay Sharma. 2015. Inferring Latent User Properties from Texts Published in Social Media. *Proceedings of AAAI-2015*, 4296–4297.
- Whalley, Zita. 2018. Why Russians are Obsessed with Mayonnaise? <https://theculture-trip.com/europe/russia/articles/why-russians-are-obsessed-with-mayonnaise/>.
- Wiegand, Michael, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andres Montoyo. 2010. A Survey on the Role of Negation in Sentiment Analysis. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, 60–68. Association for Computational Linguistics.
- Wilson, Theresa, and Dan Sperber. 2007. On Verbal Irony. *Irony in Language and Thought*: 35–56.
- Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 347–354.
- Zefirova, Tatyana, and Natalia Loukachevitch. 2019. Irony and Sarcasm Expression in Twitter. *EPiC Series in Language and Linguistics. Proceedings of Third Workshop “Computational Linguistics and Language Science”*, vol. 4, 45–49.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Social Network Analysis in Russian Literary Studies

*Frank Fischer and Daniil Skorinkin*

## 29.1 INTRODUCTION

Network analysis has come to be an essential method in the Digital Humanities. A network can be described, in brief, as “a collection of points joined together in pairs by lines.” Terminologically, “a point is referred to as a node or vertex and a line is referred to as an edge” (Newman 2018, 1). If you can meaningfully describe a dataset with such nodes and edges, it is network data you deal with. Nodes can be entities like airports, cities, or devices connected to the Internet, linked to each other (or not) via edges. In the case of social networks, nodes represent people or, more generally, social entities, which easily extends to fictional characters. The edges between them describe their relations. While these relations can be of many types, literary network analysis at this stage is usually looking into communicative relations: Who is talking to whom and to what extent? This formal approach usually neglects the content of these interactions but can reveal larger structural patterns that would otherwise stay invisible as we will see in the use cases presented below. Network analysis is meant to complement other quantitative and qualitative approaches when it comes to interpreting literary texts.

Once we established a set of network data, the broad range of algorithms and methods developed within network theory becomes available to make the material “speak” in different ways. The visualization of network data often

---

The original version of this chapter was revised. A correction to this chapter can be found at [https://doi.org/10.1007/978-3-030-42855-6\\_33](https://doi.org/10.1007/978-3-030-42855-6_33)

---

F. Fischer (✉) • D. Skorinkin  
Higher School of Economics (HSE University), Moscow, Russia  
e-mail: [ffischer@hse.ru](mailto:ffischer@hse.ru)



comes first but is usually only the starting point of a more precise analysis, because the underlying data can be interpreted more meaningfully by help of literally hundreds of different algorithms. The nature of questions around network data can roughly be divided into graph- and node-related questions. The former allow for an analysis of the structural evolution of texts, while the latter allow for new ways of categorizing character types.

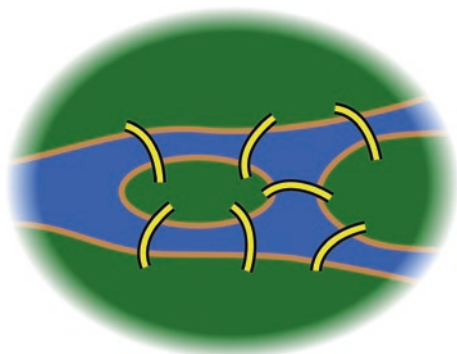
This chapter is structured as follows. A short look into the origins of (social) network analysis in general and literary network analysis in particular will be followed by a methodology section which will explain how to extract and formalize network data before introducing basic graph- and node-related measures. We then present exemplary use cases for literary network analysis, for both drama and novels.

The data for the subsection on drama originates from the Russian Drama Corpus (RusDraCor, see <https://dracor.org/rus>), a Text Encoding Initiative (TEI)-encoded collection of Russian drama from 1747 to the 1940s (Skorinkin et al. 2018). In the words of the Text Encoding Initiative, TEI is “a standard for the representation of texts in digital form” (<http://tei-c.org>). It is usually “expressed using a very widely-used formal encoding language called XML” (Burnard 2014). The data for the subsection on the novel consists in an annotated version of Tolstoy’s *War and Peace* (for more on other corpora, see Chap. 17).

## 29.2 THE ORIGINS OF SOCIAL NETWORK ANALYSIS

When talking about methods and tools, it is always insightful to look at their historicity, that is, the circumstances which led to their invention. In the case of graph theory, we have to go back to the year 1736 and Swiss mathematician Leonhard Euler. He was confronted with the seven bridges of the back then Prussian city of Königsberg and a question: Is it possible to cross all seven bridges reaching across river Pregel one after another without crossing a bridge twice? By finding an abstraction of the problem, Euler was able to proof that this, in fact, is impossible. He understood the four involved landmasses as nodes and the bridges as edges (see Fig. 29.1). The number of bridges and their endpoints were key for the solution of the problem: all four landmasses

**Fig. 29.1** The seven bridges of Königsberg.  
Wikimedia Commons,  
[https://commons.wikimedia.org/wiki/File:7\\_bridges.svg](https://commons.wikimedia.org/wiki/File:7_bridges.svg), licence:  
CC BY-SA 3.0



are reached by an odd number of bridges, but for it to work there should be a maximum of two landmasses (nodes) with an odd number of bridges (edges); these two landmasses could then serve as starting and end point, whereas the other two would have to feature an even number of bridges leading to them.

From this historical anecdote, we only take with us the idea of abstracting interconnected entities as graphs and jump two centuries ahead on the timeline, to April 3, 1933. On that very day, an article appeared in *The New York Times* reporting about a new method called “psychological geography” (later renamed to “sociometry”), which was developed by psychosociologist Jacob Levy Moreno. This method promised to visualize attraction and repulsion between individuals within communities showing “the strange human currents that flow in all directions from each individual in the group toward other individuals” (McCulloh et al. 2013). Moreno was one of the first to use network visualizations to describe social phenomena.

Another jump on the timeline and we are in the 1960s at Harvard, where scholars such as Harrison White achieved the so-called “Harvard Breakthrough,” which through methodological innovations “firmly established social network analysis as a method of structural analysis” (Scott 2000, 33). Looking at these developments, Linton Freeman lists “four defining properties” of social network analysis:

1. It involves the intuition that links among social actors are important.
2. It is based on the collection and analysis of data that record social relations that link actors.
3. It draws heavily on graphic imagery to reveal and display the patterning of those links.
4. It develops mathematical and computational models to describe and explain those patterns (Freeman 2011, 26).

We will find all these properties in literary network analysis, too. So, when did the literary studies start to become interested in network analysis? At first, this was not driven by inherent research questions, but by the mere fact that literature is an entertaining use case for social network analysis. Computer scientist Donald Knuth, author of *The Art of Computer Programming* and creator of the TeX typesetting system, needed example data for the Stanford GraphBase, a program and dataset collection for the generation and manipulation of graphs and networks (Knuth 1993). The list of datasets featured character interactions in the chapters of *Anna Karenina*, *David Copperfield*, and *Les Misérables* (<https://people.sc.fsu.edu/~jburkardt/datasets/sgb/sgb.html>)

The files for these three novels contain data on the co-occurrence of literary characters per chapter, which makes for genuine network data. Interestingly, anyone who has ever opened an example file in the number-one network analysis tool in the Humanities, Gephi, will have seen the very network graph of *Les Misérables*, because it is very prominently provided as a Gephi example file (Bastian et al. 2009).

After some more individual approaches to the network analysis of novels (Schweizer and Schnegg 1998 on the post-1989 novel *Simple Stories* by East-German author Ingo Schulze), the network analysis of dramatic texts started out with Shakespeare (Stiller et al. 2003; Stiller and Hudson 2005). Yet these first incentives did not come from literary scholars, and it took some more years until that eventually happened with the studies of Franco Moretti in 2011 and Peer Trilcke in 2013.

These two papers were the starting signal for a broad application of the network paradigm in digital literary studies, leading to several dozen papers in this field within the following five years. The main focus was on dramatic texts, as under normal circumstances they are easier to segment than novels, given their clear division into acts, scenes, and speech acts. While earlier works revolved around the network analysis of just a few individual texts, now hundreds or thousands of texts were examined, following the “Distant Reading” paradigm, which sets out to complement the close reading of texts. In the practice of Distant Reading, digital methods are used to analyze a number of texts that can be orders of magnitude larger than what an individual can read.

One result of this development was the “Distant-Reading Showcase” (Fischer et al. 2016), which put 465 German-language plays on a poster in chronological order, visually illustrating the structural transformation of German drama between 1730 and 1930. Using the same method, we can plot the extracted social networks of the 144 plays contained in the Russian Drama Corpus to date (Fig. 29.2). This unusual view from the digital stratosphere can reveal macrostructures: what is visible from such a distance are general shifts from simple network structures to more complex ones throughout the two centuries between Sumarokov’s tragedy “Horev” (1747) and Mayakovsky’s and Bulgakov’s plays of the 1920s and 1930s.

## 29.3 METHODOLOGY

### 29.3.1 *Formalizing Literary Network Data: The “Digital Spectator”*

In order to extract network data from fictional texts, we have to define a consistent way to formalize character interaction. A relation between two characters as we define it is established if both characters are performing a speech act in a given segment of a play, usually a scene. Following this definition, if character A and character B are speaking in the same scene, they are linked to each other.

This formalization is inspired by Romanian mathematician Solomon Marcus who in his book *Mathematical Poetics* (1973) suggested a formalization of a theater play undertaken by an “unusual spectator,” one who is only capable to observe the entrances and exits of the actors and monitor their co-occurrences on stage without listening to what they say. In the digital age, it is very simple

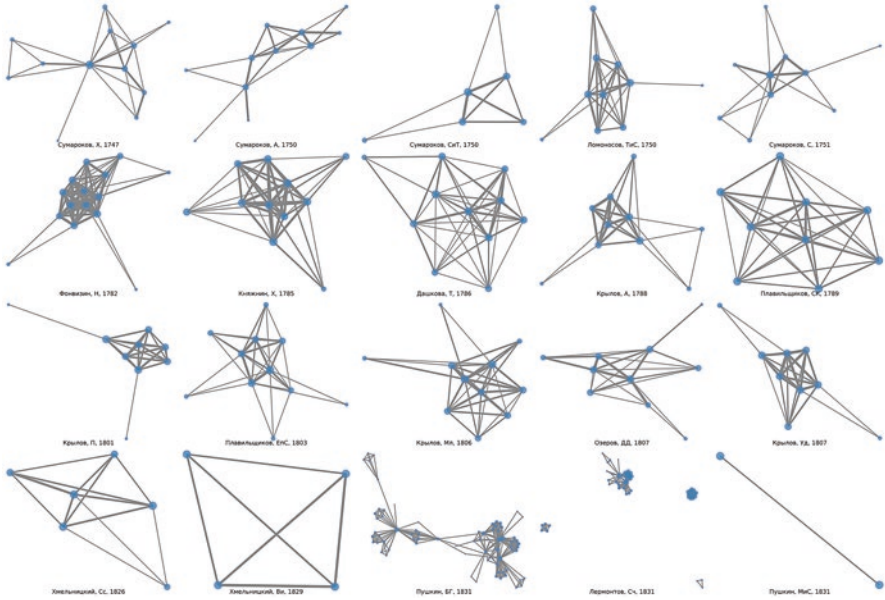


Fig. 29.2 Extracted social networks of 20 Russian plays. Excerpt (left-upper corner) from a larger poster displaying 144 plays in chronological order (1747–1940s). Version in full resolution: <https://doi.org/10.6084/m9.figshare.12058179>

Table 29.1 Number of co-occurrences of characters in A. Ostrovsky’s *Groza* (abbreviated)

Source	Target	Weight
Varvara	Katerina	12
Kabanova	Kabanov	10
Kabanov	Katerina	7
Boris	Varvara	6
Kabanov	Varvara	5
Kudrâš	Boris	5
Kabanova	Varvara	5
Kabanova	Katerina	5
...	...	...

to operationalize such formalization on a large scale, which is why we could rename the concept and call this method “the digital spectator.” Put in action, the digital spectator extracts the co-occurrences of speaking characters. Let us take Ostrovsky’s play “Groza” (“The Storm,” 1859) as an example, one of the pivotal Russian plays of the nineteenth century, which caused a scandal with its clear implication of adultery. The number of co-occurrences between characters looks as shown in Table 29.1.

This (abbreviated) table simply collects the number of co-occurrences between all characters of the play (in the “Weight” column). The table headers

“Source” and “Target” are interchangeable in our example since we are not collecting the direction of information flows.

This is already everything we need for a network analysis of “Groza.” A visualization of this simple formalization is shown in Fig. 29.3. It comprises all characters of the play (including side characters of acts 4 and 5 lacking proper names), and we can clearly see the core of the network, the Kabanov family with mother (Kabanova) and daughter (Varvara), son (Kabanov) plus wife (Katerina). Without involving one line of the actual text, we arguably found the protagonists of the play just by looking at their position in the network. It is important to note that the “epistemic thing” of our analysis is different than that of traditional textual analyses of literary texts (Trilcke and Fischer 2018). We are not analyzing the actual text of the play, but a strict formalization of it. There, it cannot hurt to stress once more that a formal approach like network analysis does not set out to replace more traditional approaches, but to complement them.

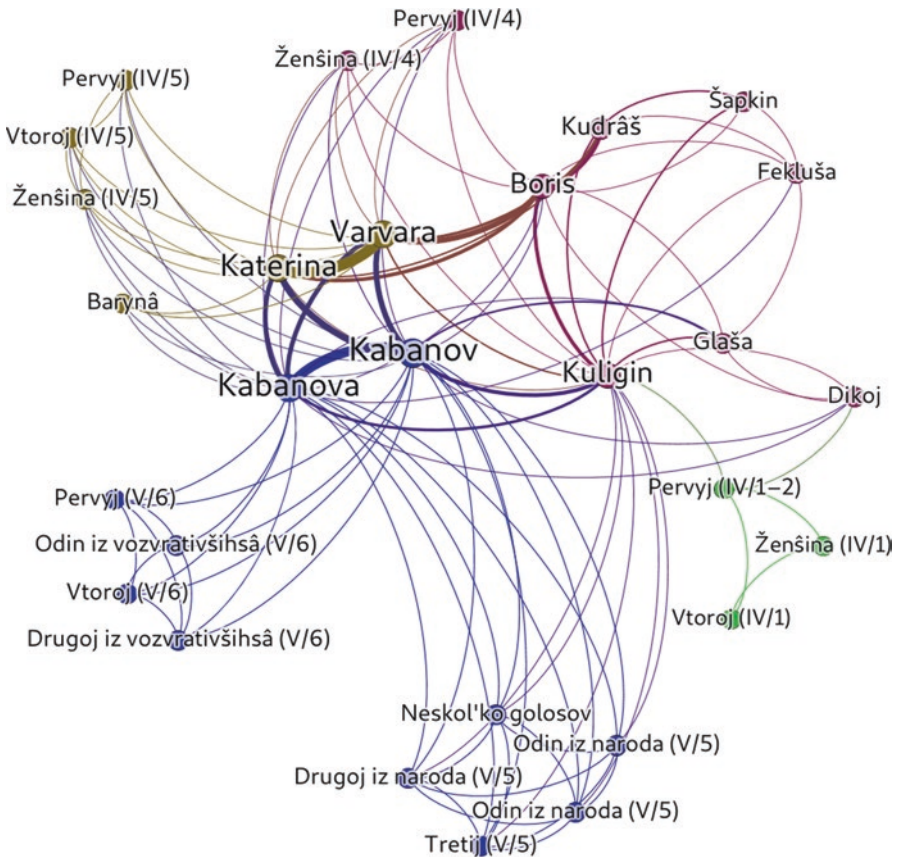


Fig. 29.3 Network graph for Ostrovsky’s *Groza*

Since the formalization step is so crucial, we developed an easy-entry tool to acquaint literary scholars with the problem and enable them to extract literary network data by hand. The tool Easy Linavis (ezlinavis)—an abbreviation for “Easy Literary Network Visualisation”—is available at <https://ezlinavis.dracor.org/>. The network data is generated live while entering speaking entities scene by scene:

```
# Act I
## Scene I
Kuligin
Kudrăș
Șapkin
## Scene II
Dikoj
Boris
## Scene III
Kuligin
Boris
Kudrăș
Șapkin
Feklușa
...
```

As its output, ezlinavis generates a CSV file which can be downloaded and opened in a network analysis tool like the aforementioned Gephi.

Our take on formalizing character interaction has some advantages (it can be easily automatized and, thus, scaled up), but also some shortcomings. It is important to not forget the rationale behind a formalization and to be consistent after a formalization method has been established. Following our operationalization, characters that do not speak are invisible to our “digital spectator.” For example, the blind old man playing the violin in the first scene of Pushkin’s “Mozart and Salieri” (1831) does not raise his voice, so he doesn’t appear in our formalization (in an admittedly not very interesting network with only two characters, i.e., Mozart and Salieri). While we might lose some information and dimensions of the literary piece, we accept this limitation in order to gain something, namely scale. By being able to automatize the extraction of character relations, we can look at a larger number of texts and distill patterns that would otherwise remain invisible.

Since we already introduced Gephi as one of the most popular tools for analysis, we should take the opportunity to mention alternative software. Other Graphical User Interface (GUI)-driven programs like Pajek, Cytoscape, and NodeXL are complemented by the two programming libraries NetworkX and igraph, which are usually used from within higher programming languages such as Python or R. These libraries have in common that most of the established network algorithms are already implemented and well documented so that they can directly be put to use.



### 29.3.2 *Graph-Related Measures*

From the abundance of graph-related measures that can be used to describe the properties of a network, we want to introduce six basic ones:

- *Network size*: The number of nodes of a network; in our case, the number of (speaking) characters in a play.
- *Network diameter*: The highest value among all shortest distances between two nodes. For example, the shortest distance between two directly connected nodes is 1. If node A is connected to nodes B and C, but B and C are not directly connected, then the shortest distance between them (through node A) is 2, and so forth.
- *Network density*: A value between 0 and 1 indicating the ratio between all realized to all possible connections. In average, comedies are denser than tragedies (one reason for this is that, at the end of comedies, the majority of the cast often appears on stage to witness the resolution of the comic conflict, thereby establishing relationships between characters that are reflected in a higher network density).
- *Clustering coefficient*: Another value between 0 and 1, measuring the number of transitive relations: if node A is related to node B and B is related to node C, then A is also related to C. The value is determined by the number of such closed triplets over the total number of triplets.
- *Average path length*: For each pair of nodes in a connected network, there is a shortest path length. The average path length is thus the average of all shortest path lengths.
- *Maximum degree*: The degree is the number of relations of a node to other nodes. The maximum degree shows the character with the most relations (i.e., the plurality of interactions), thus playing a central role in the play.

### 29.3.3 *Node-Related Measures*

Graph-related measures are complemented by node-related ones, which allow us to zoom in on single networks and talk about individual nodes. There are literally hundreds of node-related measures, among which are these three basic ones:

- *Degree*: Like stated above, the degree is the number of relations of a node to other nodes.
- *PageRank*: A recursive algorithm, different from degree insofar that it counts not only the number of relations to other nodes, but also depends on the PageRank of these other nodes. According to PageRank, the importance of a node depends on the importance of other nodes linking to it.



- *Betweenness centrality*: A measure of centrality in a graph based on shortest paths. The betweenness centrality of a node does not value the mere number of direct connections to other nodes, but the number of shortest paths between other nodes leading through that node.

Now that we have introduced some basic terminology and measures, let us look at the network properties of some literary works.

## 29.4 USE CASES

### 29.4.1 Drama

Graph-related values for five selected plays from our Russian Drama Corpus look as shown in Table 29.2.

Just by looking at the network metrics, it becomes apparent how much the two plays by Sumarokov and Pushkin differ structurally, although they are basically revolving around the same storyline (the story of False Dimitrij during the Time of Troubles around 1600). A diameter of 6 and a network size of 79 shows how Pushkin stretches the plot in a very Shakespearean manner. This strong influence is confirmed by a letter that Pushkin wrote (in French) to his friend Raevsky, dating from July, 1825, around the time he finished “Boris Godunov” (spelling follows the original):

mais quel homme que ce Schakespeare! je n’en reviens pas. ... Voyez Schakespeare. Lisez Schakespeare (now what a man is this Shakespeare! I can’t believe it ... Look at Shakespeare. Read Shakespeare). Pushkin (1962, 178)

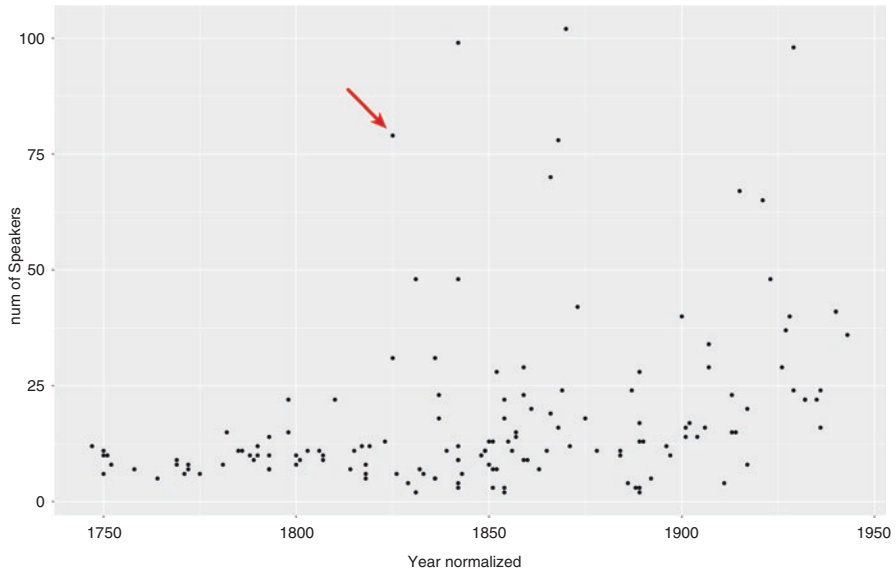
**Table 29.2** Graph-related values for five selected Russian plays

<i>Play</i>	<i>Network size</i>	<i>Network diameter</i>	<i>Network density</i>	<i>Clustering coefficient</i>	<i>Average path length</i>	<i>Maximum degree</i>
Sumarokov: <i>Dimitrij Samozvanec</i> ( <i>Dimitrij the Impostor</i> , 1771)	6	2	0.73	0.77	1.27	5 (Dimitrij)
Pushkin: <i>Boris Godunov</i> (1825)	79	6	0.11	0.89	3.03	29 (Boris)
Griboedov: <i>Gore ot uma</i> ( <i>Woe from Wit</i> , 1825)	31	3	0.44	0.8	1.57	23 (Čackij)
Gogol’: <i>Revizor</i> ( <i>The Government Inspector</i> , 1836)	31	3	0.49	0.82	1.52	26 (Gorodnichij)
Ostrovsky: <i>Groza</i> ( <i>The Storm</i> , 1859)	29	4	0.28	0.83	1.93	23 (Kabanova)

The revolutionary change that Pushkin brought to Russian drama can be shown when put into context. Figure 29.4 shows the network sizes of 144 Russian plays in chronological order. Until 1825, the network size of plays stays well below 25, but with Pushkin’s “Boris Godunov,” the network size suddenly explodes: 79 speaking entities are counted, and the diagram also shows that after Pushkin there is a broader variety of different network sizes, a changed landscape of how character networks are crafted in Russian drama after Pushkin’s initial ignition.

Without trying to overinterpret these very basic metrics, it is interesting to note that Pushkin’s play exhibits the lowest density of all plays present in the table above, but at the same time shows the highest clustering coefficient. A comparatively high clustering coefficient in a larger network with several distinguishable communities means that the individual nodes of these communities are tightly connected among each other, a property known from real-world networks, which also applies to “Boris Godunov” (cf. Fig. 29.5 below). Such real-life social networks have been called “small worlds,” building on the idea that every citizen of the world knows every other citizen over only six edges.

After looking at entire networks, let us throw a glance at node-related values and how we can use them to study literary characters. Distance and centrality measures can be used to describe and interpret the position of a node in the network. It has been suggested to use the average distance as an indicator for



**Fig. 29.4** Network sizes of 144 Russian plays in chronological order, x-axis: (normalized) year of publication, y-axis: number of speaking entities per play. Arrow indicates Pushkin’s “Boris Godunov.” Russian Drama Corpus (<https://dracor.org/rus>)



connects the two larger clusters, Gavriła Puškin. While his degree is quite low, he occupies a strategically important position. He, in fact, acts as a messenger and mediator. He is sent from Poland to Moscow to convey to Boris the terms of False Dimitrij and later convinces Boris's military chief Basmanov to change sides, which eventually helps Dimitrij win the throne. Gavriła Puškin, as a follower of Dimitrij, also announces the decrees of the new tsar to the People ("Narod"), thus becoming the only character connecting all larger clusters of the network.

A visual interpretation of this play may be fruitful already, but pinning interpretations on actual numbers adds more precision, so let us come back to the node-related metrics. We chose five characters of the play and listed some network-analytical values in the table below, contrasted by the number of words uttered by these characters (Table 29.3).

A network-based interpretation would first ascertain that Boris has connections to more characters than his opponent Grigorij. At second glance, his position is weaker, since Grigorij is connected to more nodes completely dependent on him, strengthening his position for the eventual usurpation. And last but not least, Gavriła Puškin. Like seen above, he does not excel in the mere number of connections, but he is the bottleneck through which the important information has to pass, yielding in a very high value for betweenness centrality. We can assume that the crucial role of a side character like Gavriła Puškin is not accidental. The idea that Pushkin's noble ancestors played an active part in Russian history can be pursued up to poems like "Moã rodoslovnaã" (1830).

The fact that some of the above metrics contradict each other again strengthens the importance of a multidimensional approach when it comes to the quantitative analysis of characters and character types in literary texts.

### 29.4.2 Novels

The social network analysis of novels has developed a tad slower. Unlike in the case of drama, there are usually no speaker names in front of a speech act, which is why the automated extraction of communication networks is far more

**Table 29.3** Selected network metrics for five characters in A. Pushkin's *Boris Godunov*

<i>Character</i>	<i>Number of spoken words (without stage directions)</i>	<i>Degree</i>	<i>PageRank</i>	<i>Betweenness centrality</i>
Boris	1660	29	0.038	405
Grigorij	1967	26	0.044	1501
Basmanov	303	15	0.020	629
Šujskij	770	13	0.021	197
Gavriła Puškin	424	12	0.018	1482

complicated here. The simpler approaches rely on named-entity recognition to extract character names before choosing a text window to relate characters to each other. This can happen on sentence, paragraph or chapter level and yields very different results, depending on the method chosen. Since characters are often mentioned indirectly via pronouns or other referring expressions, a lot of work has to go into coreference resolution. But despite the more difficult task, the network analysis of novels has yielded first promising results (Grayson et al. 2016; Jannidis 2017).

We made our own little foray into the network analysis of novels—with Leo Tolstoy’s *War and Peace*. With help of named-entity recognition tools and a pinch of manual markup, we identified character mentions throughout the novel. Though by no means comprehensive, our markup contains 25,600 unambiguously identified appearances of individual characters across the text of *War and Peace*. We used the markup to automatically extract the social network of the novel. Each time two characters were mentioned within one sentence, they were assumed to be interacting in some way. Figure 29.6 contains the visualization of the resulting network of 119 nodes.

Let us turn to numbers and compare character centralities using the multi-dimensional approach described above. The table below ranks the most central characters according to three different centrality measures (Table 29.4).

Overall, Pierre seems to be the most central character—hardly a surprise to anyone familiar with the novel. Differences between centrality measures are also telling. Betweenness centrality obviously assigns more importance to the historical/military characters. If we examine the military subnetwork of Tolstoy’s novel, we can see that it is less dense—and more hierarchical. Political and military figures in the novel do not have as much interaction as the main nonhistorical characters of *War and Peace*, who are constantly thrown into all sorts of social groups and circumstances. But when Kutuzov or Napoleon or Aleksandr I do get involved, they mostly interact with their inferiors (marshals, generals), who in turn convey the message down the command chain. Some actual examples from the novel include the scene in which Kutuzov, Russian commander-in-chief, talks to a regimental commander (interaction), who in turn talks extensively to his subordinate battalion commander (interaction). Yet, there is no direct conversation between Kutuzov and the battalion commander. The reader hardly ever notices this fact, but the structure of the network seems to highlight this setting-dependent difference in communication patterns.

Whether Tolstoy, himself a retired artillery officer with war experience, purposefully attempted to create an opposition of “War interaction” versus “Peace interaction” in his novel, remains an open question. But the difference in the social network structure in *War and Peace* clearly correlates with the settings. To show this, we produced separate networks for each of the 15 books (parts of volumes in the canonical Russian four-volume edition) and the epilogue of *War and Peace*. Figures 29.7, 29.8, and 29.9 present three sample networks for

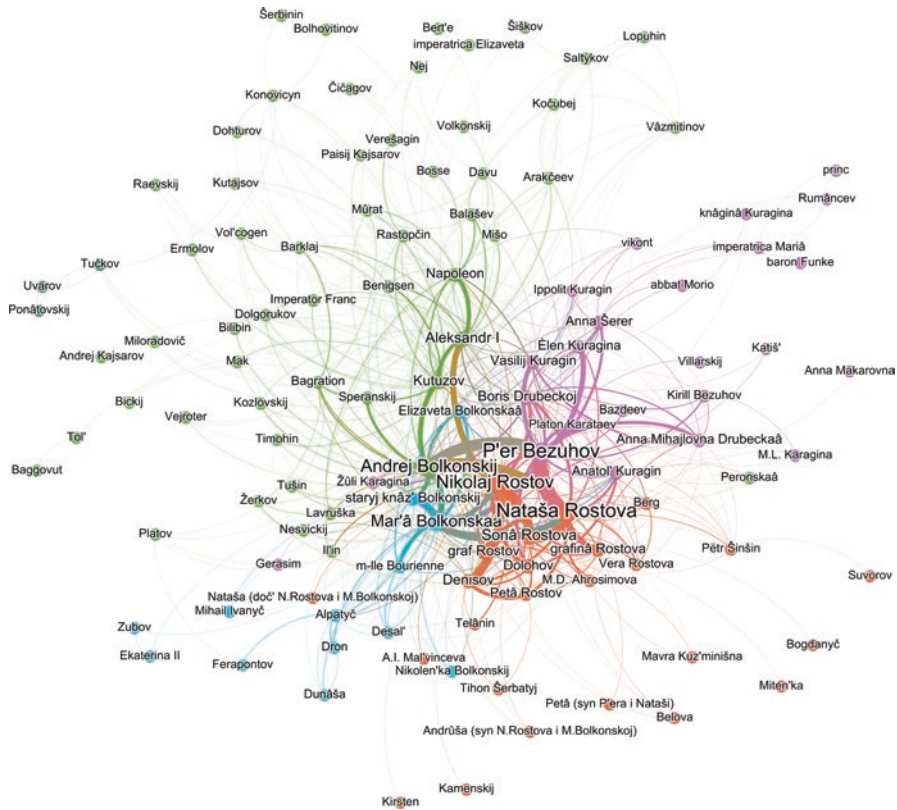


Fig. 29.6 Network visualization of L. Tolstoy’s *War and Peace*

Table 29.4 Central characters in L. Tolstoy’s *War and Peace* ranked according to three different centrality measures

<i>Degree (weighted)</i>	<i>PageRank</i>	<i>Betweenness centrality</i>
P'er Bezuhov	P'er Bezuhov	Kutuzov
Nataša Rostova	Nataša Rostova	P'er Bezuhov
Nikolaj Rostov	Nikolaj Rostov	Aleksandr I
Andrej Bolkonskij	Andrej Bolkonskij	Napoleon
Mar'ja Bolkonskaá	Mar'ja Bolkonskaá	Andrej Bolkonskij
Soná Rostova	Aleksandr I	Nikolaj Rostov
Aleksandr I	Kutuzov	graf Rostov
Kutuzov	Soná Rostova	staryj knáz' Bolkonskij
Denisov	Napoleon	Nataša Rostova
Boris Drubeckoj	Denisov	Vasilij Kuragin



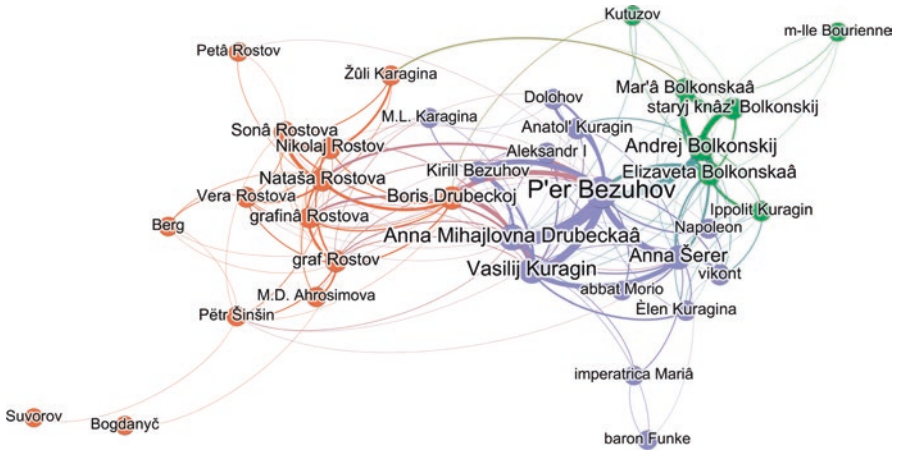


Fig. 29.7 Network visualization of L. Tolstoy's *War and Peace*, book 1 (first part of the first volume)



Fig. 29.8 Network visualization of L. Tolstoy's *War and Peace*, book 10 (second part of the third volume)



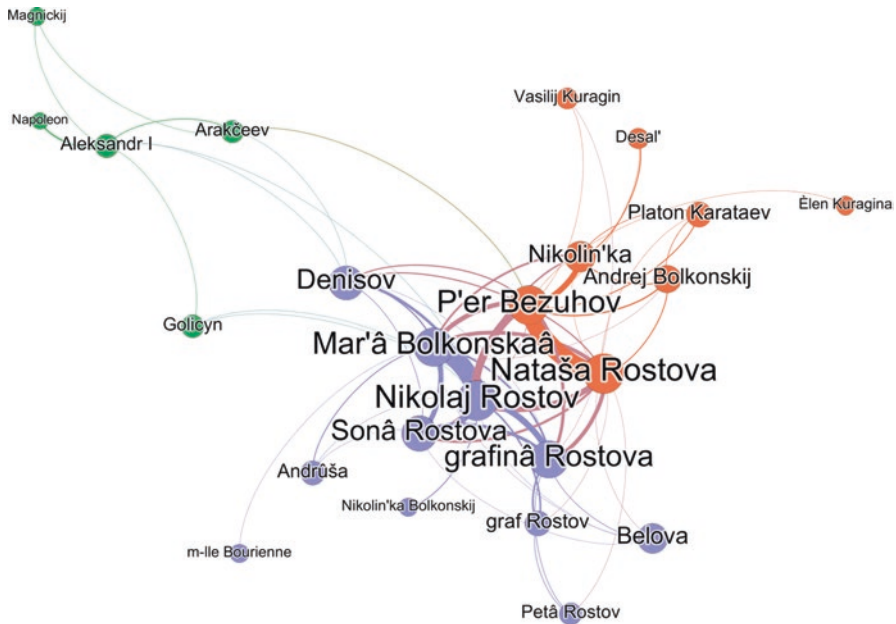


Fig. 29.9 Network visualization of L. Tolstoy's *War and Peace*, epilogue

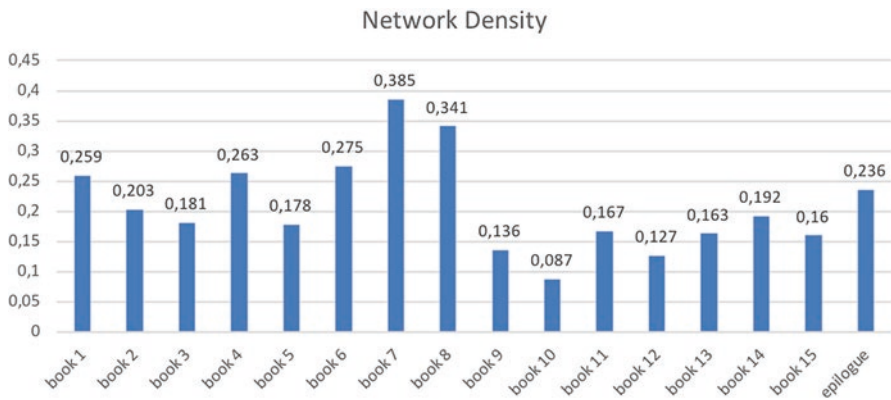


Fig. 29.10 Network densities of separate books (parts) of *War and Peace*

separate books: book 1, starting the novel; book 10, in which the Borodino battle occurs; and the epilogue that wraps up the novel.

The network in Fig. 29.8 represents Book 10 (second part of the third volume). This is one of the most battle-torn parts of *War and Peace*, as it describes the preparation and events of the Borodino battle. This network exhibits the lowest density in the whole novel—one could speculate that war and military settings disrupt human interaction.

Figure 29.10 shows the density dynamics throughout the whole novel, which can be interpreted in terms of war/peace cycles of *War and Peace*. The novel begins in book 1 with peaceful events and higher-than-average density of the character network. This is interrupted by the war of the third coalition, ending with the disastrous Austerlitz battle (books 2 and 3)—and lower-than-average density. Book 4 brings us back to the peaceful life by describing the life of the Rostov family with Nikolai Rostov on vacation from his regiment. In book 5, Nikolai, having lost a small fortune in a card game, goes back to service, the war gains momentum, Pierre breaks up with his wife completely and goes on his spiritual search—peaceful life is disrupted everywhere, and network density drops along with it. However, this time the war ends quickly in book 6 with the Treaties of Tilsit, Prince Andrej falls in love with Nataša—and the reader enters the high-density zone of peaceful life again. Book 7, the densest of all in the novel, describes the idyllic life of the Rostov family in their Otradnoe estate. The events of book 8 take place in Moscow, and this is where peace ends with Anatol's attempt to steal Nataša away. Next comes book 9—Napoleon invades Russia, not only disrupting peace, but also the social network of the novel. Then comes the Borodino battle—the watermark moment of the whole novel, and the lowest density point. The war and sorrows continue, and the density remains below the average until the very end. Only in the epilogue, which wraps up the events of the novel proper, the network density reaches the same above-average value that it had at the beginning of *War and Peace*.

## 29.5 CONCLUSION

The network analysis of literary texts has developed into a prolific subdiscipline of the Digital Humanities, a formal approach revealing hitherto invisible structures and structural changes in literary history.

The extraction and formalization of network-analytical data is the first step to gaining workable network data. It can be done manually or automatically, depending on the scale of the research question and the data available. Following data formalization, the visualization step oftentimes is a first indicator for the quality of the extracted network data. A visualization can be used for interpretation, too, but the real power of network analysis resides in the underlying numbers and available algorithms as we have demonstrated with a few examples in this chapter.

Further development will depend on whether it will be possible to establish versatile and stable infrastructures for the general digital analysis of literary texts, based on reliable text corpora and technical interfaces, like Application Programming Interfaces (APIs) or other endpoints that make it easier to access structural data. The DraCor platform (<https://dracor.org/>) is one such attempt addressing the digital research on drama (Fischer et al. 2019). By offering an interface for TEI-encoded drama corpora, it can open a comparative angle to the digital literary studies, and also help to position Russian drama within the context of other national literatures. A glance at the richness of

existing TEI-encoded drama corpora will help to understand these opportunities:

- Théâtre Classique: 1290 French plays from seventeenth and eighteenth centuries
- Shakespeare His Contemporaries: 853 English plays written between 1550 and 1700
- German Drama Corpus: 474 German-language plays from between 1730 and 1930
- Russian Drama Corpus: 144 Russian plays published between 1747 and the 1940s
- Letteratura teatrale nella Biblioteca italiana: 139 Italian plays
- Dramawebben: 68 Swedish plays
- Shakespeare Folger Library: all (37) Shakespeare plays
- Ludvig Holbergs skrifter: 36 comedies
- Biblioteca Electrónica Textual del Teatro en Español de 1868–1936 (BETTE): 25 Spanish plays
- Emothe: The Classics of Early Modern European Theatre: 113 plays including translations (Italian, English, French, Spanish)

Since all these corpora are encoded in TEI, they are comparable, although being written in different languages and stemming from different epochs. The comparative aspect is well within reach and complements similar efforts in the field of the analysis of the European novel (Schöch et al. 2018).

Beyond the added methodology for the study of literary texts, the knowledge of network metrics also sharpens the senses for the functions of other kinds of networks we are surrounded by in everyday life, be they online communities, metro lines, or highways. They are all based on the same assumptions and can be examined and understood using the same methods. The successful import of network analysis into the humanities thus leads to a broader understanding of realities beyond one's own discipline and to new opportunities for interdisciplinary cooperation.

## REFERENCES

- Algee-Hewitt, Mark. 2017. Distributed Character: Quantitative Models of the English Stage, 1550–1900. *New Literary History* 48 (4): 751–782.
- Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media*.
- Burnard, Lou. 2014. The TEI and XML. In *What is the Text Encoding Initiative? How to Add Intelligent Markup to Digital Resources*. Marseille: OpenEdition Press. <http://books.openedition.org/oep/680>.
- Fischer, Frank, Christopher Kittel, Peer Trilcke, Mathias Göbel, Andreas Vogel, Hanna-Lena Meiners, and Dario Kampkaspar. 2016. Distant-Reading Showcase. 200 Years

- of German-Language Drama at a Glance. In *Conference Proceedings of DHd 2016*, Leipzig. <https://doi.org/10.6084/m9.figshare.3101203.v2>.
- Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, and Daniil Skorinkin. 2018. To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930). In *Conference Proceedings of DH 2018*, 193–200. Ciudad de México. <https://dh2018.adho.org/?p=9362>.
- Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtel, Christopher Kittel, Carsten Milling, and Peer Trilcke. 2019. Programmable Corpora—Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor. In *Conference Proceedings of DHd 2019*, 194–197. Mainz and Frankfurt. <https://doi.org/10.5281/zenodo.2596094>.
- Freeman, Linton C. 2011. The Development of Social Network Analysis—with an Emphasis on Recent Events. In *The SAGE Handbook of Social Network Analysis*, ed. John Scott and Peter J. Carrington, 26–39. London: SAGE Publications.
- Grayson, Siobhán, Karen Wade, Gerardine Meaney, and Derek Greene. 2016. The Sense and Sensibility of Different Sliding Windows in Constructing Co-Occurrence Networks from Literature. In *2nd IFIP International Workshop on Computational History and Data-Driven Humanities*. Trinity College: Dublin.
- Jannidis, Fotis. 2017. Netzwerke. In *Digital Humanities: Eine Einführung*, ed. Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, 147–161. Stuttgart: Metzler.
- Knuth, Donald. 1993. *The Stanford Graph Base: A Platform for Combinatorial Computing*. ACM Press.
- Marcus, Solomon. 1973. *Mathematische Poetik*, 287–370. Frankfurt/M: Athenäum Verlag.
- McCulloh, Ian, Helen Armstrong, and Anthony Johnson. 2013. *Social Network Analysis with Applications*. New York, NY: Wiley.
- Moretti, Franco. 2011. *Network Theory, Plot Analysis*. Stanford Literary Lab Pamphlets 2. <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>.
- Newman, Mark. 2018. *Networks*. 2nd ed. Oxford University Press.
- Pushkin, Aleksandr. 1962. *Sobranie sočinenij v 10 tomah* [Collected Works in 10 Volumes]. Vol. 9: Pis'ma 1815–1830 [Letters 1815–1830]. Moscow 1959–1962.
- Schöch, Christof, et al. 2018. Distant Reading for European Literary History. A COST Action. In *Proceedings of DH2018*, Ciudad de México. <https://dh2018.adho.org/en/?p=11345>.
- Schweizer, Thomas, and Michael Schnegg. 1998. Die soziale Struktur der ‘Simple Stories’: Eine Netzwerkanalyse. Accessed April 30, 2019. <https://www.ethnologic.uni-hamburg.de/pdfs-de/michael-schnegg/simple-stories-publication-michael-schnegg.pdf>.
- Scott, John. 2000. *Social Network Analysis: A Handbook*. Thousand Oaks, CA: Sage Publications.
- Skorinkin, Daniil, Frank Fischer, and German Palchikov. 2018. Building a Corpus for the Quantitative Research of Russian Drama: Composition, Structure, Case Studies. In *Proceedings of the International Conference “Dialogue 2018,”* 662–682. <http://www.dialog-21.ru/media/4332/skorinkind.pdf>.
- Stiller, James, and Matthew Hudson. 2005. Weak Links and Scene Cliques Within the Small World of Shakespeare. *Journal of Cultural and Evolutionary Psychology* 3 (1): 57–73.
- Stiller, James, Daniel Nettle, and Robin I.M. Dunbar. 2003. The Small World of Shakespeare’s Plays. *Human Nature* 14 (4): 397–408.

- Trilcke, Peer. 2013. Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft. In *Empirie in der Literaturwissenschaft*, ed. Philip Ajouri, Katja Mellmann, and Christoph Rauen, 201–247. Münster: Paderborn mentis Verlag.
- Trilcke, Peer, and Frank Fischer. 2018. Literaturwissenschaft als Hackathon: Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen. In *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden*, ed. Martin Huber, and Sybille Krämer. [https://doi.org/10.17175/sb003\\_003](https://doi.org/10.17175/sb003_003).

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Tweeting Russian Politics: Studying Online Political Dynamics

*Mikhail Zherebtsov and Sergei Goussev*

## 30.1 INTRODUCTION

The popularity of social media studies in the context of Russian politics started to take off during the 2011–2012 civil protests against what are widely seen as fraudulent results of the 2011 parliamentary elections. In the absence of impartial and objective coverage of elections in the traditional media (Golos 2011) various Social Network Sites (SNS) appeared to be instrumental in circulating information among citizens, effectively serving as an alternative source of trustworthy information on the election process. The key feature of social media functionality during the 2011–2012 protest was its multi-channel and multi-hierarchical structure. Spontaneously emerging information posted online about fraud and other infringements over the course of the elections was picked up and popularized by famous bloggers and popular public channels. It helped build awareness of the magnitude of committed acts and reaffirm large public discontent regarding the validity of the election process and the pronounced winner—the pro-Kremlin *Edinaá Rossiá* (United Russia) party. Furthermore, social media were instrumental in organizing and coordinating the protest as a key means of information circulation (for more on digital activism, see Chap. 8). Earlier academic accounts were positive on the crucial role of SNS in

---

M. Zherebtsov (✉)  
Institute of European, Russian and Eurasian Studies, Carleton University,  
Ottawa, ON, Canada  
e-mail: [mikhail.zherebtsov@carleton.ca](mailto:mikhail.zherebtsov@carleton.ca)

S. Goussev  
Ottawa, ON, Canada

“mobilizing the discontent of citizens under the conditions of a semi-authoritarian political regime” (Lonkila 2012, 9).

This type of research received impetus from a chain of protests, ranging from the Occupy movement in mostly developed countries to the Arab Spring in the Middle East and North Africa region. In these cases, the internet in general and social media in particular were the key factors in the scope and magnitude of the protests. The events spurred a vigorous research into the phenomenon of social media in the context of social capital and civic engagement (Agarwal et al. 2014; Fuchs 2014), mobilization of protests (Earl et al. 2013; Greene 2013; Breuer et al. 2015), as well as methodological boundaries of political science, associated with new and growing computational methods (Tremayne 2014; Sinclair 2016; Tucker et al. 2016).

The events of the Arab Spring further reinvigorated the discourse of democratization of authoritarian states, claiming social media to be the key underlying technology promoting political change (Tufekci and Wilson 2012). These studies have fallen on the fertile soil of the Russian protest realities, determining the key theme of research. Therefore, in the context of politics, social media and digital social networks, Russian studies explored the main democratization hypothesis (Greene 2013), looking for answers as to why the Russian case did not result in a tumultuous upheaval akin to the Arab Spring (White and McAllister 2014; Reuter and Szakonyi 2013; Pallin 2017). In their answers, authors outlined several key features of social media in Russia. First of all, they agreed that up until the 2011–2012 protests the Russian digital public sphere had been developing relatively freely, without the tight oversight of the government. Secondly, recognizing the importance of Internet technologies, the authorities preferred a rather flexible model of domination over the rigid regulatory framework, which are popular in other authoritarian countries (with China being the exemplary case). On the one hand, an active and popular pro-government audience was cultivated and exhibited to the entire political spectrum, on the other, any anti-government sentiment was disrupted by various means, including the use of bots and trolls. Finally, the government undertook measures to domesticate the ownership over key social network sites and to ensure compliance of the large international ones through excessive regulation.

Following the emergence and controversy around the cyber activities of Russian government-affiliated organizations outside Russia’s territorial border, the research agenda and discourse then shifted towards a deeper study of trolls and bots (Jensen 2018; Stukal et al. 2017). Therefore, given the public interest in these specific topics, Russian social media studies have been dominated by a rather narrow research agenda, mainly referring to abnormal and critical situations. Obviously, patterns of users’ behavior would differ during these events, from their behavior under normal circumstances.

There have been earlier attempts to depict the topology of Russian digital social networks (Barash and Kelly 2012; Kelly et al. 2012) as well as the contents of key discussions (Nagornyy and Koltsova 2017; Maslinsky et al. 2013), using advanced computational methods of Social Network Analysis (SNA) and



topic modeling. Yet, apart from specific and quite narrow-focused contributions from other fields and especially computational linguistics, such studies remain on the periphery of contemporary policy research.

## 30.2 GOALS AND DATA

This chapter aims to fill this gap by analyzing the intra- and intergroup structure of politically engaged users of Russian social media and presenting the dynamics of ongoing political discussions. It builds on already conducted research and applies key SNA methods and approaches to the corpora of social media data from the Russian segment of Twitter. Therefore, the goals of the chapter are twofold: (1) to demonstrate the potential of SNA to analyze political discussions in Russian social media, and (2) to establish online political communities, determine their internal structure as well as measure their interconnectedness and detect key influencers. The chapter explores a number of hypotheses regarding the role of social media in contemporary Russian politics, which were partially inspired by the previous research in the field.

We propose and test a three-tier analytical strategy, outlining the macro-, meso-, and micro-levels of network analysis. We check whether Russian virtual society is generally divided across the same ideological lines as the public sphere, representing two scattered, yet distinct groups representing the pro-government sphere on the one hand, and, mainly, “non-systemic” opposition forces on the other (Gel'man 2015). Applying automated community-detection methods, we determine and visualize existing online communities. We further analyze their characteristics and, comparing their user structure with the contents of selected discussions, determine their ideological basis. Finally, we investigate relationships between and within communities, establish key leaders and influencers, as well as test the possibility of a dialogue between existing online political communities (for another case of network analysis, see Chap. 29).

While other social media platforms are more utilized by the wider public in Russia, such as VK (formerly VKontakte) or OK (formerly Odnoklassniki), we focus on Twitter due to three factors. Firstly, as a micro-blogging platform, it is determined by its inherently public nature, allowing the sharing and viewing of content without a restrictive permission structure, even permitting the viewing and following of all public content without a Twitter account. In part due to this, for public figures the platform has become a sort of *modus operandi*, both amongst the pro-government and “non-systemic” opposition. In this regard, Twitter, to a certain extent, has complemented LiveJournal—another very popular blogging platform as a means of reaching to a wider audience, yet with shorter announcements and “punchier” messages. Twitter remains the sixth most important social network platform in Russia with 9.9 million of unique monthly users.<sup>1</sup> In absolute numbers, Russian Twitter segment is the fifth largest in the world in terms of active user accounts.<sup>2</sup> Moreover, the Russian segment appears to be the most politicized, as a high proportion of the

top 100 most followed accounts are political figures and media accounts, compared to other countries.<sup>3</sup> Secondly, Twitter has been a platform highly utilized for political information dissemination and event coordination, including for protests, both internationally, such as during the Arab spring, and in Russia, particularly during the 2011–2012 protests (Lotan et al. 2011; Wolfsfeld et al. 2013; Spaizer et al. 2017). Thirdly, it has been argued that foreign-owned social media (Facebook and Twitter) have a greater impact on the patterns of circulation of anti-government and pro-protest information than domestically owned platforms (VK, OK), due to greater state control over the latter ones (Reuter and Szakonyi 2015). Taken together, these factors demonstrate that Twitter in Russia is an important and contested space for the politically engaged segment of the Russian population and is a relevant and valuable platform for the political analysis of the country. Therefore, given its inherently open and public nature and high nominal politicization, Twitter in Russia is regarded as a valuable object for analysis of politically active social networks and political communication in the country.

To perform the empirical assessment, we collected six samples of Twitter data on topics of international or national political importance. Given their extensive coverage in traditional media altogether with higher than average Twitter activity, each event demonstrated resonance in Russian political society (see Table 30.1 for list). Each sample was collected individually using Twitter's Search feature of the REST API, which allows the retroactive extraction of recent popular tweets containing specific keywords and returning a sample of tweets made in the preceding 7 to 9 days. The advantage of this approach is that it allows the collection of content preceding, during and following each specific informational event. To construct and evaluate user communities for each event, which are commonly understood to be based on who each user chooses to follow (Colleoni et al. 2014; Barberá et al. 2015; Halberstam and Knight 2015), we collected data on all friends/followers of users who participated in the sampled political discussions. This approach results in the collection of event data, participant users, and relationships between them. The total corpora of all six samples included 175K users and 978K tweets and retweets.

### 30.3 ASSESSMENT OF SNA METHODS

#### 30.3.1 *Macroscopic Methods: Visualizing Russian Online Political Communities*

The common starting point of network analysis is the detection of network structure and visualization of the resultant communities. Visualization allows at-a-glance assessment of the patterns present in between captured entities and the identification of which subsequent methods are relevant to investigate specific details of relationships of interest. Among various graph-visualization methods, force-directed layouts have become highly popular for practitioners in part due to the fact that they are aesthetically pleasant and intuitive (Koren

**Table 30.1** Six events and size of captured networks

<i>Event</i>	<i>Date of event</i>	<i>Timeline captured</i>	<i>Total dataset captured</i>	<i>Keywords used (translation)<sup>a</sup></i>
Two-year anniversary (2016) of the accession of Crimea (collecting pro- and anti-Russian sentiment towards the event)	Mar 18	Mar 15– Mar 25	Users (nodes): 7,772 Tweets: 12,919 Edges: 397K	#Krymnaš (#CrimeaIsOurs)
Eurovision Song Contest 2016; the victory of Ukrainian singer Džamala	May 14	May 8–May 22	Users: 118,725 Tweets: 364,734 Edges: 8.49M	Džamala (Jamala); Evrovidenie (Eurovision)
Dmitrij Medvedev’s public comment to pensioners “there is no money, but you hang in there”	May 23	May 20– May 30	Users: 17,932 Tweets: 44,324 Edges: 691K	Medvedev (Medvedev)
Release of Ukraine’s prisoner of war Nadežda Savčenko in a prisoner swap	May 25	May 21– May 30	Users: 42,625 Tweets: 200,350 Edges: 3.17M	Savčenko (Savchenko)
Liberal governor Nikita Belyh’s arrest for corruption	Jun 24	Jun 19–Jun 28	Users: 25,272 Tweets: 84,992 Edges: 1.56M	Belyh (Belykh)
Turkey’s 2016 failed military coup against president Erdogan (responses of Russian audience, following Russian-Turkish discord resulting from the shooting down Russian fighter jet in 2015)	Jul 15	Jul 16–Jul 19	Users: 44,947 Tweets: 271,391 Edges: 3.60M	Turci* (Turkey); Tureck* (Turkish); Erdogan* (Erdogan); vosstani* (rebellion); perevorot* (coup)

<sup>a</sup>Keywords and hashtags were used to collect a focused discussion sample and to minimize unrelated discussions. For keywords specified with a \*, all possible keyword inflections were utilized in the search

2003). Force-directed layouts, such as the commonly utilized ForceAtlas2, simulate a natural physical system of forces acting upon each other, with nodes repulsing each other like charged particles and edges attracting nodes like springs (Bastian et al. 2009; Jacomy et al. 2014). Applied to social networks such as the Twitter follower network, the method visually clusters

well-connected users and segregates loosely affiliated groups. It furthermore maps the network, practically visualizing the distance between users and user groups.

Modularity and community-detection methods quantify key network parameters, complementing visualization layouts. The modularity statistic measures how divided a network is into segregated groups, ranging from  $-0.5$  to  $1$ , with the upper range indicating stronger module segregation (Brandes et al. 2007). Community-detection algorithms analyze the network structure and assign nodes to communities, providing a statistical analysis that complements the visual representation of force-directed graphs. The established network parameters support the evaluation of sociological theories and research hypotheses, such as the presence of “echo chambers” in social networks. Based on the hypothesis that most engagement happens amongst likeminded and connected individuals (Bakshy et al. 2015; Colleoni et al. 2014), this phenomenon has been widely investigated in international cases (Colleoni et al. 2014; Barberá et al. 2015); however it is hitherto under researched for the Russian case.

Choosing an appropriate community-detection method depends on the network type as well as computational resources, with particularly large-sized networks presenting a challenge. For directed networks, as in this case, with edges representing follower relationships or communication patterns like retweets or mentions, the Infomap method is appropriate (Lancichinetti and Fortunato 2009).<sup>4</sup> The Infomap method (Rosvall and Bergstrom 2008) simulates a random walk along the edges of the network and categorizes communities where information can flow quickly amongst well-connected users and is unlikely to leave to another group (Rosvall et al. 2009). We apply the Infomap method to categorize user communities on the six captured political samples, calculate modularity, and visualize each using the ForceAtlas 2 (Jacomy et al. 2014) force-directed layout (see Fig. 30.1).

We observe that the political Russian Twitter space contains a highly stable community structure that parallels the real political landscape in the country, with two major political communities and a multitude of smaller ones, reacting to all political events in the country. The collected data allows us to assess various characteristics of established communities. Particularly, the basic follower method is complemented with an evaluation of network structures based on typical Twitter activities such as retweets and mentions. Both are useful as a retweet can be a symbolic representation of the consonance of opinions or importance of specific information, whereas mentions provide a wider spectrum of reactions and relationships between users. As hypothesized, there is division into two major competing political forces (Gel'man 2015), with the two major communities being (1) the pro-Kremlin (pro-government) supporters and Russian nationalists (community 0 or purple), and (2) the liberal and non-systemic opposition (community 2 or teal).

We also found that the “echo chamber” theory applies well to the Russian Twitter network, as the follower-based communities were highly polarized.

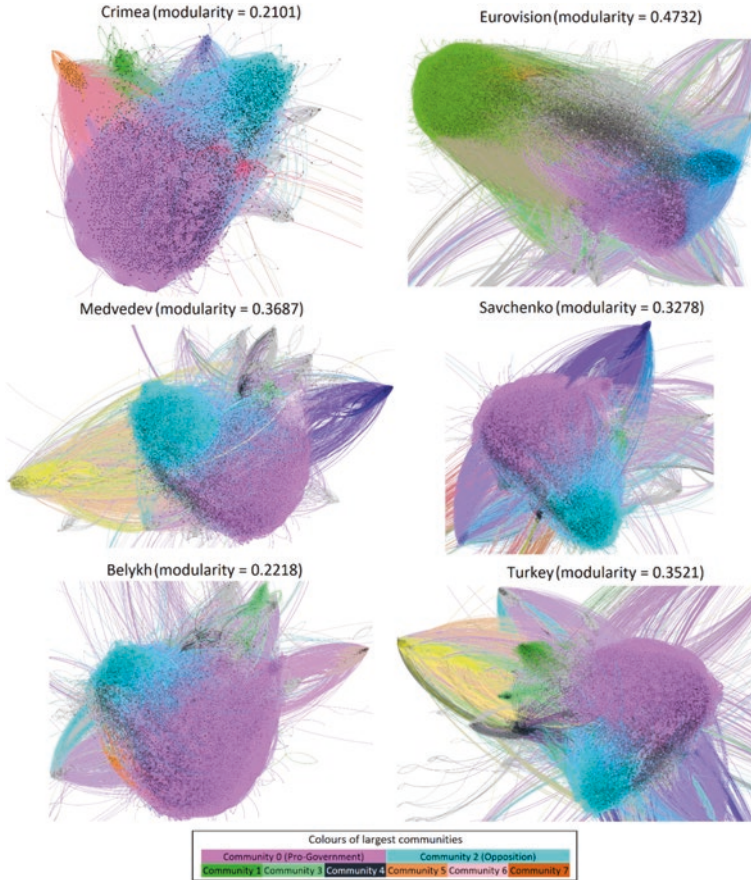


Fig. 30.1 The structure of political communities on Twitter by event

Modularity varied between the events, from a relatively low statistic of 0.2101 on the Crimea sample to a moderately strong statistic of 0.4732 for the Eurovision sample (see Fig. 30.1).<sup>5</sup> Furthermore, users in each community showed a strong preference to retweet, mention, and communicate with users in their own community and low preference to do the same for users in other communities. As Tables 30.8, 30.9, and 30.10 (Annex A) show, on average 75% of all mentions, retweets, and replies happened within, and only 25% happened between users of different groups. Specifically looking at the pro-government and Opposition communities, we see that they are very highly polarized, as they retweet on average only 5% of the content created in the rival group. Interestingly, the Opposition group is less polarized of the two, possibly due to being on average three and half times smaller than the pro-government group.

### 30.3.2 *Mesoscopic Methods and Russian Political Communities: Similar or Different?*

Upon detecting and visualizing a macro structure of the whole network, it is useful to detail each of the detected communities through the two methods of density and transitivity. Both demonstrate the compactness of each community network, showing whether a group's users are only loosely connected or highly interrelated and hence likely to be ideologically contiguous. Whereas density approaches the network holistically, measuring the proportion of connections that are present in the network against the total number of possible connections, transitivity measures the proportion of triangles (or three users connected to each other) against all possible triangles, a stronger indicator of interrelationships. Therefore, networks that have high density but low transitivity will be relatively interlinked, but not all users will know each other.<sup>6</sup> In practical terms, naturally built tight communities signify the presence of numerous multi-user interactions and the sharing of social trust and social capital within the group (Coleman 1990).

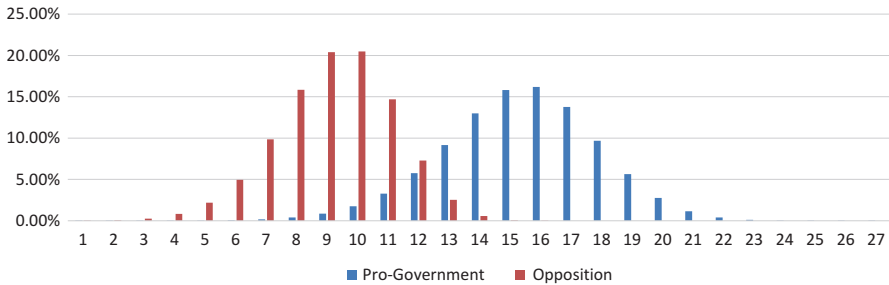
A further method is to detect cliques in a network, or a subset of nodes that are all connected to all other nodes in the clique. In social networks, cliques are sometimes referred to as clique communities, where groups of users are completely interconnected, with larger communities often containing many cliques. A benefit of clique analysis is that the prevalence and average size of cliques in a community network provides insight into the structure of the political group. A community with a large group of tri-node cliques (triads) demonstrates a relatively dispersed community, whereas the presence of several cliques with a large number of nodes in each hint at relevant sub-community structure for further analysis. Furthermore, as information is disseminated on Twitter through follower relationships, cliques represent a method of evaluating information propagation through a community, as well as a detailed analysis of the behavior of users in one versus another clique, as individuals tend to be highly influenced by the clique they belong to (Borgatti et al. 2009).

We find that the identified main political communities in Russian Twitter have vastly different characteristics and vary by event. The opposition community is a relatively dense and closely knit group, generally having stronger ties between individuals and likely sharing more meaningful interpersonal relationships. The pro-government community on the other hand is a more loosely related group of independent mini-communities, possessing more sporadic links between the sub-groups. In all six samples, the density of the opposition community exceeded that of the pro-government group (Table 30.2). Looking at transitivity, the pattern is repeated, although not as strongly and not for every sample. Clique distribution further underlies the social structure of both, as cliques in the opposition tend to be much smaller (Figs. 30.2 and 30.3). The looser amalgam of large cliques in the pro-government group also underlies the importance of public opinion leaders to reach each of these larger mini-communities.

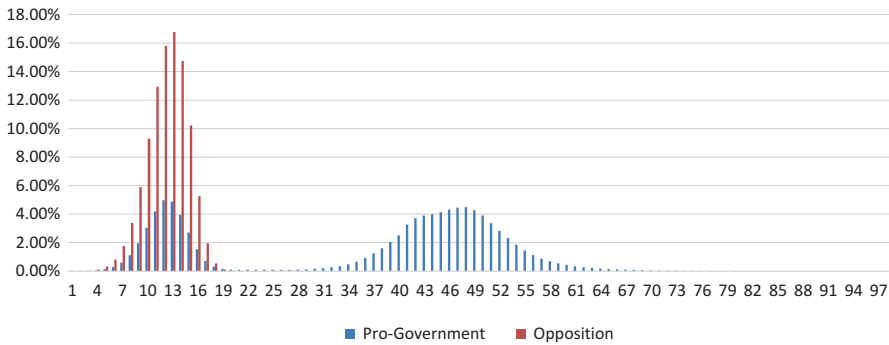
**Table 30.2** Density and transitivity of the network in its entirety and within its main communities

<i>Event</i>	<i>Density</i>			<i>Transitivity</i>		
	<i>Network</i>	<i>Pro-govt community</i>	<i>Opposition community</i>	<i>Network</i>	<i>Pro-govt community</i>	<i>Opposition community</i>
Belyh	0.41	0.92	2.64	9.85	10.73	18.12
Crimea	0.88	1.96	2.62	11.82	12.69	18.29
Eurovision	0.10	0.53	1.13	11.49	8.24	10.72
Medvedev	0.42	1.20	1.83	13.61	16.74	13.52
Savchenko	0.30	0.82	1.22	9.28	10.45	10.95
Turkey	0.23	0.70	1.48	9.20	9.49	11.80

Note: Displayed as percentage due to small scale



**Fig. 30.2** Clique size frequency distribution by community—Crimea sample



**Fig. 30.3** Clique size frequency distribution by community—Medvedev sample

An important remark concerning the utilization of meso methods that is applicable to both Russian and international contexts, has to be made, however. Both, density and transitivity could be sensitive to the quality of sample data. For instance, keyword or hashtag searches could miss statements and,



hence users, that indirectly reference the political event. This would inevitably affect the subnetwork structure. Furthermore, the rate limits and index algorithms, used by Social Media Platform APIs, could also seriously impact meso methodologies (Pfeffer et al. 2018). One way to alleviate sample issues is to use multi-sample approaches to demonstrate cross-event consistency, as done in this study. Another is sampling using only general limitations, such as language or location. While Twitter's free API does not offer location filtering, language filtering has a potential for Russian political analysis, as fewer (compared to international languages) users outside the country would engage in online political discussion.<sup>7</sup>

### 30.3.3 *Microscopic Methods: Opinion Leaders in Russian Online Political Networks*

Following the evaluation of network structure and community sub-structure, scholars often turn to the identification and measurement of the impact of network's "influencers," as well as the comparison of these influencers to offline opinion "leaders." Traditional elites, who have always had the ability to shape the political narrative, have seen their power greatly expanded with Twitter and other social media spaces (Jungherr 2014). Previous research, both internationally (Bakshy et al. 2011) and on Russia (Roesen and Zvereva 2014) has found that traditional "leaders" can be cumulatively overshadowed on social media by "ordinary influencers" (Bakshy et al. 2011, 8), or median public figures with an average "offline" influence.

#### 30.3.3.1 *Identifying and Evaluating "Influencers"*

The analysis is based on a sample of 469 accounts, which comprises public personalities and organizations as well as traditional media. These users were selected if they: (1) actively post on politically relevant events; (2) have at least ten thousand followers; and (3) either occupy positions in a government/non-government organization, or are well-known media personalities. The sampling technique adapted the "snowballing" approach but required several stages in order to improve the validity of the outcome. First, a top tier of politically relevant users was manually selected from the list of top 100 most popular accounts in the Russian segment. Secondly, from all samples collected, the 1000 most followed accounts were selected and manually sorted in order to identify politically relevant ones. These two steps together resulted in a list of 240 accounts. Among these accounts, only those that followed no more than 500 others, were selected. Subsequently, the list of friends of each was obtained, but only those who themselves had at least 10,000 followers were selected. Qualitative filtering of this list resulted in the creation of the master sample of 469 active Twitter public personalities.

The selection process inherent with this type of sampling technique can be considered as establishing a representative collection of user accounts. Given the "echo chamber" effect, it can be assumed that those who use Twitter as an

interactive platform, and not only spread but also receive information, will strategically connect with (or themselves follow) a limited amount of personalities, many themselves public figures and involved in analogous activities (politics in the case of this study). Although the 10,000-follower threshold is rather arbitrary, it allows the selection of only those accounts that have the potential to efficiently create and/or disseminate political information. Similarly, the 500-friend threshold excludes those personalities who apply a tactic of following any account that interacts with them, and whose inclusion would not improve the sample.<sup>8</sup> To support the assessment of the endurance and impact of content created by opinion leaders, the last 3200 tweets of each were downloaded using the REST API and ranked in terms of their impact on political discussions. Four types of politically relevant leader accounts can be identified in the Russian Twitter segment. The first type are personal accounts of top politicians, media, and public personalities. Many of these accounts can be regarded as official, as they are verified “*de-jure*,” while others produce content that corresponds with ideological views of their nominal owners and therefore can be regarded as “*de-facto*” genuine. The second cluster comprises of accounts of traditional media sources, which utilize the platform predominantly to reach a wider user audience. In most instances, tweets produced by these types of accounts contain links to materials issued on these media’s websites, sometimes with opinionated comments that reflect the editors’ ideological preferences. These accounts appear to be the most interconnected within as well as outside the ideologically bounded communities they belong to. The third type includes official accounts of government agencies, which were selected for analysis on the basis of multiple premises. Twitter has been actively used by private sector companies and entrepreneurs for marketing purposes. Indeed, there is a growing body of research on the subject matter, which explores and analyses strategies of efficient public relations and marketing for businesses. If used efficiently, Twitter could boost a company’s performance. The same logic is applied to political organizations (Waters and Williams 2011; Towner and Dulio 2012), who adopt advanced technologies of governance within the Government 2.0 paradigm. This approach was officially adopted in Russia in the context of the Federal Program “Information Society 2011–2020” (Zherebtsov 2019). Accounts that produce and circulate political satire and politically relevant entertainment content comprise the fourth type of accounts, which we conventionally refer to as the parody group. While they themselves are not sources of official information or representatives of certain political groups, such accounts appear at the epicenter of selected discussions and disseminate certain sentiments. Moreover, they are quite popular not only among regular users but also among top political influencers.

The analysis of content produced by the leaders reveals several remarkable trends. There is a certain consistency between the groups in terms of retweeting and liking messages. The parody group outperforms all others in the combined popularity of its messages. Needless to say, all accounts in our sample that belong to this group produce and share oppositional sentiment. Personal

accounts of political leaders comprise the second most popular group on Twitter. Interestingly, the content produced by these types of accounts is as often retweeted (or shared and thus actively endorsed) as it is liked (or passively endorsed). The Twitter activity of traditional media appears to be much lower than the first two types of accounts. To some extent, this demonstrates a quite remarkable pattern of social engagement in the Russian Twitter segment. While entertainment purposes are prevalent even in the context of political discourse (as demonstrated by the overwhelming popularity of parody accounts), users tend to get involved in political discussions and favor opinionated statements of political pundits and media personalities over factual information circulation. This being said, it was to be expected that official accounts appeared to be the least publicized in our sample; a trend best explained by the nature of content produced and shared by the accounts of this group. As official accounts tended to share links to digests and press releases, produced by the press-services of their respective agencies, this information is regarded as the least entertaining (or “infotaining”) to users.

Table 30.3 illustrates the ranking of leaders’ accounts by popularity in terms of both active (retweets) and passive (likes) endorsements, both on average for all accounts over the entire sample collected, and using a subsample of the top 10,000 most popular tweets authored by the leaders. With the former, the picture is quite consistent. The latter, however, demonstrates that for retweets, the group of official accounts ranks higher—second rather than fourth—as compared to likes, while parody accounts rank lower—fourth. A cursory evaluation of this shift, based on content analysis, revealed an unusually high activity of automated Twitter accounts (i.e. bots), indicating an evidence of selective strategy of boosting certain topics.

The types of leaders also differ from one another in terms of their capacity to influence the content and sentiment of online conversations. To perform this task, the most critical metrics of individual tweets—likes and retweets—were queried from a sample of 3200 most recent tweets, authored by the leaders. These metrics were aggregated and the average number of “likes” and retweets per leader was calculated. Used independently, it provides a good estimate of the “average power of a tweet” of the given user, although it does not consider the issue of outliers—accounts with relatively short lifespan and

**Table 30.3** Leaders’ impact metrics using all data collected or focused on the top 10k popular tweets

<i>Account type</i>	<i>All tweets</i>		<i>Sample of top 10K popular tweets</i>	
	<i>Av. Num. retweets</i>	<i>Av. Num. likes</i>	<i>Av. Num. retweets</i>	<i>Av. Num. likes</i>
Personal (302)	(2) 26.80	(2) 27.67	(3) 691.88	(2) 959.91
Media (86)	(3) 15.31	(3) 12.10	(1) 848.53	(3) 855.88
Official/Gov-t (49)	(4) 12.37	(4) 10.60	(2) 697.61	(4) 691.25
Parody (32)	(1) 98.35	(1) 150.74	(4) 664.49	(1) 1303.07

yet, quite high performance metrics. To address this, the maturity of accounts was estimated by multiplying the “tweet power” metric by the average number of tweets per day (1). Given the fact that all accounts in the leaders sample are real and used actively, the issue of automated content generation did not affect the overall calculations. Assuming that bots are less likely to be followed by leaders, the sample showed no evidence of the presence of unusually and/or suspiciously active accounts. Therefore, the average leader account generates approximately 16.03 (+/-2.3) tweets per day and the most active account, quite expectedly belonging to the media group, generates on average 170 tweets per day.

$$(1) \text{ Average tweet power of Leader}_i = \left( \overline{\text{Favorites}_i} + \overline{\text{Retweets}_i} \right) * \overline{\text{NumTweetsPerDay}_i}$$

The overall list of candidate impact obtained by (1) was sorted from most to least impactful, and the overall list of 469 was broken down into quantiles. Figure 30.4 represents the breakdown of account types per quantile. Obviously, the parody group accounts generate content ordinary citizens are eager to react to: 53.1% of such accounts in the sample appeared in the first quantile. Approximately a quarter of personal accounts demonstrate the tendency to generate highly resonant content (23.5% in the first quantile). Interestingly, this most populous group is almost evenly distributed. Official accounts follow a somewhat normal distribution, peaking in the third quantile, hence generating relatively impactful content. The discrepancy between this distribution and the high performance of these account types in the top 10,000 sub-sample (Table 30.2) raises the importance of future in-depth content analysis of messages produced by this group. On the one hand, this content could be artificially “boosted”; on the other, top “tweets” could actually discuss politically crucial issues and be genuinely shared alongside the network, which to some

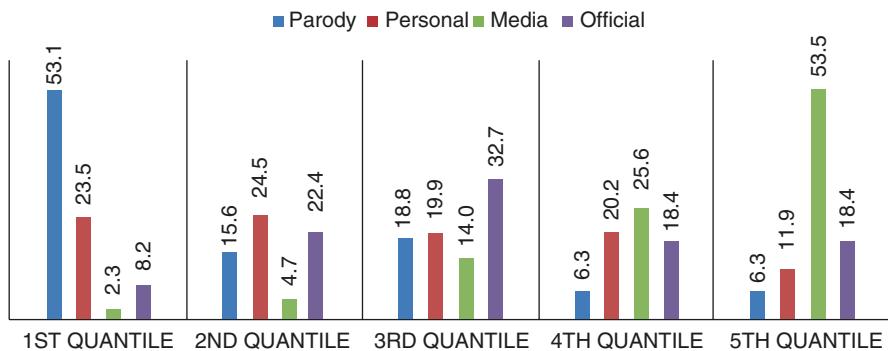


Fig. 30.4 Influence of leaders' content, distribution across quantiles

extent, supports the thesis of the bursty nature of Twitter networks (Myers and Leskovec 2014). Finally, and rather surprisingly, the most prolific group of media accounts tends to be distributed towards the lower part of the scale.

### 30.3.3.2 *Developing an Index of “Influence”*

The average tweet power metric for Twitter data, while indicative of certain patterns in ongoing and historical discussions, does not take into consideration network “influence” and the leaders’ ability to disseminate content throughout the network and in their particular communities. While the topic has gained significant attention in the research community, no established and widely accepted method of identifying Twitter influencers exists. Time-invariant approaches tend to compute influence on the basis of either centrality (network-dominated approach), or content impact (retweet-dominated approach). At the same time a combination of both methods could be quite productive. We propose a method utilizing network centrality and demonstrated ability to disseminate content.

Influence can be defined as the ability to seed discussions and spread content throughout the network. It can be seen as a derivative of two major parameters: the importance of content and its ability to meet the aspirations of ordinary users and the capacity of this content to spread through the network and be visible to a wide audience. The former is marked by users’ reaction to content, similar to the approaches taken in evaluating “influencers” above. The latter, on the other hand, evaluates the placement of the leader within a network or community, as a central placement creates a better opportunity to disseminate content amongst a wider audience. As such, we determine PageRank centrality on the “follow” relationship of Twitter, which is seen as both as an indicator of information-gathering, as well as social connection between two users, especially if it is reciprocated (Myers et al. 2014; Frederick et al. 2012).

Centrality is the most commonly used approach to determine the importance of nodes in a network (Livne et al. 2011; Romero et al. 2011). PageRank Centrality (Page et al. 1999), most famously used in Google search, assigns a probability distribution to the network, representing the chance of randomly picking a specific node. When applied to social networks, it allows the ranking of users by importance relative to each other. Centrality was combined with the aggregated average number of “likes” and “retweets” obtained from the 3200 tweets authored by each individual “leader.” Combining both parameters yields an index of identified leaders’ influence, which represents the potential to have an impact, rather than a *bona fide* substantiation of influence. Adopting the average tweet power metric (1), we multiply it by PageRank centrality of each candidate to generate “influence” index ( $\lambda$ ).

$$(2) \text{Leaders' influence}_i = \text{Average tweet power of Leader}_i * \text{Centrality}_i$$

Introducing centrality and combining the data with leaders’ assigned *InfoMap* communities alters the observed distribution considerably. Firstly, the first two quantiles of the most influential Twitter users are comprised

predominantly of opposition accounts (Fig. 30.7, Annex B). In the first quantile, two-thirds (or 60%) of accounts belong to the opposition and only one-third to the pro-government community. A similar situation is observed in the second quantile, where 71% of accounts can be referred to as belonging to the opposition, and only 29% to a pro-government group. The first quantile included such popular opposition leaders as Aleksei Navalny, Leonid Volkov, Oleg Kashin, media outlets TV Rain (*Dožd'*), *Èho, Moskvy*, and *Meduza*, as well as highly influential parody accounts. The pro-government group, although outnumbered by its opponents, is represented by its most outspoken pundits (Vladimir Solovyev and Alexei Pushkov) and notable media sources (*RIA Novosti*, *Vesti News*). Interestingly, the most followed political accounts of Vladimir Putin and Dmitry Medvedev, although appearing in the top quantile, are located in the middle and in the end of it respectively. Secondly, the distribution of account types across quantiles is more flat, with a decline in proportion of parody and personal accounts in the first quantile and an increase in media and official accounts, and a gain in parody and media accounts at the expense of personal and official accounts, in the second quantile (Fig. 30.8, Annex B). This can be understood as indication that media and official accounts, while not impactful in terms of content, are central to the network and hence have a higher ability to distribute their content. The shift of a certain proportion of parody accounts from the first to the second quantile, as well as the relative decline of personal accounts, is a further validation of the presence of the “echo chamber.” As parody and personal content is usually popular in specific audiences, these accounts are not highly followed by opposing communities and hence do not share central position in the whole network.

Furthermore, such dominance of opposition accounts in the top half of the influence index speaks of the higher importance of this form of communication for the opposition and also supports evidence of the greater structuration and network sophistication from the network analysis. The opposition not only focuses on social media as its main form of reaching the audience but also emphasizes the role of opinion leaders. In this regard, Alexei Navalny is the major actor and the greatest influencer not only within his own political community, but also in the entire network. Pro-government pundits, like Vladimir Solovyev and Alexei Pushkov outperform their own formal leaders in terms of influence in the virtual community, and accounts of traditional federal mass media are instrumental in the dissemination of the pro-government content. This establishes a new framework of evaluating Russian political Twitter, which is quite different from Kelly et al. (2012) in terms of network structure and from Greene (2018) in terms of content.

### 30.3.3.3 Cross-Validation of the Proposed “Influence” Index

The proposed index (2) requires further testing and validation. Given the nature of the research topic, where outcomes are easily predictable on the basis of traditional theories and concepts of Russian politics, the best way to test reliability of a new instrument would be the utilization of another approach. Given the fact that this new method is a derivative of major other influence indicators,

reusing them would result in unfavorable procedural overlap and, thus, similarity of outcomes. To overcome this issue, and avoid complex dynamic methods, this research adapts the principle that utilized the Hirsch index (h-index) of academic impact (3).

Hirsch is rather unexpectedly suitable and productive for measuring leaders' performance on Twitter, and even overcomes deficiencies visible in the context of scholarly work. Firstly, leaders on Twitter are akin to scholars in academia, producing content aimed at specific audiences and seek endorsement for their work in terms of citations or "likes" and "retweets." Secondly, both academic papers and blog messages increase their value through references, with the growth being well documented and easily accessible. Thirdly, academics and leading bloggers both tend to increase their visibility by producing the maximum possible high-quality content. Moreover, the ample quantity of blog posts overcomes the limitations of academic work, where the number of contributions is usually lower.

$$(3) \text{Leaders' influence with Hirsch}_i = \text{hirsch method}(\text{Favorites} + \text{Retweets}) \\ * \text{Centrality}_i$$

Therefore, the use of the h-index seems justified, as it addresses the issue of outliers (i.e. highly popular tweets) as well as the lifespan of accounts (immature, yet highly popular accounts) and provides a weighted rank of significant contributions. To put it simply, the h-index algorithm finds an "ideal point" between the number of contributions and their relative popularity, which for Twitter can be considered as the sum of "likes" and "retweets" for each users' post (*hirsch method(Favorites + Retweets)*). All leaders were ranked according to the obtained indices and the resultant list was compared with the ranked list of leaders, obtained through the index method proposed by this research. Spearman's rank correlation coefficient ( $\rho$ ) was utilized to establish whether both methods were concordant. It demonstrated a high correlation coefficient of 0.69 between the proposed influence index (2) and the modified h-index (3). Notably, this coefficient was calculated when the h-index did not refer to the centrality parameter of each leader account. Including the centrality indicator increased the correlation coefficient to 0.80.

As a ranking algorithm, the h-index provides a useful method for establishing the most influential contributors and can be used for ranking leaders. It also confirms the validity of the proposed time-invariant influence rank (2). As any other methods, the h-index for Twitter is not without deficiencies and potentially may not be used for samples where leaders are highly popular and produce a large quantity of tweets. As the Twitter REST API limits access to 3200 most recent posts, the h-index will not be able to produce an index higher than the quantity of posts. Yet in the case of current measurements of Russian Twitter, this issue was not a problem, as the most popular user—Alexei Navalny—scored only 902 points on the scale. Furthermore, both indices (2



and 3) are consistent and consonant with common wisdom that the actual disposition of actors and organizations within the political arena should be correlated with their political influence.

### 30.4 BEYOND THE SCORE: CROSS VALIDATION OF DETECTED PATTERNS

#### 30.4.1 *Further Validating “Echo Chambers”*

Focusing on intra- and cross-community conversation, we observed homophilous conversation patterns between the various communities as users tended to share content with like-minded individuals inside their own community, most especially between the pro-government and Opposition groups. Nominal homophily however can be misleading as users in small communities are more likely to converse across community lines simply because their community is small, and users in large communities are unlikely to converse outside their group due to its relative proportion. We adopt a method developed by Currarini et al. (2007) to validate the nominal homophily observed. Specifically, nominal homophily, or the proportion of conversations a community has within itself ( $H_i$ ) is compared to its relative size within the network ( $w_i$ ).

$$(4) H_i = w_i; (5) H_i > w_i; (6) H_i < w_i$$

Baseline homophily (4) occurs when the proportion of user conversations within a community equals the relative size of the community, indicating that on aggregate, users in that community show no special preference or bias for their own friends. Inbreeding homophily (5) indicates that users are biased and converse more often within their own group than is expected on the basis of its relative demographic size. Finally, if a community shows heterophilous patterns (6), the number of conversations within the group will be less than the relative size of the group.<sup>9</sup> To enable comparisons between communities of various sizes and different conversation types on Twitter, we standardize homophily indicators for (7) baseline homophily, (8) inbreeding homophily, and (9) heterophily.

$$(7) \frac{H_i}{w_i} = 1; (8) \frac{H_i}{w_i} > 1; (9) \frac{H_i}{w_i} < 1$$

Investigating standardized homophily indicators ( $\frac{H_i}{w_i}$ ), we find that each community demonstrated strong inbreeding homophily (Table 30.4). Interestingly, the non-systemic opposition is more homophilous than the pro-government community. A few communities, such as communities 3 and 4, while quite small, demonstrated excessively high standardized homophily indicators.

**Table 30.4** Relative community sizes and standardized homophily indicators

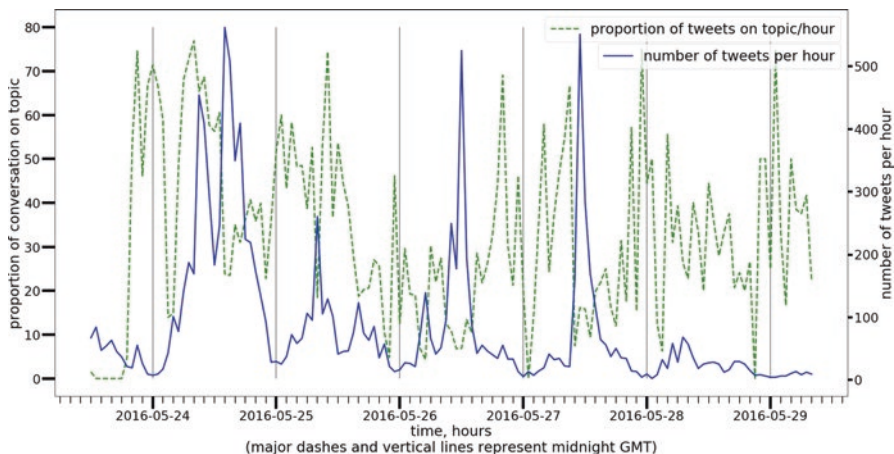
	<i>Com 0</i> [ <i>pro-government</i> ]	<i>Com 1</i>	<i>Com 2</i> [ <i>opposition</i> ]	<i>Com 3</i>	<i>Com 4</i>
Relative size	22.7%	9.3%	6.2%	0.8%	0.4%
$\frac{H_i}{w_i}$ (retweets)	3.96	8.07	12.53	95.66	175.80
$\frac{H_i}{w_i}$ (mentions)	3.56	8.81	10.74	111.16	0.00

### 30.4.2 *Makeup of Two Main Political Communities and Their Reactions to Political Events*

Individually assessing the pro-government community, we observe that it was by far the largest political group, always actively participating in all events. The community displayed strong pro-Putin, pro-government, anti-western (including anti-US [United States], anti-Europe and anti-Ukraine), and, in some instances even nationalist sentiment. While sometimes a little critical of the regime, the users in this community (community 0 or purple in Fig. 30.1 above) generally disseminated information in line with a patriotic narrative and demonstrated two patterns of Twitter use. If the informational event was not negative to Russia or the government, such as the two-year anniversary of the accession of Crimea as part of the Russian Federation, then reactions were usually event-specific and generally positive. However, if the informational event was inherently negative to the government, reaction was usually split between a certain proportion of anti-government content, and neutral or positive pro-government reaction. In certain cases, a pattern is evident whereby positive content was coordinated around specific keywords that were trending negatively in order to coopt the term and spin it positively, distracting the conversation to unrelated pro-government content.

Reactions to Prime Minister Medvedev's comment of "there is no money, but you hang in there" to Crimean pensioners, posted to YouTube on May 23, 2016, demonstrate these two patterns well (Fig. 30.5). While some users derided the Prime Minister's comments, factual and neutral reactions were quite prevalent. A large amount of disseminated content focused on unrelated positive topics to distract and mitigate the initial negative reaction inside the community. Two stories were widely mentioned on May 23 and 24 focusing on specific keywords. The first focused on the word "money" by distributing a story on the Prime Minister promising to find money for museums in Crimea. The second focused on the terms "economy" and "investment," disseminating content about the release of a government plan, approved by the Prime Minister, aiming to increase domestic demand for the products of Russian chemical and petrochemical industries. Other reactions also included the factual reporting of the Prime Minister's comments or presenting the information in a neutral fashion, with tweets such as "Medvedev admitted that there is no money to index pensions."<sup>10</sup> Interestingly, such neutral posts usually did not include links and were composed of just text.

The solid line indicates the number of tweets on an hourly basis (right axis) in the pro-government community, and the dashed line indicates the



**Fig. 30.5** Pro-government community reaction to Medvedev’s comment to pensioners in Crimea

proportion of conversation (i.e. of the tweets and retweets made during that hour) that had to do with Medvedev’s comment (left axis). Tweets containing “money,” “hang in there,” “pensioners,” “have a good day” (“*deneg*,” “*den’gi*,” “*deržites’*,” “*pensii*,” “*pensij*,” “*nastroenie*”) were used to calculate the proportion. Specifically, lemmatized words in each tweet were checked against the lemmas of desired keywords. Tweets that contained keywords known to be used in the counter strategy were excluded.

The second of the two main political groups, the Opposition community (community 2, marked teal in Fig. 30.1 above), was on average 3.5 times smaller. Its users displayed negative, ironic, and critical assessments of the Russian government and also disseminated Ukrainian-friendly, pro-US, and pro-Western content. While users in this community also shared content on liberal values, such as opposition to authoritarian government or support for democracy, a majority focused on vilifying the government, with users spreading negative memes or ridiculing government strategies or statements. Indeed, particularly virile ridicule and even contempt of the government followed Medvedev’s comments in Crimea. The community is also made up of a sizable proportion of Russian-speaking Ukrainians, which seemed to influence how the group reacts to informational events. Indeed the Savchenko affair and the Eurovision contest, both highly interrelated with Ukrainian politics, are the largest samples of captured Opposition group users, with the former being the largest by number of tweets and the latter, the largest in quantity of engaged users.

Similar to the pro-government community, content shared within the Opposition group followed a dual pattern. If the informational event was neutral or negative to the government, and hence in line with community expectation, users either discussed the topic in a neutral fashion or spread content negative to the government. However, when the informational event ran counter to community expectations, then reaction was split as users reacted in

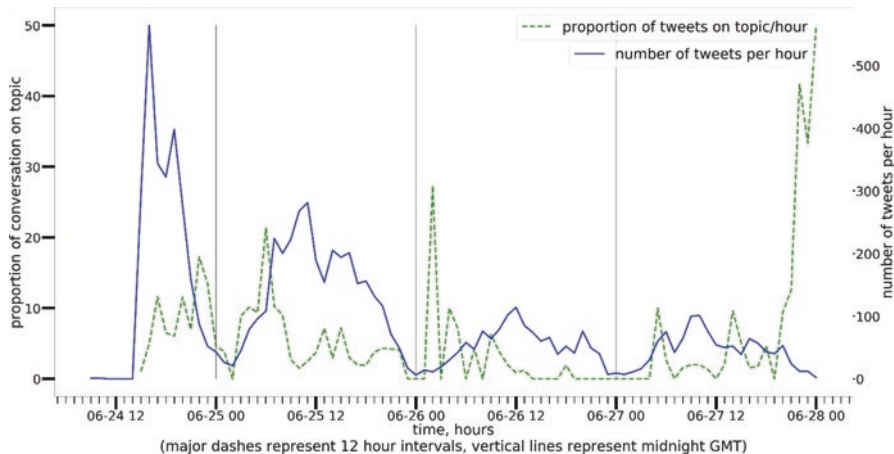


Fig. 30.6 Opposition community reaction of disbelief to Belykh’s guilt

different ways. The three-pronged reaction following the arrest of the liberal governor of the Kirov Oblast, Nikita Belykh, well demonstrates this trend. Factual and neutral tweeting was predominant; however, genuine shock and disbelief, often including statements that Belykh was set up, was also prevalent (see Fig. 30.6). Finally, a third opinion expressed by the community was that of anti-Belykh statements, believing that he betrayed the liberal movement by becoming a systemic politician.

Represents data from June 25 to June 28. The solid line indicates the number of tweets per hour (right axis) and the dashed line indicates the proportion of conversation believing Belykh was set up (left axis). Tweets containing lemmatized words including “setup,” “don’t believe” and “provocation” (“*podstava*,” “*podstavili*,” “*podstavit*,” “*podstav*,” “*ne verú*,” “*poverit*,” “*provokaciá*,” “*provocirovali*”) were used to calculate the proportion.

Comparing the standardized rate of tweets per hour between the two main communities, we see that the pro-government group reacted very differently than the Opposition group to several events, most notably during the Medvedev event (see Medvedev Chart in Annex C).<sup>11</sup> The initial spike of tweeting activity in mid-day on May 24 lagged by a few hours the initial and larger reaction by the Opposition group, indicating that the pro-government group was less likely to immediately react to the negative information. The secondary and larger spike of activity in the latter half of the day is particularly interesting, given its size and the content shared during it was mostly not about Medvedev’s comment at all. Comparing the (nominal) number of tweets every two hours with the proportion of the tweets that have to do with Medvedev’s comments, we see that the conversation shifted to discussing other topics during this second spike (see Fig. 30.5). Following this, the proportion of unrelated content on Medvedev continued to dominate discussions within the community, with the proportion mentioning the pensioner comment mostly remaining in the

20–30% range. Moreover, given that the keyword used to collect this sample was “Medvedev,” we hypothesize that this pattern of distraction specifically had to do with positively portraying the Prime Minister.

### 30.4.3 *Finding Bots Within Russian Twitter*

SNA is undeservedly neglected in the context of the mainstream topic of “bot” and “troll” impact on Russian “online” political discussions. Recent research points that the proportion of content created on Russian Twitter has a 50% probability of being produced by “bots” (Stukal et al. 2017). As stated elsewhere (Murthy et al. 2016), bots can have an impact on simple indicators such as follower counts or hashtag boosting. This may impact users who follow others indiscriminantly and network metrics as a whole, which assess all tweets in the network without taking into account the structure of social networks (Ferrara 2018). As ordinary users tend not to follow “bots” and segregate themselves into isolated “echo-chambers,” bots are likely to be segregated to isolated communities that have little influence on real politically engaged users.

To evaluate the potential impact and ensure the validity of our findings, we apply a three-method strategy to measure the prevalence of bots within the identified network structure. First, we evaluate the proportion of duplicate and highly similar content created in each community, as bots are known to repeat (not retweet) identical tweets (Lawrence 2015). As tweets can have very minor purposeful variation, applied to them by bot designers, such as adding an extra hashtag or changing the beginning of the tweet, we compare the similarity of tweets by excluding tweet extremities. Secondly, to validate the results of the first method, we qualitatively assess a sub-sample of tweets shared in suspect communities. Finally, we apply the popular *Bot or Not* method, also known as botometer, to score the likelihood of each user in our samples to be a “bot” (Davis et al. 2016), a feature-based method that evaluates a set of behaviors of a Twitter account and assigns it a score (probability) of being a “bot” (Ferrara 2018). A tried and relatively accurate approach (Yang et al. 2019), it is appropriate for cross-validating other methods utilized.

We find that in the two main political communities, the proportion of duplicate and similar tweets varied by event; however, the pro-government community demonstrated a much larger proportion of both in all events (Table 30.5). For instance, in the Medvedev sample, 43% of tweets are very similar in the pro-government community, many spreading the positive new story of government support of Russia’s petrochemical industry on May 24. The low proportion of duplicate content in the opposition community is intriguing, as is the relatively high proportion of similar content. A possible explanation could be that the opposition is a more dynamic group of users who follow more sophisticated bots that utilize more complex natural language or image methods to spread content. We propose this fascinating puzzle as a question for further research in the area.

Outside the main political communities, qualitative and duplicate/similarity analysis revealed that many visually segregated communities were made up to a very large degree of “bots,” or at least accounts sharing very similar content

**Table 30.5** Duplication and similarity of content in two main communities (% identical; % similar)

<i>Event</i>	<i>Pro-government community</i>	<i>Opposition community</i>
Belyh	12.4%; 28.2%	0.1%; 6.6%
Crimea	1.2%; 5.1%	NA; 2.2%
Eurovision	10.8%; 27.9%	0.3%; 13.3%
Medvedev	11.9%; 43.1%	1.1%; 19.5%
Savčenko	11%; 31.5%	0.2%; 24.5%
Turkey	12.7%; 33%	0.1%; 16.4%

**Table 30.6** Sample of “bot” communities detected (% identical; % similar)

<i>Event</i>	<i>Community 3 (light green)</i>	<i>Community 7 (dark orange)</i>	<i>Community 8 (dark purple)</i>	<i>Community 21 (yellow)</i>	<i>Community 43 (red)</i>
Belyh	53%; 54%	24%; 24%	NA; NA	60%; 60%	NA; 27%
Crimea	NA	NA	NA	NA; 96%	NA; NA
Eurovision	43%; 46%	83%; 83%	3%; 16%	73%; 74%	NA; 11%
Medvedev	31%; 36%	61%; 61%	3%; 19%	44%; 47%	NA; 13%
Savčenko	27%; 32%	64%; 64%	3%; 12%	71%; 72%	2%; 12%
Turkey	46%; 49%	24%; 24%	NA; NA	57%; 57%	8%; 36%

(Table 30.6). The more unsophisticated groups posted identical or very similar content, including as high as 83% of all tweets for an event (community 7, dark orange). Others showed more complex approaches, such as tweeting news headlines or factual statements, with or without a corresponding link in the tweet. Interestingly, when links were present, they often pointed to Yandex News or even more commonly to heterodox news or blogging sites. While the information captured in the samples of the six events for these communities were often political, the public timelines of the “bot” accounts often included completely unrelated content, such as for commercial purposes and advertising. We assume that these communities of “bots” are owned by marketing or consulting firms and tweet specific content based on the requirements of their clients, without any particular impact on actual political discussions.

Evaluating the average automation probability of users in each community, as reported by botometer, reinforces our findings (Table 30.7). Communities of real users had on average low probability of being automated, with a relatively small proportion of users removed or suspended. By the same token, communities, previously identified as highly likely as being “botnets” or having a large prevalence of “bots” (Communities 3, 7, 8 and others) had much higher average probability of being automated. Moreover, a large proportion of accounts in these communities have since been suspended by Twitter. Recently the company expanded its activities to diminish the impact of bots and trolls by suspending multiple accounts.<sup>12</sup> The results of these actions reinforced our findings, as the proportion of suspended accounts in the communities we identified as botnets was much higher than in real user communities. Indeed, entire

**Table 30.7** Botometer (Bot or Not) results (average universal probability of automation; proportion of accounts no longer present two years after samples were originally collected)

<i>Event</i>	<i>Pro-government (community 0)</i>	<i>Opposition (community 2)</i>	<i>Community 3</i>	<i>Community 7</i>	<i>Community 8</i>
Belykh	11.2%; 15%	5.8%; 10.1%	49.8%; 37.1%	50.4%; 4.8%	NA; 100%
Crimea	7.8%; 12.4%	7.6%; 9.7%	54.7%; 50%	NA	NA
Medvedev	15.8%; 17%	7.8%; 10.5%	53.6%; 38.1%	NA	37.8%; 97%
Savchenko	11.7%; 15.1%	8.8%; 9.8%	57.2%; 32.6%	45.6%; 3.6%	28.6%; 98.2%

“botnets” have all but been suspended by Twitter in the two years since the samples were originally collected (see Table 30.7). On the other hand, some remain relatively untouched. From all above indicators, we conclude that “bots,” while undeniably highly numerous and often verbose on Russian Twitter, are often segregated to isolated mini communities that have little impact on real politically engaged users. Real political communities do likely possess a certain proportion of bots, however, as identified in the literature (Kollanyi 2016; Murthy et al. 2016), these bots are likely to be complex and highly sophisticated, making their study challenging but their potential impact on shifting real conversations greater.

### 30.5 CONCLUSION

Data on public and political engagement of Russian citizens, active political discussions and debates, and even protest coordination activities, are readily available to researchers studying Russian politics. This chapter illustrates how SNA can be instrumental and ineluctable in evaluating key research hypotheses utilizing such data. Using six resonant political discussions collected from Twitter over the summer of 2016, we validate multiple political and sociological theories important for Russian studies. Firstly, we find that Russian online society is divided among the same ideological lines as the public sphere, representing two distinct and consistent communities of users, one supporting the Kremlin, and a “non-systemic” opposition that opposes it. Secondly, we validate the presence of “echo-chambers” in Russian social networks, identifying polarization between the two main political groups. Thirdly, we observe that “influencers” on Russian Twitter are not generally traditional political elites, but “interesting” and highly informative users such as that of famous pundits, parody accounts, or news sources. Furthermore, given regular users’ isolation into self-created separate ideological communities as well as further mini-communities, we comment that the expected impact of information control strategies by the government are likely quite limited.

We obtain our results through the application of a thorough and holistic approach in evaluating the network structure, focusing on three levels of analysis. Macroscopic methods, such as community detection and network visualization, supports the evaluation of the “overall” picture of online Russian political



society. Mesoscopic methods validate the detected structures and provide insight into the sub-structure of each detected community. Finally, microscopic methods identify the “influential” users who are able to widely disseminate content and impact political discussions.

We find that SNA is also an economical method to detect “bots” in social networks and evaluate their impact on real political users. Currently a hotly debated topic both internationally (Murthy et al. 2016; McKelvey and Dubois 2017; Ferrara 2018) and in Russia (Stukal et al. 2017; Lawrence 2015), our findings demonstrate that the impact of “bots” on Russian social networks is likely quite negligible. We find that while numerous and often verbose, “bots” are mostly isolated to mini-communities far removed from real politically engaged users, and as such are unlikely to impact real political discussions. We conclude by noting that the efficiency of SNA in extracting real and valid structures in Russian social networks makes it a prerequisite and fundament for the application of further advanced methods, such as topic or sentiment analysis, when studying Russian politics (for more on sentiment analysis, see Chap. 28).

### ANNEX A: POLARIZATION OF COMMUNITIES

Polarization of communities is taken from an aggregate of all six events. Note that as over 1000 groups were detected, only the most numerically relevant are shown. Percentages represent proportion of the total for each community (the “All” category).

**Table 30.8** Polarization of communities: retweets

<i>Retweets</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>8</i>	<i>All</i>
0	89.7%	0.4%	3.0%	0.0%	0.0%	0.0%	0.0%	194,393
1	2.7%	75.4%	0.3%	0.0%	0.0%	0.0%	0.0%	33,392
2	6.8%	0.0%	77.9%	0.0%	0.0%	0.0%	0.0%	54,672
3	15.3%	0.4%	0.5%	75.3%	0.0%	0.0%	0.0%	837
4	11.9%	1.4%	0.7%	1.4%	75.9%	0.0%	0.0%	286
5	50.5%	1.0%	7.5%	0.5%	0.0%	13.0%	0.0%	200
8	1.1%	0.0%	0.0%	0.0%	0.0%	0.0%	19.5%	87
All	196,233	33,698	53,563	793	232	138	22	350,616

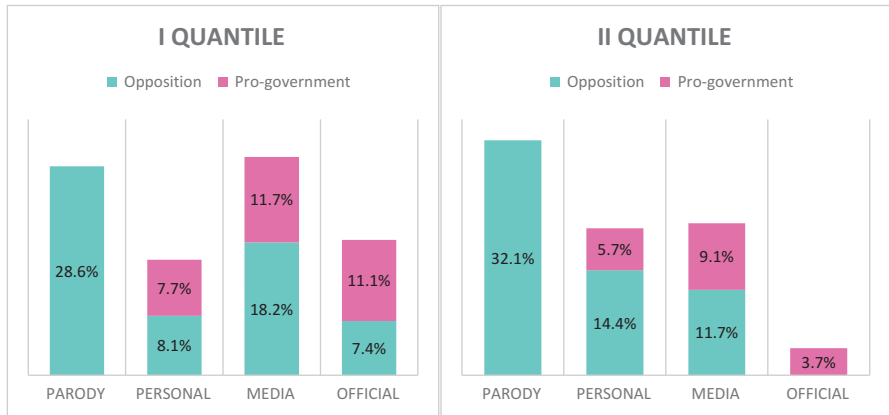
**Table 30.9** Polarization of communities: mentions

<i>Mentions</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>5</i>	<i>All</i>
0	80.7%	0.3%	8.8%	0.0%	0.0%	12,114
1	2.9%	82.3%	0.5%	0.0%	0.0%	1405
2	21.0%	0.3%	66.8%	0.0%	0.0%	3312
3	7.5%	0.0%	1.3%	87.5%	0.0%	80
5	35.0%	0.0%	15.0%	0.0%	0.0%	20
All	11,584	1456	3671	70	1	21,691

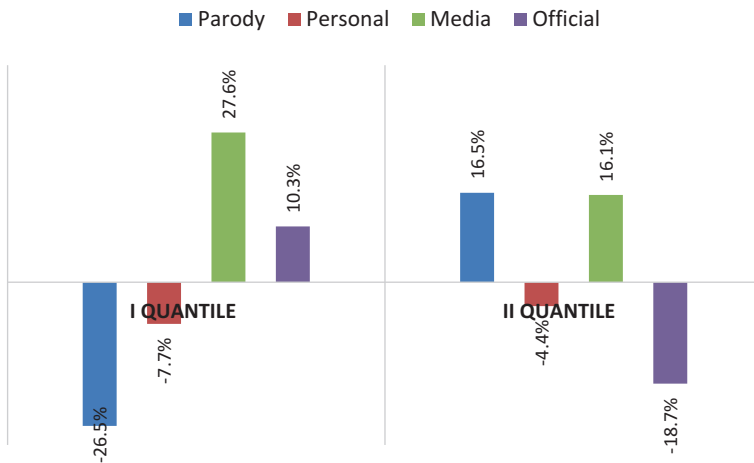
**Table 30.10** Polarization of communities: replies

<i>Replies</i>	<i>0</i>	<i>1</i>	<i>2</i>	<i>All</i>
0	85.3%	0.1%	6.5%	5354
1	2.0%	83.7%	0.0%	245
2	24.3%	0.1%	63.7%	1172
All	5423	250	1206	8083

ANNEX B: MICROSCOPIC



**Fig. 30.7** First and second quantile breakdown of account types by political communities



**Fig. 30.8** Change in influence of leaders' accounts when centrality complements the impact of their distributed content

### ANNEX C: TWEET PATTERNS FOR THE TWO MAIN COMMUNITIES, TWEETS PER HOUR BY EVENT

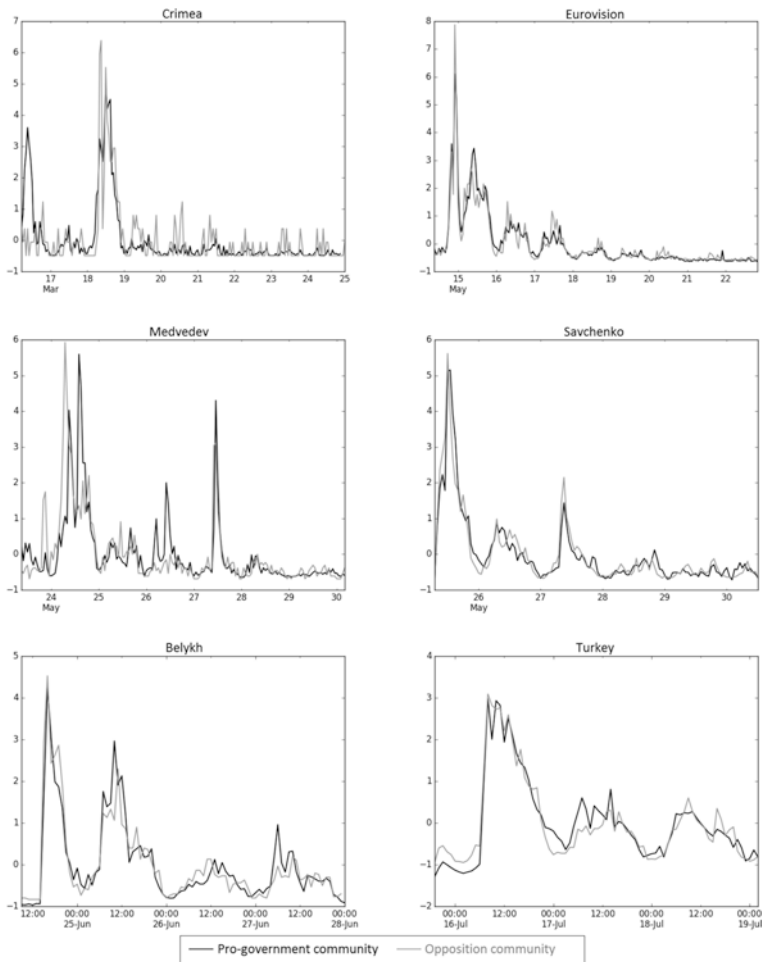


Fig. 30.9 Tweet patterns for the two main communities, tweets per hour by event

### NOTES

1. Compared to 46.6M users a month for VK and an internet penetration of 109.5M. Source: *Translate Media* (<https://www.translatemedia.com/translation-services/social-media/russian-social-media/>).
2. According to the *Statista.com*. Source: <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>.
3. In 2016, amongst the top 100, 17 accounts belonged to politicians, government organizations or media personalities who comment mostly on political

- issues, and a further 17 were of news media. In 2017, the pattern was similar with 15 and 16 respectively. Prime Minister Dmitry Medvedev has remained the most followed politician and always either tops or is at the top of the list. Although the trend has shifted towards popularization of non-political accounts, as of 2019, the nominal politicization of Twitter still remains quite high with 28 politically relevant accounts among the top 100. In comparison, German, French, UK, and American politicians' positions are less representative in the top 100, as these segments are largely dominated by the non-political (celebrities, media and sport personalities) opinion leaders.
4. Lancichinetti and Fortunato (2009) provides a good overview of available methods for various data and network constraints.
  5. As a test case, we amalgamated the six samples into one master network, which yielded a modularity of 0.551—a relatively strong statistic of network stratification.
  6. For introduction and application of methods in Network Analysis, see Zinoviev (2018).
  7. See Twitter's Developer documentation for available streaming parameters and their limitations: <https://developer.twitter.com/en/docs/tweets/filter-real-time/guides/basic-stream-parameters.html>.
  8. For instance, the #followback hashtag is commonly used by users to increase their number of followers.
  9. See Currarini et al. (2007) for in depth explanation.
  10. Note that this type of tweet was repeated by many users with slight variation in language.
  11. Given the difference in the sizes of the pro-government and Opposition communities, a comparison of their participation in events was performed by transforming the nominal tweets per hour scale into a standardized one using corresponding t-scores.
  12. In July 2018, the Washington Post reported that Twitter suspended 70M accounts in May and June. See Timberg and Dwoskin (2018) for more information.

## REFERENCES

- Agarwal, S.D., W.L. Bennett, C.N. Johnson, and S. Walker. 2014. A Model of Crowd Enabled Organization: Theory and Methods for Understanding the Role of Twitter in the Occupy Protests. *International Journal of Communication* 8: 646–672.
- Bakshy, Eytan, Jake M. Hofman, Winter Mason, and Duncan J. Watts. 2011. Everyone's an Influencer: Quantifying Influence on Twitter. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining SE – WSDM '11*: 65–74. <https://doi.org/10.1145/1935826.1935845>.
- Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 348 (6239): 1130–1132. <https://doi.org/10.1126/science.aaal160>.
- Barash, Vladimir, and John Kelly. 2012. Saliency vs. Commitment. Dynamics of Political Hashtags in Russian Twitter. *The Berkman Center Research Publication* 9: 0–16. <https://ssrn.com/abstract=2034506>.
- Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. 2015. Tweeting from Left to Right: Is Online Political Communication More Than

- an Echo Chamber? *Psychological Science* 26 (10): 1531–1542. <https://doi.org/10.1177/0956797615594620>.
- Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Third International AAAI Conference on Weblogs and Social Media*, 361–362. <https://doi.org/10.1136/qshc.2004.010033>.
- Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. Network Analysis in the Social Sciences. *Science* 323 (5916): 892–895. <https://doi.org/10.1126/science.1165821>.
- Brandes, Ulrik, Daniel Delling, Marco Gaertler, Robert Görke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. 2007. On Finding Graph Clusterings with Maximum Modularity. In *Graph-Theoretic Concepts in Computer Science*, 121–132.
- Breuer, Anita, Todd Landman, and Dorothea Farquhar. 2015. Social Media and Protest Mobilization: Evidence from the Tunisian Revolution. *Democratization* 22 (4): 764–792. <https://doi.org/10.1080/13510347.2014.885505>.
- Coleman, James. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap.
- Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson. 2014. Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication* 64 (2005): 317–332. <https://doi.org/10.1111/jcom.12084>.
- Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2007. An Economic Model of Friendship: Homophily, Minorities and Segregation. *Ssrn* 20. <https://doi.org/10.2139/ssrn.1021650>.
- Davis, Clayton A., Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. *BotOrNot: A System to Evaluate Social Bots*, 4–5. <https://doi.org/10.1145/2872518.2889302>.
- Earl, Jennifer, Heather McKee Hurwitz, Analicia Mejia Mesinas, Margaret Tolan, and Ashley Arlotti. 2013. This Protest Will Be Tweeted: Twitter and Protest Policing During the Pittsburgh G20. *Information Communication and Society* 16 (4): 459–478. <https://doi.org/10.1080/1369118X.2013.777756>.
- Ferrara, Emilio. 2018. Measuring Social Spam and the Effect of Bots on Information Diffusion in Social Media. In *Complex Spreading Phenomena in Social Systems*, 229–255. Springer. [https://doi.org/10.1007/978-3-319-77332-2\\_13](https://doi.org/10.1007/978-3-319-77332-2_13).
- Frederick, Evan L., Choong Hoon Lim, Galen Clavio, and Patrick Walsh. 2012. Why We Follow: An Examination of Parasocial Interaction and Fan Motivations for Following Athlete Archetypes on Twitter. *International Journal of Sport Communication* 5 (4): 481–502.
- Fuchs, Christopher. 2014. *OccupyMedia!: The Occupy Movement and Social Media in Crisis Capitalism*. Alresford, UK: Zero Books.
- Gel'man, Vladimir. 2015. Political Opposition in Russia: A Troubled Transformation. *Europe-Asia Studies* 67 (2): 177–191. <https://doi.org/10.1080/09668136.2014.1001577>.
- Golos. 2011. *Pervoe Zaávlenie asociacii 'GOLOS' po rezul'tatam dolgosrochnogo nablúdeniá hoda izbiratel'nyh kampanij mestnyh, regional'nyh i vyborov deputatov Gosudarstvennoj dumy Rossii, naznačennyh na 4 dekabrá 2011 g. Étapy vydviženíá i registracii*.
- Greene, Samuel. 2013. Beyond Bolotnaia. *Problems of Post-Communism* 60 (2): 40–52. <https://doi.org/10.2753/PPC1075-8216600204>.

- . 2018. Twitter and the Russian Street: Memes, Networks and Mobilization. *Center for New Media and Society Working Paper 1*.
- Halberstam, Yosh, and Brian Knight. 2015. Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter. *Journal of Public Economics* 143: 73–88. <https://doi.org/10.1016/j.jpubeco.2016.08.011>.
- Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. 2014. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PloS One* 9 (6): e98679.
- Jensen, Michael. 2018. Russian Trolls and Fake News: Information or Identity Logic? *Journal of International Affairs* 71 (1.5): 115–124.
- Jungherr, Andreas. 2014. The Logic of Political Coverage on Twitter: Temporal Dynamics and Content. *Journal of Communication* 64 (2): 239–259. <https://doi.org/10.1111/jcom.12087>.
- Kelly, John, Vladimir Barash, Karina Alexanyan, Bruce Etling, Robert Faris, Gasser Urs, and John G. Palfrey. 2012. Mapping Russian Twitter. *The Berkman Center Research Publication* 760 (2010): 1–60. <https://doi.org/10.5465/amr.2011.0193>.
- Kollanyi, Bence. 2016. Where Do Bots Come from? An Analysis of Bot Codes Shared on GitHub. *International Journal of Communication* 10 (June): 4932–4951.
- Koren, Yehuda. 2003. On Spectral Graph Drawing. In *International Computing and Combinatorics Conference*, 496–508. Berlin.
- Lancichinetti, Andrea, and Santo Fortunato. 2009. Community Detection Algorithms: A Comparative Analysis. *Physical Review E* 80 (9): 1–12.
- Lawrence, Alexander. 2015. Social Network Analysis Reveals Full Scale of Kremlin's Twitter Bot Campaign. *Globalvoices* 2015. <https://globalvoices.org/2015/04/02/analyzing-kremlin-twitter-bots/>.
- Livne, Avishay, Matthew P. Simmons, Eytan Adar, and Lada A. Adamic. 2011. The Party Is Over Here: Structure and Content in the 2010 Election. *Association for the Advancement of Artificial Intelligence* 161 (3): 201–208. <https://doi.org/10.1007/s00024-003-2459-0>.
- Lonkila, Markku. 2012. Russian Protest On- and Offline: The Role of Social Media in the Moscow Opposition Demonstrations in December 2011. *FIIA BRIEFING PAPER* 98.
- Lotan, Gilad, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, and Danah Boyd. 2011. The Arab Spring | The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communication* 5 (0): 31. <https://ijoc.org/index.php/ijoc/article/view/1246>.
- Maslinsky, Kirill, Sergey Koltcov, and Olessia Koltsova. 2013. Changes in the Topical Structure of Russian-Language LiveJournal: The Impact of Elections 2011. *Ssrn*. <https://doi.org/10.2139/ssrn.2209802>.
- McKelvey, Fenwick, and Elizabeth Dubois. 2017. Computational Propaganda in Canada: The Use of Political Bots. *The Computational Propaganda Project Working Paper Series*.
- Murthy, Dhiraj, Alison B. Powell, Ramine Tinati, Nick Anstead, and Susan J. Halford. 2016. Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital. *International Journal of Communication* 10 (June): 4952–4971.

- Myers, Seth, and Jure Leskovec. 2014. The Bursty Dynamics of the Twitter Information Network. In *Proceedings of the 23th International Conference on World Wide Web (WWW)*, 913–924. <https://doi.org/10.1145/2566486.2568043>.
- Myers, Seth A., Aneesh Sharma, Pankaj Gupta, and Jimmy Lin. 2014. Information Network or Social Network? The Structure of the Twitter Follow Graph. *International World Wide Web Conference*: 493–498. <https://doi.org/10.1145/2567948.2576939>.
- Nagornyy, Oleg, and Olessia Koltsova. 2017. Mining Media Topics Perceived as Social Problems by Online Audiences: Use of a Data Mining Approach in Sociology. *Srvn*. <https://doi.org/10.2139/ssrn.2968359>.
- Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. “The PageRank Citation Ranking: Bringing Order to the Web.”
- Pallin, Carolina Vendil. 2017. Internet Control Through Ownership: The Case of Russia. *Post-Soviet Affairs* 33 (1): 16–33.
- Pfeffer, Jürgen, Katja Mayer, and Fred Morstatter. 2018. Tampering with Twitter’s Sample API. *EPJ Data Science* 7 (1): 50. <https://doi.org/10.1140/epjds/s13688-018-0178-0>.
- Reuter, Ora John, and David Szakonyi. 2013. Online Social Media and Political Awareness in Authoritarian Regimes. *British Journal of Political Science* 45 (1): 29–51. <https://doi.org/10.1017/S0007123413000203>.
- . 2015. Online Social Media and Political Awareness in Authoritarian Regimes. *British Journal of Political Science* 45 (1): 29–51.
- Roesen, Tine, and Vera Zvereva. 2014. Social Network Sites on the Runet. In *Digital Russia: The Language, Culture and Politics of New Media Communication*, ed. Ingunn Lunde, Michael Gorham, and Martin Paulsen, 72–87.
- Romero, Daniel M., Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. 2011. Influence and Passivity in Social Media. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 18–33.
- Rosvall, Martin, and Carl T. Bergstrom. 2008. Maps of Random Walks on Complex Networks Reveal Community Structure. *PNAS* 105 (4): 1118–1123. <https://doi.org/10.1103/PhysRevLett.100.118703>.
- Rosvall, Martin, D. Axelsson, and Carl T. Bergstrom. 2009. The Map Equation. *European Physical Journal: Special Topics* 178 (1): 13–23. <https://doi.org/10.1140/epjst/e2010-01179-1>.
- Sinclair, Betsy. 2016. Network Structure and Social Outcomes: Network Analysis for Social Science. In *Computational Social Science*, ed. R. Michael Alvarez, 121–139. Cambridge University Press.
- Spaiser, Viktoria, Thomas Chadefaux, Karsten Donnay, Fabian Russmann, and Dirk Helbing. 2017. Communication Power Struggles on Social Media: A Case Study of the 2011–12 Russian Protests. *Journal of Information Technology and Politics* 14 (2): 132–153. <https://doi.org/10.1080/19331681.2017.1308288>.
- Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2017. Detecting Bots on Russian Political Twitter. *Big Data* 5 (4).
- Timberg, Craig, and Elizabeth Dwoskin. 2018. Twitter Is Sweeping out Fake Accounts Like Never Before, Putting User Growth at Risk. *The Washington Post* 2018.
- Towner, Terri L., and David A. Dulio. 2012. New Media and Political Marketing in the United States: 2012 and Beyond. *Journal of Political Marketing* 11 (1–2): 95–119.
- Tremayne, Mark. 2014. Anatomy of Protest in the Digital Era: A Network Analysis of Twitter and Occupy Wall Street. *Social Movement Studies* 13 (1): 110–126.



- Tucker, Joshua A., Jonathan Nagler, Megan Macduffee Metzger, Pablo Barberá, Richard Penfold-Brown, and Duncan Bonneau. 2016. Big Data, Social Media, and Protest: Foundations for a Research Agenda. In *Computational Social Science Discovery and Prediction*, ed. R. Michael Alvarez, 199–224.
- Tufekci, Zeynep, and Christopher Wilson. 2012. Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square. *Journal of Communication* 62 (2): 363–379.
- Waters, Richard D., and Jensen M. Williams. 2011. Squawking, Tweeting, Cooing, and Hooting: Analyzing the Communication Patterns of Government Agencies on Twitter. *Journal of Public Affairs* 11 (4): 353–363. <https://doi.org/10.1002/pa.385>.
- White, Stephen, and Ian McAllister. 2014. Did Russia (Nearly) Have a Facebook Revolution in 2011? Social Media’s Challenge to Authoritarianism. *Politics* 34 (1): 72–84.
- Wolfsfeld, G., E. Segev, and T. Sheafer. 2013. Social Media and the Arab Spring: Politics Comes First. *The International Journal of Press/Politics* 18 (2): 115–137. <https://doi.org/10.1177/1940161212471716>.
- Yang, Kai-Cheng, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. Arming the Public with Artificial Intelligence to Counter Social Bots. *ArXiv Preprint ArXiv:1901.00912* 2019: 48–61. <https://doi.org/10.1002/hbe2.115>.
- Zherebtsov, Mikhail. 2019. Taking Stock of Russian EGovernment. *Europe-Asia Studies*: 1–29.
- Zinoviev, Dmitry. 2018. *Complex Network Analysis in Python: Recognize-Construct-Visualize-Analyze-Interpret*. Pragmatic Bookshelf.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# The State of the Art: Surveying Digital Russian Art History

*Reeta E. Kangas*

## 31.1 INTRODUCTION

Digital methodologies can be used to complement more traditional approaches to art history. Yet, mainly due to the difficulties of analyzing visual material with computer-assisted methods, digital art history and visual analysis are areas that have arguably developed slower than other branches of digital humanities (see Drucker 2013, 5; Klinke 2016, 16; Lozano 2017, 2). With regard to training in digital methods, the efforts in the humanities are rather scattered and digital training in art history is especially lacking (Zorich 2013, 16). However, the field of digital humanities could benefit immensely from the traditionally strong expertise of art historians in curating and creating cultural data (Schelbert 2017, 5). Indeed, digital art history is not quite as new a field as is often thought. Some even argue that art history is not behind other disciplines in the development of the digital (Zweig 2015, 40–41). Nonetheless, despite the recent advances in image recognition, digital methods have not been as widely adopted in the analysis of the visual as in some other fields of humanities, making it easy to overlook the efforts that do exist.

Though Soviet and Russian visual materials have been studied quite extensively, they have not been analyzed much within the framework of digital art history. This is partly because many of the widely used digital art repositories contain only Western European and Renaissance art, leaving other art in a marginalized role (see, e.g., Münster et al. 2018, 380). However, scholars are now

---

R. E. Kangas (✉)  
University of Turku, Turku, Finland  
e-mail: [reeta.kangas@utu.fi](mailto:reeta.kangas@utu.fi)

debating the advantages of creating and making accessible digitized Russian visual material in collections and archives in- and outside of Russia (see Kizhner et al. 2018; Bridgers and Blood 2010; Kain 2018). For example, Biryukova et al. (2017) discuss how virtual cultural storages and virtual museums can be used to preserve the Russian cultural heritage. Other researchers have analyzed the possibilities and problems associated with making 3D models of cultural heritage objects and Russian architectural monuments, such as churches and monasteries, and presenting them online (see e.g. Borodkin et al. 2015; Zherebyatiev and Ionova 2014). Indeed, as Biryukova et al. (2017, 157) show, many of Russia's most popular virtual museums contain churches and monasteries reconstructed in a virtual space. Some researchers, like Anna Sanina (2019), Olha Korniienko (2014), and S. Polovinets and E. Baranova (2018), have even applied digital methods to the study of Russian and Soviet satirical visual material. However, the majority of research using digital methods is based on content analyses, which in turn have relied on the coding of the images by the researcher or research assistants. To the best of my knowledge, no machine learning or computer vision methods have been used in visual studies of Russian and Soviet material.

In this chapter, I chalk out some options and possibilities to expand and apply new digital research methods and visual analyses, in order to complement the more traditional approaches to Russian and Soviet art history. As an example of a field of art historical research that may benefit from such digital methods, I use my own research on Soviet political cartoons published during the "Great Patriotic War," as the years of war between the Soviet Union and Nazi Germany, 1941–1945, of the Second World War are known in Russian. During these years, the official party newspaper *Pravda* published 185 political cartoons, bearing nine different artists' signatures. In the course of my research, I manually collected these political cartoons and assembled an Excel spreadsheet with detailed annotations that essentially functioned as a database and allowed me to conduct a qualitative analysis on them.

To interpret a political cartoon, it is necessary to understand the contextual, textual, and visual features and information contained in them. This requires the researcher to have a certain amount of background knowledge. I employ Ernst Gombrich's (2002, 142–154) ideas, according to which a communicating image consists of three components—context (the environment within which the cartoon exists), caption (the verbal elements of the image), and code (the visual language the artist uses). This chapter thus ultimately discusses how digital methods could facilitate a Gombrichian analysis of a communicative image, such as a Soviet political cartoon.

Before getting into the use of digital methods to enhance the research of visual material, it is first necessary to give a brief overview of the situation in Russia regarding the digitization of material, copyright laws, and open access. Next, I look at recent developments in digital methods for art history and their potential application to Russian and Soviet art, especially with regard to

visualizing data and the use of machine-learning algorithms to help analyze databases of relevant textual and visual material. One of the benefits of using such algorithms is that they could facilitate piecing together the cultural context for an art historical research project. However, as I discuss in the following section, the vast amount of cultural knowledge that is required for a machine to properly understand the representations within an image is where machine-learning algorithms reach their limits. As is laid out in the final section, these limits could be overcome by combining the new digital methods with traditional art historical research methods. Ideally, larger research projects, featuring both Information Technologies (IT) professionals and trained art historians, would enable us to create more useful art historical databases that would allow for a more effective use of the new digital methods while also combining the strengths of both digital and traditional art historical research.

### 31.2 DIGITIZATION, COPYRIGHT LAWS, AND OPEN ACCESS

It is a common trend that archival material and cultural artifacts are being digitized at an increasing rate. However, the level of digitization is not universal, and its organizational forms differ. Certain cultures, mainly Western European and North American ones, are making bigger investments in their digitization projects and are thus overrepresented, while others remain in a marginalized position (Rodríguez Ortega 2013, 131). Some countries, like Russia, have government involvement driving the digitization, while in others it remains the task of individual organizations.

It is often difficult for art historians to find relevant, available, open-access, and good quality visual material in digital repositories (Münster et al. 2018, 369–371; *ibid.*, 380). And with digital archives that are not exclusively devoted to Russian data, it is occasionally difficult to search for Russian material, because not all archives attach keywords such as “Russia” or “Slavic” to their documents and objects (cf. Bridgers and Blood 2010, 78; for more, see Chap. 20). These problems with accessing digital databases often lead to researchers creating their own personal collections (Münster et al. 2018, 371). Thus, Russian art history remains very much a question of the researcher knowing where to look for accessible and relevant material, and in many cases still traveling to the destination to retrieve it.

Online resources of Russian digitized art are rather limited. However, some resources do exist. For instance, some Russian art museums have now made parts of their collections available online, and some have even created virtual tours of their museums (see, e.g., Virtual Visit, the State Hermitage Museum). A number of museums and galleries, including the State Russian Museum, are also collaborating with the Google Arts and Culture project to digitize and put parts of their collections online (see Virtual Russian Museum). In addition to such classic art resources, there are also newspaper, journal, and photography repositories that may be of interest to art historians. For instance, a

collaboration of the Russian search engine Yandex with museums and private collectors has resulted in a large online photo archive (see *Istoriâ Rossii v fotografiâh*).

The National Library of Finland has also recently subscribed to East View's digital collections (<http://www.eastview.com>), sidelining their microfilm collections. However, the East View interface offers only a text-based search option, which makes looking for images in the newspaper difficult. Furthermore, compared to the library's microfilms, the digital archive's image quality is worse and some issues of *Pravda* that were available on microfilm are missing from the digital archive. Nonetheless, these digital copies of *Pravda* provide an easier option for studying the textual environment—the Gombrichian caption—within which the image exists. But while digital repositories like East View make accessing digitized material easier for those who have online access, they do not provide the services for free (for more, see Chap. 20). They offer researchers material that would otherwise require them travelling to archives to retrieve, but they do not make the information openly available to everyone. Furthermore, the digitization of textual sources is generally much more common than that of, for example, art objects.

An ongoing Ministry of Culture led program aims to have all the Russian Federation Museum Collections cataloged, digitized, and available online at <https://goskatalog.ru> by 2026 (see *Gosudarstvennyj katalog Muzejnogo fonda Rossijskoj Federacii*). That is, it aims to make available metadata and pictures of all the items in the public museums. The participation of private museums in this project is voluntary (Kizhner et al. 2018, 351–352). In 2018—eight years before the project was supposed to be completed—only 14% of the objects were digitized and 9%, that is 7,034,904 objects, were included in the database (ibid., 352–354). By May 2019, the number of objects in the online catalogue was 11,017,513, which means that the digitization and cataloging process advanced by about four million objects within the past year or so.

This digitization project by the Russian Ministry of Culture, however, does not currently grant complete open online access to the cultural heritage objects. For example, in St. Petersburg the number of digitized items is higher than on average in Russia, and even higher than digitization on average in European museums, but the number of items available online is lower than the average in Russia (Kizhner et al. 2018, 356–358). Furthermore, the quality of the photographs is not necessarily an aspect to which much attention has been paid. This becomes evident when scrolling through the various images of the catalog. A more thorough “digital museumification,” that is, a proper transformation of the object into digital form with full metadata, would be needed to make the objects in the catalog more useable to the researcher (cf. Biryukova et al. 2017, 153). This could be achieved by using crawlers or appropriate scripts. If the Russian Ministry of Culture's project's aim is achieved by 2026, and if the Russian policies allow for open licenses on cultural heritage objects, this would

provide substantially easier access to the Russian cultural heritage to a wider audience, including international researchers (Kizhner et al. 2018, 363). It remains to be seen where the digitization project will lead and what it will in the end provide to the researchers of Russia.

Copyright laws largely limit what digitized material remains closed and what becomes public, which influences the research and other projects connected to cultural objects (Arditi 2018, 54; Roued-Cunliffe 2018, 288). Internationally, copyright laws largely accept the so-called fair use policy of images, which means that they can be used without obtaining permission from the copyright holder in certain cases, such as in research papers where the image is the direct subject of the analysis rather than merely an illustration. However, the Russian laws are more restrictive. Here, the state legislation supports the so-called “permissions culture,” which works counter to the fair use policy. Accordingly, museums can retain the rights to all their material even in a case that would be regarded as fair use in a research publication. In practice, this varies from museum to museum and the researcher needs to figure out the museum’s practices. For instance, the State Hermitage allows their material to be used, for example, for educational purposes, in conference presentations, and in PhD theses. But permission is required to use images for commercial purposes or in research publications, or to publish conference slides online (Kizhner et al. 2018, 359). The fact that visual material is by nature copyright heavy, when compared, for example, to text sources, hinders the work of individual researchers as well as the building of digital repositories that would benefit the field more broadly.

The complexities of the copyright laws and the “permission culture” that prevail in Russia make it unfeasible for an individual researcher to make their personal databases of primary material open to other researchers. While nothing prevents me from publishing my metadata, the fear of litigation or of being denied further access to the archival material has kept me from making my collection of *Pravda* political cartoons accessible to the public. Instead, I have the database stored on personal devices and the images saved in accessible formats, such as PNG and TIFF. Indeed, when thinking about the storage of data, it is necessary to consider whether the data can be made open and who could benefit from it. For data to be openly accessible, it is necessary to use file formats that are possible to use with a variety of non-commercial programs and are likely to stay in use for a long period of time (Roued-Cunliffe 2018, 292). Such formats include, but are not limited to, JSON, XML, and IIIF. With regard to Russian images, a more widespread use of the annotation ready IIIF format by the heritage institutions and in the Russian museum catalogs would provide researchers better access and more possibilities to present the images with stable Uniform Resource Identifiers (URIs).

### 31.3 NEW DIGITAL APPROACHES TO VISUAL ANALYSIS AND ART HISTORY

Digital methods provide new approaches to art history, such as the visualization and display of data and research results, the digitization and digital rendering of art, and most recently the use of convolutional neural networks (CNNs) for simple and even more complex recognition tasks. The increasing computational power we have at our disposal now enables rather more complex visualizations than the traditional bar chart or pie chart. For instance, one can create complex visualizations that consist of a large quantity of individual images that, when combined, provide an overall picture (Schelbert 2017, 4). New visualization techniques also allow for more dynamic “moving” charts in electronic form. The use of such visualizations in digital visual research has been criticized for its lack of accompanying interpretation (see Bishop 2017, paragraph 9). However, when approached with care, new modes of visualization can be very powerful at revealing tendencies that might otherwise be missed. Thus, with the *Pravda* political cartoons, one could create visualizations to exemplify their structure, their connections to historical events, their intertextuality, and other aspects, in the spirit of Gombrich’s notion of contextualizing an image. One could, for example, place the cartoons on a map of Europe, showing where each one takes place, or do a cross-referencing of countries and animal characters to see the significance of animal symbolism in the cartoons.

In addition to such visualizations of data, digital methods also offer other options for representing research findings. Digitized art and the digital rendering of art artifacts enable the researchers to bring the art to a wider audience. For example, the University of Nottingham’s project *Windows on War: Soviet Posters 1943–1945* (see Windows on War), which was conducted by a multidisciplinary team, allows the visitor to look at the images while at the same time reading about culturally specific information and the historical contextualization of the images. In a sense, this makes the communication of art history independent of both location and time, allowing people to immediately access art from around the world and even to view a virtual reconstruction of an already destroyed artwork, such as an old building (Kellaway 2013, 95–96; Borodkin et al. 2015, 5–7). Furthermore, contemporary digital online spaces offer us the possibility to reconstruct old exhibitions, of which we have photographic evidence, such as the “godless corners” of the early twentieth-century Soviet Union (Kain 2018, 219). Thus, the use of digital methods is not limited to the actual process of conducting the research or disseminating the results within the academic environment; they can be employed in researchers’ popular outreach efforts as well.

One of the difficulties of digital humanities is to turn the primary material into useful data: to quantify a body of material that is not traditionally handled in such a way and to combine this quantification with humanities methods and theories of enquiry (Otty and Thomson 2016, 135). Manually annotating images is perhaps still the most common way of approaching the problem of



turning an image into a format that is possible to analyze with the use of a computer. Here, researchers annotate the images with appropriate keywords, that is, metadata, which are then used as a basis to build a database and conduct a computer-assisted analysis to find underlying tendencies of the material (see, e.g., Korniienko 2014; Sanina 2019). I followed this procedure when I made my database of the 185 *Pravda* political cartoons. My metadata included, when relevant, information on the cartoon's date of publication, page, position on the page, artist, title, captions, text in image, quote, poet, poem, characters, countries represented, symbols, combinations of a symbol and a country/person, combinations of attributes and a person/country, and the roles of the characters. This allowed me to analyze the cartoons on the basis of the assigned keywords and to make cross-references between them. Such studies rely on the human to do the coding, instead of employing machine vision, which to this date is not yet at a level where most researchers would completely rely on it or know how to use it for the principal annotation of the primary visual material.

The possibility that a machine could take over such basic art historical analyses would help immensely with metadata extraction and other rather mechanical work. The extraction of this metadata, in turn, would directly facilitate the analysis of Gombrich's caption and code—text, quotes, and title being part of the caption and characters, symbols, and attributes part of the code. Naturally, using machines to do this would enable researchers to process much larger datasets. And furthermore, a machine would assign keywords more consistently than a group of coders, who are each assigning keywords based on their varying interpretations (cf. Rose 2007, 60–61; Bell 2001, 22). By combining this computer analysis of a vast body of imagery with an art historian's analysis of specific images from that same body, one could also create a two-sided database. The first side would comprise simple computer-assigned characteristics of large amounts of images, while the second would consist of the art historian's keywords and would address the more semantic notions of the image (Dressen 2017, 8). This would allow the researcher to conduct a qualitative analysis with specific images as examples, while the bulk of the images serve as a contextualizing device.

The computer vision technology that would facilitate such analyses is in a process of constant development. For some time now, computers have been able to reliably detect the colors and textures of an artwork, which does not help us to make any semantic interpretations but does facilitate more precise quantitative analyses of colors and shadings used by various artists as well as to identify artworks and attribute them to artists (Manovich 2015, 22; Schelbert 2017, 5). According to Emily L. Spratt, the image analysis capabilities of machines is now approaching the second of Erwin Panofsky's three levels of art historical analysis. That is, they can not only identify basic elements within the image, such as animals or people, but also detect conventional cultural representations, such as religious motifs (Spratt 2017, 12). In the case of the Union of Soviet Socialist Republics (USSR), these could include ideological motifs, such as depictions of revolutionary events, or certain types of characters, such

as the archetypal worker or peasant. Applying such a tried and true art historical theory as Panofsky's to these new developments is, of course, not straightforward. But the fact that the capabilities of computer image analysis are now being directly compared to the image analysis skills of people is, on its own, very telling.

At present, a number of research groups are working to push the limits of what computer vision can do with comics (cf. Laubrock and Dubray 2019; Laubrock and Dunst 2019; Young-Min 2019). Any such research is, of course, heavily dependent on the availability of appropriate training sets. For instance, the Digital Comic Museum (DCM) hosts a set of nearly 200,000 pages of American comics published before 1959, segmented into panels and text bubbles by machine, and transcribed using optical character recognition (OCR), which can be downloaded at <https://github.com/miyyer/comics> (see Digital Comic Museum). Due to the imperfections of having the segmentation done by a machine, Nguyen et al. (2018) have also produced a subset of 772 pages from the DCM that has been fully annotated by humans. With the help of this training set, among others, researchers have achieved good results in identifying various elements of a comic, such as speech bubbles, panels, and captions. And they are now moving on to more advanced recognition tasks, such as getting machines to recognize recurrent characters, image-text relations, and simple narrative structures (Laubrock and Dunst 2019, 11–20; *ibid.* 28). It is conceivable, that similar datasets could be created of recent Russian comics. But unfortunately, many other areas of art history do not yet benefit from such vast and high-quality datasets, and as discussed in the following section, Russian art history is no exception to this rule.

### 31.4 THE CURRENT LIMITS OF MACHINE LEARNING

Perhaps the biggest challenge of using machine learning for analyzing visual imagery is that it requires very large datasets to train the algorithms. For large corpora of visual material with repetitive elements, such as medieval manuscripts, it has already proven to be especially feasible to use computer vision to annotate the images, saving the researchers countless working hours (Bell et al. 2013, 27). As with the medieval manuscripts, if it were possible to construct a sufficiently large training set, it now seems entirely possible that a machine could be trained to help analyze political cartoons. For instance, a machine could learn to identify certain characters by distinguishing the exaggerated physical attributes that make a caricature look like its target, such as the moustache and hair of Hitler or the big mouth and short stature of Goebbels. However, care would have to be taken to include enough features so that the computer would not, for example, mistake Chaplin for Hitler on the basis of his moustache. Furthermore, conventional facial recognition currently works by identifying the dimensions of the face. So, for this to work, either the facial recognition software would have to be expanded or a separate recognition

algorithm would have to be developed to identify exaggerated and satirized physical features (for more on machine learning, see Chap. 26).

The question in this specific case is whether the Soviet political cartoons, or visual propaganda more generally, are repetitive enough that training computers to do the annotation would be feasible. While features like Hitler's moustache certainly repeat, each cartoonist has their own style and the themes and topics vary. The only way to find out for sure would be to gather a sufficiently large dataset and try it out. Saurav Jha et al. (2018) point out that the training sets that are currently available are too small to train a neural network to recognize cartoon characters. Hence, the training sets need to be supplemented with the inclusion of photographs of the people who appear in the cartoons. However, the inclusion of large amounts of photographic material decreases the feature recognition of the cartoons. Additionally, the more exaggerated the features of the character, the more the machine has trouble identifying the face. Going even further, one wonders whether a machine could learn to detect satire and ridicule. Or make the connection and find the similarities and differences between a caricatured and a photographed Hitler. If a computer could effectively learn to examine the Gombrichian code of an image, it would enable the faster analysis of large visual corpora of propaganda imagery and, possibly, provide us with a more complete picture of the ways in which visual propaganda functions.

In addition to the size of the training set, the quality of its images and their similarity to the material that is to be investigated are also crucial, lest the neural network end up interpreting images in different ways than the trainer's intention. As the training set of convolutional neural networks influences the way the network interprets other images shown to it, the training set needs to be carefully planned so that it will not cause erratic results (Spratt 2017, 4). When an image is different from the images of the training set, the machine may end up facing difficulties. For example, in one image-to-image translation project, the machine was taught to transform images so that a winter scenery became a summer scenery, a photograph into a painting by a famous artist, or a horse into a zebra. However, the training set did not contain images of horses, which resulted in the machine transforming not only the horse's coat but also the skin of the bare-chested President of Russia riding the horse into a zebra-patterned being (Zhu et al. 2017). This exemplifies how the computer interprets visual material only on the basis of the training set that has been used. Thus, the machine does not have the contextual information and interpretation capabilities that a human in a similar situation would have.

Images have a high level of cultural coding. So even if a computer can extract large amounts of data from an image, it cannot understand the semantic side of an image as well as a human does. Current developments in computer vision aim to bridge this "semantic gap," to allow a machine to detect basic semantic meanings based on the information it can obtain from an image (Manovich 2015, 22). However, in the same way that computer vision needs to be trained to recognize images, humans have been trained by culture and society to

recognize and interpret them correctly (Spratt 2017, 7). In other words, for a machine to be able to correctly analyze the significant elements of an image, it essentially needs the same training and cultural knowledge that a human has. The almost incomprehensibly vast amount of information that forms this cultural context of an image is where machine learning and computer vision reach their limits and where the guidance and supervision by trained art historians will for the foreseeable future remain essential for any research project.

Our interpretations are always dependent on our spatial, temporal, and cultural contexts. Any interpretation by an art historian—or anyone else for that matter—is dependent on their background (Gaehtgens 2013, 23–24). For example, given an image of a miserable tiger, a contemporary of the wartime political cartoons in the Soviet Union would have understood its significance as a play on the German heavy tank *Tiger* getting stuck in the muddy spring of the Eastern Front, as would someone familiar with the fate of *Tiger* tanks. However, without the contextual knowledge, the symbolism of the animal could end up signifying something else, such as the characteristics of the Germans as defeated wild animals. That is, the interpretation I make might differ largely from the interpretation someone else makes—can a computer make such semantic interpretations?

In the same way that the interpretation of data is dependent on the background of the researcher, the way that the data are organized depends on the interpretations of the researcher. Thus, the way I organize data might differ largely from how someone else does it (see Otty and Thomson 2016, 115). In other words, when making interpretations or organizing data, one needs to remember one's own contextual situation and not blindly trust digital methods and believe that they will provide completely replicable and authoritative results (for more, see Chap. 21). And until machine-learning algorithms can be trained to take into account a reasonable proportion of the cultural context of its target material, we must bear in mind that any interpretations made by such algorithms will be based on a considerably narrower background than that of any human researcher.

### 31.5 HOW HUMANS AND MACHINES CAN WORK TOGETHER

The advantages of the new digital methods and of traditional art historical research are conveniently complementary. Indeed, by combining the strengths of a trained researcher with the capabilities of machine-learning algorithms, it should be possible to cancel out any limitations of either. The digitization of visual imagery enables researchers to conduct contextual analyses of images that would not be feasible without access to digital resources. Thus, it facilitates an even wider contextual analysis than Gombrich (2002, 142–154) could have had in mind when writing about the context of the image. As has been discussed, developments in the digital methods are currently on the cusp of making this possible. Well-designed databases with easy accessibility and properly annotated images would help researchers to examine the intertextuality

and connections between different works of art and other cultural, social, and historical phenomena (Brandhorst 2013, 72–73). For instance, it would help the researchers if computers could do a search and comparison within a database for art works that are similar to the one that they are examining. Of course, for this to be practical and feasible, it will first be necessary for the machines to be able to reliably identify and catalog certain elements within the visual artifacts. In this way, a computer could go through large corpora and their meta-data considerably faster than a human (Klinke 2016, 16). Furthermore, the comparison of many images with each other when it comes to composition and aesthetics could provide new insights into how different artists composed their images (Pfisterer 2018, 138). After all, it is impossible to compare as many images in person as it would be with a computer.

The high level of intertextuality of political cartoons would also become more evident with such comparative computational methods. Their connections to various areas of culture, including the Soviet visual propaganda imagery, which had the tendency to repeat and borrow ideas from previous images, illustrate propaganda's dependency on the cultural context within which it operates (see Kangas 2017, 46–47). For a contextual analysis of the political cartoons, the pages on which they were published—or even whole issues of *Pravda*—could be processed with OCR for a cross-referencing of the news text with cartoon. The comparison of the text surrounding the image with the actual image could provide additional contextualization, complementing the researcher's efforts to place the image within the context of the war events. The computer could also assign a value to the similarity between specific features of the cartoons and other images and cultural artifacts, war events, or their geographical location. These values could then be mapped onto a graph in which all the variables would be presented together in a dynamic visualization. However, due to the complexity and the wide variety of cultural representations, this is currently still beyond the capabilities of machine-learning.

More generally, with the help of computer programs that could search for such open access information—this requires open access as well—many projects could benefit from the information as a part of the contextualization of art (Dressen 2017, 4–5). And the emergence of large text databases of art historical material will enable a more thorough and accessible contextual analysis of the visual, which has traditionally been slow and cumbersome due to the large amount of background material needed (Drucker 2013, 10). Thus, with digital methods, it is possible for human researchers to take into account ever larger amounts of background information when conducting their research.

With regard to training a machine-learning algorithm to detect conventional cultural representations and thus establish connections to the broader cultural context, here too it is a question of having sufficient information available. Thus, apart from computers with the necessary processing power to do the analysis, the digitization and accessibility of cultural artifacts is essential for a full analysis that takes into account both the cultural context and the formal properties of the object (Schelbert 2017, 5). But is it possible to train a machine

to see what is not presented? Indeed, in images, and texts too, what has been left out conveys semantic information. That is, in an image what you cannot see is often as important—or even more important—than what you can see (Rose 2007, 72). Even if it is possible to train a machine-learning algorithm to “see” what is missing in an image, it is difficult for it to do an analysis of the meaning of the omissions. For example, in Soviet political cartoons, the Soviet Union or its allies are rarely shown. However, their omission does not mean that their presence is not implied. Once again, here, the interpretative skills and supervision by a trained art historian is necessary to fully understand what is going on.

The increasing use of digital methods in the study of the visual does not necessarily mean the overthrow of art history’s more traditional methodologies. In combining digital and traditional quantitative methods, the researcher can draw a range of conclusions from their datasets that would be difficult to manage without the machine’s computational power. But at the same time, through qualitative analyses, the researcher can make interpretations and evaluations that a computer cannot (see Klinke 2016, 28; Lozano 2017, 6; Rose 2007, 70–71). To employ such a wide methodological oeuvre calls for interdisciplinarity and/or collaboration between specialists from varying backgrounds. There have been several calls for such collaborations (e.g., Glinka et al. 2016, 209; Klinke 2016, 31; Mercuriali 2018, 149). Having teams that employ people with expertise from different fields and with different skill sets would further the goal of creating large, accessible databases, as well as the planning of new complex methods of analysis.

### 31.6 CONCLUSION

In this chapter I looked at some of the advantages and challenges that the digital study of visual material will encounter. My starting point was to consider these issues in the light of a previous research project which was conducted mainly with more “traditional” methods, such as archival work on microfilms, digitization of material, and conducting a “manual” qualitative analysis of the primary material. I used my earlier analysis of Soviet visual data as an example and discussed the possibilities of digital methods that I could have used in the project.

In some ways, researchers of Russian and Soviet visual material face many of the same challenges that any other academics face when using machine-learning methods to enhance their research. For instance, it is important to train the machine to be “intelligent” and to learn to “see” appropriately and ethically—we would not want the machine to learn to tamper with the data to make the researcher happy. Additionally, the training process is still too slow and complex to be used in a small-scale research project, but the development of machine and deep learning might change the situation and make these methods more approachable for a wider base of researchers.

In other ways, researchers of Russian art history face their own unique set of challenges in adopting the new digital methods. For instance, the problem of the “semantic gap,” that is, that computers are not able to handle the semantic side of the objects they are analyzing, is especially pertinent when analyzing visual imagery, which is heavily reliant on a large amount of contextualizing information. And the collection of such contextualizing information in digital databases, so as to make it useful for machine-learning algorithms, is further confounded in Russia by their restrictive copyright laws and permission culture.

These restrictive copyright laws are especially detrimental, as the lack of openly accessible, large-scale databases of visual material is the primary bottleneck preventing the use of new digital methods for conducting art historical research in Russia. As a result, these digital methods have not yet found a secure foothold within Russian visual studies. Some research has been conducted, but it has mainly relied on rather traditional computational methods, such as content analysis. Considering the breadth of visual material that Russia—which is generally considered to be a very visual culture—and the Soviet Union have produced, it would be extremely advantageous to employ some of the more recent digital methodologies to that material.

Nonetheless, larger projects featuring interdisciplinary teams and collaborations within art history and other fields that employ digital methods for visual analysis could yield considerable results. In many cases, a digital project would benefit from the participation of people with varying backgrounds and skills, such as IT, quantitative, and qualitative methods. Through the co-operation of people with all of these different skill sets, it would be possible to employ the digital methods more fully and find new creative solutions to, for instance, create suitable databases that better serve the researchers or more dynamic visualizations of the research results. So, despite the challenges some of them still face, the new digital methods provide many new possibilities for the study of the visual, facilitating an easier examination of the images’ context, caption, and code, in the spirit of Ernst Gombrich.

## REFERENCES

- Arditi, David. 2018. MusicDetour: Building a Digital Humanities Archive. In *Research Methods for the Digital Humanities*, ed. Lewis Levenberg, David Rheams, and Tai Neilson, 53–61. Cham: Palgrave Macmillan.
- Bell, Philip. 2001. Content Analysis of Visual Images. In *Handbook of Visual Analysis*, ed. Theo van Leeuwen and Carey Jewitt, 10–34. London: Sage.
- Bell, Peter, Joseph Schlecht, and Björn Ommer. 2013. Nonverbal Communication in Medieval Illustrations Revisited by Computer Vision and Art History. *Visual Resources* 29 (1–2): 26–37.
- Biryukova, Marina, Elena Gaevskaya, Antonina Nikonova, and Marina Tsvetaeva. 2017. Interdisciplinary Aspects of Digital Preservation of Cultural Heritage in Russia. *European Journal of Science and Technology* 13 (4): 149–160.
- Bishop, Claire. 2017. Against Digital Art History. *CUNY Graduate Center*, March 15. <https://humanitiesfutures.org/papers/digital-art-history/>.



- Borodkin, Leonid I., Timur Ya. Valetov, Denis I. Zherebyatiev, Maxim S. Mironenko, and Vyacheslav V. Moor. 2015. Rerezentaciá i vizualizaciá v onlajne rezul'tatov virtual'nojrekonstrukcii [Representation and Visualisation of Virtual Reconstruction Results on a Website]. *Istoričeskaá informatika* 3–4: 3–18.
- Brandhorst, Hans. 2013. Aby Warburg's Wildest Dreams Come True? *Visual Resources* 29 (1–2): 72–88.
- Bridgers, Jeffrey, and Katherine Blood. 2010. Not So Hidden: Slavic and East European Collections Ready for Study through the Library of Congress Prints & Photographs Division. *Slavic and East European Information Resources* 11 (2–3): 77–90. <https://doi.org/10.1080/15228886.2010.480965>.
- Digital Comic Museum. n.d. Accessed 24 Nov 2019. <https://digitalcomicmuseum.com>.
- Dressen, Angela. 2017. Grenzen und Möglichkeiten der digitalen Kunstgeschichte und der Digital Humanities—eine kritische Betrachtung der Methoden [The Limits and Possibilities of Digital Art History and Digital Humanities—A Critical Observation on the Methods]. In *Critical Approaches to Digital Art History*, ed. Angela Dressen and Lia Markey. *kunsttexte.de* 4.
- Drucker, Johanna. 2013. Is There a 'Digital' Art History? *Visual Resources* 29 (1–2): 5–13.
- Gaehetgens, Thomas W. 2013. Thoughts on the Digital Future of the Humanities and Art History. *Visual Resources* 29 (1–2): 22–25.
- Glinka, Katrin, Christopher Pietsch, Carsten Dilba, and Marian Dörk. 2016. Linking structure, texture and context in a visualization of historical drawings by Frederick William IV (1795–1861). *International Journal for Digital Art History* 2: 198–213.
- Gombrich, Ernst. 2002. *The Image and the Eye: Further Studies in the Psychology of Pictorial Representation*. London: Phaidon.
- Gosudarstvennyj katalog Muzejnogo fonda Rossijskoj Federacii [State Catalog of the Museum Fund of the Russian Federation]. n.d.. <https://gokatalog.ru>. Accessed 24 May 2019.
- Istoriá Rossii v fotografiáh [History of Russia in Photographs]. n.d.. <https://russiain-photo.ru/>. Accessed 20 Nov 2019.
- Jha, Saurav, Nikhil Agarwal, and Suneeta Agarwal. 2018. *Bringing Cartoons to Life: Towards Improved Cartoon Face Detection and Recognition Systems*. <https://arxiv.org/pdf/1804.01753.pdf>.
- Kain, Kevin M. 2018. Early Soviet Visual Antireligious Propaganda: The Display of Print Images in the Past, Present and Digital Future. *Slavic & East European Information Resources* 19 (3–4): 216–241. <https://doi.org/10.1080/15228886.2018.1539608>.
- Kangas, Reeta. 2017. *Cartoon Fables: Animal Symbolism in Kukryniy's Pravda Political Cartoons, 1965–1982*. PhD Diss., University of Turku.
- Kellaway, Brooke. 2013. Cataloging Contemporary Art in the Digital Age. *Visual Resources* 29 (1–2): 89–96.
- Kizhner, Inna, Melissa Terras, Maxim Rumyantsev, Kristina Sycheva, and Ivan Rudov. 2018. Accessing Russian Culture Online: The Scope of Digitization in Museums Across Russia. *Digital Scholarship in the Humanities* 34 (2): 350–367.
- Klinke, Harald. 2016. Big Image Data Within the Big Picture of Art History. *International Journal for Digital Art History* 2: 14–37.
- Korniienko, Olha. 2014. Sovetskaá moda čerez prizmu satiričeskogo žurnala Perek' (1964–1991 gg.): Baza dannyh, kontent-analiz karikatur [Soviet Fashion in the

- Prism of the Satirical Magazine *Perec'* (1964–1991): Database, Content Analysis of Caricatures]. *Istoričeskaá informatika* 4: 50–67.
- Laubrock, Jochen, and David Dubray. 2019. CNN-Based Classification of Illustrator Style in Graphic Novels: Which Features Contribute Most? In *MultiMedia Modeling. MMM 2019. Lecture Notes in Computer Science*, ed. Ioannis Kompatsiaris, Benoit Huet, Vasileios Mezaris, Cathal Gurrin, Wen-Huang Cheng, and Stefanos Vrochidis, vol. 11296, 684–695. Cham: Springer. [https://doi.org/10.1007/978-3-030-05716-9\\_61](https://doi.org/10.1007/978-3-030-05716-9_61).
- Laubrock, Jochen, and Alexander Dunst. 2019. Computational Approaches to Comics Analysis. *Topics in Cognitive Science*: 1–37. <https://doi.org/10.1111/tops.12476>.
- Lozano, Jorge Sebastián. 2017. Digital Art History at the Crossroads. In *Critical Approaches to Digital Art History*, ed. Angela Dressen and Lia Markey. *kunsttexte.de* 4.
- Manovich, Lev. 2015. Data Science and Digital Art History. *International Journal for Digital Art History* 1: 12–35.
- Mercuriali, Giacomo. 2018. Digital Art History and the Computational Imagination. *International Journal for Digital Art History* 3: 140–151.
- Münster, Sander, Christina Kamposiori, Kristina Friedrichs, and Cindy Kröber. 2018. Image Libraries and Their Scholarly Use in the Field of Art and Architectural History. *International Journal on Digital Libraries* 19 (4): 367–383. <https://doi.org/10.1007/s00799-018-0250-1>.
- Nguyen, Nhu-Van, Christophe Rigaud, and Jean-Christophe Burie. 2018. Digital Comics Image Indexing Based on Deep Learning. *Journal of Imaging* 4 (7): 1–34. <https://doi.org/10.3390/jimaging4070089>.
- Otty, Lisa, and Tara Thomson. 2016. Data Visualisation and the Humanities. In *Research Methods for Creating and Curating Data in the Digital Humanities*, ed. Matt Hayler and Gabriela Griffin, 113–139. Edinburgh: Edinburgh University Press.
- Pfisterer, Ulrich. 2018. Big Bang Art History. *International Journal for Digital Art History* 3: 133–139.
- Polovinets, S., and E. Baranova. 2018. Kontent-analiz karikatur ‘Kalinigradskoj pravdy’ [Content Analysis of the Caricatures from ‘Kalinigradskaya Pravda’]. In *Estestvennonaučnye metody v cifrovoj gumanitarnoj srede. Materialy Vserossijskoj naučnoj konferencii s meždunarodnym učastiem (g. Perm', 15–18 maá 2018 g.)* [Methods of the Natural Sciences in the Field of Digital Humanities: Materials from the All-Russian Scientific Conference with International Participation (Perm, May 15–18, 2018)], ed. S. I. Kornienko, 51–55. Perm: Perm University Press.
- Rodríguez Ortega, Nuria. 2013. Digital Art History: An Examination of Conscience. *Visual Resources* 29 (1–2): 129–133.
- Rose, Gillian. 2007. *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*. 2nd ed. London: Sage.
- Roued-Cunliffe, Henriette. 2018. Developing Sustainable Open Heritage Datasets. In *Research Methods for the Digital Humanities*, ed. Lewis Levenberg, David Rheams, and Tai Neilson, 287–307. Cham: Palgrave Macmillan.
- Sanina, Anna. 2019. ‘Who Are You Kidding?’: Visual Political Irony in Contemporary Russia. *Qualitative Inquiry* 25 (4): 432–444. <https://doi.org/10.1177/1077800418790292>.
- Schelbert, Georg. 2017. Art History in the World of Digital Humanities. Aspects of a Difficult Relationship. In *Critical Approaches to Digital Art History*, ed. Angela Dressen, and Lia Markey. *kunsttexte.de* 4.

- Spratt, Emily L. 2017. Dream Formulations and Deep Neural Networks: Humanistic Themes in the Iconology of the Machine-Learned Image. In *Critical Approaches to Digital Art History*, ed. Angela Dressen and Lia Markey. *kunsttexte.de* 4.
- Virtual Russian Museum. n.d.. <http://rusmuseumvrm.ru/projects/google/?lang=en>. Accessed 22 Nov 2019.
- Virtual Visit, the State Hermitage Museum. n.d.. <https://www.hermitagemuseum.org/wps/portal/hermitage/panorama>. Accessed 22 Nov 2019.
- Windows on War: Soviet Posters 1943–1945. n.d.. <http://windowsonwar.nottingham.ac.uk/>. Accessed 7 Jan 2019.
- Young-Min, Kim. 2019. Feature Visualization in Comic Artist Classification using Deep Neural Networks. *Journal of Big Data* 6 (1): 1–18. <https://doi.org/10.1186/s40537-019-0222-3>.
- Zherebyatiev, Denis I., and Polina A. Ionova. 2014. Virtual'naâ rekonstrukcii [Sic!] usad'by Peršino—unikal'nogo pamâtnika dvorânskoj kul'tury konca XIX—načala XX v. [Virtual Reconstruction of Peršino Estate—A Unique Monument of Noble Culture in the late XIX—Early XX Centuries]. *Istoričeskaâ informatika* 4: 31–49.
- Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. *Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks*. 2017 IEEE International Conference on Computer Vision (ICCV). <https://arxiv.org/pdf/1703.10593.pdf>.
- Zorich, Diane M. 2013. Digital Art History: A Community Assessment. *Visual Resources* 29 (1–2): 14–21.
- Zweig, Benjamin. 2015. Forgotten Genealogies: Brief Reflections on the History of Digital Art History. *International Journal for Digital Art History* 1: 38–49.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Geospatial Data Analysis in Russia's Geoweb

*Mykola Makhortykh*

## 32.1 INTRODUCTION

The rise of digital technologies has led to the emergence of new ways in which physical spaces are perceived, experienced, and mapped. The availability of high-quality satellite imagery amplified by the unprecedented possibilities for crowdsourcing geospatial data (Crampton 2009) has enabled the emergence of multiple platforms dealing with geographic information. It was followed by the integration of geographically aware computing in the architecture of major social media platforms (Crampton et al. 2013) and the growing capabilities for location tracking embedded into mobile devices (Sansurooah and Keane 2015). Together, these changes have given rise to a global collection of services which use the geographic data for different domains' applications. These services are currently known as “geospatial Web” (Lake and Farley 2009) or simply “geoweb” (Crampton 2009).

The emergence of geoweb and associated “neographic” (Haklay et al. 2008) practices of publishing, sharing, and visualizing information about places and people has significant implications for academic research. In the large-scale review of studies, which use geospatial data, Stock (2018) demonstrates these data's applicability to a wide range of research fields, including recreation, crisis management, and environment studies. The reasons for the growing adoption of geospatial data vary from the emergence of geographic datasets of unprecedented size and granularity (Elwood 2010) to the transformation of citizens into geospatial subjects able to produce and employ geospatial data (Wilson 2011). Their use is amplified by innovative possibilities for identifying and

---

M. Makhortykh (✉)  
University of Bern, Bern, Switzerland  
e-mail: [mykola.makhortykh@ikmb.unibe.ch](mailto:mykola.makhortykh@ikmb.unibe.ch)

mapping spatial relationships enabled by artificial intelligence and big data (VoPham et al. 2018).

Russia is not an exception from this trend as shown by the increasing number of studies applying geospatial data to study subjects varying from electoral fraud (Kobak et al. 2016) to Silk Road tourism (Tikunov et al. 2018) to Second World War remembrance (Bernstein 2016). Yet, the use of geospatial data in the context of Digital Russian Studies has its own specifics attributed both to the general role of digital media in Russia's media ecologies and to the particular importance of geoweb in this geopolitical context. The explosive growth of Internet use in Russia in 2000s has led to profound changes in the language and communication in multiple domains, including politics (Gorham et al. 2014). The importance of the digital sphere increased even further since the beginning of the Ukraine crisis in 2014, which marked the unprecedented level of state-sponsored cynicism toward the media sphere and its growing instrumentalization for propaganda and disinformation (Roudakova 2017). In this "post-truth" (Surowiec 2017) environment, geolocation data that allow to (dis)prove the existence of specific phenomena emerge as a pivotal factor for making and refuting knowledge claims (e.g. about the presence of Russian troops in Ukraine (Shim 2018)).

To further contextualize the features of Russian geoweb and examine how recent studies address opportunities and challenges provided by it, I will start by reviewing different sources of geospatial data available in the Russian context, varying from social media platforms to crowdsourced databases. I will then move toward discussing possible ways of extracting location information; these ways vary from mapping location names provided through metadata to specific geographic coordinates to extracting location from verbal or visual texts or inferring it from users' activity on social media. Then, I will explore different ways to use geospatial data, such as mapping spatial distribution of socioeconomic phenomena and analyzing mediatization of cultural practices. Additionally, I will briefly discuss the ethical aspects of some of these uses, in particular privacy-related issues. Finally, I will conclude by recapping the main arguments of the chapter and scrutinizing possible directions for future uses of geospatial data in Digital Russian Studies.

## 32.2 DATA ACQUISITION

The first question to address in research using geoweb analysis is what kind of geospatial data is to be used. As I mentioned in the introduction, the distribution of location tracking devices and geographic crowdsourcing gave rise to multiple platforms dealing with geospatial data; however, the format, scope, and quality of these data vary significantly depending on the platform. To illustrate these differences, I will review below three categories of geospatial data sources, which are of particular relevance for Digital Russian Studies: crowdsourced databases, open datasets, and social media.

### 32.2.1 *Crowdsourced Databases*

The availability of digital technology allowing to collect, visualize, and share geospatial data led to the emergence of multiple projects focused on crowdsourcing “volunteered geographic information” (Goodchild 2007). Unlike established sources of geographic information (e.g., open datasets produced by national mapping agencies), crowdsourced databases rely on the assumption that geospatial content produced and edited by multiple individuals will eventually converge on a consensus (Elwood et al. 2012, 575). While this assumption does not guarantee the same quality of data as in the case of sources produced by certified experts, crowdsourced projects are able to account for attributes which are usually omitted by traditional mapping agencies and capture fast-changing phenomena (e.g., natural disasters).

The scope and focus of volunteered geographic projects vary significantly. Some of them, such as Open Street Map (OSM) (<https://www.openstreetmap.org>), HERE Maps (<https://mapcreator.here.com/>), or Yandex People’s Map (<https://n.maps.yandex.ru/>), pursue the goal of creating and sustaining free digital maps or gazetteers. Other projects have limited temporal and thematic focus. Both in Russia and in the West,<sup>1</sup> the latter projects often arise as part of the volunteered reporting in the context of natural disasters<sup>2</sup> or armed conflicts.<sup>3</sup>

Both categories of crowdsourced databases can be of use in the context of Digital Russian Studies. Many global initiatives provide relevant geospatial information, which can be used for Russia-centered research. For instance, Quinn and Tucker (2017) used OSM and Wikimapia (<https://wikimapia.org/>) to trace how crowdsourced maps are used to represent disputed areas such as Crimea and found substantial differences in the ways geopolitical disagreements were visualized and addressed. These differences were attributed to the OSM hosting more contributions from Western editors, whereas Wikimapia was more eager to transmit the Russian official discourse. Other examples include the study by Kulakov, Petrina, and Pavlova (2016), who used Wikimapia for evaluating digital smart services utilized for cultural heritage tourism planning, and the research by Karbovskii et al. (2014), who employed Wikimapia for simulating the process of decision making based on 2012 Krymsk flooding.

Additionally, the Russian digital landscape features a number of crowdsourced projects dealing with specific domains or topics. Despite their variety and rich data, these projects have so far received limited acknowledgement in academic scholarship. A few exceptions include, for instance, *Pomnite nas* (Remember Us) (<http://www.pomnite-nas.ru/>), a project devoted to collecting geospatial data about Second World War monuments devoted to Soviet soldiers (Bernstein 2016). Another example is *RosYama* (Russian Pit) (<https://rosyama.ru/>), a civic project initialized by Alexei Navalny, a Russian anti-systemic opposition leader and activist, who created an online crowdsourced service for reporting road potholes (Ermoshina 2014). Many of these projects are not necessarily designed as sources of geolocation data for academic research

and, instead, intended to facilitate social activities (e.g. collective remembrance of the Second World War in the case of *Pomniti nas*). Despite these non-academic goals, these projects can still be a valuable asset to the researcher who would creatively approach their data. For instance, geolocation data offered by *RosYama* can be used not only for research focused on the quality of Russian roads but also for visualizing geographic networks of activists or detecting the misappropriation of funds planned by specific regions for repairing the roads (for more projects like this, see Chap. 8).

The major challenge of using crowdsourced databases is related to the quality of data provided through them. Because of the lack of authoritative control over their content, the possibility of encountering errors or conscious distortions of geographic facts is higher than in the case of open datasets. In the larger crowdsourced databases such as Wikimapia or Yandex's People Map, such probability is lower because of the large number of contributors, which leads to faster error correction. The situation with small databases is more challenging: often, these projects are curated by small groups of users with limited time and financial resources. While the data offered by them can still be valuable (or even unavailable by other means), it is important to critically assess their quality and identify (as much as possible) who contributes to the database and for what ends.

### 32.2.2 *Open Datasets*

Besides the rise of volunteered geographic initiatives, the unprecedented ease of accumulating and sharing geospatial data resulted in the distribution of open datasets produced by certified actors such as state institutions and mapping agencies. Generated using authoritative geographic sources, these datasets are characterized by higher data quality when compared with crowdsourced databases. While the turn toward open data that are made available through official portals (for instance, [data.gov](http://data.gov) or [europeandataportal.eu](http://europeandataportal.eu)) originated in the West, where these datasets are often employed in academic research on the subjects varying from earthquakes to government institutions' budgets (Ding et al. 2010; Shadbolt et al. 2012), Russia increasingly joins the open data movement.

A number of Russian official agencies make their data available through online portals, such as Russian Open Data Portal (RODP) ([data.gov.ru](http://data.gov.ru)) or Open Data Portal of Moscow City Government ([data.mos.ru](http://data.mos.ru)) (Bundin and Martynov 2015; Koznov et al. 2016; Repponen 2018). A selection of Russian portals, where open datasets are published, is provided in Table 32.1. Despite being subjected to a number of drawbacks, including often limited data pre-processing, absence of unified data standards for different organizations, and the lack of application programming interfaces (APIs) (fifin 2017), these portals provide access to a variety of unique geospatial datasets from different domains, varying from culture (e.g., the dataset on the geospatial distribution of places related to Russian poetess Anna Akhmatova in Moscow [Data.gov



**Table 32.1** Open datasets in Russian geoweb

<i>Dataset name</i>	<i>Web address</i>	<i>Description</i>	<i>English interface</i>
Russian Open Data Portal	<a href="https://data.gov.ru/">https://data.gov.ru/</a>	Collection of datasets related to Russian state agencies	Yes
Open Data Portal of Moscow City Government	<a href="https://data.mos.ru/">https://data.mos.ru/</a>	Collection of datasets produced by the Moscow City Government	Yes
Open Data Portal of Russian Ministry of Culture	<a href="https://opendata.mkrf.ru/">https://opendata.mkrf.ru/</a>	Collection of datasets produced by the Russian Ministry of Culture	No
Open Data Hub	<a href="https://opengovdata.ru/">https://opengovdata.ru/</a>	Collection of datasets related to Russian state agencies, commercial organizations, and non-governmental organizations (NGOs)	Yes

2016]) to crime (e.g., data about the number of committed, resolved, and unresolved crimes by region in Russia [Data.gov 2014]; for more on government data, see Chap. 23).

Two platforms which are of particular interest in this context are Russian Open Data Portal (RODP) and Open Data Hub (ODH). Both platforms provide a large number of datasets (22,233 for RODP and 8151 for ODH) from multiple Russian organizations (1102 and 42 organizations, respectively). These organizations vary from the federal organizations (e.g., the Ministry of Justice or the Federal Statistics Service) to the local ones (e.g., Tomsk Oblast administration). Not all of these datasets deal with geospatial information, but many of them do and can serve as a valuable source of data for geospatial research.

### 32.2.3 *Social Media*

As noted in other chapters of the handbook (see Chapters 20 and 30 on social media use in the context of Digital Russian Studies), social media platforms constitute a major source of digital data. Geospatial data are not an exception as the majority of social media platforms provide in one form or another information about the location of their users and/or content produced. Stock (2018) notes that the majority of studies focus on a few Western platforms, such as Twitter and Flickr, which have accessible APIs and contain geotagged content.<sup>4</sup> This combination allows both identifying the location in which some content available through the platforms is produced and also searching and retrieving data for the specific geographic range (e.g., for collecting messages and images produced within recreational areas to trace visitors' numbers [Tenkanen et al. 2017] and behavior [Sessions et al. 2016]).

In addition to Western social media platforms, Russian geoweb includes several major local platforms, such as VK (also known as VKontakte),

Odnoklassniki, and Moj Mir. Among these platforms, however, only VK provides easy access to its API, which allows retrieving a wide range of geospatial data (Tikunov et al. 2018). Specifically, VK API includes a number of functions also known as methods, which can be used for data extraction (for more on social networks, see Chaps. 19 and 30).<sup>5</sup>

The most common type of geospatial data provided by VK is the one on the country and the city/town of residence, which constitutes part of user profile (Zamyatina and Yashunsky 2018). In the case of publicly available profiles, these data can be retrieved using `users.get` method. The method takes as its input user ids which are of interest for the researcher and the list of fields that have to be retrieved (“country” and “city” are a common choice). These data can be further enriched and/or verified via other profile fields available on VK such as the ones on employment and education.

Besides data available as part of user profiles, VK also provides access to check-in data, which can be retrieved via `places.getCheckins` method. The method takes as input latitude and longitude coordinates and returns posts made within the specified area together with ids of users who published them. Similarly, VK allows retrieving images uploaded by users together with these images’ geographic coordinates using `photos.get` method. The method returns geographic coordinates of retrieved images if these coordinates are provided by the user. Using this method, it is possible to retrieve a sample of images from specific geographic regions in order to, for instance, examine the ways in which these regions are represented online (Tikunov et al. 2018).

### 32.3 LOCATION EXTRACTION

After choosing the specific data source(s) and acquiring actual data, the next step is to process these data. In the case of geospatial data, the major purpose of processing involves the extraction of specific location(s) to which the data refer to or represent. Depending on the data format and available metadata, the process of location extraction can be as simple as retrieving exact geotags present in the metadata or mapping the location name to data from a geographic information system. In other cases, it can be more complex and involve the use of machine learning techniques to recognize the names of geographic entities in visual or verbal texts or to infer the location based on online user activity.

*Geographic coordinates extraction from document metadata.* The easiest—and most common (Stock 2018)—way of detecting location is by using geographic coordinates included in the document (meta)data. Such an approach is particularly applicable for data available from open datasets as well as crowd-sourced databases, which often include specific geographic coordinates. Additionally, some platforms such as Twitter and VK provide geographic coordinates for some types of their content.<sup>6</sup> The question of validity of these data, however, is an open one: especially in the case of geotagged content from social

media platforms, there is also a need to differentiate between the place in which the content was published and the place to which it actually refers.

*Location name extraction from document metadata.* In the cases when geographic coordinates are not provided, one of the alternatives is to extract place names from the metadata. This process usually consists of two steps: (1) toponym recognition: that is, identification of the toponym in the body of the metadata (Sagcan and Karagoz 2015), and (2) toponym resolution: that is, assigning of geographic coordinates to the recognized toponyms (Lieberman and Samet 2012). An example of the platform for which this approach can be highly beneficial is VK, which allows users to report their place of residence in their profiles. While the platform itself does not connect these data to a geographic information system, the location names can be retrieved via VK API and then connected to a geocoding service (e.g., Google Maps) to generate geographic coordinates (Lee et al. 2013; Baucom et al. 2013).

The most popular approach to location name extraction from the metadata is the gazetteer-based one, where the extracted location names are matched with the list of geographic named entities such as the ones provided by GeoNames (<https://www.geonames.org/>). Because of the limited number of gazetteers for the Russian language, such lists are often taken from Wikipedia or from a few training datasets such as FactRuEval (Starostin et al. 2016). At the same time, this approach suffers from a number of issues, including, for instance, intended or unintended mispronunciation (such as *Maskva* instead of *Moskva*) or instances of double naming (e.g. *Sankt-Peterburg* and *Piter*). To address these limitations, more complex approaches were proposed (for reviews, see Leidner 2007; Leidner and Lieberman 2011); a recent study comparing different approaches to the task indicates that approaches using lexical context of toponyms and their importance (e.g., by solving typonym-related ambiguity by always preferring options with the largest population) perform particularly well (Weissenbacher et al. 2019).

*Location name extraction from raw text.* This approach is similar to the location name extraction from document metadata and involves the same two steps: toponym recognition and toponym resolution. However, unlike the former approach which relies on the document's metadata, the latter one takes as input raw text data. Stock (2018, 219) notes that a major benefit of this approach is that it can be used for any text-based message (e.g. photo/video descriptions or blog posts). This approach tends to be less accurate than the one relying on supplied geotags, especially as geographic names are often ambiguous. However, it is often the only way to extract location in the cases when geographic coordinates are not provided.

The usual way of extracting location from raw texts employs the named entity recognition approach: that is, automatic detection of the words which refer to certain geographic locations. The process of detection is based on named entity recognition tools, such as Stanford or GATE, which combine machine learning techniques with pre-made geographic gazetteers, such as

GeoNames or OpenStreetMap (Stock 2018, 220; for practical examples see Jaiswal et al. 2013; Inkpen 2016; Bassi et al. 2016).

While most of the research on named entity recognition approach is tailored to the English language, in recent years the growing number of works employs this technique for the Russian context.<sup>7</sup> Because of the limited number of pre-made Russian gazetteers, a number of studies (see, for instance, Sysoev and Andrianov 2016) employ Wikipedia as a source of information. Additionally, there are several training datasets which include geographic data. An example of such a dataset is FactRuEval, an open annotated corpus of Russian texts.<sup>8</sup> The paper by Ivanitskiy et al. (2016) discusses in more details how FactRuEval can be used for geographic named entity retrieval from Russian sources.

*Location inference from user activity.* In some cases, the documents in question do not provide explicit references to the geographic entity; however, even under these conditions, it is still possible to infer the location based on earlier user activity. Jurgens et al. (2015) summarizes several approaches based on user networks which can be applied for dealing with this task. The majority of these approaches involve identification of users sharing the closest connections with the user in question and then using data from them to infer the user's location.

Another approach is based on content produced by the user online. A number of studies (Cheng et al. 2010; Chang et al., 2012; Han et al. 2014) discuss the possibility of inferring geographic location from local terms also known as location indicative words (LIWs) (Han et al. 2014). LIWs are terms which are particularly representative for specific places, either because of being indicative of certain locations (e.g. "rockets" for Houston) or language practices (e.g. "howdy" for Texas). Consequently, LIWs can be used to predict the location of a user who uses these terms through machine learning techniques.

Several studies (Han et al. 2014; Mourad et al. 2017) apply the latter approach to detect location based on Russian LIWs. The main idea behind it is to acquire textual data produced by users at certain geographic locations (Twitter was used in the above-mentioned studies, but the same principle can be employed for Instagram or VK) and then create separate text corpora for each location in question. Then, for each location LIWs are extracted and the model is trained. Han et al. (2014) offer a detailed discussion of different approaches toward LIWs extraction and show that information gain ratio approach provides the best performance.

*Location name extraction from image.* While location extraction from images is more challenging than from textual data, several techniques allow addressing this task. The first of them is based on the use of geographic information, in particular geotags, embedded in the image metadata. Usually provided in EXIF format (Stock 2018, 222), these metadata are created by the camera and include data about the image creation date, camera settings, and geolocation. Some platforms, such as Flickr, provide API access to these metadata, thus allowing to search these platforms' contents for images from specific areas and specific time span (McDougall and Temple-Watts 2012).

The second technique can be employed in the cases where no metadata is provided and involves the comparison of image similarity. Stock (2018, 222) identifies a number of approaches used to address this task, varying from the use of scale-invariant feature transformation (SIFT) for comparing selected image features (Crandall et al. 2009) to color and texture histograms employed in the domain of computer vision (Gallagher et al. 2009). After identifying these features for the image in question, they can then be compared with large image datasets (e.g., coming from Flickr) to identify similarities.

*Location name extraction from video.* Similar to location extraction from image, several other major approaches for location extraction can be identified. The first of them involves the use of video metadata (e.g., geographic coordinates produced by Global Positioning System [GPS] and compass sensors, which are embedded into video descriptions). This information can be used to identify the region in which the video was produced. Then geoinformation services (e.g., OSM) can be used to extract data about visible objects in the region (e.g., monuments or office buildings) in 2D or 3D.<sup>9</sup> Using OSM data, the descriptive tags can be generated for different objects in the area (e.g., their addresses and names), and then the object models can be compared with objects from the videos. Then, the relevance of each tag for specific video frame is calculated (i.e., to detect if a specific tag is present or absent on the frame) (Shen et al. 2011). While currently there are no papers applying this approach to the Russian context, such an approach is language-agnostic and can be implemented for any video independently of the language in which it is produced, until there is some metadata available.

The second approach can also be employed in the cases where no video metadata is present and combines audio and visual features of videos for identifying the location shown in them. For this purpose, a geotagged collection of videos is required; this collection is then used for calculating the audiovisual similarity with non-geotagged content. Specifically, visual frames and soundtrack are extracted from the videos, and then visual and acoustic features are computed for each one of them. Following the extraction, k-nearest neighbor algorithm (a classification algorithm, which classifies the unknown objects according to the classes of k closest neighbors) is employed to identify geotagged videos which look and sound more similar to the non-geotagged content (Sevillano et al. 2015).

### 32.4 LOCATION USE

After the location is extracted and identified, it can be used for actual analysis. As I noted earlier, the advantage of geospatial data is their versatility and applicability for addressing a wide range of research questions. In this section, I scrutinize some of the uses of geoweb in the context of Digital Russian Studies, from mapping the spatial distribution of phenomena and specifying actors' identities and relationships to scrutinizing the role of location in online cultural practices.

*Mapping the spatial distribution of phenomena.* An important feature of using geospatial data is its rich potential for mapping socioeconomic and (geo) political phenomena. These phenomena vary from tourist mobility (e.g., spatial and temporal dimensions of tourist flows [Lu and Stepchenkova 2015; Kirilenko and Stepchenkova 2017]) to electoral fraud during Russia's federal elections (Kobak et al. 2016) and migration patterns (Zamyatina and Piliarov 2013). Geotag data can be also used for mapping contested phenomena, when official reports are often subjected to censorship or disinformation, such as the involvement of Russian troops in the conflict in Eastern Ukraine based on Instagram data (Czuperski et al. 2015). While the use of geospatial data for studying such contested cases often raises multiple concerns (e.g., concerning the reproducibility and the quality of available data), it can still provide valuable insights for researchers.

*Specifying actor identities and relationships.* Another common use of geospatial data is for identifying specific actors and tracking connections between them. Such tasks are particularly common for studies in political communication and/or disinformation online: for instance, Zelenkauskaitė and Balducci (2017) used geospatial data to specify the origins of users commenting on Russian language news portals in Lithuania, whereas Helmus et al. (2018) employed geoweb to track the identities of users involved in Russian propaganda and counter-propaganda efforts on Twitter. Disinformation, however, is not the only subject which can be investigated in this context as shown by Smirnov et al. (2016), who used geospatial data for identifying friendship networks between youngsters on VK.

*Scrutinizing digitization of cultural practices.* The use of geospatial data increasingly becomes part of the mediatization of cultural practices, varying from war remembrance to tourism. Bernstein (2016) in his research on Second World War memory in Russia showed how the formation of a geotagged database of Soviet monuments enriches existing memory practices by producing virtual embodiments of existing memorials and re-iterating the mainstream Soviet narrative of the war. Another example is the use of geotagged images as part of sharing—and shaping—travel experiences as shown by several studies focused on the use of geospatial information to examine vacation culture in Russia (Kirilenko and Stepchenkova 2017; Tikunov et al. 2018).

*Exploring identity narration.* Besides extensive possibilities for tracking phenomena, digital platforms also enable new ways of (re)-imagining individual and collective identities. A number of studies (Stefanidis et al. 2013; Croitoru et al. 2015) suggest that geospatial data can serve as a strong identifier of group belonging and individual self-expression. Examples of such identifications are, for instance, elements of individual user profiles on Wikipedia, where userboxes are employed for declaring individuals' interests, preferences, and personal details (Neff et al. 2013). In the context of Digital Russian Studies, these means of self-expression often deal with geospatial data (e.g., place of residence [Dounaevsky 2014]) or geopolitical aspects of territoriality (e.g., belonging of the Southern Ossetia to Georgia). Another example is the use of geolocation

data for producing digital maps of the conflict in Eastern Ukraine (e.g., MilitaryMaps or Liveuamap), which are used to visualize the borders of imagined communities (e.g., of the self-declared confederation of Novorossiia [Makhortykh 2018]).

### 32.5 GEOSPATIAL DATA AND RESEARCH ETHICS

The advent of big data research opens unprecedented possibilities for studying different phenomena, but it also raises multiple ethical concerns. Some of these concerns are related to the general considerations of using big data for research purposes (e.g., acquiring proper permissions for data use [Richards and King 2014]), but some are rather specific for geospatial data, in particular in the Russian context. In this pre-final section, I will briefly discuss three of these concerns: validity, privacy, and reliability.

*Privacy.* Security and privacy are two key concerns of using geospatial data for research purposes (Li et al. 2016). The use of portable GPS receivers in mobile devices together with the enrichment of social media data with geospatial information raise concerns about the use of these data for tracking individuals' actions and movements (Loebel 2012). While such data can be beneficial for many types of research, their use also requires the researcher to recognize the potential consequences for the privacy of users. Such consequences are particularly important in cases dealing with highly sensitive and/or polarizing subjects, where the use of geotag data can cause material or immaterial harm for research participants.

The privacy risks are even greater when geotag data is used for studying phenomena occurring in authoritarian states. An example of a highly privacy-sensitive subject is research on anti-government protests, where geospatial data can be (ab)used to identify the location of individual protesters and expose their involvement in the protests, thus bringing legal repercussions by the state. To address this concern, the use of personal data should be minimized and (pseudo)anonymization techniques should be used. On the official level, however, Russian legislation is still catching up with the notion of big data and their uses for research purposes (for an overview, see Zharova and Elin 2017). Consequently, the protection of the data rights of individuals in Russia is still significantly less strict than in the European Union (EU) countries, where it is regulated by the EU General Data Protection Regulation (GDPR).

*Validity.* Sheppard (2005, 74) defines validity as the degree to which the use of a specific instrument or finding is sound, defensible, and well-grounded for the issue at hand. The question of validity is of particular relevance for the use of geospatial data, because of their significant potential for being used for manipulation: both through the data and their visualization (Sheppard and Cizek 2009). In some cases, the use of data can be invalidated by their wrong interpretation (i.e., when geospatial information is used to prove a point which is incorrect), whereas in other cases obscure visualizations of data can mislead the public.



An example of the invalid use of geographic data is the contrasting reporting of the 2018 clashes near Chigari village in Eastern Ukraine. Both the Ukrainian authorities and pro-Russian insurgents produced video records showing them controlling certain landmarks, which were claimed to be related to the village in question. Despite these claims, not all of the shown landmarks were related to Chigari and eventually it was proven that the village was controlled by the Ukrainian army, but not before causing significant confusion. A possible way of increasing validity according to Sheppard and Cizek (2009, 2112) is to use more flexible and interactive approaches for geospatial data analysis, thus allowing end users more control over results' reporting.

*Reliability.* Sheppard (2005) argues that reliability is another major concern of using geospatial data. Unlike validity, which focuses on the possible (ab)uses of geospatial data for drawing invalid conclusions, reliability concerns the internal consistency of analysis and the possibility to produce the same results under similar conditions. The issue of reliability is of particular importance for analyses produced via crowdsourced databases and social media as both data sources are subjected to frequent changes and often provide limited possibilities for consistent data access.

An example of reliability issues which accompany the use of geospatial data is MilitaryMaps mentioned earlier. This crowdsourced database aggregates updates from conflicts in the post-Soviet space as well as in the Middle East and provides geotags indicating the movement of troops and outbursts of violence. From September 2018, however, the previously open project switched toward paid subscription, which made it harder to recreate analyses based on MilitaryMaps data. Another reliability-related limitation of the project is its reliance on the GoogleMaps framework, which stores markers that are added to the map only for a one-year period. Sheppard and Cizek (2009) suggest that the main way to amend these and other reliability issues is the use of more prescriptive approaches to data analysis and presentation based on recognized quality standards.

## 32.6 CONCLUSIONS

In this chapter, I discussed the possible uses of data available through geoweb, the integrated and discoverable collection of geographically related web services and data (Lake and Farley 2009), in the context of the Digital Russian Studies. Increasingly employed for academic studies worldwide, geoweb data are of particular importance for Russia-centered digital research, serving both as a pivotal factor for making and verifying knowledge claims by regional actors and an integral means of producing individual and collective narratives on subjects varying from international conflicts (Shim 2018) to presidential elections (Kobak et al. 2016).

The use of geoweb for Digital Russian Studies is facilitated by the large volume of geospatial data available today. As I discussed above, these data can be divided into three broad categories according to their source: (a) crowdsourced databases, (b) open datasets, and (c) social media. Out of these three, social

media data are the hardest to get and often require extensive pre-processing; however, they are also applicable to a wide range of research questions, in particular the ones related to inter-user interactions. Furthermore, the largest Russian social media platform, VK, provides public access to multiple forms of geospatial data (e.g. users' self-declared place of residence/work and check-in data), thus enabling more possibilities for data collection than many Western platforms.

The research possibilities provided by geospatial data are amplified by the quickly developing toolkit of analytical techniques used to extract geographic location from different data formats. The complexity of techniques varies depending on the data format. In the simplest scenarios, geographic coordinates or the location's administrative address are provided in the metadata and only has to be matched with data from existing geographic information systems. In the more difficult scenarios, the location has to be extracted from the content or inferred from the user's earlier activity using a combination of machine learning and geographic gazetteers. Much still can be done to better adapt these techniques to the Russophone context, in particular in terms of improving named entity recognition techniques and developing better gazetteers. Yet, even in the current state of research, there are plenty of possibilities for using the mentioned techniques for different types of Russia-centered studies.

The importance of location extraction techniques is exemplified by the wide range of research questions to which Russian geospatial data are applicable. These research questions vary from the spatial distribution of socioeconomic and political phenomena, such as migration and electoral fraud, to the verification of knowledge claims about the presence of Russian troops in Eastern Ukraine to the analysis of mediatization of cultural practices of war remembrance and the exploration of narrative uses of geospatial data for communicating individual and collective identities.

Despite their significant potential for Digital Russian Studies, the future of geospatial data is not fully clear. The existing concerns about complex interrelations between privacy and geospatial data are amplified by the current calls for tightening the government's control over the Internet in Russia, leading to increasing restrictions on data retrieval from Russian platforms' APIs, including VK. These limitations might curb the amount of geospatial data available from social media; however, the growing number of open datasets and crowd-sourced databases suggests that Russia's geoweb will remain a valuable research venue for Digital Russian Studies for years to come.

## NOTES

1. For Western examples see Goodchild and Glennon (2010) and Cavelty and Giroux (2011).
2. See, for instance, Virtual Bell (<http://russian-fires.ru/>) project, which was used from 2010 till 2013 to collect reports of wildfires in Russia.

3. An example of such a project is MilitaryMaps (<https://militarymaps.info/>), which crowdsources the recent developments in conflict zones where Russian geopolitical interests are involved.
4. See, for instance, Cheng, Caverlee, and Lee (2010), Chang et al. (2012), Van Canneyt, Schockaert, and Dhoedt (2016), Tenkanen et al. (2017).
5. For more information see the description provided by VK documentation (<https://vk.com/dev/methods>; also available in English).
6. In the case of social media, however, geotagged content is often subjected to numerous limitations, which have to be recognized: in the case of Twitter, only a small number of users (i.e., less than one percent [Cheng, Caverlee, and Lee 2010]) enable geolocation in their profiles, whereas in VKontakte the use of the check-in option is used by the limited number of users and often confined to specific social situations (e.g., tourist visits).
7. See, for instance, Gareev et al. (2013), Sysoev and Andrianov (2016), Anh, Arkhipov, and Burtsev (2017).
8. For the description, see Starostin et al. (2016); the dataset is available in open access at <https://github.com/dialogue-evaluation/factRuEval-2016>.
9. For an overview of tools which can use OSM data to make 3D models see <https://wiki.openstreetmap.org/wiki/3D>.

## REFERENCES

- Anh, Le T., Mikhail Y. Arkhipov, and Mikhail S. Burtsev. 2017. Application of a Hybrid Bi-LSTM-CRF Model to the Task of Russian Named Entity Recognition. In *AINL 2017: Artificial Intelligence and Natural Language*, 91–103. Cham: Springer.
- Bassi, Jonathan, Sukanya Manna, and Sun Yu. 2016. Construction of a Geo-Location Service Utilizing Microblogging Platforms. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, 162–165. Piscataway, NJ: IEEE.
- Baucom, Eric, Azade Sanjari, Xiaozhong Liu, and Miao Chen. 2013. Mirroring the Real World in Social Media: Twitter, Geolocation, and Sentiment Analysis. In *Proceedings of the 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing*, 61–68. New York: ACM.
- Bernstein, Seth. 2016. Remembering War, Remaining Soviet: Digital Commemoration of World War II in Putin’s Russia. *Memory Studies* 9 (4): 422–436.
- Bundin, Mikhail, and Aleksei Martynov. 2015. Russia on the Way to Open Data: Current Governmental Initiatives and Policy. In *Proceedings of the 16th Annual International Conference on Digital Government Research*, 320–322. New York: ACM.
- Chang, Hau-wen, Dongwon Lee, Mohammed Eltaher, and Jeongkyu Lee. 2012. @Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*, 111–118. Piscataway, NJ: IEEE.
- Cheng, Zhiyuan, James Caverlee, and Kyumin Lee. 2010. You are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 759–768. New York: ACM.
- Crampton, Jeremy. 2009. Cartography: Maps 2.0. *Progress in Human Geography* 33 (1): 91–100.

- Crampton, Jeremy W., Mark Graham, Ate Poorthuis, Taylor Shelton, Monica Stephens, Matthew W. Wilson, and Matthew Zook. 2013. Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb. *Cartography and Geographic Information Science* 40 (2): 130–139.
- Crandall, David, Lars Backstrom, Daniel Huttenlocher, and Jon Kleinberg. 2009. Mapping the World's Photos. In *Proceedings of the 18th International Conference on World Wide Web*, 761–770. New York: ACM.
- Croitoru, Arie, N. Wayant, A. Crooks, Jacek Radzikowski, and Anthony Stefanidis. 2015. Linking Cyber and Physical Spaces Through Community Detection and Clustering in Social Media Feeds. *Computers, Environment and Urban Systems* 53: 47–64.
- Czuperski, Maksymilian, John Herbst, Eliot Higgins, Alina Polyakova, and Damon Wilson. 2015. *Hiding in plain sight: Putin's war in Ukraine*. Washington, DC: Atlantic Council of the United States.
- Data.gov. 2014. *Statističeskaja informacija 'Informacija o zaregistrirrovannyh, raskrytyh i neraskrytyh prestuplenijah'* [Statistical data 'Information about Registered, Resolved and Unresolved Crimes']. <https://data.gov.ru/opendata/7727739372mvdgiac38>. Accessed 17 Aug 2019.
- . 2016. *Ahmatovskie mesta v Moskve* [Ahmatova's Places in Moscow]. <https://data.gov.ru/opendata/7708660670-ahmatova>. Accessed 17 Aug 2019.
- Ding, Li, Dominic DiFranzo, Alvaro Graves, James R. Michaelis, Xian Li, Deborah L. McGuinness, and Jim Hendler. 2010. Data-Gov Wiki: Towards Linking Government Data. In *2010 AAAI Spring Symposium: Linked Data Meets Artificial Intelligence*, 38–43. Menlo Park, CA: AAAI.
- Dounaevsky, Helene. 2014. Building Wiki-History: Between Consensus and Edit Warring. In *Memory, Conflict and New Media: Web Wars in Post-Socialist States*, ed. Ellen Rutten, Julie Fedor, and Vera Zvereva, 130–143. London and New York: Routledge.
- Dunn Cavelty, Myriam, and Jennifer Giroux. 2011. Crisis Mapping: A Phenomenon and Tool in Emergencies. *CSS Analyses in Security Policy* 103: 1–4.
- Elwood, Sarah. 2010. Geographic Information Science: Emerging Research on the Societal Implications of the Geospatial Web. *Progress in Human Geography* 34 (3): 349–357.
- Elwood, Sarah, Michael F. Goodchild, and Daniel Z. Sui. 2012. Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers* 102 (3): 571–590.
- Ermoshina, Ksenia. 2014. Democracy as Pothole Repair: Civic Applications and Cyber-Empowerment in Russia. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 8(3). <https://cyberpsychology.eu/article/view/4315/3365>. Accessed 17 Aug 2019.
- filfin. 2017. Počemu otkrytye dannye nikomu ne nužny [Why Nobody Uses Open Data]. July 17. *Habr*. <https://habr.com/post/331036/>. Accessed 17 Aug 2019.
- Gallagher, Andrew, Dhiraj Joshi, Jie Yu, and Jiebo Luo. 2009. Geo-Location Inference from Image Content and User Tags. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 55–62. Piscataway, NJ: IEEE.
- Gareev, Rinat, Maksim Tkachenko, Valery Solovyev, Andrey Simanovsky, and Vladimir Ivanov. 2013. Introducing Baselines for Russian Named Entity Recognition. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, 329–342. Berlin: Springer.

- Goodchild, Michael. 2007. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal* 69 (4): 211–221.
- Goodchild, Michael, and Alan Glennon. 2010. Crowdsourcing Geographic Information for Disaster Response: A Research Frontier. *International Journal of Digital Earth* 3 (3): 231–241.
- Gorham, Michael, Ingunn Lunde, and Martin Paulsen. 2014. Introduction. In *Digital Russia: The Language, Culture and Politics of New Media Communication*, ed. Michael Gorham, Ingunn Lunde, and Martin Paulsen, 1–11. New York: Routledge.
- Haklay, Muki, Alex Singleton, and Chris Parker. 2008. Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass* 2 (6): 2011–2039.
- Han, Bo, Paul Cook, and Timothy Baldwin. 2014. Text-Based Twitter User Geolocation Prediction. *Journal of Artificial Intelligence Research* 49: 451–500.
- Helmus, Todd, Elizabeth Bodine-Baron, Andrew Radin, Madeline Magnuson, Joshua Mendelsohn, William Marcellino, Andriy Bega, and Zev Winkelman. 2018. *Russian Social Media Influence: Understanding Russian Propaganda in Eastern Europe*. Santa Monica: RAND Corporation.
- Inkpen, Diana. 2016. Text Mining in Social Media for Security Threats. In *Recent Advances in Computational Intelligence in Defense and Security*, 491–517. Cham: Springer.
- Ivanitskiy, Roman, Alexander Shipilo, and Liubov Kovriguina. 2016. Russian Named Entities Recognition and Classification Using Distributed Word and Phrase Representations. In *Proceedings of the 3rd Annual International Symposium on Information Management and Big Data*, 150–156. Aachen: CEUR.
- Jaiswal, Anuj, Wei Peng, and Tong Sun. 2013. Predicting Time-Sensitive User Locations from Social Media. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 870–877. New York: ACM.
- Jurgens, David, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*, 188–197. Menlo Park, CA: AAAI.
- Karbovskii, Vladislav, Daniil Voloshin, Kseniia Puzyreva, and Aleksandr Zagarskikh. 2014. Personal Decision Support Mobile Service for Extreme Situations. *Procedia Computer Science* 29: 1646–1655.
- Kirilenko, Andrei, and Svetlana Stepchenkova. 2017. Sochi 2014 Olympics on Twitter: Perspectives of Hosts and Guests. *Tourism Management* 63: 54–65.
- Kobak, Dmitry, Sergey Shpilkin, and Maxim Pshenichnikov. 2016. Integer Percentages as Electoral Falsification Fingerprints. *The Annals of Applied Statistics* 10 (1): 54–73.
- Koznov, Dmitriy, Olga Andreeva, Uolevi Nikula, Andrey Maglyas, Dmitry Muromtsev, and Irina Radchenko. 2016. A Survey of Open Government Data in Russian Federation. In *Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 173–180. Porto: SCITEPRESS.
- Kulakov, Kirill, Oksana Petrina, and Anastasiya Pavlova. 2016. Smart Service Efficiency: Evaluation of Cultural Trip Planning Service. In *Proceedings of the 19th Conference of Open Innovations Association*, 106–112. Piscataway, NJ: IEEE.
- Lake, Ron, and Jim Farley. 2009. Infrastructure for the Geospatial Web. In *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*, Scharl, Arno, and Klaus Tochtermann, eds, 15–27. London: Springer Science & Business Media.

- Lee, Ryong, Shoko Wakamiya, and Kazutoshi Sumiya. 2013. Urban Area Characterization Based on Crowd Behavioral Lifelogs Over Twitter. *Personal and Ubiquitous Computing* 17 (4): 605–620.
- Leidner, Jochen. 2007. *Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names*. Irvine, CA: Universal-Publishers.
- Leidner, Jochen, and Michael Lieberman. 2011. Detecting Geographical References in the Form of Place Names and Associated Spatial Natural Language. *SIGSPATIAL Special* 3 (2): 5–11.
- Li, Songnian, Suzana Dragicevic, Francesc Antón Castro, Monika Sester, Stephan Winter, Arzu Coltekin, Christopher Pettit, et al. 2016. Geospatial Big Data Handling Theory and Methods: A Review and Research Challenges. *ISPRS Journal of Photogrammetry and Remote Sensing* 115: 119–133.
- Lieberman, Michael, and Hanan Samet. 2012. Adaptive Context Features for Toponym Resolution in Streaming News. In *Proceedings of the 35th international ACM SIGIR Conference on Research and Development in Information Retrieval*, 731–740. New York: ACM.
- Loebel, Jens-Martin. 2012. Is Privacy Dead?—An Inquiry into GPS-Based Geolocation and Facial Recognition Systems. In *IFIP International Conference on Human Choice and Computers*, 338–348. Berlin: Springer, Berlin.
- Lu, Weilin, and Svetlana Stepchenkova. 2015. User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. *Journal of Hospitality Marketing & Management* 24 (2): 119–154.
- Makhortykh, Mykola. 2018. *Charting the Conflicted Borders: Narrating the Conflict in Eastern Ukraine Through Digital Maps*. Paper presented at The Power of Borderland(s): In Media's Res. June 28–29, Greifswald.
- McDougall, Kevin, and P. Temple-Watts. 2012. The Use of LIDAR and Volunteered Geographic Information to Map Flood Extents and Inundation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1: 251–256.
- Mourad, Ahmed, Falk Scholer, and Mark Sanderson. 2017. Language Influences on Tweeter Geolocation. In *Proceedings of the European Conference on Information Retrieval*, 331–342. Cham: Springer.
- Neff, Jessica J., David Laniado, Karolin E. Kappler, Yana Volkovich, Pablo Aragón, and Andreas Kaltenbrunner. 2013. Jointly They Edit: Examining the Impact of Community Identification on Political Interaction in Wikipedia. *PLOS One* 8 (4). <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0060584>. Accessed 17 Aug 2019.
- Quinn, Sterling, and Doran A. Tucker. 2017. How Geopolitical Conflict Shapes the Mass-Produced Online Map. *First Monday* 22 (11). <https://journals.uic.edu/ojs/index.php/fm/article/view/7922/6558>. Accessed 17 Aug 2019.
- Repponen, Ilona. 2018. Linked Ontology on Metadata on Russian Open Governmental Data. <http://urn.fi/urn:nbn:fi:csc-kata20181221153713379920>. Accessed 17 Aug 2019.
- Richards, Neil, and Jonathan King. 2014. Big Data Ethics. *Wake Forest Law Review* 49: 393–432.
- Roudakova, Natalia. 2017. *Losing Pravda: Ethics and the Press in Post-Truth Russia*. Cambridge: Cambridge University Press.



- Sagcan, Meryem, and Pinar Karagoz. 2015. Toponym Recognition in Social Media for Estimating the Location of Events. In *2015 IEEE International Conference on Data Mining Workshop*, 33–39. Piscataway, NJ: IEEE.
- Sansurooah, Krishnun, and Bradley Keane. 2015. The Spy in Your Pocket: Smartphones and Geo-Location Data. In *Proceedings of the 13th Australian Digital Forensics Conference*, 95–103. <https://ro.ecu.edu.au/adf/154/>. Accessed 17 Aug 2019.
- Sessions, Carrie, Spencer A. Wood, Sergey Rabotyagov, and David M. Fisher. 2016. Measuring Recreational Visitation at US National Parks with Crowd-Sourced Photographs. *Journal of Environmental Management* 183: 703–711.
- Sevillano, Xavier, Xavier Valero, and Francesc Alías. 2015. Look, Listen and Find: A Purely Audiovisual Approach to Online Videos Geotagging. *Information Sciences* 295: 558–572.
- Shadbolt, Nigel, Kieron O’Hara, Tim Berners-Lee, Nicholas Gibbins, Hugh Glaser, and Wendy Hall. 2012. Linked Open Government Data: Lessons from data.gov.uk. *IEEE Intelligent Systems* 27 (3): 16–24.
- Shen, Zhijie, Sakire Arslan Ay, Seon Ho Kim, and Roger Zimmermann. 2011. Automatic Tag Generation and Ranking for Sensor-Rich Outdoor Videos. In *Proceedings of the 19th ACM International Conference on Multimedia*, 93–102. New York: ACM.
- Sheppard, Stephen. 2005. Validity, Reliability, and Ethics in Visualization. In *Visualization in Landscape and Environmental Planning: Technology and Applications*, ed. Ian Bishop and Eckart Lange, 79–97. London: Taylor and Francis.
- Sheppard, Stephen, and Petr Cizek. 2009. The Ethics of Google Earth: Crossing Thresholds from Spatial Data to Landscape Visualisation. *Journal of Environmental Management* 90 (6): 2102–2117.
- Shim, David. 2018. Satellites. In *Visual Global Politics*, ed. Roland Bleiker, 265–271. New York: Routledge.
- Smirnov, Ivan, Elizaveta Sivak, and Yana Kozmina. 2016. In Search of Lost Profiles: The Reliability of VKontakte Data and its Importance in Educational Research. *Voprosy Obrazovaniya* 4: 106–122.
- Starostin, Anatoli, Dmitry Granovsky, Svetlana Alexeeva, Anastasia Bodrova, Alexander Chuchunkov, et al. 2016. FactRuEval 2016: Evaluation of Named Entity Recognition and Fact Extraction Systems for Russian. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2016”*, 1–19. Moscow: RSUH.
- Stefanidis, Anthony, Andrew Crooks, and Jacek Radzikowski. 2013. Harvesting Ambient Geospatial Information from Social Media Feeds. *GeoJournal* 78 (2): 319–338.
- Stock, Kristin. 2018. Mining Location from Social Media: A Systematic Review. *Computers, Environment and Urban Systems* 71: 209–240.
- Surowiec, Pawel. 2017. Post-Truth Soft Power; Changing Facets of Propaganda, Kompromat, and Democracy. *Georgetown Journal of International Affairs* 18 (3): 21–27.
- Sysoev, Andrey, and Ivan Andrianov. 2016. Named Entity Recognition in Russian: the Power of Wiki-Based Approach. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2016”*, 1–10. Moscow: RSUH.
- Tenkanen, Henriikki, Enrico Di Minin, Vuokko Heikinheimo, Anna Hausmann, Marna Herbst, Liisa Kajala, and Tuuli Toivonen. 2017. Instagram, Flickr, or Twitter:



- Assessing the Usability of Social Media Data for Visitor Monitoring in Protected Areas. *Scientific reports* 7 (1): 1–11.
- Tikunov, Vladimir S., Vitaly S. Belozarov, Aleksander N. Panin, and Stanislav Antipov. 2018. Geoinformation Monitoring of Key Queries of Search Engines, and Geotagging Photos in the North-Caucasian Segment of the Tourist Route 'Great Silk Road'. *Annals of GIS* 24 (4): 255–260.
- Van Canneyt, Steven, Steven Schockaert, and Bart Dhoedt. 2016. Categorizing Events Using Spatio-Temporal and User Features from Flickr. *Information Sciences* 328: 76–96.
- Vopham, Trang, Jaime E. Hart, Francine Laden, and Yao-Yi Chiang. 2018. Emerging Trends in Geospatial Artificial Intelligence (GeoAI): Potential Applications for Environmental Epidemiology. *Environmental Health* 17 (1): 1–6.
- Weissenbacher, Davy, Arjun Magge, Karen O'Connor, Matthew Scotch, and Graciela Gonzalez. 2019. Semeval-2019 task 12: Toponym Resolution in Scientific Papers. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, 907–916. Stroudsburg, PA: ACL.
- Wilson, Matthew. 2011. 'Training the Eye': Formation of the Geocoding Subject. *Social & Cultural Geography* 12 (4): 357–376.
- Zamyatina, Nadezhda, and Aleksandr Piliarov. 2013. Sever, Social'nye Seti i 'Diaspora Naoborot'. *Demoskop* 547–548. <http://www.demoscope.ru/weekly/2013/0547/analit07.php>. Accessed 17 Aug 2019.
- Zamyatina, Nadezhda, and Alexey Yashunsky. 2018. Virtual Geography of Virtual Population. *Monitoring of Public Opinion: Economic and Social Changes* 143 (1): 117–137.
- Zelenkauskaitė, Asta, and Marcello Balducci. 2017. 'Information Warfare' and Online News Commenting: Analyzing Forces of Social Influence Through Location-Based Commenting User Typology. *Social Media + Society* 3 (3): 1–13.
- Zharova, Anna, and Vladimir Mikhailovich Elin. 2017. The Use of Big Data: A Russian Perspective of Personal Data Security. *Computer Law & Security Review* 33 (4): 482–501.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





# Correction to: The Palgrave Handbook of Digital Russia Studies

*Daria Gritsenko, Mariëlle Wijermars, and Mikhail Kopotev*

## **CORRECTION TO:**

D. Gritsenko et al. (eds.), *The Palgrave Handbook of Digital Russia Studies*,  
<https://doi.org/10.1007/978-3-030-42855-6>

This book was inadvertently published with an incorrect affiliation for the Authors Ekaterina Artemova, Anastasia Bonch-Osmolovskaya, Frank Fischer, Arto Mustajoki, Polina Kolozaridi and Daniil Skorinkin. The affiliations have now been amended in the book.

The correct affiliation is  
Higher School of Economics  
(HSE University)  
Moscow, Russia

---

The updated original online version for this book can be found at  
<https://doi.org/10.1007/978-3-030-42855-6>

© The Author(s) 2021  
D. Gritsenko et al. (eds.), *The Palgrave Handbook of Digital Russia  
Studies*, [https://doi.org/10.1007/978-3-030-42855-6\\_33](https://doi.org/10.1007/978-3-030-42855-6_33)

C1

# INDEX<sup>1</sup>

## A

Academic activity, 485, 490  
Academic corruption, 492  
Academic culture, 484  
Academic genre, 485  
Academic plagiarism, 9, 483–497  
Activism, 87, 135–150, 162, 206, 207,  
209–214, 255, 261, 272, 537  
Affordance analysis, 372  
Amateur literature, 265–267  
Antiplagiat, 498n6  
Archive, 2, 8, 9, 158, 353–366,  
371–385, 391, 392, 438,  
570–572  
Artificial intelligence (AI), 7, 45, 46, 53,  
55, 60, 62, 65, 82, 169n4, 180,  
181, 270, 586  
Authenticity, 353, 366, 374, 375, 484,  
485, 490, 492–497  
Authorship, 166, 245, 246, 485  
Automated indexing, 321  
Automated text analysis, 338, 417  
Automatic document processing, 320,  
322, 324, 328, 331  
Automatic sentiment analysis,  
9, 501–511  
Automatic text categorization, 330  
Average distance, 526, 527

## B

Betweenness, 525, 528, 529  
Big data, 2, 5, 8, 23, 37, 60, 62, 65, 108,  
234, 302, 303, 347, 348, 439, 440,  
586, 595  
Blogger, 19, 21, 118, 124, 208, 209,  
211, 227, 268, 284–286, 343,  
537, 552  
Bot, 85, 118, 124, 139, 143, 270, 291,  
340–342, 538, 548, 549, 557–560  
Bot-generated content, 340

## C

Centrality, 48, 373, 525, 526, 528, 529,  
550, 552  
Character network, 533  
Civil activity, 338  
Civil law, 83, 392  
Clique, 544, 545  
Clique analysis, 545  
Close reading, 124, 427, 433–435, 443,  
455, 520  
Collaborative consumption, 230–233  
Community network, 544  
Complex affordance, 374, 375, 382–384  
Computational analysis, 427, 430,  
436, 438

<sup>1</sup>Note: Page numbers followed by ‘n’ refer to notes.

- Computer-assisted research, 302  
 Computer program, 243, 430, 579  
 Computer-readable text, 438, 439  
 Computer technology, 77, 174, 243, 258  
 Conceptual indexing, 330  
 Connective activism, 149  
 Consumer behavior, 221–234  
 Contemporary art, 160, 242, 244, 250, 251  
 Contextualization, 574, 579  
 Convention hypothesis, 493, 494, 496  
 Corpus, 157, 260, 300–306, 309, 310, 312, 314, 365, 378–380, 415, 420, 421, 433, 437, 448–455, 457, 459, 460, 467, 474, 476, 477, 592  
 Corpus-based analysis, 300  
 Corpus family, 303  
 Corpus-informed analysis, 314  
 Corpus linguistics, 303, 469  
 Correlation coefficient, 552  
 Critical approach, 190, 372  
 Cross-border shopping, 225, 228–230  
 Crowdsourcing, 28, 37, 139, 142, 288, 290, 509, 511, 585–587  
 Cultural heritage, 570, 572, 573, 587  
 Cultural industry, 248  
 Cultural knowledge, 571, 578  
 Cultural vector, 281, 284–285  
 Cyber army, 115  
 Cyberfeminist, 211, 215  
 Cyberfiction, 262, 264, 266, 267  
 Cyberspace, 85, 86, 116, 117, 124, 126–129, 206, 207, 215, 279, 292  
 Cyber superpower, 116, 128  
 Cyrillic domain, 127
- D**
- Data activism, 139, 140, 142, 148  
 Data agenda, 404, 405  
 Data collection, 107, 172, 178, 179, 234, 336, 342, 348, 389, 597  
 Data-driven education, 179  
 Data ecosystem, 395, 396, 404  
 Datafication, 5, 167, 172, 173, 175, 177–181  
 Data infrastructure, 47, 395, 397  
 Data publication, 389, 390, 393, 395, 397–399, 404, 405, 429, 575  
 Dataset, 3, 108, 124, 139, 306, 308, 309, 343, 357, 389–391, 395–400, 402, 409–419, 428, 429, 431, 435–440, 450, 465, 466, 474, 476, 504, 510, 511, 517, 519, 575–577, 580, 585–593, 596, 597  
 Deep learning, 9, 465–479, 508, 580  
 Democratization, 3, 28, 36, 206, 256, 260, 286, 372, 373, 447, 538  
 Diary, 258, 364, 365, 371–385  
 Digital archive, 8, 9, 250–252, 354–366, 366n1, 371–385, 571, 572  
 Digital art, 241–252, 571  
 Digital art history, 569  
 Digital artwork, 242, 244, 252  
 Digital citizenship, 79, 210  
 Digital consumption, 6, 221, 222, 233, 234, 287, 292  
 Digital criticism, 160  
 Digital economy, 53–71, 78, 82–85, 122, 248–250, 445  
 Digital Economy Program, 53–71  
 Digital education, 171, 172, 175, 176, 180, 243  
 Digital era, 5, 157, 160, 169n4, 190, 196, 202, 244  
 Digital form, 2, 37, 56, 157, 187, 242, 249, 518, 572  
 Digital format, 39, 43, 44, 187, 256, 265, 382  
 Digital gender, 205–215  
 Digitalization, v, 2–5, 7, 10, 15–29, 33, 35–46, 48, 54, 57, 63–70, 78–82, 86, 87, 115, 135, 148, 150, 156, 157, 159, 164, 171–182, 187–202, 206, 209, 221–234, 243, 251, 444, 445, 447–449, 453, 454, 456, 457, 460, 484, 485  
 Digital journalism, 155–168, 196, 287  
 Digital literacy, 82, 177, 243, 251  
 Digital literature, 255–272  
 Digital network, 6, 155, 165, 200, 201  
 Digital platform, 47, 59, 61, 78, 79, 82, 180, 181, 192, 195, 198, 200, 210, 215, 279, 287–289, 372, 375, 594  
 Digital revolution, 55, 70, 364

Digital security, 115, 121, 126, 128, 129, 251  
 Digital space, 116, 161, 188, 193, 194, 200, 205–215  
 Digital sphere, 196, 256, 259, 260, 586  
 Digital technology, 5–7, 16, 19, 23, 29, 35, 37, 44, 53, 54, 57, 59–63, 70, 71n5, 135, 136, 138, 148–150, 155, 157, 160–164, 169n16, 171, 172, 174, 175, 178, 181, 182, 187, 188, 200, 202, 206, 207, 210, 230, 242, 244, 245, 248, 257, 365, 372, 447, 585, 587  
 Digital television, 164  
 Digital text, 266  
 Digital text analysis, 439  
 Digital transformation, 2, 4, 35, 53–71, 222, 225, 234, 391, 447  
 Digitization, 9, 17, 34, 37–39, 41–43, 48, 77–87, 271, 353–360, 362, 364–366, 381, 382, 438, 439, 570–574, 578–580, 594  
 Digressive narration, 260, 265  
 Disambiguation, 324, 325, 416, 472  
 Discursive activism, 214  
 Dissnet, 484–486, 490–494  
 Document processing, 78, 320, 322, 324, 328, 331  
 Dogmatic corruption, 199  
 Domain-specific text collection, 505, 508, 510, 511  
 Domestic distribution, 358, 359  
 Drama, 246, 304, 518, 520, 525–528, 533  
 Duma, 18, 21, 83–85, 95–97, 99, 103, 105, 106, 148, 150n3, 151n9, 162, 191

**E**  
 Education datafication, 177, 178, 181  
 Education digitalization, 171–182  
 Education quality control, 179  
 Electoral fraud, 25, 261, 290, 339, 586, 594, 597  
 Email categorization, 473  
 Embedding, 9, 270, 347, 443–461, 466–468, 470–477, 511  
 Embedding space, 451–454, 457

Entity recognition approach, 591, 592  
 European history, 438, 439  
 Experience economy, 227  
 Experimental digital art, 252

**F**  
 Fake news, 199, 340, 469, 470, 478, 502  
 Fan fiction, 129n2, 265–267, 271  
 Federal anti-terrorism, 119  
 Federal executive authority, 393, 394  
 Federal executive body, 65, 118, 392–395, 400, 402  
 Foreign trade, 438  
 Formal approach, 517, 522, 533  
 Friendship network, 342, 594

**G**  
 Gender, 7, 146, 161, 205–215, 228, 247, 250, 258, 341, 342, 344, 379, 420, 453  
 Gender capital, 206  
 Gender identification, 474  
 Geographical city, 344  
 Geographic information, 587, 590, 592, 597  
 Geolocation, 10, 336, 342, 343, 586–588, 592, 594, 598n6  
 Geospatial information, 587, 589, 594, 595  
 Geotag, 590–592, 594–596  
 Geoweb, 10, 585–597  
 Geoweb analysis, 586  
 Global Internet, 2, 6, 69, 117, 126, 127, 259, 261, 267, 270, 271, 300  
 Governance, 2, 7, 8, 17, 33, 36, 37, 48, 49, 61, 63–65, 69, 70, 78, 86, 117, 126, 127, 136, 149, 171–182, 249, 389, 547

**H**  
 Hackathon, 396, 400–402  
 Hashtag activism, 139, 142  
 Hashtag campaign, 213, 214  
 High-tech sector, 66  
 H-index algorithm, 552  
 Historical analysis, 190, 278, 292, 575

Historical experience, 359, 363, 364, 366  
 Historical knowledge, 9, 353–355, 357, 383, 384  
 Historical research, 428, 570, 571, 578  
 Humanities, v, 4–6, 8, 10, 300, 302, 360, 428, 485, 487, 495, 519, 534, 569, 574  
 Hyperfiction, 255, 258, 264–267  
 Hyperlink, 158, 246, 255, 258, 265  
 Hypertext, 159, 246, 256, 257, 259, 264–265, 267, 272, 284  
 Hypertextuality, 159

**I**

Image analysis, 10, 575, 576  
 Inappropriate borrowing, 494  
 Indexing, 167, 319, 321, 327, 330, 377, 381, 382  
 Industrial internet, 60, 62  
 Information exchange, 69, 200, 201, 243  
 Information gain ratio approach, 592  
 Information retrieval (IR), 319–328, 331, 336, 411  
 Instagram, 19, 20, 29n5, 137, 142, 208, 209, 213, 225, 232, 249, 269, 283, 285, 346, 592, 594  
 Institutional subscription, 359  
 Internet access, 1, 16, 17, 22, 257, 287  
 Internet art, 246, 247  
 Internet infrastructure, 116, 120  
 Internet regulation, 115, 119, 289, 292  
 Intertextuality, 574, 578, 579

**J**

Journalistic autonomy, 199  
 Journalistic practice, 155–168  
 Justified borrowing, 496

**K**

Knowledge representation, 320

**L**

Language generation, 468, 470  
 Language model, 468, 469, 472, 478  
 Large-scale digitization, 355, 356, 362

Large-scale plagiarism, 484, 490–492, 494, 497  
 Large-scale research, 341, 347  
 Law entity, 392  
 Learning competence, 172  
 Legal document, 78, 391  
*Lemma*, 305, 310, 555  
 Lemmatization, 302, 411, 414, 430  
 Lesbian, gay, bisexual, and transgender (LGBT), 146–148, 205, 208  
 Lexicon, 211, 323, 487, 505, 507–511  
 Liberalization, 189, 444–448, 454, 457, 460  
 Linguistic behavior, 346  
 Linguistic creativity, 260, 267  
 Linguistic information, 302  
 Linguistic research, 314  
 Linguistics, 8, 9, 127, 188, 211, 212, 260, 266, 267, 300–304, 306, 311, 315n2, 321, 323, 324, 331, 336, 346, 415, 416, 418, 452, 465, 469, 474–476, 478, 487, 495, 506, 508, 539  
 Literary communication, 260, 261, 268, 272  
 Literary network, 271, 517–523, 527  
 LiveJournal, 19, 22, 136, 137, 164, 165, 196, 269, 271, 283–286, 289, 338, 343, 345, 413, 419, 420, 539  
 Location extraction, 590–593, 597  
 Long-term data, 431

**M**

Machine translation, 411, 468, 475  
 Mainstream culture, 249, 282, 285  
 Manual text analysis, 338  
 Market configuration, 335  
 Masculinity, 205–208, 210, 211  
 Mediatization, 190–192, 586, 594, 597  
 Metadata, 103, 108, 119, 305, 336, 397, 438, 439, 572, 573, 575, 579, 586, 590–593, 597  
 Mobilization, 16, 23, 128, 144, 146, 149, 150n2, 151n9, 166, 211, 213, 214, 261, 269, 282, 288–291, 338, 339, 538



- Modeling algorithm, 409, 413, 416, 417, 428–430, 432, 433
- Modernization, 18, 37–40, 43, 44, 55, 68–71, 175, 311–314, 445–450, 453, 455–458, 460
- Modernization agenda, 46, 444–446, 454, 455, 460
- Monograph, 484, 490
- Multiword expression, 305, 320, 321, 323–325, 328, 330, 331
- Museufication, 252
- MyStem, 430
- N**
- Name extraction, 591–593
- National corpus, 303, 304
- National economy, 223
- Natural language, 468, 557
- Natural language processing (NLP), 9, 319–331, 409, 414, 429, 431, 465–478, 501, 511
- Network analysis, 9, 10, 271, 517–520, 522, 523, 529, 533, 534, 539, 540, 551
- Network density, 524, 533
- Network-dominated approach, 550
- Network size, 524–526
- Network structure, 201, 520, 529, 540, 542, 546, 557, 559
- Network visualization, 519, 527, 530–532, 559
- Neural net modeling, 487
- Neural network, 466–471, 478, 508, 511, 577
- Nominal homophily, 553
- Non-geotagged content, 593
- O**
- Odnoklassniki, 2, 142, 198, 224, 268, 283, 285, 337, 339, 539, 590
- Official discourse, 146, 149, 447, 587
- Offline reality, 335, 342–344
- Omnichannel retailing, 226, 227, 234
- Online activism, 135–149, 150n4, 212–214
- Online community, 19, 207, 208, 401
- Online consultation, 26, 27
- Online discussion, 118, 149, 209, 214
- Online religion, 187
- Online retailer, 224, 226, 229
- Online shopping, 221–230, 234
- Ontological dependence, 326–328, 331
- Open data, 18, 289, 389, 390, 393–405, 588
- Open data ecosystem, 395, 396, 404
- Open dataset, 397–399, 402, 586–590, 596, 597
- Open government, 9, 17–20, 29, 78–80, 389–405
- Orthodox Christianity, 188, 193, 198
- P**
- Padonki*, 260, 267, 272
- Paraphrase, 484, 487, 492
- Participatory culture, 257, 266, 270, 281
- Path length, 524
- Patriarchal discourse, 212, 213, 215
- Perestroika, 162, 243, 251, 260, 356, 357, 427
- Personal archive, 359, 360
- Plagiarism, 9, 483–497
- Plagiarism detection, 483–497
- Political activism, 138, 148
- Political agenda, 9, 314, 355, 357, 360, 445
- Political change, 23, 27, 404, 446, 448, 452, 455, 457, 458, 460, 538
- Political communication, 16, 17, 19–21, 29, 270, 460, 540, 594
- Political debate, 136, 137, 269, 443
- Political history, 151n12
- Political innovation, 290
- Political language, 449, 452
- Political leadership, 446, 448, 456, 460
- Political participation, 15–29
- Political program, 23, 149, 311, 446, 447, 453, 456
- Political vector, 281, 282, 289–291
- Post-Internet literature, 258, 268, 272
- Probability distribution, 410, 428, 550
- Protest movement, 20, 23, 26, 116, 261, 290
- Prozbito*, 364, 365, 371–385, 439
- Public administration, 33–35, 39–41, 43–45, 47–49, 60, 63, 393

- Public discourse, 24, 147, 181, 206, 209, 211, 212, 413, 444, 445, 448, 454, 457, 460
- Public spending, 403
- Public sphere, 34, 118, 189, 191, 195, 197, 198, 200, 206, 269, 280, 282, 290, 538, 539, 559
- Public transport, 29n4, 404
- Python, 431, 473, 523
- Q**
- Qualitative analysis, 309, 314, 444, 445, 447, 456, 459, 460, 557, 570, 575, 580
- Quality assessment, 80, 182, 410–412, 417–419, 433
- Quantile, 549–551, 561
- Quantitative text analysis, 450
- Quiet activism, 209, 210
- R**
- Regularization, 413–415
- Regulatory framework, 56, 398, 404, 538
- Religious identity, 189, 193
- Religious life, 187, 189–192, 199, 202
- Religious sphere, 187
- Research community, 336, 344, 478, 550
- Retailing, 225–227, 234, 234n1
- Retweet, 540, 542, 543, 548, 550, 552, 555, 557, 560
- Retweet-dominated approach, 550
- Roskomnadzor, 85, 87n2, 95–109, 118–120, 122, 140, 145–147
- Ruble depreciation, 225
- Runaway object, 280–281, 292
- Runet, 6, 7, 85–87, 106–107, 109, 115–129, 136, 137, 141, 164, 165, 207, 210, 212, 215, 255–272, 277–292, 300–302, 469
- Russian activism, 140–142
- Russian archives, 2, 353, 354, 356–360, 362, 363, 371
- Russian art, 241, 250–252, 569–581
- Russian blogging, 283
- Russian blogosphere, 268
- Russian corpus, 302–303
- Russian culture, 6, 10, 248, 249, 270, 284
- Russian digital history, 439
- Russian education, 171–182
- Russian fan fiction, 129n2, 266, 271
- Russian financial crisis, 223
- Russian Formalism, 259, 270
- Russian geoweb, 585–597
- Russian government, 2, 7, 21, 37–46, 63, 70, 78, 82, 85, 97, 102, 107, 109, 115–118, 120, 122–128, 164, 167, 189, 251, 341, 360, 362, 366, 393, 397, 398, 449, 555
- Russian history, 9, 198, 355, 363–366, 427–440, 528, 570
- Russian Internet, 1, 6, 7, 96, 106, 115–117, 145, 147, 176, 228, 277–292, 336
- Russian language, 6, 8, 102, 123, 136, 161, 211, 212, 256, 269, 271, 272, 277, 280, 284, 285, 301, 302, 310, 311, 320, 337, 338, 345, 347, 348, 359, 411–419, 421, 430, 448, 465–479, 486, 506, 509, 511, 591, 594
- Russian legislation, 83, 96, 104, 105, 107, 108, 595
- Russian market, 101, 161, 224, 229, 233
- Russian media, 156, 158, 161, 164, 169n12, 177, 200, 285, 286, 310, 311, 313, 314, 339, 445, 448, 454, 455, 460
- Russian opposition, 135, 149, 167
- Russian religious landscape, 188–190
- Russian Revolution, 375
- Russian science, 64, 489
- Russian security service, 7, 144–145
- Russian sentiment lexicon, 508, 509
- Russian society, 2, 5, 10, 39, 78, 192, 198, 205, 207, 208, 212, 215, 268, 311, 335, 336, 484
- Russian state, 16, 20, 77, 87, 121, 125, 127, 137, 139, 141, 144, 148, 149, 193, 209, 279, 286, 354, 355, 360, 365, 438

Russian tradition, 197, 198  
 RuThes, 8, 319–331, 509  
 Rutube, 156, 270

## S

Sacralization, 197  
 Samizdat, 255–272, 280  
 Sberbank, 62, 63, 82, 169n12, 176, 177,  
 224, 283, 507  
 Securitization, 87, 117, 123, 129,  
 282, 291  
 Security service, 7, 122, 144–145, 158,  
 165, 169n18, 283  
 Semantic analysis, 409  
 Semantic feature, 306  
 Semantics, 305–307, 309, 321, 325,  
 327, 331, 409, 410, 415, 416, 418,  
 421, 445, 451–454, 456–458, 460,  
 461, 465, 467, 469, 470, 473,  
 474, 487, 511, 575, 577, 578,  
 580, 581  
 Sentiment analysis, 9, 411, 412, 415,  
 416, 466, 472–474, 501–511, 560  
 Sentiment classification, 472, 473  
 Sentiment lexicon, 505, 507–511  
 Sequential affordance, 374, 380,  
 382, 384  
 Sharing economy, 221, 222, 230–234  
 Simple affordance, 374, 375,  
 378–380, 384  
 Small-scale plagiarism, 485, 492–496  
 Social capital, 538, 544  
 Social movement, 116, 149, 151n9,  
 212, 339  
 Social network analysis (SNA), 336, 338,  
 342, 345, 517–534, 538–553, 557,  
 559, 560  
 Social research, 343, 348, 501  
 Source investigation, 167  
 Source text, 486–490  
 Sovereign internet, 95, 120,  
 126–128, 284  
 Sovereignization, 282, 291  
 Soviet history, 366, 384  
 Soviet Internet, 279  
 Soviet journalism, 162

Spatial distribution, 10, 586, 593,  
 594, 597  
 State archives, 359, 364, 380, 384, 438  
 State procurement industry, 403  
 State violence, 357, 358  
 Statistical calculation, 428  
 Statistical inference, 339, 348  
 Statistical information, 394  
 Stop-word list, 429, 431  
 Strategic development, 54, 63, 164

## T

Tax border, 229  
 Technical affordance, 284  
 Technoactivism, 139–141, 144–146, 148  
 Technological development, 3, 6, 64, 65,  
 68, 71, 280, 281, 446, 447, 449,  
 456, 460  
 Technological infrastructure, 233, 234,  
 284, 291  
 Technological revolution, 54, 174  
 Technological vector, 281–284  
 Telecommunication network, 69, 398  
 Term extraction, 416  
 Text analysis, 338, 417, 428, 439, 444,  
 461, 522  
 Text annotation, 416  
 Text classification, 409, 465, 473–476  
 Text generation, 469, 478  
 Text mining, 428, 460  
 Text2vec, 449, 454, 456  
 Textual analysis, 444, 448, 450, 460, 522  
 Textual authenticity, 484, 485, 492–496  
 Textual corpus, 314, 380, 437  
 Thematic relevance, 416  
 Thesaurus, 8, 319–331, 509  
 Thesaurus construction, 323  
 Topicality, 410, 415, 417, 420  
 Topic detection, 410, 415, 421  
 Topic interpretability, 410, 414,  
 417–419, 421  
 Topic model/topic modeling (TM), 9,  
 271, 338, 343, 344, 347, 409–421,  
 427–440, 443–461, 473, 539  
 Topic prevalence, 456  
 Topic quality, 410, 412, 418

Topic saliency, 417, 420  
 Toponym recognition, 591  
 Traditional archive, 366  
 Traditional art, 571, 578  
 Training collection, 330, 508, 510, 511  
 Transformation strategy, 54  
 Transitivity, 325, 327, 330, 544, 545  
 TreeTagger, 431, 434  
 Troll army, 124  
 Tweet, 22, 289, 338–340, 412, 415, 417,  
 503, 507, 510, 511, 540, 547–550,  
 552, 554–558, 562, 563n10, 563n11  
 Twitter, 9, 19, 20, 22, 101–103, 108,  
 120, 128, 142, 149, 161, 200, 213,  
 252, 258, 259, 269, 283, 284, 289,  
 338–341, 412, 417, 419, 503, 507,  
 509–511, 539–544, 546–548,  
 550–554, 557–559, 563n3, 563n12,  
 589, 590, 592, 594, 598n6

## U

Unauthorized borrowing, 485, 495  
 Uniform analysis, 439  
 Urban civic activity, 344  
 User vector, 282, 287–288  
 Utopianism, 162

## V

Vernacular archive, 364–366  
 Virtual census, 341  
 Virtual demography, 336, 337,  
 341–342, 344

Visual analysis, 569, 574–576  
 Visual imagery, 576, 578, 581  
 Visualization, 167, 243, 324, 473, 517,  
 522, 527, 529, 531–533, 540, 542,  
 559, 574, 579, 595  
 Visual propaganda, 577, 579  
 VKontakte (VK), 2, 104, 105, 119, 120,  
 136, 142, 147, 198, 209, 210, 225,  
 232, 252, 268, 269, 283, 285, 286,  
 288, 337–339, 341–347, 413, 539,  
 540, 562n1, 589–592, 594, 597,  
 598n5, 598n6

## W

Weakness hypothesis,  
 493–496  
 Web-based activity, 138  
 Webcam, 25, 26  
 Wikimapia, 587  
 Word embedding model, 470–473  
 WordNet, 320–325, 330–331  
 Word representation, 418, 467,  
 475, 476  
 Word2vec, 416, 418, 454, 471, 473  
 Word2vec modeling, 488

## Y

Yandex, 2, 5, 7, 62, 104, 105, 120,  
 156, 164, 165, 169n4, 176, 224,  
 229, 283, 403, 478, 490,  
 510, 572  
 YandexTaxi, 403