

Digital Methods and Research Data Management in the Humanities and Social Sciences

Veranstalter: Discuss Data Project; German Historical Institute Moscow; Centre for Digital Humanities at the National Research University Higher School of Economics

Datum, Ort: 07.10.2019–08.10.2019, Moskau

Bericht von: Mark Schwindt, Seminar für Slavistik / Lotman-Institut für russische Kultur, Ruhr-Universität Bochum

The Workshop organized by the *Discuss Data* project¹, the German Historical Institute Moscow (DHI Moskau) and the Center for Digital Humanities at the National Research University Higher School of Economics, Moscow, brought together scholars and researchers from various fields of the humanities and the social sciences who are interested in the study of the post-Soviet region, using digital methods and tools. The primary objective of the workshop was to create an opportunity for exchange about challenges, implications, and solutions regarding applications of innovative digital methods and the research data management in the respective scientific fields.

In the first session *Social Science Data* the team of *Discuss Data* consisting of FELIX HERRMANN (Bremen), EDUARD KLEIN (Bremen), DANIEL KURZAWA (Göttingen) and UBBO VEENTJER (Göttingen) introduced an open platform for interactive publishing, discussion, and management of scientific datasets. Following the FAIR principles of research data management (Findability, Accessibility, Interoperability and Reusability) and utilizing interactive social media functionalities of the platform the project aims to realize the crowdsourcing potential of the scientific community. Through the integration with the existing digital infrastructure for the Humanities (DARIAH-DE), *Discuss Data* will allow open access publication of datasets with increased visibility and long-term data preservation options allowing researchers to examine, validate and reproduce data-based scientific results. The official launch of the platform is scheduled for March 2020.

ANDREI SEMENOV (New Haven) contin-

ued the session by evaluating the state of the art regarding event catalogs in the social sciences. Reflecting on his experience of compiling the catalog of protest events in Russia (CPR – Contentious Politics in Russia) at the Center for Comparative History and Political Studies in Perm, Semenov outlined opportunities and limitations of algorithmic data aggregation, enrichment, and integration with geospatial information as well as network analysis and clustering of the event data, effectively extending the scope of the underlying analytical method. Through the querying of large datasets of Russian mass media offered by *Integrum*² Semenov managed to identify more than 7,000 instances of protest activities between 2012 and 2015. Spread across 1,500 locations this semi-automatically generated event catalog offers an overview of the protest activities in contemporary Russia.

The second and third sessions focused on *Historical Data*, starting with STEPHEN WHEATCROFT (Melbourne) who emphasized the importance of preservation, open distribution, and analysis of Russian historical and economical statistical data. Though main state libraries, as well as independent enthusiasts, are actively digitizing archival materials, only a small portion of the historical statistical data has been made publicly available. Following the example of modern state statistical systems, Wheatcroft expressed the need for a more systematic reproduction of secret historical statistical materials in an easily convertible operational (digital) form. Some of the materials are, in fact, still exclusively available only on microfilm. The problem of availability is, however, also closely tied to the problem of reliability of historical data since some of it might have been censored or falsified for political means. Which, according to Wheatcroft, could be mitigated

¹Discuss Data is a collaboration between the Research Center for East European Studies at the University of Bremen and the State and University Library, Göttingen, funded by the German Research Foundation (DFG), <https://discuss-data.net> (03.03.2020).

²The information agency Integrum World Wide is engaged in the global monitoring of the post-Soviet information space and development of research tools for Russian and post-Soviet studies providing access to the Integrum databases, <http://integrumworld.com> (03.03.2020).

through the much-needed mechanism for annotation of statistical data sets. Another concern raised throughout previous discussions was the general uncertainty about copyright laws that could theoretically apply to such data materials in general and in Russia, Germany or the US in particular.

The general concern with copyright issues and the implementation of good scientific practices in data-driven textual studies were more broadly addressed by STEFAN HEßBRÜGGEN-WALTER (Moscow) and MARK SCHWINDT (Bochum) who provided an outline of their research project on the use of digital methods in Russian conceptual history. While building a text corpus of essay collections to diachronically study the construction and semantic transformations of the concept of freedom in Russian intellectual history, the researchers encountered a newly adopted German text mining regulation that contradicts the research data management requirements of the German Research Foundation. According to that regulation, text mining data derived from copyright-protected sources would need to be deleted after the publication of the study results. To at least partly solve this issue, an additional prosopographic study was proposed to help distinguish between still protected and already publicly available source materials.

Concerned with the preservation of cultural heritage by post-Soviet institutions NADEZHDA POVROZNIK (Perm) presented a project that focuses on the study of the history and the development of Russian virtual museums in the digital age. The analysis of snapshots of museum informational resources stored in web archives³ to determine key characteristics as well as the development stages of both structural and functional features of the museums in a virtual environment was presented as a significant methodological improvement.

SEBASTIAN KINDLER (Moscow) introduced an ambitious research project regarding the fate of both Soviet and German prisoners of war. Initiated through a mutual declaration of the Russian and German foreign ministers on June 22, 2016, the project aims to collect and digitize sources documenting captivity of soldiers and civilians and make that

data available to researchers and the general public. The database will provide biographical (name, date of birth, nationality) and geospatial (place of captivity, transports, etc.) information about individual persons. This way it will be easier to determine what happened to larger groups of prisoners through pattern recognition.

The last presentation in this session, shared between SERGEI KORNIENKO (Perm) and DINARA GAGARINA (Perm), was dedicated to the research of Russian parliamentary tradition. Gagarina introduced a platform⁴ containing exhaustive verbatim records of parliamentary sessions of the Russian pre-revolutionary State Duma, as well as meta-data about individual deputies and the structure of parliamentary institutions, fractions, and commissions. Serving as an important source for researchers of Russian parliamentary history, the platform allows the conduction of both statistical and prosopographical studies, building models of socio-cultural appearances and activities in pre-soviet Russia.

The second day began with the session on *Archive Data*. MICHAEL MELNICHENKO (St. Petersburg) presented *Prozhito*⁵ – a publicly accessible digital database of historical personal diaries that launched 2014 as an independent initiative and was recently embedded into a larger Center for the Study of Ego-Documents founded by the European University, St. Petersburg. The current text corpus consists of more than 4,000 diaries by approximately 1,700 authors (a total of 450,000 entries). The newly founded center, however, aims to extend the digital archive with memoirs, autobiographies and personal correspondence. According to Melnichenko, the Prozhito team (including seven hundred volunteers and three hundred students) works closely with living diarists and their heirs to bring intimate historical experiences into the public domain.

Another project concerned with conceptual history was presented by VLADISLAV RJÉOUTSKI (Moscow). It is guided by the

³Way Back Machine, <https://web.archive.org> (03.03.2020).

⁴The Parliamentary History of Late Imperial Russia, <http://parlament.psu.ru> (03.03.2020).

⁵Prozhito Project, <https://prozhito.org> (03.03.2020).

main hypothesis that modern political terminology in Russian was created in the process of translation of European political texts of the 18th century. This led to the creation of a corpus of annotated translation examples of basic political concepts in form of a publicly accessible database. That database includes archiographic and bibliographic descriptions of translated texts as well as translations of key concepts alongside with fragments of the originals. According to Rjéoutski, this resource will allow for a diachronic study of the cultural transfer of main political concepts and expand the view on the political discourse in 18th century Russia.

The last session on *Literature and Arts* started with a video call from YAKOV KLOTS (New York) who introduced a virtual environment that traces the history of circulation, first publications and reception of contraband Russian literature by the Russian diaspora outside the USSR. *The Tamizdat Project*⁶ is an ever-growing public bibliographical database of archival materials for researchers and enthusiasts alike. The main aim of the project, according to Klots, is to explore the historical and socio-cultural climate in which masterpieces of Russian literature first appeared abroad indicating the political atmosphere inside the Soviet Union at the same time.

Followed by the talk by ANNA NIZHNIK (Moscow) who spoke about the creation of an archive of microhistorical data like interviews, meetings (i.e. salons, museums, poetry readings) and informal connections of individual and institutional actors of the literary 1990s in Russia. Combining old methods of traditional literary criticism and new research instruments like graph theory for social interactions, discourse analysis and geographic information system (GIS) this data will certainly enrich our understanding of the everyday life of contemporary Russian writers.

The session concluded with the presentation by FRANK FISCHER (Moscow) on Programmable Corpora – a new term Fisher and his associates are trying to establish within digital humanities for projects with a wide range of programmable functionalities. One of those projects is *DraCor*⁷, a research infrastructure for European drama that provides extended research options like network anal-

ysis, multilevel repository connections (TEI, API, R, Python, SPARQL, Excel, etc.) as well as rich markup and metadata to various TEI encoded drama collections (German Drama Corpus, Russian Drama Corpus, Spanish Drama Corpus, etc.). Not only does this approach allow for comprehensive research in digital literary studies, but it also assures a better reproducibility by following the aforementioned FAIR principles.

The Workshop ended with a discussion focusing on research data management, potential copyright challenges in Russia and Germany and good scientific practices. Felix Herrmann expressed his hopes for an easier data management in the humanities that would hopefully be at least partly achieved through the digital infrastructure and service platforms like *Discuss Data*. Another problem emphasized by Nadezhda Povroznik was the lack of the overall visibility of introduced projects both inside and outside academia. In that regard, the workshop provided an important step towards a more open, intersectional discussion about digital methods in humanities and social sciences.

Conference overview:

Session I: Social Science Data

Chair: Christian Fröhlich (HSE, Moscow)

Felix Herrmann, Eduard Klein (University of Bremen), Daniel Kurzawe, Ubbo Veenster (SUB Göttingen): Research Data Management with *Discuss Data*

Andrei Semenov (Yale University): Compiling Event Catalogues on Russian Politics: Scope and Limits

Session II: Historical Data #1

Chair: Thomas Skowronek (GWZO, Leipzig)

Stephen Wheatcroft (University of Melbourne, via Skype): Critical Thoughts about Research Data Management in History

Stefan Heßbrüggen-Walter (HSE, Moscow), Mark Schwindt (Ruhr University Bochum): Discourses of Freedom: the Uses of Digital

⁶ Tamizdat Project, <http://tamizdatproject.org> (03.03.2020).

⁷ Drama Corpora Project (DraCor), <https://dracor.org> (03.03.2020).

Conceptual History

Nadezhda Povroznik (Perm State University):
Digital History of Virtual Post-Soviet Museums

Session III: Historical Data #2

Chair: Alexei Kouprianov (HSE, St. Petersburg)

Sebastian Kindler (GHI Moscow): Soviet and German Prisoners of War and Internees

Sergey Kornienko, Dinara Gagarina (HSE, Perm): Parliamentarians of Pre-Revolutionary Russia at the Beginning of the 20th Century

Session IV: Archive Data

Chair: Anastasiya Bonch-Osmolovskaya (HSE, Moscow)

Michail Melnichenko (University of Bremen):
Electronic Archive of Ordinary Soviet Citizens

Vladislav Rjéoutski (GHI Moscow): Project
Presentation „Korpus russkikh perevodov“

Session V: Literature and Arts

Chair: Katrin Neumann (Max Weber Stiftung, Bonn)

Yakov Klots (City University of New York, via Skype): Tamizdat Online Project

Anna Nizhnik (Moscow State University):
Microhistorical Data of Literary 90-s: Between Self- and Meta-Interpretation

Frank Fischer (HSE, Moscow): DraCor: A Data-Driven View on Russian Drama

Concluding Discussion: Research Data and its Challenges for Scholars in the Humanities and Social Sciences

Tagungsbericht *Digital Methods and Research Data Management in the Humanities and Social Sciences*. 07.10.2019–08.10.2019, Moskau, in: H-Soz-Kult 11.03.2020.