

What the Scientific Community Needs from Data Access under Art. 40 DSA: 20 Points on Infrastructures, Participation, Transparency, and Funding

Klinger, Ulrike; Ohme, Jakob

Erstveröffentlichung / Primary Publication

Stellungnahme / comment

Empfohlene Zitierung / Suggested Citation:

Klinger, U., & Ohme, J. (2023). *What the Scientific Community Needs from Data Access under Art. 40 DSA: 20 Points on Infrastructures, Participation, Transparency, and Funding*. (Weizenbaum Policy Paper, 8). Berlin: Weizenbaum Institute for the Networked Society - The German Internet Institute. <https://doi.org/10.34669/WI.WPP/8.2>

Nutzungsbedingungen:

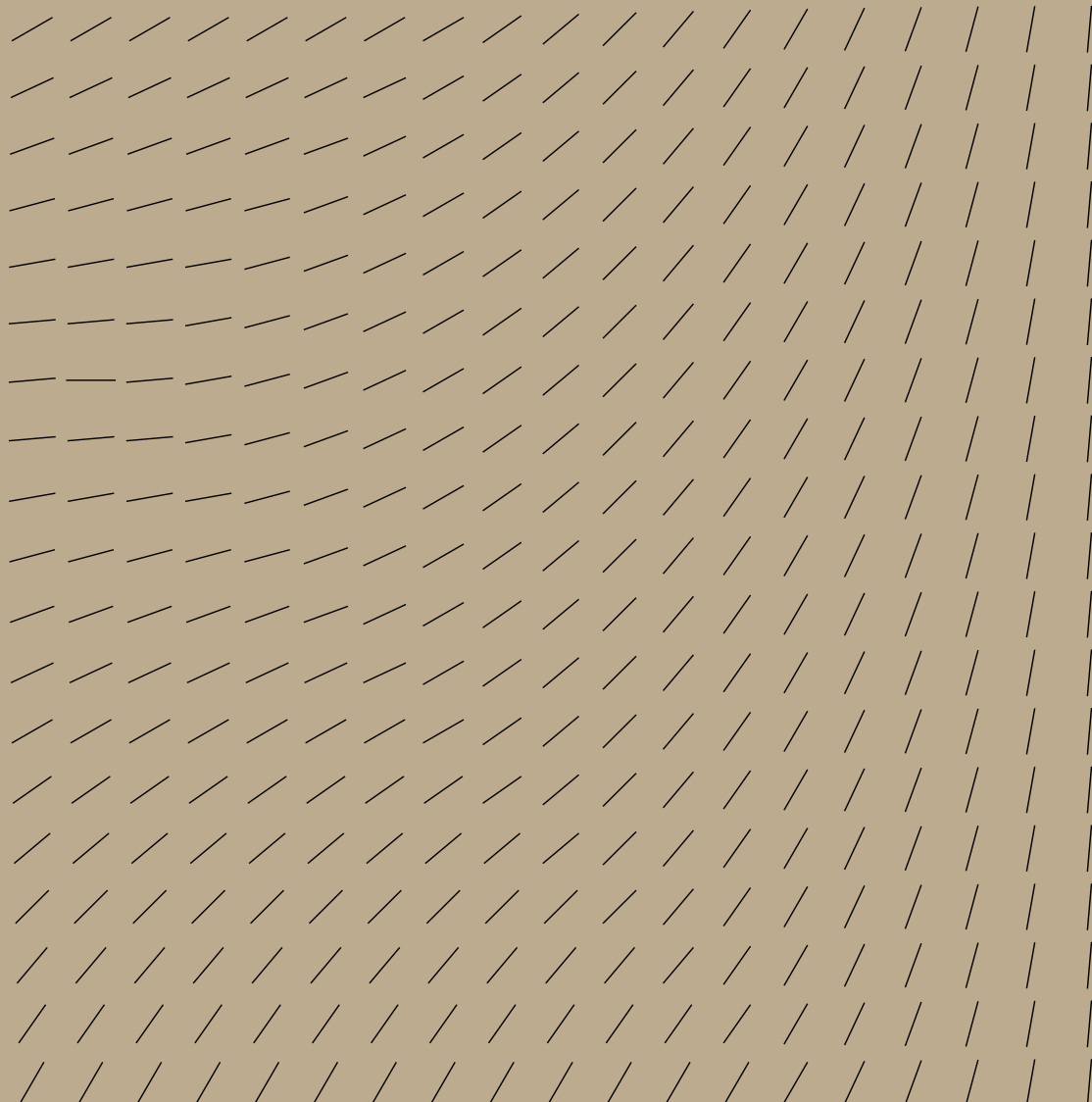
Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

weizenbaum
institut

ens  European New School
of Digital Studies



OCTOBER 2023

Weizenbaum Institute for the Networked Society

What the Scientific Community Needs from Data Access under Art. 40 DSA

**20 Points on Infrastructures, Participation, Transparency,
and Funding**

Imprint

What the Scientific Community Needs from Data Access under Art. 40 DSA.
20 Points on Infrastructures, Participation, Transparency, and Funding,
Weizenbaum Policy Paper 8, October 2023

AUTHORS:

Ulrike Klinger \ European University Viadrina \ klinger@europa.uni.de
Jakob Ohme \ Weizenbaum Institute \ jakob.ohme@weizenbaum-institut.de

PUBLISHER:

Weizenbaum Institute for the Networked Society
Hardenbergstraße 32 \ 10623 Berlin \ Tel.: +49 30 700141-001
info@weizenbaum-institut.de \ www.weizenbaum-institut.de

EDITORIAL TEAM:

Moritz Buchner \ Weizenbaum Institute
Christina Dinar \ Leibniz-Institut für Medienforschung | HBI
Robert Peter \ Weizenbaum Institute
Kaltrina Shala \ Weizenbaum Institute

DOI: [10.34669/WI.WPP/8.2](https://doi.org/10.34669/WI.WPP/8.2)

ISSN: 2940-8490

LICENSE:

This paper is available open access and is licensed under Creative Commons Attribution 4.0 (CC BY 4.0): <https://creativecommons.org/licenses/by/4.0/>

For inquiries regarding this paper please contact Robert Peter at:
robert.peter@weizenbaum-institut.de

This policy paper has been funded by the Stiftung Mercator.

The Weizenbaum-Institut e.V. adheres to the Code of Conduct of the EU Transparency Register, register ID 194885445254-71.

The Weizenbaum Institute is funded by the Federal Ministry of Education and Research of Germany (BMBF).

About the Weizenbaum Institute

The Weizenbaum Institute – The German Internet Institute analyzes, evaluates and shapes relevant aspects of the digitally networked society. With basic, interdisciplinary and problem-oriented research projects on the ethical, legal, economic, political and social dimensions of digital change and the exploration of concrete solutions, it provides politicians, business and civil society with evidence- and value-based options for action in order to shape digitalization in a sustainable, self-determined and responsible manner. The Institute is supported by a research network from Berlin and Brandenburg, which includes Freie Universität Berlin, Humboldt-Universität zu Berlin, Technische Universität Berlin, Berlin University of the Arts, and the University of Potsdam, as well as the Fraunhofer Institute for Open Communication Systems (FOKUS) and the WZB Berlin Social Science Center. The Weizenbaum Institute is funded by the German Federal Ministry of Education and Research (BMBF) and the State of Berlin. For more information, visit www.weizenbaum-institut.de.

About the authors

Ulrike Klinger is Professor for Digital Democracy at the European New School of Digital Studies, European University Viadrina, Frankfurt (Oder).

Contact: klinger@europa-uni.de

Jakob Ohme is Research Group Lead of the “Digital News Dynamics” group at the Weizenbaum Institute for the Networked Society, Berlin.

Contact: jakob.ohme@weizenbaum-institut.de

About this paper

The paper underwent an iterative development process stemming from two workshops held in March and June 2023. These workshops were organized by the authors and supported by Stiftung Mercator. Approximately 70 scientists and platform researchers from various fields actively contributed to the paper’s development. Concurrently, informal discussions occurred with scientists hailing from different European Union nations, as well as regulatory authorities in Germany. In response to the EU Commission’s Call for Evidence issued in May 2023, the authors submitted an interim status report, see Klinger & Ohme (2023). *Delegated Regulation on Data Access Provided for the Digital Services Act: Response to the Call for Evidence DG CNECT-CNECT F2 by the European Commission*. <https://doi.org/10.34669/WI.WPP/7>

What the Scientific Community Needs from Data Access under Art. 40 DSA

20 Points on Infrastructures, Participation, Transparency, and Funding¹

To ensure free and objective research on social media platforms and their impact on systemic risks to the economy and society, reliable and independent access to platform data with high data quality is essential.

Article 40 of the Digital Services Act (DSA) creates for the first time a clear regulation that grants science independence from individual platforms and improved **data quality**, thus ensuring that socially relevant aspects of digitization can be investigated appropriately, consistently, and **independently**. It makes it possible to respond more quickly and accurately to new issues and developments in an evidence-based manner, thus contributing to a **fair, digital public sphere** that considers societal risks and opportunities.

This policy paper aims to inform the expected Delegated Act of the EU Commission² as well as the legislative process for the German Digital Services Act (Digitale Dienste Gesetz) and to formulate necessities from the perspective of platform researchers. This perspective is of utmost importance, as research on systemic risks depends on the expertise of scientific actors. In the implementation and further legislative process at national and European level, the following 20 points are particularly important from a scientific perspective:

Who should have access?

- 1. Enable equal access for platform researchers:** Research on digitization and its impacts has been characterized by unequal access to resources and data. The goal must be to grant all platform researchers (if accredited) equal access to data. This will lead to greater diversity and independence in research. One-stop shops and research data centers can support this.

¹ We acknowledge the work of the contributors who helped shaping this paper with their valuable inputs and comments: Alexander Wehde, Andrea Lorenz, Daniela Stockmann, Erik Tuchtfield, Johannes Breuer, Judith Möller, Julia Niemann-Lenz, Katharina Kaesling, Katharina Kleinen-von Königslöw, Katrin Weller, Marie-Therese Sekwenz, Mario Haim, Matthias C. Kettemann, Matthias Spielkamp, Mia Berg, Michael Meyer-Resende, Philipp Darius, Philipp Lorenz-Spreen, Pia Sombetzki, Richard Kuchta, Simon Munzert, Sonja Schimmler, Valerie Hase.

² Cf. the detailed proposals on data access in Klinger & Ohme (2023). Delegated Regulation on Data Access Provided for the Digital Services Act: Response to the Call for Evidence DG CNECT-CNECT F2 by the European Commission. <https://doi.org/10.34669/WI.WPP/7>

2. **Allow access to non-EU researchers and journalists:** The DSA must not lead to the compartmentalization of European science. Therefore, access for non-European researchers in cooperation with European researchers must be possible. The access criteria described in Art. 40 para. 8 lit. b DSA should be interpreted broadly for access according to DSA 40 para. 12 for journalists who act in the sense of public education and information.

How can fair access work?

3. **Access must be free of charge and independent:** The current monetary exploitation of X's (formerly known as Twitter) API access clearly shows how important free access is for making research egalitarian and diverse. The DSA clearly puts the onus on the platforms to provide data access for science; the platforms should also bear the costs for this.
4. **Comprehensive data access for accredited researchers:** The following information should be available for vetted researchers as defined in Art. 40 para. 4, 8 DSA:
 - Information on data structures, sample (representativeness), and available platform data, e.g., as a list of variables, collected metadata, and operationalizations at data access
 - Information about recent or soon-planned changes to the data structures and the available data
 - Information about the exact processing, aggregation, and anonymization of the raw data
 - Data that includes internal weighting of data features for algorithmic sorting/selection and other internal decisions such as A/B testing or new platform functionality
 - Platform-specific, publicly accessible data in machine-readable form, in real-time and historically, and available at any time
 - Platform-specific, personal data in machine-readable form, historically, only after a successful vetting process of researchers and, if applicable, consent of the users (for personal, non-public data)
 - Cross-platform data prepared in a standardized, machine-readable form

5. **Custom-fit access modes**³: Access to historical and real-time data (streaming APIs) should be as complete as possible via data interfaces (APIs) provided free of charge. Different levels of data sensitivity⁴ should offer differently formulated accesses and procedures⁵. At high sensitivity levels in so-called data clean rooms⁶, lower level in virtual laboratory environments (VLEs)⁷. Other possibilities for access modes are closed collections, ad libraries, or data donations, which can also be integrated via APIs or other interfaces, for example. For experimental studies and algorithmic audits, it should be possible to set up so-called sock puppets (i.e., online identities explicitly created for research purposes).
6. **Flexible access design**: Art. 40 para. 8 DSA so far only provides for individual access, i.e., by individual researchers. However, the application procedure should be more flexible. An institutional anchoring with a longer term and the possibility of adding new team members would be optimal. Data access should also be usable for academic teaching as well as early-career researchers who ensure the future of platform research. A digital self-disclosure of the research project's independence from commercial interests (DSA 40. para. 8, b) and funding of the research project (DSA 40. para. 8, c) should be sufficient for the approval process under DSA 40. para. 12. Conditions DSA 40. para. 8, d) (data security and confidentiality requirements) and e) (access necessary and proportionate) disproportionately restrict access, as this concerns publicly accessible data. It must be clarified to what extent or how collected data can be shared (e.g., for replication studies) or reused (e.g., in consortia).

How should the approval process be designed?

7. **Conduct of the approval process preferably by the local Digital Services Coordinator (DSC)**: the approval process should be conducted in a timely manner (within 7 days), ideally by (or in close coordination with) the DSC of the applicant's home country, to allow for relief of busy DSCs (likely in Ireland). An appeal mechanism to this process should be offered through an independent advisory board attached to the local DSC (see draft bill for the German Digitale Dienste Gesetz, para. 22), due to greater knowledge of national institutions and processes.

³ See note 1.

⁴ See recital 51 GDPR <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679&qid=1694877783373>

⁵ See the Hertie School Data Science Lab's detailed proposal, Implementing Data Access of the DSA, https://hertie-school-f4e6.kxcdn.com/fileadmin/2_Research/2_Research_directory/Research_Centres/Centre_for_Digital_Governance/5_Papers/Implementing_Data_Access_Darius_Stockmann_2023.pdf, last accessed 04/24/2023.

⁶ Data clean rooms are safe, secure environments where personally identifiable information (PII) is cleaned and processed so that it can be made available for a variety of data analysis purposes.

⁷ The virtual laboratory environment is an interactive environment for creating and performing simulated experiments and analyses.

- 8. Distinguishing data sensitivity for access requests:** In the approval process, a distinction should be made between access to medium- and highly-sensitive data as defined in Art. 40 para. 4 DSA and prioritized access to publicly available data as defined in Art. 40 para. 12 DSA.
- 9. Prioritize publicly available data:** Publicly accessible data from the platforms must be made available at all times and without a vetting process, according to Art. 40 (12) DSA. This refers to data not restricted by the user and can be viewed by freely browsing the platform. Technically, this means, above all, the provision of real-time access through APIs (data interfaces) and dashboards. A definition of semi-public data is needed, e.g., “private” Facebook or Telegram groups with thousands of members. Art. 40(12) DSA should be implemented as a priority (also concerning upcoming EU and US elections) and will already cover a large part of scientific data needs. This requires early, clear guidance from the EU Commission on the implementation of Art. 40(12).
- 10. Low-threshold, modular application procedure:** For simplification and standardization, we propose a modular principle in which data and usage interests, as well as data protection measured, are queried in a standardized manner. Users should only need to specify general research interests, similar to the former Twitter Academic API, and also allow exploratory research. Information about what specific data/variables are available should be easily accessible. In addition to German, the application should be available at least in English to enable inclusive data access. Access should be granted via login to an online tool where data from multiple platforms can be accessed simultaneously, also to ensure much-needed cross-platform research on systemic risks.
- 11. Expand infrastructures and resources, build competence through close exchange:** Uniform and clear infrastructure requirements must be formulated for research institutions so that appropriate preparations can be initiated, such as a secure server infrastructure. There should be contact persons for questions and support in access for those who do not have the necessary technical skills. There is a need for targeted research funding programs at the European and national levels that explore systemic risks based on requested platform data. These highly specialized and resource-intensive projects cannot be carried out under existing funding programs.
- 12. Involvement of the scientific community in peer review processes:** Applications that have already been submitted in this way in a similar form should be reviewed primarily for formalities, ethics, and feasibility of data sharing before direct, timely approval is given. Applications for which there is no precedent should be reviewed more comprehensively, in terms of content, in a peer review system (and may also be reviewed only formally in the future). This reduces review cases and resources. Adequate compensation for researchers involved in the peer review process should be guaranteed.

How can data quality be safeguarded?

- 13. Advisory board structure for the DSC:** The planned, independent advisory board of the German DSC should advise on overarching strategic issues with regard to the implementation of the DSA and allow the inclusion of scientific issues. In order to fulfill these tasks, researchers who can demonstrate expertise in the field of empirical research with platform data should be represented on the advisory board. Duplicate structures with a planned Intermediary Advisory Mechanism according to Art. 40(13) DSA should be avoided.
- 14. Conditions for the quality of data preparation:** The preparation of the data should be the main responsibility of the platforms and the form of preparation should meet uniform standards in order to meet the listed requirements. The catalog of interfaces and data points should not be defined unilaterally by the platforms but should be determined in a standardized manner through a process independent of the platforms and involving the researchers (e.g., independent advisory mechanism (IAM) to support data sharing under Art. 40(13) DSA). Standardization is necessary to ensure that platform-comparative and cross-platform studies are methodologically valid⁸. This also applies to the provision of text, image/video and audio to enable platform comparisons. Direct forwarding of research questions to the platforms should be prevented to exclude any influence on the part of the platforms.
- 15. Documentation and transparency of variables, measurements, and data collection:** Information about when collected data and measurements from platforms were changed over time should be retrievable. Research needs more transparency about metrics, e.g., what does a “like” or “views” mean on different platforms? As uniform as possible variable naming and metadata structure should be strived for – also for “public” data (Art. 40(12) DSA)- here the opportunity for cross-platform simplification and harmonization of indicators exists.
- 16. Right to scraping for scientific quality assurance:** Indispensable for scientifically sound quality control of data is the “right to scrap”. Without the right to scrap, the data access offered by platforms cannot be validated. Incomplete data sets, such as those repeatedly provided by platforms in the past (e.g., the Facebook Ad Archive), are hardly useful for research and, in the worst case, are harmful because they falsify the results. Data donations by users are also essential for research and quality control. The possibilities for this should be expanded and simplified.

⁸ Cf. also item 1 in https://algorithmwatch.org/en/wp-content/uploads/2023/05/Open_letter_DSA.pdf

- 17. Platform data quality reports:** Platforms should publish regular reports that include aspects of the data quality provided (e.g., validation by external bodies, own data cleansing initiatives, completeness of data, etc.).
- 18. Insight into the platform data pool for research purposes, including non-European:** Art. 40 para. 4 DSA is intended to provide access to the entire platform data pool, as systemic risks can only be identified if there is clarity about the totality of all data available to the platforms. Viewing and pre-structuring of existing platform data should be done centrally by the DSCs. Non-European platform data must be accessible, e.g., in the context of manipulated social media data and disinformation. The market location principle (*lex loci solutionis*) must be applied here.⁹
- 19. Oversight:** An authority must be appointed that can independently check the quality of the data preparation (possibly an Intermediary Advisory Mechanism, which would have to be provided with extra resources for this purpose, cf. Klinger & Ohme). Highly qualified data scientists are needed to assess the data provided by the platforms in terms of quality and potential uses, who can provide a professional assessment, including on the open question of what the data queries of approved researchers may even contain in terms of societal risks. In the event of poor data quality, a short-term and transparent (complaint) procedure should be available for readjustment; in case of doubt, fine procedures should also be able to follow.
- 20. Serious penalty mechanisms are good, but incentive structures are better:** Deliberate delays in data access by a platform should result in penalty proceedings. However, we recommend not only to work with serious sanction mechanisms, but also to create incentives for platforms in the medium term by providing sufficient resources for data access (e.g., special compliance). If the platforms refuse to grant access, research institutions should be able to proceed by means of an action for failure to act or a refusal counterclaim, so that mechanisms and corresponding resources should also be provided for this.

Signatories (institutions)

⁹ Since the scope of the DSA applies according to the market location principle (regardless of the platform's place of establishment), non-European data from non-European platforms are also included, as long as these platforms offer services in the EU or to EU citizens.

OKTOBER 2023

Weizenbaum-Institut für die vernetzte Gesellschaft

Was die Wissenschaft im Rahmen des Datenzugangs nach Art. 40 DSA braucht

**20 Punkte zu Infrastrukturen, Beteiligung, Transparenz
und Finanzierung**

Impressum

Was die Wissenschaft im Rahmen des Datenzugangs nach Art. 40 DSA braucht.
20 Punkte zu Infrastrukturen, Beteiligung, Transparenz und Finanzierung.
Weizenbaum Policy Paper 8, Oktober 2023

AUTOR:INNEN:

Ulrike Klinger \ Europa-Universität Viadrina \ klinger@europa.uni.de
Jakob Ohme \ Weizenbaum-Institut \ jakob.ohme@weizenbaum-institut.de

HERAUSGEBER:

Weizenbaum-Institut für die vernetzte Gesellschaft
Hardenbergstraße 32 \ 10623 Berlin \ Tel.: +49 30 700141-001
info@weizenbaum-institut.de \ www.weizenbaum-institut.de

REDAKTION:

Moritz Buchner \ Weizenbaum-Institut
Christina Dinar \ Leibniz-Institut für Medienforschung | HBI
Robert Peter \ Weizenbaum-Institut
Kaltrina Shala \ Weizenbaum-Institut

DOI: [10.34669/WI.WPP/8.1](https://doi.org/10.34669/WI.WPP/8.1)

ISSN: 2940-8490

COPYRIGHT:

Diese Veröffentlichung ist unter der Creative-Commons-Lizenz „Namensnennung 4.0 International“ (CC BY 4.0) lizenziert: <https://creativecommons.org/licenses/by/4.0/>

Für Rückfragen zu diesem Positionspapier wenden Sie sich bitte an Robert Peter unter robert.peter@weizenbaum-institut.de

Dieses Policy Paper wurde durch die Stiftung Mercator gefördert.

Der Weizenbaum-Institut e.V. hält sich an den Verhaltenskodex für Interessensvertreterinnen und Interessensvertreter im Rahmen des Lobbyregistergesetzes. Die Registernummer des Weizenbaum-Institut e.V. im Lobbyregister des Bundestages lautet: R003857.

Das Weizenbaum-Institut wird vom Bundesministerium für Bildung und Forschung (BMBF) gefördert.

Das Weizenbaum-Institut für die vernetzte Gesellschaft erforscht interdisziplinär und grundlagenorientiert den Wandel der Gesellschaft durch die Digitalisierung und entwickelt Gestaltungsoptionen für Politik, Wirtschaft und Zivilgesellschaft. Ziel ist es, die Dynamiken, Mechanismen und Implikationen der Digitalisierung besser zu verstehen. Hierzu werden am Weizenbaum-Institut die ethischen, rechtlichen, ökonomischen und politischen Aspekte des digitalen Wandels untersucht. Damit wird eine empirische Grundlage geschaffen, die Digitalisierung verantwortungsvoll zu gestalten. Um Handlungsoptionen für Politik, Wirtschaft und Zivilgesellschaft zu entwickeln, verknüpft das Weizenbaum-Institut die interdisziplinäre problemorientierte Grundlagenforschung mit der Exploration konkreter Lösungen und dem Dialog der Gesellschaft. Weitere Informationen: www.weizenbaum-institut.de

Über die Autor:innen

Ulrike Klinger ist Professorin für Digitale Demokratie an der European New School of Digital Studies, Europa-Universität Viadrina, Frankfurt (Oder).

Kontakt: klinger@europa-uni.de

Jakob Ohme ist Leiter der Forschungsgruppe „Dynamiken digitaler Nachrichtenvermittlung“ am Weizenbaum-Institut für die vernetzte Gesellschaft, Berlin.

Kontakt: jakob.ohme@weizenbaum-institute.de

Über dieses Policy Paper

Das Papier ist in einem iterativen Prozess auf Basis von zwei Workshops im März und Juni 2023 entstanden, organisiert durch die Autor:innen und gefördert von der Stiftung Mercator. Insgesamt waren ca. 70 Wissenschaftler:innen und Plattformforschende aus unterschiedlichen Disziplinen an der Entstehung beteiligt. Gleichzeitig fanden Hintergrundgespräche mit Wissenschaftler:innen aus anderen EU-Ländern und Regulatoren in Deutschland statt. Einen Zwischenstand haben die Autor:innen als Response zum Call for Evidence der EU-Kommission im Mai 2023 eingereicht: Klinger & Ohme (2023). *Delegated Regulation on Data Access Provided for the Digital Services Act: Response to the Call for Evidence DG CNECT-CNECT F2 by the European Commission*. <https://doi.org/10.34669/WI.WPP/7>

Was die Wissenschaft im Rahmen des Datenzugangs nach Art. 40 DSA braucht

20 Punkte zu Infrastrukturen, Beteiligung, Transparenz und Finanzierung¹⁰

Um eine freie und objektive Forschung über Social-Media-Plattformen und deren Auswirkungen auf systemische Risiken für Wirtschaft und Gesellschaft zu gewährleisten, ist ein verlässlicher und unabhängiger Zugang zu Plattform-Daten mit hoher Datenqualität unabdingbar.

Artikel 40 des Digital Services Act (DSA) schafft erstmals eine klare Regelung, die der Wissenschaft **Unabhängigkeit** von den einzelnen Plattformen und eine verbesserte **Datenqualität** gewährt und so sicherstellt, dass gesellschaftlich relevante Aspekte der Digitalisierung angemessen, konsistent und unabhängig untersucht werden können. Er ermöglicht, schneller und passgenau auf neue Fragestellungen und Entwicklungen evidenzbasiert zu reagieren und so zu einer **fairen, digitalen Öffentlichkeit** beizutragen, die sowohl gesellschaftliche Risiken als auch ihre Chancen in den Blick nimmt.

Dieses Policy Paper zielt darauf, den erwarteten Delegated Act der EU-Kommission¹¹ als auch das Gesetzgebungsverfahren zum deutschen Digitale-Dienste-Gesetz zu informieren und Notwendigkeiten aus Sicht von Plattformforschenden zu formulieren. Diese Sichtweise ist von größter Bedeutung, da von der Expertise wissenschaftlicher Akteure die Erforschung der systemischen Risiken abhängt. Bei der Umsetzung und im weiteren Gesetzgebungsverfahren auf nationaler und europäischer Ebene sind aus Sicht der Wissenschaft folgende 20 Punkte besonders wichtig:

Wer soll Zugang haben?

21. Gleiche Zugänge für Plattformforschende ermöglichen: Die Forschung zur Digitalisierung und ihren Auswirkungen ist bisher von ungleichem Zugang zu Ressourcen und Daten geprägt. Ziel muss es sein, allen Plattformforschenden (sofern akkreditiert) die gleichen Datenzugänge einzuräumen. Dies führt zu größerer Vielfalt und

¹⁰ Wir bedanken uns bei den Mitwirkenden, die mit ihren wertvollen Beiträgen und Kommentaren zur Gestaltung dieses Papiers beigetragen haben: Alexander Wehde, Andrea Lorenz, Daniela Stockmann, Erik Tuchtfeld, Johannes Breuer, Judith Möller, Julia Niemann-Lenz, Katharina Kaesling, Katharina Kleinen-von Königslöw, Katrin Weller, Marie-Therese Sekwenz, Mario Haim, Matthias C. Kettemann, Matthias Spielkamp, Mia Berg, Michael Meyer-Resende, Philipp Darius, Philipp Lorenz-Spreen, Pia Sombetzki, Richard Kuchta, Simon Munzert, Sonja Schimmler, Valerie Hase.

¹¹ Vgl. die ausführlichen Vorschläge zu den Datenzugängen in Klinger & Ohme (2023). Delegated Regulation on Data Access Provided for the Digital Services Act: Response to the Call for Evidence DG CNECT-CNECT F2 by the European Commission. <https://doi.org/10.34669/WI.WPP/7>

Unabhängigkeit in der Forschung. Zentrale Anlaufstellen und Forschungsdatenzentren können dabei unterstützen.

22. Nicht-EU-Forschenden und Journalist:innen Zugang gewähren: Der DSA darf nicht zur Abschottung der europäischen Wissenschaft führen. Daher müssen Zugänge für nicht-europäische Forschende in Kooperationen mit europäischen Forschenden ermöglicht werden. Die in Art. 40 Abs. 8 lit. b DSA beschriebenen Zugangskriterien sollten für Zugänge nach DSA 40. Abs. 12 bei Journalist:innen, die im Sinne der öffentlichen Aufklärung und Information handeln, weiträumig interpretiert werden.

Wie kann ein fairer Zugang funktionieren?

23. Der Zugang muss kostenfrei und unabhängig sein: Die aktuelle monetäre Verwertung des X (vorher bekannt als Twitter) API-Zugangs zeigt deutlich, wie wichtig ein kostenfreier Zugang ist, um Forschung egalitär und vielfältig gestalten zu können. Der DSA nimmt die Plattformen klar in die Pflicht, Datenzugang für Wissenschaft bereitzustellen; die Plattformen sollen auch die Kosten dafür tragen.

24. Umfänglicher Datenzugang für zugelassene Forschende: Für akkreditierte Forschende (*vetted researchers*) i.S.d. Art. 40 Abs. 4, 8 DSA sollten folgende Informationen zur Verfügung stehen:

- Informationen über Datenstrukturen, Stichproben(-repräsentativität) und verfügbare Plattform-Daten, z.B. als Liste von Variablen, erhobenen Metadaten und Operationalisierungen beim Data Access
- Informationen über kürzlich getätigte oder demnächst geplante Änderungen an den Datenstrukturen und den verfügbaren Daten
- Informationen über die genaue Aufbereitung, Aggregation und Anonymisierung der Rohdaten
- Daten, die das interne gewichtete von Datenfeatures für die algorithmische Sortierung/Auswahl und andere interne Entscheidungen wie A/B testing oder neue Plattformfunktionalitäten beinhalten
- Plattformspezifische, öffentlich zugängliche Daten in maschinenlesbarer Form, in Echtzeit und historisch sowie jederzeit verfügbar
- Plattformspezifische, personenbezogene Daten in maschinenlesbarer Form, historisch, nach Prüfverfahren (*vetting process* und ggf. Zustimmung der Nutzenden bei personenbezogenen, nicht öffentlichen Daten)
- Plattformübergreifende Daten in maschinenlesbarer Form standardisiert aufbereitet

25. Passgenaue Zugangsmodi:¹² Über kostenlos bereitgestellte Datenschnittstellen (APIs) soll Zugang zu möglichst vollständigen historischen Daten und Echtzeit-Daten (Streaming APIs) bestehen. Verschiedene Stufen der Daten-Sensibilität¹³ sollten unterschiedlich ausformulierte Zugänge und Verfahren bieten.¹⁴ Bei hohen Sensibilitätsstufen in sogenannten Daten-Reinräumen (*data clean rooms*),¹⁵ niedrigstufiger in virtuellen Laboratorien.¹⁶ Weitere Möglichkeiten für Zugangsmodi: Abgeschlossene Sammlungen, Ad Libraries oder Datenspenden, die z.B. auch über APIs oder andere Schnittstellen integriert werden können. Für Experimentalstudien und Algorithmic Audits sollten sog. *sock puppets* (d.h. eigens für Forschungszwecke kreierte Online-Identitäten) eingerichtet werden können.

26. Flexible Zugangsgestaltung: Art. 40 Abs. 8 DSA sieht bisher nur individuelle Zugänge, also von einzelnen Forschenden, vor. Das Antragsverfahren sollte aber flexibler gestaltet sein. Optimal wäre eine institutionelle Verankerung mit längerer Laufzeit und der Möglichkeit, auch neue Teammitglieder hinzuzufügen. Der Datenzugang sollte auch für die akademische Lehre sowie für Early-Career-Researchers nutzbar sein, die die Zukunft der Plattformforschung sicherstellen. Für die Zulassungsprüfung unter DSA 40. Abs. 12 sollte eine digitale Selbstauskunft zur Unabhängigkeit des Forschungsprojekts von kommerziellen Interessen (DSA 40. Abs. 8, b) und die Finanzierung des Forschungsprojekts (DSA 40. Abs. 8, c) ausreichen. Die Konditionen DSA 40. Abs. 8, d) (data security and confidentiality requirements) und e) (access necessary and proportionate)“ schränken den Zugang unverhältnismäßig ein, da es sich hier um öffentlich zugängliche Daten handelt. Zu klären ist, inwiefern bzw. wie erhobene Daten geteilt (z.B. für Replikationsstudien) oder weiterverwendet werden können (z.B. in Konsortien).

Wie sollte der Zulassungsprozess gestaltet sein?

27. Durchführung des Zulassungsverfahrens bevorzugt durch den örtlichen Digital Services Coordinator (DSC): Der Zulassungsprozess sollte zeitnah (innerhalb von 7 Tagen) durchgeführt werden, idealerweise vom (oder in enger Abstimmung mit dem)

¹² Ebd.

¹³ Siehe Erwägungsgrund 51 DSGVO <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679&qid=1694877783373>

¹⁴ Vgl. den ausführlichen Vorschlag des Hertie School Data Science Lab, Implementing Data Access of the DSA, https://hertieschool-f4e6.kxcdn.com/fileadmin/2_Research/2_Research_directory/Research_Centres/Centre_for_Digital_Governance/5_Papers/Implementing_Data_Access_Darius_Stockmann_2023.pdf, zuletzt abgerufen am 24.04.2023.

¹⁵ Daten-Reinräume sind sichere, geschützte Umgebungen, in denen Daten zur persönlichen Identifizierung (PII) bereinigt und verarbeitet werden, damit sie für eine Vielzahl von Datenanalysezwecken zur Verfügung gestellt werden können.

¹⁶ Das virtuelle Laboratorium ist eine interaktive Umgebung zur Erstellung und Durchführung simulierter Experimente und Analysen.

DSC des Heimatlandes der Antragstellenden, um eine Entlastung hochbeanspruchter DSCs (voraussichtlich in Irland) zu ermöglichen. Ein Beschwerdemechanismus zu diesem Verfahren sollte durch einen am örtlichen DSC angegliederten unabhängigen Beirat (siehe Referentenentwurf für das Digitale-Dienste-Gesetz § 22) angeboten werden, aufgrund von größerer Kenntnis nationaler Institutionen und Vorgänge.

28. Sensitivitätsgrad bei Datenzugang unterscheiden: Im Zulassungsprozess sollte zwischen dem Zugang zu mittel- und hochsensiblen Daten i.S.d. Art. 40 Abs. 4 ff. DSA und dem priorisierten Zugang zu öffentlich zugänglichen Daten i.S.d. Art. 40 Abs. 12 DSA unterschieden werden.

29. Öffentlich zugängliche Daten priorisieren: Öffentlich zugängliche Daten der Plattformen sind jederzeit und ohne Zulassungsverfahren (*vetting process*) bereitzustellen, nach Art. 40 Abs. 12 DSA. Damit gemeint sind Daten, die nicht nutzerseitig eingeschränkt und durch freies Browsing der Plattform einsehbar sind. Technisch bedeutet dies vor allem die Bereitstellung von Echtzeit-Zugängen durch APIs (Datenschnittstellen) und Dashboards. Es bedarf einer Definition halb-öffentlicher Daten, z.B. „privater“ Facebook- oder Telegram-Gruppen mit tausenden Mitgliedern. Art. 40 Abs. 12 DSA sollte prioritär (auch mit Hinblick auf anstehende Wahlen in der EU und den USA) umgesetzt werden und wird einen Großteil der wissenschaftlichen Datenbedarfe bereits abdecken. Hierzu benötigt es baldige, klare Leitlinien der EU-Kommission zur Umsetzung von Art. 40(12).

30. Niedrigschwelliges Baukastenprinzip in der Antragsstellung: Zur Vereinfachung und Standardisierung schlagen wir ein Baukasten-Prinzip vor, bei dem Daten- und Nutzungsinteressen sowie die Datenschutzvorkehrungen standardisiert abgefragt werden können. Nutzende sollten nur generelle Forschungsinteressen angeben müssen, ähnlich der ehemaligen Twitter Academic API, um auch explorative Forschung zu ermöglichen. Informationen darüber, welche konkreten Daten/Variablen verfügbar sind, sollten leicht zugänglich sein. Neben Deutsch sollte der Antrag mindestens auf Englisch zur Verfügung stehen, um einen inklusiven Datenzugang zu ermöglichen. Der Zugang sollte über den Login zu einem Online-Tool erfolgen, in dem gleichzeitig Daten von mehreren Plattformdaten zugänglich gemacht werden können, auch um dringend benötigte plattformübergreifende Forschung zu systemischen Risiken zu gewährleisten.

31. Infrastrukturen und Ressourcen erweitern, Kompetenzaufbau durch engen Austausch: Es müssen einheitliche und klare Infrastrukturanforderungen an Forschungsinstitutionen formuliert werden, damit entsprechende Vorbereitungen eingeleitet werden können, so z.B. eine gesicherte Serverinfrastruktur. Es sollte Ansprechpersonen für Fragen geben sowie Unterstützung im Zugang für diejenigen, die nicht die erforderlichen technischen Kompetenzen besitzen. Es bedarf gezielter Forschungsförderprogramme auf europäischer und nationaler Ebene, die systemische

Risiken anhand von beantragten Plattformdaten erforschen. Diese hochgradig speziellen und ressourcenintensiven Projekte lassen sich nicht im Rahmen bestehender Förderprogrammen durchführen.

- 32. Beteiligung der Wissenschaft an Begutachtungen durch Peer-Review-Verfahren:** Anträge, die so in ähnlicher Form schon einmal gestellt worden sind, sollten vor allem auf Formalien, Ethik und Umsetzbarkeit der Datenweitergabe geprüft werden, bevor eine direkte, zeitnahe Genehmigung erfolgt. Anträge, für die es keine Präzedenz gibt, sollten in einem Peer-Review-System umfassender, inhaltlich begutachtet werden (und können in Zukunft ebenfalls nur noch formal geprüft werden). Dies reduziert Prüffälle und Ressourcen. Eine angemessene Aufwandsentschädigung für beteiligte Forschende am Peer-Review Verfahren sollte gewährleistet sein.

Wie kann Datenqualität gewährleistet werden?

- 33. Beiratsstruktur für den DSC:** Der geplante, unabhängige Beirat des deutschen DSC soll bei übergeordneten strategischen Fragestellungen hinsichtlich der Umsetzung des DSA beraten und wissenschaftliche Fragestellungen einbringen. Um diesen Aufgaben gerecht zu werden, sollten Forschende, die Expertise im Bereich empirischer Forschung mit Plattformdaten ausweisen können, im Beirat vertreten sein. Doppelstrukturen mit einem geplanten Intermediary Advisory Mechanism nach Art. 40(13) DSA sind zu vermeiden.
- 34. Bedingungen für die Qualität der Datenaufbereitung:** Die Aufbereitung der Daten sollte in der Hauptverantwortung der Plattformen liegen und die Form der Aufbereitung sollte einheitliche Standards erfüllen, um den aufgeführten Anforderungen zu entsprechen. Der Katalog der Schnittstellen und Datenpunkte sollte nicht einseitig von den Plattformen definiert, sondern standardisiert über einen von den Plattformen unabhängigen Prozess ermittelt werden, an dem die Forschenden beteiligt sind (z.B. unabhängiger Beratungs-Mechanismus (IAM) zur Unterstützung der Datenweitergabe nach Art. 40 Abs. 13 DSA). Die Standardisierung ist notwendig, damit plattformvergleichende und plattformübergreifende Studien methodologisch valide sind.¹⁷ Dies gilt auch für die Bereitstellung von Text, Bild/Video und Audio, um Plattformvergleiche zu ermöglichen. Eine direkte Weiterleitung der Forschungsfragen an die Plattformen sollte unterbunden werden, um eine Einflussnahme seitens der Plattformen auszuschließen.
- 35. Dokumentation und Transparenz der Variablen, Messungen und Datenerhebung:** Informationen darüber, wann erhobene Daten und Messungen von Plattformen

¹⁷ Vgl. auch Punkt 1 in https://algorithmwatch.org/en/wp-content/uploads/2023/05/Open_letter_DSA.pdf

über die Zeit verändert wurden, sollten abrufbar sein. Die Forschung benötigt mehr Transparenz über Metriken, z.B. was bedeutet ein “Like” oder “Views” auf verschiedenen Plattformen? Eine möglichst einheitliche Variablenbenennung und Metadatenstruktur ist – auch für „öffentliche“ Daten – anzustreben, hier besteht die Möglichkeit der plattformübergreifenden Vereinfachung und Harmonisierung von Indikatoren.

- 36. Recht auf Scraping zur wissenschaftlichen Qualitätssicherung:** Unerlässlich für die wissenschaftlich fundierte Qualitätskontrolle der Daten ist das “Recht auf Scraping”. Ohne das Recht auf Scraping können die von Plattformen angebotenen Datenzugänge nicht validiert werden. Unvollständige Datensätze, wie sie in der Vergangenheit immer wieder von den Plattformen zur Verfügung gestellt wurden (z.B. das Facebook Ad Archive), sind für die Forschung kaum sinnvoll nutzbar und im schlimmsten Fall schädlich, weil sie die Ergebnisse verfälschen. Auch Datenspenden durch Nutzende sind für Forschung und Qualitätskontrolle unerlässlich. Die Möglichkeiten dafür sollten ausgebaut und vereinfacht werden.
- 37. Berichte der Plattformen zur Datenqualität:** Die Plattformen sollten regelmäßig Berichte veröffentlichen, die auch Aspekte der bereitgestellten Datenqualität (z.B. Validierung durch externe Gremien, eigene Initiativen der Datenbereinigung, Vollständigkeit von Daten, etc.) beinhalten.
- 38. Einblick in den Plattform-Datenpool für Forschungszwecke, auch außereuropäisch:** Mit Art. 40 Abs. 4 DSA soll Zugang zum gesamten Plattform-Datenpool gewährt werden, da systemische Risiken nur bei Klarheit über die Gesamtheit aller den Plattformen zur Verfügung stehenden Daten erkannt werden können. Das Einsehen und die Vorstrukturierung existierender Plattform-Daten sollten durch die DSCs zentral erfolgen. Außereuropäische Plattform-Daten müssen zugänglich sein, z.B. im Kontext von manipulierten Social-Media-Daten und Desinformation. Das Marktortprinzip muss hier Anwendung finden.¹⁸
- 39. Oversight:** Es muss eine Instanz eingesetzt werden, die die Qualität der Datenaufbereitung unabhängig prüfen kann (evtl. Intermediary Advisory Mechanism, der dafür mit extra Ressourcen ausgestattet werden müsste, vgl. Klinger & Ohme¹). Für die Einschätzung der von den Plattformen gelieferten Daten in Bezug auf die Qualität und die Nutzungsmöglichkeiten sind hochqualifizierte Data Scientists nötig, die eine fachgerechte Einschätzung abgeben können, auch in der offenen Frage, was die Datenabfragen der zugelassenen Forschenden in Bezug auf gesellschaftliche Risiken überhaupt beinhalten können. Bei mangelhafter Datenqualität sollte ein kurzfristiges und transparentes (Beschwerde-)Verfahren zum Nachjustieren bereitstehen.

¹⁸ Da der Geltungsbereich des DSA nach dem Marktortprinzip (unabhängig vom Niederlassungsort der Plattform) Anwendung findet, sind auch außereuropäische Daten von außereuropäischen Plattformen inbegriffen, solange diese Plattformen Dienste in der EU oder für EU-Bürger anbieten.

40. Ernsthafte Bußgeldverfahren sind gut, Anreizstrukturen aber besser: Bewusste Verzögerungen beim Datenzugang durch eine Plattform sollten ein Bußgeldverfahren zur Folge haben. Jedoch empfehlen wir nicht nur, mit ernstzunehmenden Sanktionsmechanismen zu arbeiten, sondern auch mittelfristig Anreize für die Plattformen zu schaffen, indem ausreichend Ressourcen für den Datenzugang zur Verfügung gestellt werden (z.B. besondere Compliance). Bei Verweigerung der Plattformen, Zugriff zu gewähren, sollten Forschungseinrichtungen mittels Untätigkeitsklage oder Versagungsklage vorgehen können, sodass auch hierfür Mechanismen und entsprechende Ressourcen bereitgestellt werden sollten.

Signatories (institutions)

Weizenbaum-Institut für die vernetzte Gesellschaft, Berlin

European New School, Europa-Universität Viadrina, Frankfurt (Oder)

AlgorithmWatch

Bayerisches Forschungsinstitut für Digitale Transformation | bidt

Bundesarbeitsgemeinschaft »Gegen Hass im Netz«

Center for Advanced Internet Studies (CAIS)

Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPK)

D64 – Zentrum für Digitalen Fortschritt

Democracy Reporting International (DRI)

GESIS Leibniz-Institut für Sozialwissenschaften e.V. Stiftung Neue Verantwortung (SNV)

Signatories (individuals)

Prof. Ulrike Klinger, European New School, Europa-Universität Viadrina, Frankfurt (Oder)

Dr. Jakob Ohme, Head of Research Group, Weizenbaum Institute, Berlin

Mia Berg, Research Associate, Ruhr-Universität Bochum

Dr. Johannes Breuer, Senior Researcher, GESIS – Leibniz-Institut für Sozialwissenschaften & Center for Advanced Internet Studies (CAIS)

Kilian Bühling, Postdoctoral Researcher, Freie Universität Berlin, Weizenbaum-Institut

Dr. Mike Cowburn, Postdoctoral Researcher, European University Viadrina

Dr. Philipp Darius, Postdoctoral Researcher, Center for Digital Governance, Hertie School

Prof. Mario Haim, Professor, Ludwig-Maximilians-Universität, München

David Hartmann, Weizenbaum-Institut

Dr. Valerie Hase, Postdoctoral Researcher, Ludwig-Maximilians-Universität München

JProf. Katharina Kaesling, Technische Universität Dresden

Peter Kahlert, European New School of Digital Studies

Prof. Matthias C. Kettmann, Professor, Universität Innsbruck, Alexander von Humboldt Institut für Internet und Gesellschaft, Leibniz-Institut für Medienforschung | Hans-Bredow-Institut

Prof. Katharina Kleinen-von Königslöw, Professorin, Universität Hamburg

Julian Kohne, GESIS - Leibniz Institute for the Social Sciences, Department of Computational Social Science, Team Designed Digital Data

Richard Kuchta, Analyst and Researcher, Democracy Reporting International

Prof. Jana Lasser, RWTH Aachen, Lehrstuhl für Methodik und Theorie computerbasierter Geistes- und Sozialwissenschaften

Andrea Lorenz, Research Associate, Universität Hamburg

Dr. Philipp Lorenz-Spreen, Research Scientist, Max-Planck-Institut für Bildungsforschung, Berlin

Marie-Therese Sekwenz, PhD Candidate, TU Delft/ Technology Policy and Management/ AI Futures Lab

Prof. Klaus, Meier, Katholische Universität Eichstätt-Ingolstadt

Michael, Meyer-Resende, Co-founder & Executive Director, Democracy Reporting International (CRI), Berlin

Prof. Judith Möller, Professorin, Leibniz-Institut für Medienforschung | Hans-Bredow-Institut, Universität Hamburg

Prof. Simon Munzert, Professor, Data Science Lab, Hertie School, Berlin

Prof. Christoph Neuberger, Wissenschaftlicher Geschäftsführer & Direktor, Weizenbaum-Institut, Berlin; Professor, Freie Universität Berlin

Dr. Julia Niemann-Lenz, Senior Research Associate, Universität Hamburg

Dr. Douglas Parry, Stellenbosch University

Prof. Jan-Hendrik Passoth, European New School, Viadrina Universität Frankfurt/Oder

Dr. Sonja Schimmler, Forschungsgruppenleiterin, Weizenbaum-Institut, Berlin

Lukas Seiling, Student Researcher, Digital News Dynamics, Weizenbaum Institute for the Networked Society

Pia Sombetzki, Policy & Advocacy Manager, AlgorithmWatch

Matthias Spielkamp, Co-founder & Executive Director, AlgorithmWatch

Prof. Daniela Stockmann, Professorin, Center for Digital Governance, Hertie School, Berlin

Christian Strippel, Forschungsgruppenleiter, Weizenbaum-Institut, Berlin

Erik Tuchtfield, Max-Planck-Institut für ausländisches öffentliches Recht und Völkerrecht, Heidelberg

Alexander Wehde, Student Assistant, Forschungsstelle für Rechtsfragen neuer Technologien sowie Datenrecht (ForTech) e.V.

Dr. Katrin Weller, Senior Researcher, GESIS Leibniz-Institut für Sozialwissenschaften