

IRDT PAPERSERIES Nr. 14

Opinion on the delegated act on Art. 40 DSA by the research projects DIGIT's EU and TWON and the Digital Law Institute Trier

Dr. Max Dregelies; Pia Diemath; Julian Gschneidner; Fabian Hoffmanns; Prof. Dr. Pascal Jürgens; Prof. Dr. Christian Nuernbergk; Cosima Pfannschmidt; Prof. Dr. Benjamin Raue; Prof. Dr. Achim Rettinger; Prof. Dr. Damian Trilling; Prof. Dr. Antje von Ungern Sternberg

Version 1.0 (12.12.2024), CC BY 4.0.

The opinion has been prepared by the two interdisciplinary research projects DIGIT's EU and TWON and the Digital Law Institute (IRDT) of the Trier University (Germany).

'Digital Sovereignty of Europe' (DigitS EU) is an interdisciplinary research project of Trier University, funded by the research initiative Rheinland-Pfalz 2024-2028. It aims to accompany the implementation of the EU's new digital legal order. The interdisciplinary research network draws on a wide range of expertise from the fields of law, political science and media science as well as computational linguistics, business administration and sinology.

The project 'Twin of Online Social Networks' (TWON) is a Horizon Europe project that develops methods to study the impact of platform mechanics on the quality of public debate. We are developing a so-called digital twin to create a simulation of online social networks and to study such effects. However, to calibrate such a digital twin and to verify conclusions, access to platform data is of paramount importance.

Access to data from VLOPs and VLOSEs is essential for research. It is also essential for providing guidance on how to deal with systemic risks. DigitS EU, TWON, and the Digital Law Institute Trier expressly welcome the access to data granted by Art. 40 DSA and the adoption of a Delegated Act. We appreciate the opportunity to comment on the draft Delegated Act.

Institute for Digital Law Trier (IRDT), Directors: JProf. Dr. Lea Katharina Kumkar, Prof. Dr. Benjamin Raue, Prof. Dr. Thomas Rübner, Prof. Dr. Antje von Ungern-Sternberg
Universität Trier, Campus II, Gebäude H, 54286 Trier, Behringstraße 21, 54296 Trier, <https://irdt.uni-trier.de>

Citation:

Dregelies et al., Opinion on the delegated act on Art. 40 DSA by the research projects DIGIT's EU and TWON and the Digital Law Institute Trier, IRDT PAPERSERIES Nr. 14

In order for Art. 40(4) et seqq. DSA to be effective, it is necessary that the barriers to data access are not artificially raised and that the specific needs of the scientific community are taken into account.

We strongly recommend providing

- simplified procedures to access data for peer review and follow-up research (see 1)
- more specific details to prevent misunderstandings in the application process (see 2-8).

I. Peer Review and repeatability

Sciences requires the possibility to review and repeat the original research that was made based on access to VLOP/VLOSE data. Thus, it should be possible for other researchers to access the relevant data, for example for the purpose of a peer review process or to review research results in general. A simplified procedure should be open to reviewers in a peer review process and researchers who want to repeat the vetted research project.¹

The scientific community needs to be able to peer review and replicate the original research that was conducted based on access to VLOP/VLOSE data. It should be possible for other researchers to access the relevant data, for example for the purpose of a peer review process or to review research results in general. A simplified procedure should be available to reviewers in a peer review process and to researchers who wish to repeat the research project that has been reviewed.

Scientific publications go through a peer review process to ensure that scientific standards have been met and that the results are reproducible. In order to check this reproducibility, reviewers in particular need access to the data on which the scientific work is based. Scientific experiments must also be repeatable by other researchers and allow results to be verified, modified, or falsified. A large amount of studies show the importance of access to raw data, as researchers can independently come to very different conclusions. Therefore, other researchers need access to the original data to verify the results.

This is particularly important because the data base of VLOPs and VLOSEs changes extremely quickly. Research results are not comprehensible if the subsequent data access concerns a new data base.

Neither Art. 40 DSA nor the Delegated Act provide a specific framework for peer review or the use of follow-up data. We suggest that a new article be created for this purpose.

Such an article could be based on the following principles. Research projects that have already been approved should be published in the DSA data access portal. Peer reviewers and other research groups should be given the opportunity to link themselves to a particular research project, i.e. they can apply to carry out the same or similar research. The DSC should not be required to re-examine the research project as such and should only check that the investigators or new follow-on investigators meet the other requirements. Persons who can demonstrate that they are acting within

¹ Also *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, p. 6.*

the framework of a peer review process should be regularly approved, unless there are indications of a risk of abuse. It is also important to ensure that the names of reviewers are not visible in the DSA data access portal, to allow for blind review.

In the Q&A of 19 November, it was reported that if all safeguards are met, access to the data may be possible at a later date. This should be clarified in the Delegated Act.

We therefore propose to insert the following article:

Article 9a

- (1) Reviewers in a peer review process shall be granted access to the data of a vetted research project upon request if they can demonstrate that this is necessary for a peer review procedure regarding the research project of the vetted researcher.*
- (2) The Digital Service Coordinator of establishment shall verify that the request includes the following elements:
 - a. the identity and contact details of the reviewer;*
 - b. the proposed safeguards to mitigate possible risks in terms of confidentiality, data security and personal data protection corresponding to the data requested, including as regards the modalities of access to and processing of the data;*
 - c. an estimation of the required duration of the access.**
- (3) The request can be rejected if there are indications of a risk of abuse.*

II. Renewed access for vetted researchers

Vetted researchers also need renewed access to the data, even if the research project has been provisionally closed. It should be possible to reopen access to the data.

It may also be necessary for authorised researchers, who have previously accessed data for their original research, to access the data again at a later date, for example if changes are proposed in a peer review process. Even when research projects are (provisionally) closed, it may also be necessary to re-access the data, for example if research results are challenged. Therefore, it must be possible to re-access the data again at a future date.

We therefore propose to insert the following article:

Article 9b

A vetted researcher shall be granted renewed access to the data of an already approved research project if this is necessary in the context of the research. The vetted researcher shall therefore submit a request to the Digital Service Provider of establishment.

III. Enabling pre-inquiries on preconditions to research

Since researchers do not know in advance what kind of data the platforms actually possess, researchers should have the opportunity to ask the VLOP/VLOSE for specific information.

Systemic risks and their potential causes cannot always be specified with sufficient clarity in an application under Art. 40 DSA, even if they seem very likely and plausible. For example, it may be reasonable to assume that foreign powers are conducting disinformation campaigns to undermine and manipulate democratic opinion forming in the EU, even if the methods remain unknown (e.g. dissemination of illegal or harmful content, boosting of popularity through inauthentic behaviour). Very often, it will be difficult to assess whether a systemic risk exists and what its causes and consequences are.

In order to formulate well-founded inquiries to assess particular systemic risks, researchers require some a priori knowledge of the problem they want to examine and the relevant data sets. Identifying which data is required, however, may require insight into broader dynamics and trends derived from social scientific theories which require data.

It can be problematic for researchers to make a sufficiently specific request to the DSC in order to gain access to this type of data. This is all the more relevant because researchers do not know which parameters the platforms change and which groups are affected. It should therefore be possible to link the research project to questions that the platforms would have to answer.

To take an example from foreign interference in elections: Malevolent actors may try to prevent specific groups of citizens from voting. Such a suspicion is in and of itself not sufficient to formulate an Art. 40 DSA request, because both the messaging and target population remain unclear. Targeted messages may, for example, aim to suppress voter groups through threats, deception, persuasive messages, polarization or other, even currently unknown strategies. They could also attempt to render official communication inefficient through denial-of-service attacks or large volumes of irrelevant content (“noise”, as employed in the PRC). Target populations could consist of ordinary voters, opinion leaders, election officials etc. If Art. 40 DSA is to be effective in enabling research on systematic risks, it needs to facilitate exploratory research aimed at understanding the defining features that are necessary for sufficiently clear requests (including, but not limited to message content, involved user groups, and interaction patterns).

IV. Verification of the data

For reliable research results, it is essential that the data is correct. A control mechanism is therefore needed to ensure that the data has not been manipulated.

Neither the DSA nor the Delegated Act explicitly entail a control mechanism which makes sure that the data provided is correct, i.e. complete, accurate and without errors (e.g. illegible data). Mistakes when providing data have already occurred in the past.² As a minimum requirement, VLOPs/VLOSEs should be obliged to provide corrected datasets to all researchers involved if it is verified that the relevant data was wrong. Researchers should be allowed to check the data provided themselves, for example via scraping, and this right should be stipulated in an explicit provision.

V. Systemic risks

The reference to systemic risks in Art. 34 DSA should not be understood too narrowly. Only a broad understanding can enable basic research capable of determining whether a VLOP/VLOSE concept leads to a systemic risk in the long term.

Research only falls within the scope of Art. 40 DSA if it refers to systemic risks as understood in Art. 34(1), Art. 35 DSA. However, throughout the process of identifying systemic risks, it may occur that data needs to be accessed to enable the verbalisation of a particular systemic risk in the first place. Hate and incivility, especially online, are complex and multi-layered. As a result, it is difficult to find new variations of systemic risks if they have not been previously recognised. This would require that more general enquiries could be made.³ It should be clarified in the form of a recital that basic research on systemic risks is also covered by the provision, as well as research based on initial suspicion that a systemic risk may arise.

We therefore propose to insert the following recital:

(XX) It is not always possible to predict in advance which algorithms, interfaces or behaviours could lead to a systematic risk. However, corresponding this basic research is particularly important for understanding systematic risks. As a consequence, research projects that aim to investigate systematic risks that were not previously recognised should also be authorised.

² For example <https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists>.

³ Thomas, K., Akhawe, D., et al (2021). SoK: Hate, Harassment, and the Changing Landscape of Online Abuse. 2021 IEEE Symposium on Security and Privacy (SP), 247–267.

VI. Research groups

Research groups should be provided with data access for every research member.

It is not clear from the Delegated Act whether only one access to the data must be provided for a research group or whether there is simultaneous access for all researchers. In the Q&A on 19 November, it was reported that all members of the research group would be granted access. This should be clarified in the form of a recital.

We therefore propose to insert the following recital:

(xx) Research is mainly carried out in groups. To enable effective research, not only the principal researcher, but all members of the research group should have access to the data.

VII. Scope of the obligation to justify according to Art. 8

It should be clarified in a recital that no excessive requirements are placed on the individual proofs within the framework of Art. 8.

Researchers must meet certain requirements when applying for data access. It has already been criticised in the context of Art. 40 DSA that there are no clear specifications as to how extensive the individual proofs must be.⁴ Neither does Art. 8 explain how extensive or detailed the justifications or evidence needs to be. A recital should provide examples of what evidence may be sufficient and emphasise that the requirements are not set too high. Excessively high requirements would only result in unnecessary bureaucracy. A certain basic trust should be placed above all in state or publicly funded research organisations.

We therefore propose to insert the following recital:

(XX) The evidence that researchers must provide in accordance with Article 8 should not be unreasonably demanding. The identity of a researcher can usually be proven by a digital copy of an ID or a passport. To prove a formal relationship between the researcher and the research organisation, a simple declaration by the research organisation that the researcher is associated is sufficient.

VIII. A/B testing

Researchers need access to (results of) A/B tests.

To assess systemic risks on VLOPs/VLOSEs, researchers do not only need observational data but also experimental data. VLOPs/VLOSEs continuously conduct so-called A/B tests, in which a specific feature is changed for one group but not another. In doing so, they can learn how changes in platform mechanics effect outcomes like user engagements. Although A/B tests are often the only way to identify influencing factors that put users at risk, the results of such tests are almost never

⁴ Denga, in Denga/Heinze/Steinrötter, EU Platform Regulation, § 6 mn. 107.

published. It should be explicitly possible for researchers to request access to lists of the tests that are conducted, the description of these tests, and their results. Preferably, going beyond this, an additional route should be created that allows researchers to initiate a request to have future specific A/B tests conducted, as far as this is ethically and practically feasible.

It should be noted that due to specific affordances and user bases, insights into tests on one VLOP/VLOSEs cannot be generalized for others. It is therefore mandatory to gain access to the specific VLOP/VLOSEs under investigation.