

weizenbaum
institut

ens  European New School
of Digital Studies

 Stiftung
 Neue
 Verantwortung

APRIL 2024

Weizenbaum Institute

Enabling Research with Publicly Accessible Platform Data: Early DSA Compliance Is- sues and Suggestions for Improvement

Position paper

Imprint

Enabling Research with Publicly Accessible Platform Data: Early DSA
Compliance Issues and Suggestions for Improvement, Weizenbaum Policy Paper 9,
April 2024

AUTHORS:

Julian Jaursch \ Stiftung Neue Verantwortung \ jjaur@stiftung-nv.de
Jakob Ohme \ Weizenbaum Institute \ jakob.ohme@weizenbaum-institut.de
Ulrike Klinger \ European University Viadrina \ klinger@europa.uni.de

PUBLISHER:

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

EDITORIAL TEAM:

Moritz Buchner \ Weizenbaum Institute
Elja Krütsfeld \ Weizenbaum Institute

DOI: [10.34669/WI.WPP/9](https://doi.org/10.34669/WI.WPP/9)

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/>

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 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

Julian Jaursch is project director, “Policy | Platform regulation” at Stiftung Neue Verantwortung, Berlin. Contact: jjjaursch@stiftung-nv.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

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

Enabling Research with Publicly Accessible Platform Data: Early DSA Compliance Issues and Suggestions for Improvement

Position paper

The EU's Digital Services Act (DSA) requires very large online platforms and search engines (VLOPs) to provide publicly accessible data to researchers meeting certain requirements ([Article 40\(12\)](#)). While some platforms have taken laudable steps to implement such data access opportunities, serious concerns remain about full compliance with the DSA in this regard. This is reflected in the European Commission's effort to [request information](#) from 17 VLOPs on how they comply with Article 40(12).

Currently, researchers from academia and civil society still face significant hurdles when trying to request publicly accessible data from VLOPs. To provide the Commission and Digital Services Coordinators with insights for their oversight work, this paper offers an analysis of early experiences with Article 40(12) data access requests and suggestions for improvement. It is based on conversations with researchers and on data collected through the [DSA 40 Data Access Tracker](#).

Compiled by Julian Jaurisch (Stiftung Neue Verantwortung), Jakob Ohme (Weizenbaum Institute) and Ulrike Klinger (Europa Universität Viadrina).

1. Application forms

Status quo: Application forms are often hard to find and unclear.

Currently, forms to **apply for platform data access under Article 40(12) can be hard to find** on platform websites and often come in a format hosted by third-party platforms (such as Google). For example, a link to X's DSA data access application is linked in a "snippet" banner at the very bottom of a single page on their website. However, the subpages, including the subpage specifically for academic researchers, do not include the form or any information about the program. They also provide little information on the desired length and detail of responses. These aspects impede researcher access and trust in the data processing of application data.

Suggestion: Application forms need to be easy to find, accessible, and secure.

Given that application data contain information about research projects and institutions (such as technical-operational measures), more **official access to application forms and hosting is necessary** across all VLOPs. Platforms should offer a website informing researchers about possibilities of data access and provide professional, standardized application forms and information on how platforms handle data from access requests. Researchers should be given a timeline of when they can expect to have their requests dealt with. Additional support could come from academic and other organizations that provide a collection of access points.

2. Eligibility criteria

Status quo: Eligibility criteria are interpreted too strictly.

The DSA does not narrow the eligibility criteria for Article 40(12) applications to only universities or other academic institutions. On the contrary, the wording on eligible researchers is rather broad, with Article 40(12) particularly mentioning not-for-profit organizations. However, at some VLOPs, **requesting data can be difficult or impossible** for researchers who are not affiliated with “traditional” academic institutions such as universities. For instance, TikTok’s Research API and the YouTube Researcher Program state that researchers must be affiliated with an eligible “academic institution”. Even if this is not explicitly mentioned, requirements such as naming a “principal investigator” can be exclusionary since such structures are uncommon outside of universities.

Applicants, particularly those from not-for-profit organizations and other non-academic institutions, also face **hurdles in applying for these programs because they are structured to grant access on an individual project basis rather than an organizational basis**. Not-for-profit researchers frequently design more general social media monitoring operations as the starting point for in-depth investigations. Under the current project-based design of many data access programs, researchers would need to reapply for access precisely at each point when uninterrupted data access becomes critical, even for research that is related or similar to the original purpose under which they applied for access.

Suggestion: Applications should be open to eligible non-academic researchers.

Various VLOPs do explicitly **gear their application forms towards non-governmental organizations**. This approach should be followed by all VLOPs, for example, by allowing a self-identification of researchers’ affiliations in the application form and by not disregarding applications from non-academic institutions and journalists. Researchers from outside the EU should be allowed to request data, if they study systemic risks in the EU.

To still ensure the possibility for some public cross-checking and to spot potentially abusive behavior, applicants and companies could voluntarily commit to creating a public transparency database with research conducted using Article 40(12). This could be an expansion of what some VLOPs are already doing based on voluntary commitments under the EU's Code of Practice on Disinformation (see sample case 2 below).

Furthermore, it would be helpful if VLOPs better **allowed for organizational applications instead of granting access on a project-by-project basis**, as this would align more closely with the operational means by which not-for-profit research institutions leverage public data. It would also better allow monitoring of ongoing societal issues, potentially in real time, as recital 98 mentions, without having to file multiple requests for the same research topic.

3. Types of research questions

Status quo: Requirements for suitable research questions are interpreted too strictly.

At times, applications have been **rejected because the proposed research questions supposedly do not meet the requirements** of the DSA, which call for research to be conducted on systemic risks (see sample case 1). This is often based on companies' strict interpretation of Article 40(12), which is not in line with its overall goal of enabling wide-ranging and timely studies by a variety of researchers.

Similarly, platforms often **ask for proof of technical-operational measures** to secure data that are outsized for the type of data requested under Article 40(12). This can be an unnecessary obstacle for researchers, especially for organizations that do not have the personnel and/or resources to use separate research infrastructure just for Article 40(12) studies.

Sample case 1: Article 40(12) requests to X

As a designated VLOP under the DSA, X has been required to comply with Article 40(12) since August 2023. Various researchers who have tried to gain access to publicly accessible data have reported rejections from X. The reason that was given often was that "it does not appear that your proposed use of X data is solely for performing research that contributes to the detection, identification, and understanding of systemic risks in the EU." While this wording is found in Article 40(12), it is an unnecessarily narrow interpretation of the provision to restrict data access by claiming a request is not specific enough without further explanations.

Suggestion: VLOPs need to differentiate between requirements for Articles 40(4) and 40(12) DSA.

Given the public nature of the requested data (for instance, publicly available videos or engagement metrics such as “likes”), the eligibility criteria for Article 40(12) should not be mistaken with access criteria under Article 40(4) in connection with Article 40(8). It is important to clarify this difference to platforms. For instance, Article 40(12) does not foresee the same vetting process that is required by Article 40(4) and 40(8) - which is to be conducted by regulators and not companies, nonetheless. In neither case, VLOPs are tasked with checking the feasibility of research questions. The point of Article 40(12) is precisely not to have an elaborate vetting process because the data in question is already public.

In this line, a **restrictive understanding of the connection to systemic risks is not proportional**, considering that Article 40(12) is also meant to enable monitoring of topics in real time, if technically possible. More generally, it is arguably prohibitive to demand that requests for public platform data can be *solely* used for only one highly specific research purpose. Similarly, while researchers certainly need to adhere to data protection law when handling Article 40(12) data, the privacy risks associated with research using such data are much lower than for non-public data (which “vetted researchers” will be able to request under Article 40(4) and 40(8)).

4. Responses to requests

Status quo: Responses take too long and explanations about rejections are too vague.

Researchers often have to wait too long to hear back from platforms about their requests. For example, preliminary data from the DSA 40 Data Access Tracker show that for X, the **average response time to data access requests is 1.5 months**. Given the small number of requirements checked in this application and the public nature of the data, this timeframe is severely delaying data access requests.

Moreover, **rejections are often not well explained** (see also point 2 above). For instance, access requests to X have gotten rejected with statements such as “application is incomplete or lacks sufficient detail” or “until we receive any contrary legislation or guidance, X will limit API access under Article 40(12) to organizations located in the EU” (see sample case 2 below; also sample case 1 above). Both are not sufficiently explained or helpful for researchers so they can adjust future requests.

Sample case 2: “Empowering researchers” under the Code of Practice on Disinformation

While not directly related to Article 40(12), the voluntary commitments various VLOPs have made under the EU’s Code of Practice on Disinformation still shine a light on various approaches to researcher access to data. For example, the latest [report on the Code from March 2024](#) shows that some companies publish the number and approvals of requests to their application programming interfaces (APIs), others do not. The report also shows that documentation on APIs and data access varies widely. Overall, it becomes clear that some of the open questions related to Article 40(12) are recurring themes from other debates in need of clarification via cross-platform, EU-wide guidance.

Suggestion: Actionable responses should come within three weeks and provide substantive explanations.

An important goal of Article 40(12) is enabling fast, data-driven research as well as monitoring of current societal topics (as recital 98 mentions). The DSA also states that data should be provided in real time, if technically possible. Achieving this goal is severely **hindered if responses take longer than three weeks**. Especially in relation to current events, but also in the planning of resources for research projects, response times longer than three weeks carry the risk of sinking some research projects.

Moreover, researchers **require substantive and actionable explanations** for why their requests were rejected. While VLOPs cannot and should not rephrase research questions for researchers, the applicants need to at least understand what caused the rejection, so they can improve and revise or re-submit access requests.

5. Types of data accessed

Status quo: Documentation of accessible data is lacking.

Some VLOPs **do not document the functionality** and accessible data researchers can apply for under Article 40(12). This complicates the application process as researchers either cannot be specific in their application texts or researchers try to be specific but platforms can reject an application based on the mismatch between proposed research and actual platform access (see sample case 3).

Sample case 3: Lack of understanding what data TikTok does and does not make available

TikTok promotes its Research API as a way to access various types of data, along with dedicated documentation. This is generally a helpful approach but the documentation previously offered conflicting information on certain issues such as queries. Moreover, it is unclear why some data is available via the Research API and other data is not (as [researchers have found](#) that other sources provide deeper insights).

Suggestion: Platforms should provide clear access mode documentation.

It would be highly beneficial for researchers if VLOPs provided **clear documentation on the access to public data**, as some companies do based in part on their voluntary commitments under the EU's Code of Practice on Disinformation. This could be achieved through a dedicated overview of public data that can be requested through their APIs and/or by highlighting the data fields considered public data in their API documentation. Information should also cover any rate limits or other restrictions of access. This is similar to what is already in place for access to commercial APIs. Any restrictions need to be transparent, justified and in accordance with Article 40(12).

6. Changes once data access has been granted

Status quo: Restrictions occur after data access requests have been granted.

In some of the few known cases where data access has been granted, **access has been changed and restricted after it was granted** (see sample case 4). This complicates researcher data access, as a change in the empirical basis can reduce the explanatory power of data and with this, the assessment of systemic risks in general.

Sample case 4: Granted access to TikTok API is attempted to be revoked.

After a researcher was granted access to TikTok's Research API, the company sent an email to the researcher stating that "After careful review of your application, it appears your research project as proposed would violate the Research API Terms of Service (...) Could you please confirm that there will be no breach of our Terms of Service and provide clarification of your research design if you believe that there will be no breach?" There is no information provided that is specific to the research project, explaining where TikTok sees a breach of its terms of service. Crucially, there is no referral to the researcher's alleged non-compliance with the DSA under which access had been granted. The reference to corporate terms of service to revoke legally protected data access seems to be at odds with the DSA.

Suggestion: Any changes to data access should be announced in advance and explained.

While changes in technical infrastructure are inevitable, such **changes need to be proactively announced** and communicated by platforms for researchers who have received data access, including specific and substantive explanations. Planned changes need to be made public in the documentation, preferably on dedicated researcher websites.

7. Costs

No change to status quo necessary: Access requests under Article 40(12) are free.

So far, it seems that researchers are not being charged for requesting data under Article 40(12). This should not change. The whole point of the article is to allow fast, free, low-barrier access to data that is already publicly available. Any financial barriers erected by VLOPs would run counter to this. This is especially pertinent since some replacements to CrowdTangle (see point 9 below) or to Article 40(12) access touted by platforms are fee-based social media monitoring services and as such not fitting alternatives.

8. Terms and Conditions

Status quo: Policies regarding data retention, refresh, sharing, and pre-publication review are often at odds with how research is conducted.

The terms and conditions of several VLOP's data access rules **require researchers who want to access public data to agree to submit any publications they generate using that data to the VLOP for review**. This significantly hinders researchers from carrying out and publishing research in the public interest, especially when that research may run counter to the interest of the company. Moreover, academic research norms often require the long-term retention of data for analysis, open data for replication and the independent sharing of results in peer-reviewed journals, yet some VLOP data access programs, such as TikTok's Research API, require that researchers refresh data every fifteen days, bars researchers from sharing data and requires that researchers send their papers to TikTok thirty days in advance of publication.

Suggestion: Abolish pre-publication review.

VLOPs should **refrain from having any pre-publication reviews** as a requirement for obtaining access to public data under Article 40(12). Terms and conditions should not be overly burdensome (or counter-intuitive) to producing public-interest research. They

should also clearly establish how data can be shared within a project team. In turn, researchers using Article 40(12) should generally commit to make their study results publicly and freely available, even though this is not legally required.

9. Existing Access

Status quo: Data access modes that have existed prior to Article 40(12) are about to be discontinued.

Meta has announced plans to **discontinue access to platform data via Crowdtangle** on August 14, 2024. This creates serious issues for researchers because it leads to a cap of existing and longitudinal research projects. It is especially unfortunate that this happens during a year with many elections across the world, including the European Parliament and US presidential elections, which leaves a gap in public data access for pre- and post-election monitoring that the new Meta content library is not able to fill.

Suggestion: Existing data access modes should be continued until a satisfactory new data access model under Article 40(12) is in place.

As mentioned in the [open letter initiated by the Mozilla Foundation](#), **existing data access modes need to be continued** until all VLOPs fully comply with public data access under Article 40(12). At minimum, this should be considered as developing sufficiently robust data access tools and granting access to a variety of public interest researchers, particularly those who have requested access to public data and have not yet heard back from VLOPs. Companies should establish (or continue and expand their existing) consultations with the wider research community, not only academic researchers, to ensure that their programs meet the ambitions of the regulation as laid out in Article 40(12).

We encourage VLOPs to ensure compliance with Article 40(12) and regulators at the European and national levels to continue their oversight work on this important provision. Specifically, we encourage the Commission and the European Board for Digital Services to develop guidelines for Article 40(12) that address some of the challenges and open questions described here.