weizenbaum
POLICY PAPER

DECEMBER 2024
11

Lukas Seiling, Jakob Ohme, Ulrike Klinger

Response to the Consultation on the Delegated Regulation on Data Access provided for in the Digital Services Act

ABOUT THE AUTHORS

Lukas Seiling is responsible for the coordination of the #DSA40 Data Access Collaboratory, a joint project by the European New School of Digital Studies (ENS) and the Weizenbaum Institute. Since 2020 they have been working for the Weizenbaum Institute where they co-led the Privacy Icons Project and are maintaining a framework for the collection of behavioral data on simulated newsfeeds.

Jakob Ohme leads the "Digital News Dynamics" group at the Weizenbaum Institute, studying digital journalism's impact versus influencers and AI. He is a Co-Principal Investigator in the #DSA40 Collaboratory, focusing on collaborative access to platform data under the EU's Digital Services Act.

Ulrike Klinger is a Professor of Digital Democracy at the European New School for Digital Studies at the European University Viadrina in Frankfurt (Oder) and an Associated Researcher at the Weizenbaum Institute in Berlin. She also is a Co-Principal Investigator in the #DSA40 Collaboratory, focusing on collaborative access to platform data under the EU's Digital Services Act.

Contact: contact@dsa40collaboratory.eu

ABOUT THIS PAPER

Weizenbaum Policy Papers provide scientifically grounded statements, position papers, and briefings on current political topics and decision-making processes.

This paper was prepared by the DSA 40 Data Access Collaboratory as a response to the public consultation on the rules for researchers to access online platform data under the Digital Services Act launched by the European Commission. The DSA 40 Data Access Collaboratory is a joint project by the Weizenbaum Institute and the European University Viadrina. It monitors and evaluates the implementation of DSA Art. 40 to ensure that scientists and non-profit organizations have access to data for studying systemic risks. This requires the effective coordination of various actors (from science, civil society, politics, and platforms) on different levels (national, international, and transnational) with regards to different kinds of data (from non-sensitive public data to sensitive non-public data) available through various interfaces (such as APIs, archives, or sandboxes).

ABOUT THE WEIZENBAUM INSTITUTE

The Weizenbaum Institute is a joint project funded by the German Federal Ministry of Education and Research (BMBF) and the State of Berlin. It conducts interdisciplinary and basic research on the digital transformation of society and provides evidence- and value-based options for action in order to shape digitalization in a sustainable, self-determined and responsible manner.

www.weizenbaum-institut.de

Weizenbaum Policy Paper

Response to the Consultation on the Delegated Regulation on Data Access provided for in the Digital Services Act

Lukas Seiling, Jakob Ohme, Ulrike Klinger

Abstract

This response provides feedback on the Delegated Regulation on Data Access provided for in the Digital Services Act. It is informed by a variety of exchanges with empirical platform researchers across Germany and Europe. The first section highlights clarifications and proposed procedures for non-public data access which are practical, workable or welcomed by scientists for other reasons. The second section outlines further opportunities for clarification, additions, or modifications to the draft text, particularly regarding the data access procedure, the data access portal, the types of data, as well as the documentation and modalities of data access.

Contents

1 General Remarks

This response builds on several gatherings and exchanges with empirical platform researchers in Germany and Europe. We value the clarifications from the draft delegated act (DDA). Overall, the response from the research community was positive; the general procedure and pipeline are feasible for researchers to request non-public platform data. Some details have caused additional questions or suggestions to amend the Delegated Act. Below we outline both - the positively received provisions and suggestions for clarifications, additions, or changes.

2 Notable Provisions Supporting Research Needs

As mentioned above, the DDA includes various additions to the Digital Services Act (DSA), such as specifying the general procedure for non-public data access and introducing the DSA data access portal. Many new provisions address ambiguities of Art. 40 DSA and have apparently considered the previous feedback from the scientific community. The provisions listed below are milestones in scientific data access and should remain in the final version of the Delegated Regulation on Data Access.

- We appreciate that the DDA does not mention fees or other costs for researchers. This creates an even ground for all researchers by not privileging researchers associated with well-resourced institutions. Data access free of cost is fundamental to ensure the general accessibility of the process for accessing public and non-public data.
- We also welcome the introduction of the DSA data access portal as a one-stop shop for all stakeholders involved in non-public data access, as outlined in Art. 3 of the DDA. As a central point for managing information about the non-public data access process, it promises to further increase accessibility, clarity, and accountability for researchers, supervisory and regulatory authorities, and data providers.
- Key provisions of the delegated act introduce foundational transparency obligations for data providers, namely a data inventory (Art. 6, Rec. 6) and data documentation (Art. 15, Rec. 26). Both provisions are essential for researchers to understand what data could be requested and to make proper use of the data provided. As highlighted in Art. 15(3) and Rec. 27, little to no limitations should apply when using said data, ensuring the flexibility necessary for researchers to study systemic risk in the European Union meaningfully.
- As part of Rec. 12, the DDA also includes an extensive overview of data that could be accessed using the non-public data access process. The named categories create a solid foundation for data access in the future (for recommendations on how these examples could be expanded, see point 3.2.1 below). We also want to positively highlight the recital's recognition of time as a relevant dimension to be considered in the study of systemic risk by acknowledging that the categories of data required may change over time (unfortunately, this acknowledgment of dynamics is not consistent throughout the DDA as reflected by the lack access amendment possibilities for researchers, for details see point 3.1.4 below).

- By differentiating between the "principal researcher" and the other applicant researchers, Art. 2 of the DDA indicates that groups of researchers can apply for data access. It also introduces no limitations about the researchers' location. We welcome this clarification as it acknowledges science as a collaborative process often involving researchers from various nations, inside and outside the European Union.
- We appreciate the publication of *all* reasoned requests upon their formulation, as outlined in Art. 11 of the DDA. This enables researchers to learn from the applications underpinning each request while laying the foundation for accountability and transparency during the remaining non-public data access process.
- We acknowledge the timing of the non-public access process outlined in Art. 7 and Rec. 8 of the DDA to be ambitious. While we welcome a standardized (Rec. 19) and swift processing of data access applications as it enables flexibility for researchers and enables a quick response to potential systemic risks, we are also aware of the workload this means for the Digital Service Coordinators tasked with vetting the applicants and engaging in dispute settlement. A suggestion for a potential change to the draft text that would address this issue is included in point 3.1.3. below.
- Another positive introduction is the option for DSCs to consult independent advisory mechanisms (Art. 14) and experts (Rec. 23), increasing their capacity for reviewing access applications and reducing the chance of regulatory capture.

Lastly, we want to stress that we strongly support the DDA's approach of making the required access modalities for non-public data relative to the sensitivity of the data received as well as the possibility to attain research objectives, as described in recitals 16 and 18. We also welcome the explicit mention of data transfer as this access modality is the preferred option among many researchers, most likely allowing for the necessary flexibility in data processing to effectively detect, identify, and understand systemic risk in the European Union. For further suggestions around access modalities, please see point 3.2.2. below.

3 Potential for Clarifications, Additions, or Changes

As highlighted above, many of the provisions in the DDA create a promising foundation for an efficient and effective data access process, enabling researchers to get the data needed to contribute to risk governance under the DSA. Still, we want to use this feedback opportunity to highlight further potential for clarifications, additions, or modification to the draft text which would further the DSA's goal of informing the public and relevant stakeholders and "bridging information asymmetries and establishing a resilient system of risk mitigation" through data access (Rec. 96, DSA).

The remainder of our response contains three parts: 3.1. feedback to the steps in the data access **procedure** and the data access portal, 3.2. feedback on the **types of data**, **documentation** and the **modalities of data access**, and 3.3. **additional comments** on the draft text.

3.1 Procedure

3.1.1 Researcher Vetting

- To decrease unnecessary friction and the communication of redundant information in the application process, we suggest granting the "vetted researcher" status for a specified time period based on the provided documents. The outcome of a successful vetting process should be definitive, meaning that data providers should not have the option to impose other last-minute obligations.
- We propose to clearly identify that researcher vetting is part of the project-based access application process that will be possible on the DSA data access portal. If applicable, the vetting for the researchers and the technical operational measures they need to provide is done by the national DSCs, as they have more opportunities and take their knowledge of the institutional landscape into account. Details on this option are included in Art. 40(9) of the DSA, which is not referenced in the DDA.
- The DDA could profit from additional specifications on what constitutes "a formal relationship between the applicant researcher and the research organization of affiliation", as stated in Art. 8(2a) DDA, expanding the formulation found in Rec. 9 ("such as employment contracts or any other form of legal association"). It is unclear if and how this provision would apply to students, non-permanent positions, or affiliated researchers. Therefore, we propose adding additional examples to Rec. 9, which include enrollment and affiliation agreements as examples of legal affiliation.

3.1.2 Access Application

- We suggest unambiguously specifying in a corresponding recital that applicant researchers and principal researchers as in Art. 2(3-4) cover all researchers independent of their location and that applications can therefore be submitted by any researcher attempting to identify, detect, or understand systemic risk in the EU (Art. 40(4) DSA) that meets the requirements set out in Art. 40(8) of the DSA.
- We also suggest that the Commission and the DSCs prepare a transparent priority mechanism for access requests, should there be too many, since the chronological processing of requests does not allow for a flexible response to emergent phenomena and favours quick rather than diligent applicants.

3.1.3 Reasoned Request

- Art. 10(1d) of the DDA requires the DSCs to include a summary of the research project in the reasoned request sent to the data providers. There is considerable concern among researchers that this provision will allow platforms to anticipate research questions before data access is granted, which would allow for the manipulation of research results (as recently shown in *Science*) or data tampering. This quality assurance issue is exacerbated by the fact that, unlike for public data, no standard exists to which non-public data can be compared for validation. We suggest sharing research questions with providers only after data delivery to ensure a minimum of quality standards. Beforehand, the general reference to systemic risk included in the summary should be sufficient to validate the legitimacy of the request.

- Regarding the harmonised time frames in which access applications have to be checked and reasoned requests have to be formulated, it would be more helpful for researchers to know the maximum timeframe. Therefore, the DSCs should suggest a realistic

timeframe that does not require an extension. Consulting independent experts can extend the process considerably. Should these experts be other researchers, the minimum time for a review should be 4 weeks.

3.1.4 Amendment Process

- Based on the current version of Art. 12 of the delegated act, only platforms have the option to request amendments to the reasoned requests. We argue that an equivalent option for updating or revising the conditions of data access should also be open to researchers to account for a variety of cases:
    - Given that research teams are dynamic, requests should be amendable to allow researchers to be added should research teams change.
    - Research questions might evolve on the basis of discoveries based on existing data access, which might require the amendment of the requested data.
    - New funding for existing projects might also require the expansion of the time frame in which researchers have access to non-public data.

3.1.5 Mediation Process

- The current specification of the mediation process in Art. 13 DDA only includes researchers if DSCs decide to reach out to them (Art. 13(5) DDA). We argue that researchers should at the least be informed if mediation starts and should generally have the opportunity to provide input to the mediation.
- The DDA should clarify what happens if mediation is unsuccessful. The current version allows to game the system if all requests are sent to unsuccessful mediation.
- Most pressingly, the draft text currently also lacks any information on options researchers can take if the data received does not conform to quality standards or does not allow for the research as intended. The Delegated Act should provide an additional dispute settlement or enforcement mechanism open to researchers. A failure to mediate should not stop researchers' particular request altogether.

3.1.6 DSA Data Access Portal

- For transparency, information on any mediation procedures should be published in the DSA access portal. Such information would allow researchers to evaluate the amount of friction associated with specific data access requests and would allow for the systematic analysis of data providers' responses to reasoned requests.
- Considering Art. 40(8g) of the DSA, the DSA data portal should also allow for the upload or referencing of publications associated with specific access requests.
- To improve findability, the data access portal should include easily accessible links to the data inventories specified in Art. 6 and Rec. 6 of the DDA.
- Considering that technical changes on the side of the data providers, like changes to the user interface, the data structure, or the algorithms on the platforms, can have an outsized effect on research results, we suggest that the final text of the DDA requires platforms to announce such changes on the data access portal which would allow for easy communication to the affected researchers.

3.2 Types of Data and Access Modalities

3.2.1 Types of Data and Documentation

- While, as stated above, Rec. 12 already contains many good examples of non-public data that could be accessed based on the DSA and acknowledged that these examples might be expanded or changed, we suggest adding additional examples to the recital in order to decrease the number of necessary amendments in the future. Thus, Rec. 12 should additionally refer to:
    - not only data related to users but all other persons (including non-users) the data providers have collected information on (e.g., through accessing all of a user's contacts)
    - internal data, which is not bound to or created by the users on the platform but to the VLOPs as organisations (e.g., to understand organisational processes and decision-making)
    - data inferred by the platforms (to this end, it could be specified that "temporary" in Art. 12(2b) DDA means that data requested could be generated from existing data)
    - data related to the governance of user attention via platform design implementations, such as content recommendations or contact ("friend" or "follow") suggestions
    - data related to any form of modification, reporting, or deletion of individual elements by their author, the other users, or the data provider itself
    - data generated as a result of usage of other access modalities (see 3.2.2.)
    - data on specific moderation decisions, which expand the data in the DSA Transparency / Statements of Reason database
    - data on individual content monetisation status and corresponding history as well as specific monetisation programs, including account participation and platforms' monetization policies
    - data on the geographic origin of specific pieces of content
    - data on the temporal development of content distribution and interaction metrics (e.g., views or shares) to allow for the historical analysis and tracing of content distribution
- Since replicability is a fundamental pillar of empirical science, data availability should allow for replication through increased access duration or through archiving by specific archiving institutions or data providers. This could also significantly speed up the access process for researchers requesting access to data that has already been accessed previously.
- Regarding the data inventory specified in Art. 6 and Rec. 6 of the DDA, we suggest further clarification to ensure that such inventories should not just detail known public data sources but indicate the non-public data accessible.
- There also needs to be a mechanism in place to validate information presented in data inventories.
- Art. 15 of the DDA currently mentions "data formats" in its heading but includes no information on the required formats. We suggest that any data shared should be accessible in easily machine-readable and open formats.
- The documentation specified in Art. 15(2) and Rec. 26 should also include metadata about the time and means of export, as well as querying, filtering, or aggregation methods applied to create the provided dataset.

- Decisions on the vulnerability posed by the publication of such documentation should not be made by the data provider but by the DSC of establishment.
- Art. 15(3) and Rec. 27 should additionally allow for merging or adding external data and applying methods (including algorithms or packages) developed by the researchers that are not publicly accessible.

3.2.2 Access Modalities

- We suggest adding wording to recital 13 of the DDA to specify that the safeguards set out in Article 40(8)(d) DSA must be assessed on a case-by-case basis and only required where strictly necessary to tackle identified risks and to be required in a proportionate manner. The obligations in Article 40(8)(d) must not be used to circumvent the obligation by providers of VLOPs and VLOSEs to grant access to data.
- We suggest using a layered approach for DSCs to generally decide about the sensitivity of data types and the necessary data access modes. This is crucial for researchers to know in advance, as they will need to prepare within their institutions to handle data and a data management plan. To this end, Art. 9(2) should include clearer information on how data sensitivity relates to safeguarding requirements and the scope of research as set out in Art. 8.
- Art. 9 of the DDA currently only mentions secure processing environments and data transfer as specific data access modalities, without any additional information on "other access modalities to be set up or facilitated by the data provider" (Rec. 16). We think the final text should at a minimum include a specific reference to additional access modalities, such as researcher sandboxes that allow for A/B testing or surveys if participants opt in. We also want to stress that any access modality should allow for the linkage of the accessed data with data from external data sources (such as surveys, libraries, etc.)

3.3 Additional Comments

- Additional clarification is needed about the attestation of a conflict of interest (as referenced in Art. 14(3c) and Rec. 23), as it is currently unclear what kind of conflicts of interest are meant by the current provisions.

Signees

Lukas Seiling (DSA 40 Data Access Collaboratory, Weizenbaum Institute, Berlin)
Jakob Ohme (DSA 40 Data Access Collaboratory, Weizenbaum Institute, Berlin)
Ulrike Klinger (DSA 40 Data Access Collaboratory, European University Viadrina, Frankfurt/Oder)
Juliane Mendelsohn (Technische Universität Ilmenau)
Judith Möller (Leibniz-Institute for Media Research, University of Hamburg)
Daniel Schnurr (Universität Regensburg)
Philipp Lorenz-Spreen (TUD Technische Universität Dresden / Dresden University of Technology)
Katrin Weller (GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany)
Johannes Breuer (GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany)
Mike Cowburn (European University Viadrina)
Christiane Eilders (Center for Advanced Internet Research, Bochum)
Christina Elmer (Institut für Journalistik, Technische Universität Dortmund)
Valerie Hase (LMU Munich)
Philipp Hacker (European University Viadrina, Frankfurt/Oder)
David Hartmann (Weizenbaum Institute, Technische Universität Berlin)
Dirk Heckmann (Bayerisches Forschungsinstitut für Digitale Transformation, München)
Tobias Mast (Leibniz Institute for Media Research, Hamburg)
Julia Niemann-Lenz (German Center for Higher Education Research & Science Studies, DZHW, Hannover)
Katharina Kaesling (TUD Technische Universität Dresden / Dresden University of Technology)
Jan-Hendrik Passoth (European University Viadrina, Frankfurt/Oder)
Alexander Pretschner (Bayerisches Forschungsinstitut für Digitale Transformation, München)
Jan Rau (Leibniz Institute for Media Research, Hamburg)
Jessica Walter (Datalab – Center for Digital Social Research, Aarhus University, Denmark)

Supporting Institutions

Weizenbaum Institute, Berlin
European New School of Digital Studies, European University Viadrina, Frankfurt/Oder
BEUC – The European Consumer Organisation, Brussels

Imprint

Lukas Seiling, Jakob Ohme, Ulrike Klinger

Response to the Consultation on the Delegated Regulation on Data Access provided for in the Digital Services Act

Weizenbaum Policy Paper # 11
Berlin, 12 \ 2024

ISSN 2748-5587 \ DOI 10.34669/WI.PP/11

Weizenbaum-Institut e.V.
Hardenbergstraße 32 \ 10623 Berlin \ Tel.: +49 30 700141-001
info@weizenbaum-institut.de \ www.weizenbaum-institut.de

COORDINATION: Moritz Buchner

DSA 40 DATA ACCESS COLLABORATORY

ens European New School of Digital Studies

weizenbaum institut

SPONSORED BY THE
Federal Ministry of Education and Research

BERLIN
Senatsverwaltung für Wissenschaft, Gesundheit und Pflege