SSOAR

Open Access Repository
www.ssoar.info

Delegated Regulation on Data Access Provided for the Digital Services Act: Response to Call for Evidence from Nordic Observatory for Digital Media and Information Disorder (NORDIS)

Walter, Jessica; Bechmann, Anja

Stellungnahme / comment

Empfohlene Zitierung / Suggested Citation:
Walter, J., & Bechmann, A. (2023). *Delegated Regulation on Data Access Provided for the Digital Services Act: Response to Call for Evidence from Nordic Observatory for Digital Media and Information Disorder (NORDIS).* Aarhus. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-86773-5

Nutzungsbedingungen:
Dieser Text wird unter einer Basic Digital Peer Publishing-Lizenz zur Verfügung gestellt. Nähere Auskünfte zu den DiPP-Lizenzen finden Sie hier:
http://www.dipp.nrw.de/lizenzen/dppl/service/dppl/

Terms of use:
This document is made available under a Basic Digital Peer Publishing Licence. For more Information see:
http://www.dipp.nrw.de/lizenzen/dppl/service/dppl/

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

NORDIS

nordishub.eu

Delegated Regulation on Data Access Provided for the Digital Services Act: Response to Call for Evidence from Nordic Observatory for Digital Media and Information Disorder (NORDIS)

This work has received funding from the European Union under Contract number: INEA/CEF/ICT/A2020/2394203

Authors:
Jessica Gabriele Walter
Anja Bechmann

Acknowledgments:
We thank Minna Aslama Horowitz, Matteo Magnani and Mathias Holm Tveen for comments provided to this document.

Dear Commission,

In response to the initiative “Delegated Regulation on data access provided for in the Digital Services Act” the Nordic Observatory for Digital Media and Information Disorder (NORDIS) as the Nordic EDMO Hub wants to contribute with joint feedback from all academic research partners in the consortium. NORDIS is a consortium of researchers and fact-checkers from Denmark, Norway, Sweden, and Finland. NORDIS is led by Aarhus University’s DATALAB with professor Anja Bechmann as principal investigator.

Our aim is to develop theories, practices and models that can help counteract digital information disorders – the spreading of misinformation, disinformation and other forms of harmful information online – and to help empower citizens in the Nordic welfare states to resist such information by enabling them to enhance their media literacy.

In this context NORDIS has also published policy recommendations: NORDIS policy recommendations

Data access needs

What types of data, metadata, data governance documentation and other information about data and how it is used can be useful to DSC's for the purpose of monitoring and assessing compliance and for vetted researchers for conducting research related to systemic risks and mitigation measures?

The research in NORDIS relies on access to all kinds of data from VLOPs and VLOSEs - from exposure data, engagement information, content, and network data to metadata that provides context to the data. As NORDIS, we have faced difficulties accessing high quality data that provides relevant information for answering pressing research questions such as: Who is exposed to and engaging with which content on social media, and how content and network structures predict spread? What do users' networks look like regarding specific content and influence in content flows, and how can we analyze content and networks across events, countries, and platforms? Research access to social media data is crucial for an understanding of the conditions for the spread and impact of harmful digital content and also for evaluating potential threats to the well-functioning and wellbeing of citizens and democratic societies, the effectiveness of and conditions for counter measures such as fact-checking and media literacy initiatives.

Even though digital methods for analyzing these questions have improved, we still face difficulties accessing relevant data but also of utilizing scalable methods for minority countries and languages: for example, with research access through social science one, the Condor dataset does not allow for the analysis of small countries due to the use of differential privacy. Also minority languages in the Nordic context are difficult to handle - due to a focus on US context by existing research (see our published article from de Place Bak et al. 2022; de Place Bak et al. 2022) and limited geolocation information in the data (from Twitter). Challenges are also described in a report addressing the war in Ukraine (NORDIS report on academic research in times of crises). One of the main conclusions drawn regarding data access is that even in case of Twitter that has in comparison to the Meta platforms and other VLOPs until recently provided better access to data we still face challenges due to limited geolocation information and information of the extent and systematics of content deletion or algorithmic curation. With regard to the other VLOPs the limited access to other content than

public content is the most restricting one. With great concern we also observe the tendency of the VLOPs to not provide easier and more comprehensive access to data but instead to even further restrict access for academic research. This goes for Twitter as well as Facebook, Instagram and Youtube (Tiktok has always been strict).

VLOSEs are important from a research perspective as data sources (for example as basis for review articles as the one published by de Place Bak et al. 2022 (see above)) and as data analysis unit. For example, NORDIS researchers have analyzed inherent characteristics of the Google Fact Check Explorer that is a search database for fact-checking articles (see Nissen et al. 2022, Nissen et al. 2022), which allows researchers assessing implications for research based on content of these platforms, for example for detection of information disorders on social media.

Metadata enriches the data created by users of VLOPs and VLOSEs and helps to understand patterns and relations in the data, and to analyze data across platforms and countries. Metadata also refers to a broad range of information that potentially and factually is relevant for the research conducted in NORDIS - such as geolocation and further demographic information of users, network information such as follower/friend/group member information, platform governance information such as deletion/downgrading of and guidelines (manual and algorithmic rules) for handling of potential harmful content such as mis- and disinformation or hate speech. This kind of information is for example crucial in understanding who are users of VLOPs and VLOSEs and why users of VLOPs and VLOSEs are exposed to and engaging with potentially harmful content.

We see transparency of data governance as the basis for reliable and independent research and we suggest to protect access to nuanced datasets in the most strict way instead of protecting data units (SOMA report on data access). In order to understand, analyze and document the data for scientific purposes (following scientific integrity e.g. accounting for what is analyzed and how) access is needed on the lowest level possible without third party interference without unclear and partisanship interference.

What sort of analysis and research might DSC's and vetted researchers conduct for the purposes of monitoring and assessing compliance and conducting research related to systemic risks and mitigation measures?

NORDIS considers and focuses on different kinds of analyses for mitigating information disorders and analyzing systematic risks - from detection to spread, exposure and engagement with a strong focus on at scale analyses based on machine learning techniques and natural language processing methods such as sentiment analysis, network analysis or topic modeling - see as an example Charquero Ballester et al. 2021 (Charquero Ballester et al. 2021). Furthermore, AI tools are of interest to the research conducted in NORDIS.

To conduct this kind of research and to ensure the high quality of it, secure data storage, high computational capacity, individual and flexible implementation of methods and tools, as well as quick, stable and easy access to the data is key. Here it is also relevant to highlight that a DSI needs to provide researchers with flexible access to data in context of unforeseen events and circumstances that will require access to different forms of platform data.

In order to increase the impact of the DSA, NORDIS also considers it relevant to conduct research on digital democracies and digital wellbeing and was in this context also actively involved in the development and publishing of the recommendations of the Nordic Think Tank for Tech and Technology: Recommendations Nordic Think Tank. NORDIS wants to point out that access to data needs is not restricted to the monitoring of compliance with the DSA and the provision of the structural and performance indicators designed by the EC-ERGA-EDMO for the Code of Practice but reaches into the understanding of the cultural fabric of democratic societies and citizens' wellbeing ranging from socializing, culture, privacy, identity, democracy and core values of trust, equality, non-transparent manipulation, power and influence logics.

Data access application

Digital Services Coordinators (DSCs) in the Member States will play a key role in assessing researchers' applications and they will act as intermediaries with the platforms. How should the application process be designed in practice? How can the vetting process ensure efficient exchanges between researchers and platform providers?

For researchers, applying for data access can be facilitated by providing transparent information about the procedure, requirements, assessment criteria and implementation. We consider research design as a crucial part of the application. Research access should be granted in a timely fashion and with a minimum of bureaucratic requirements but with strict requirement to the documentation of university full-time affiliation vetted by head of university, school or faculty (see SOMA report on data access) to secure against violation of data subjects but instead full access to data should be provided. NORDIS would favor an assessment by independent researchers (an association could be established by the DSCs). Furthermore, it should be independent of the seniority of a researcher. It can be based on a research data agreement that establishes the legal foundation for data access.

Furthermore, once access is granted, access should also allow researchers to comply with research integrity requirements - such as providing information about data, methods and analyses that build the basis of publications. This information along with data for scrutiny should continue to be available years after the granted access to secure research integrity.

To ensure that researchers are aware of the characteristics of the data they apply for, data documentation is crucial.

Article 40(8) exhaustively defines criteria for vetting researchers. How can a consistent assessment across DSCs be ensured, while still taking into consideration the specificities of each request?

Consistent assessment across DSCs will likely best be ensured by standardization of data access, storage and tools, procedures, requirements, assessment criteria and implementation and by choosing institutions and organizations as DSCs that have the expertise, financial and personnel resources for managing the process, services and communication; and ideally storing data outside

and across platforms (see SOMA report on data access). Furthermore, it is of importance that the work of the DSCs is contextually informed and consistent in terms of expertise and resources to ensure consistent and equal access to data and equal opportunity to analyze data for researchers, independent of country as object of study (small or large country in population and native language).

What additional provisions or specifications could be useful to help balance the new data access rights and the protection of users' and business' rights, e.g. related to data protection, confidential information, including trade secrets, and security?

Legal foundations should of course respect the independence of academic research and be implemented based on control mechanisms. In projects preceding NORDIS - namely SOMA - researchers from NORDIS have already discussed how data access to social media data can potentially be implemented - see: SOMA report on data access

What kind of safeguards can be put in place to assure that data gathered under Article 40 is used for the purposes envisaged and to minimise the risk of abuses?

One of the measures that could be implemented to minimize the risk of abuses is to establish agreements that require transparency regarding data sources in publications related to projects that are based on the data access and also the notification about outcomes that are based on the data. Also to have a strict entrance procedure that requires institutional clearance as suggested in the SOMA report.

Article 40(13) introduces the possibility of an independent advisory mechanisms to support the management of data access requests and vetting of researchers. What would be the added value of such a mechanism?

One of the potential benefits could be that applications can be processed quicker, and it is a potential mechanism to ensure fair assessments.

Data access formats and involvement of researchers

What technical specifications could be considered for data access interfaces, which takes into account security, data protection, ease of use, accessibility, and responsiveness (e.g. APIs, data vaults and other machine-readable data exchange formats)?

Indeed, the technical specifications can be manifold - NORDIS researchers have for example good experiences with the academic Twitter API (from 2022) and the FACEBOOK graph API from before 2015 - and besides the necessity to take aspects such as ease of use, accessibility, capacity, responsiveness etc. into account - it is important that the solution builds on resources available for academic researchers across different disciplines and contexts. NORDIS would recommend establishing an API-based access, where data can be stored at local servers at the researchers' institutions or at least outside the social media platforms as this would allow for multiple access to the same data by a research team and would prevent closing off access to data in case of unwanted results (for the platforms). Furthermore, VLOPs and VLOSEs should provide data access to data that has not been preprocessed by the VLOPs and VLOSEs.

What capacity building measures could be considered for the research community to take advantage of the opportunities provided by Article 40?

In order for the research community to take advantage of the opportunities provided by Article 40, visible and easily accessible knowledge about rights and options is crucial.

Would it be desirable and feasible to establish a common and precise language for DSCs, vetted researchers, VLOPs and VLOSEs to use when communicating about data access, e.g. by formulating a standard data dictionary and/or business glossary? How might this be implemented?

An inclusive and precise language for the DSCs would potentially benefit cross-cultural cooperations and prevent violations of agreements based on misunderstandings. Ways for implementing standards are manifold and could for example be established as a taskforce with representatives of relevant stakeholders, via surveys or expert interviews. NORDIS recommends that this would involve

regular and open communication of updates and allow for a knowledge exchange also with researchers from different disciplines without current data access needs.

Not only vetted researchers will have greater opportunities for accessing data, all researchers meeting the conditions set out in Article 40(12) will be able to get direct access to publicly available data. What processes and mechanisms could be put in place to facilitate this access in your view?

One of the steps that could facilitate the access to publicly available data would be to establish a common access point.

List of signatories from NORDIS (academic members):

Dr. Marina Charquero Ballester, DATALAB- Center for Digital Social Research, Aarhus University, Denmark

Professor Anja Bechmann, DATALAB - Center for Digital Social Research, Aarhus University, Denmark, PI NORDIS

Dr. Minna Horowitz, Docent, Media and communications, Faculty of Social Sciences, University of Helsinki

Carl-Gustav Lindén, Professor, University of Bergen & Docent, University of Helsinki

Matteo Magnani, Professor, InfoLab, Dept. of Information Technology, Uppsala University

Professor Mervi Pantti, Media and communications, Faculty of Social Sciences, University of Helsinki

Davide Vega, Assistant Professor, InfoLab, Dept. of Information Technology, Uppsala University

Dr. Jessica Walter, DATALAB- Center for Digital Social Research, Aarhus University, Denmark

Oscar Westlund, Professor Oslo Metropolitan University & Associate Professor University of Gothenburg

Yang Xu, doctoral researcher, Faculty of Social Sciences, University of Helsinki

Victoria Yantseva, Postdoctoral research fellow, InfoLab, Dept. of Information Technology, Uppsala University

Additional Nordic signatories:

Senior Lecturer, Docent Marko Ampuja, Media and Communication Studies, Faculty of Social Sciences, University of Helsinki

PhD Fellow, Petra de Place Bak, Department of Media and Journalism Studies, Aarhus University

Professor Niels Brügger, Centre for Digital Methods and Media, Aarhus University, Denmark

Postdoctoral researcher Dr. Ali Faqeeh, Complex Systems Group, Department of Computer Science, Aalto University, Finland

Senior Lecturer Nils Gustafsson, Department of Strategic Communication, Lund University

Associate professor Lone Koefoed Hansen, Department of Digital Design and Information Studies, Aarhus University.

Associate Professor Juha Herkman, Media and Communication Studies, Faculty of Social Sciences, University of Helsinki

Senior Lecturer Nils Holmberg, Department of Strategic Communication, Lund University

Associate professor and Director of Centre for Internet Studies, Jakob Linaa Jensen, School of Communication and Culture, Aarhus University.

Associate Professor Stine Liv Johansen, Center for Children's Literature and Media, School of Communication and Culture, Aarhus University.

Professor Bente Kalsnes, Department of Communication, Kristiania University College and founding partner of the SCAM project (Source Criticism and Mediated Communication).

Professor Anders Olof Larsson, Department of Communication, Kristiania University College

PhD Fellow, Amanda Skovsager Mouritsen, Department of Media and Journalism Studies, Aarhus University

Associate Professor Anders Kristian Munk, Techno-Anthropology Lab, Department of Culture & Learning, and scientific director of MASSHINE, Aalborg University.

Senior University Lecturer of Public Law Riku Neuvonen, Public Law, Faculty of Management and Business, Tampere University.

Associate Professor in Digital Media History, Janne Nielsen, Department of Media and Journalism Studies, Aarhus University

Professor of Media and Communications Policy, Hannu Nieminen, Faculty of Social Sciences, University of Helsinki, and Visiting Professor, Department of Media and Communications, London School of Economics and Political Science

Associate professor Lene Pettersen, Department of Communication, Kristiania University College

Associate professor, Søren Pold, Department of Digital Design and Information Studies, Aarhus University.

Doctoral researcher, Outi Puukko, Media and Communication Studies, Faculty of Social Sciences, University of Helsinki

Associate Professor in Journalism Studies and Digital Methods, Christoph Raetzsch, Aarhus University

Assistant Professor Michael Bossetta, Department of Media and Communication, Lund University

Associate Professor Anamaria Dutceac Segesten, Lund University (Sweden)

Professor Trine Syvertsen, Dept of media and communication, University of Oslo, Norway

Associate Professor, Jesper Tække, School of Communication and Culture, Aarhus University.