



# Researcher access to platform data under the DSA: Questions and answers

Julian Jaurisch and Philipp Lorenz-Spreen | [About this page](#) | [Sources and acknowledgements](#)

*Originally published in August 2023; last updated in November 2024*

[Researcher access to platform data under the DSA: Questions and answers](#)

[Abbreviations](#)

[Background and timeline](#)

[Why should researchers care about the DSA? What's in it for them?](#)

[Aren't there already data access rules in the EU?](#)

[Are the DSA's data access rules all in place now already?](#)

[How does the process to request data work?](#)

[Isn't the "trade secrets" exemption a huge loophole?](#)

[Very large online platforms and search engines \(VLOPs\)](#)

[What platforms have to provide data?](#)

[Is cross-platform research enabled?](#)

[Does the data access regime cover smaller platforms at all?](#)

[Vetted researchers](#)

[What are the conditions to become a vetted researcher?](#)

[Is the data access request per researcher or per research proposal?](#)

[Can research teams/consortia apply for data access? Is data sharing among researchers allowed?](#)

[Can researchers from outside the EU request data from platforms?](#)

[Can non-academic researchers request data from platforms?](#)

[Can researchers lose their “vetted researchers” status?](#)

#### [Research topics and data](#)

[Can researchers request data for any topic they want? What are some study examples?](#)

[What data \(formats\) can researchers receive?](#)

[Do the data access rules cover experiments and A/B testing?](#)

#### [Digital Services Coordinators \(DSCs\)](#)

[What is a DSC and why is important for the data access rules?](#)

[Do researchers need to file their requests to the DSCs in their country?](#)

[Isn't the Irish DSC a bottleneck for researchers?](#)

[Can researchers receive any resources for their applications or research from regulators?](#)

[How do researchers even know what DSCs look for in a “good” data access request?](#)

[What remedies are there if platforms don't comply or act nefariously?](#)

#### [Other opportunities for researchers in the DSA](#)

[Does the DSA allow/legalize scraping?](#)



[What other rules in the DSA might be useful for researchers?](#)

**!** [After all these questions, now what? Next steps for researchers](#)

## Abbreviations

Before getting started with the questions and answers, three abbreviations are good to know (they'll all be explained in greater detail later):

- **DSA: Digital Services Act**, a European Union law applying to online platforms throughout the EU
- **DSC: Digital Services Coordinator**, a new regulator each EU member state has to designate that has a big role to play in enforcing the DSA, particularly regarding data access for researchers
- **VLOP: Very large online platform** or search engine, i.e., platforms with more than 45 million users in the EU per month, which have to abide by special rules under the DSA, including the data access provisions. When speaking of platforms here, it mostly means VLOPs.

One more note: The EU flag label  denotes a link to the DSA's legal text and the information label  offers suggestions for more in-depth reading or other sources.

## Background and timeline

# Why should researchers care about the DSA? What's in it for them?

Independent **regulators in the EU need expertise from academic and civil society researchers** to understand online platforms. Without this, strong and sensible enforcement of new EU rules for platforms will not work. The new rules in the [EU Digital Services Act \(DSA\)](#) apply to a range of online platforms, with the goal of creating a “safe and transparent online environment” for consumers.

[i Analysis of various aspects of the DSA](#), edited by Joris van Hoboken et al. (Verfassungsblog ebook, 2023)

However, for years, researchers have faced obstacles when they needed online platforms' data to explore platforms' functioning, potential risks and effects. A plan for a one-off cooperation between a platform and researchers [failed miserably](#), tech companies have allegedly [pressured researchers](#) to stop their work and a key tool to study platforms has been [closed](#). In early 2023, these challenges have come to the fore again as one of the more open big platforms, Twitter, made it much [harder](#) to access data. Even when [industry-academia collaboration work very well](#), [questions around researcher independence](#) come into play.

Against this backdrop, European policymakers – pushed by academia and civil society – included a data access provision for researchers in DSA. This article ([Article 40](#)) **could potentially provide a huge opportunity for researchers to better understand platforms** and, ultimately, help independent regulators oversee platforms based on scientific evidence.

Yet, despite the promises held by the DSA, many important questions remain open. Details on data access requests, the type of data to be requested, the timelines for full application and the communication exchanges between researchers, regulators and platforms will be hashed out over the coming months. **That is why preparation and coordination by researchers and regulators is crucial.**

[i Overview of the legal/political developments](#) on researcher data access in the EU by Mathias Vermeulen (journal article, September 2022)

[i Why researcher data access is a key provision of the DSA](#) by Alex Engler (Brookings blog post, January 2021)

[i Guide to the data access rules](#) by John Albert (AlgorithmWatch blog post, December 2022)

[i Impediments to privacy research and how the DSA might help](#) by Konrad Kollnig and Nigel Shadbolt (Technology and Regulation, October 2023)

Some more sources regarding data access beyond the DSA:

[i Analysis and ideas on platform-to-researcher data access](#) by a working group of the European Digital Media Observatory (EDMO report, May 2022)

[i Dealing with GDPR concerns](#) on platform-to-researcher data access by Mathias Vermeulen (journal article, 2020)

[i Analysis of researcher access in other industries](#) by Jef Ausloos, Paddy Leerssen and Pim ten Thije (AlgorithmWatch policy paper, June 2020)

## **Aren't there already data access rules in the EU?**

No, there are no mandates for platforms to share data with researchers.

The DSA is the **first piece of EU legislation to include a legally binding data access regime** for online platforms. This makes current efforts to get enforcement right all the more vital.

So far, two of the major avenues for researchers to get platform data are voluntary and flawed:

### **Voluntary measures by individual platforms**

Some platforms provide application programming interfaces (APIs) or other one-off voluntary cooperation with researchers. These measures have overall not been successful (see [above](#)).


[i Overview of platforms' own data access regimes](#) by Richard Kuchta, Beatriz Almeida Saab and Lena-Maria Böswald (DRI policy paper, March 2023)

### **Multi-platform pledge to cooperate with researchers on disinformation**

The voluntary “Code of Practice on Disinformation” is a pledge led by the European Commission to get platforms and other stakeholders to tackle issues stemming from the spread of disinformation. Big tech companies like Instagram and TikTok have signed up to (parts of) this voluntary code. One pillar of the Code is called “Empowering the Research Community”, which includes commitments to provide automated access to public data and to support the research community.

[i Background on the Code of Practice on Disinformation](#) by the European Commission (press release, June 2022)

However, while a step in the right direction, the benefits for researchers have limits: Most importantly, the pledge is voluntary. There are evaluation indicators for each commitment and measure, but fundamentally, platforms are free to join and leave the Code as they please (for instance, [Twitter is said to have left the Code in May 2023](#)). Even if they nominally adhere to the Code, platforms' efforts to “empower the research community” have been underwhelming, as this analysis shows:

 [Analysis of platforms' reports on the Code](#) by John Albert (AlgorithmWatch blog post, February 2023)

In addition, the Code only covers research on disinformation. This is certainly a big field but the DSA's scope covering systemic risks (more on those [below](#)) is much wider.

## Are the DSA's data access rules all in place now already?

Not all of them.

The DSA has been a fully applicable law since February 2024. However, in practice, the data access regime is not fully functional yet. What needs to happen for that is a “delegated act” to be drawn up and published. This is a secondary piece of legislation that fleshes out the details of Article 40, especially those around the vetting process in Article 40(4) (see [below](#)). **The delegated act is a crucial additional piece of legislation that can hopefully clarify many of the open points on data access.** The European Commission is expected to present this in late 2024.

Compliance with the DSA rules covering researchers' use of “publicly accessible data” (mentioned in Article 40(12), which is discussed [later on](#)) have faced scrutiny by the European Commission. It has sent requests for information to 17 VLOPs to figure out how they follow the rules.

 [Announcement](#) by the European Commission (press release, January 2024)

Moreover, for the data access rules to work, EU member states need to have their Digital Services Coordinators (DSCs) in place (see [below](#)). Not all member states met the February 2024 deadline to do so.

 Here's an overview of the **timeline**:


- April 2023: Commission designates first group of “very large online platforms” (VLOPs), from which researchers can request data

 [Announcement](#) by the European Commission (press release, April 2023)

 [Context and analysis](#) by Martin Husovec (journal article, February 2023)

 [Overall DSA timeline](#) with deadlines for reports and evaluations

- May 2023: Commission holds consultation on future delegated act on data access

 All [responses](#) to the consultation (from scientific organizations, individual researchers and civil society groups)

- August 2023: DSA applies to all VLOPs
- February 17, 2024: DSA applies in full for all platforms; deadline for member states to designate their DSCs
- October 2024: Commission publishes [draft delegated act](#) on data access and opens it for consultation
- 2025: Draft delegated act on data access to be in force

While a data access request is technically possible from February 17, 2024 on, it's probably useful to wait for the delegated act and to allow the DSCs to set up a bit. This is an unfortunate delay but might allow for a speedier process later on. A contrary approach could be to test out the system as early as possible to detect potential flaws. This, ideally, should be a coordinated effort, so as not to drain resources from regulators and researchers early on.

So, there's still ample opportunity to prepare. At the end of this page is a list of suggestions on what to do when.

## How does the process to request data work?

The main process for researchers to request data from platforms goes **through a new platform regulator called Digital Services Coordinator** (DSC; see [below for more details](#)).

🇪🇺 Process for researcher data access: [Article 40\(4\) to 40\(11\) DSA](#)

Here's a **simplified step-by-step guide** on how the process works if the application is successful:

1. Researchers write an application for data access and send it to the DSC
  - Either they send it to the DSC where the platform is based
  - Or they send it to the DSC of their own country (which forwards it to the DSC where the platform is based; see [below for details](#))
2. The DSC (where the platform is based) approves the access request
  - Here's where the DSC could reject a proposal
3. Within 15 days after the DSC's approval, the platform can ask for amendments to the data access request if
  - The platform doesn't have the data
  - The platform sees security risks or risk for confidentiality, including trade secrets
4. The platform provides the data to the researchers

This figure sums up the process with a fictitious sample case:

## Simplified process of a data access request under

Imaginary sample case: A German research team with data from a very large online platform (VLOP)

**Researchers:**  
Preparation

Re  
detailing propos  
as well as

Option A  
Sends application to become  
“vetted researchers” to G  
(because the team is based in

**Digital Services  
Coordinator (DSC):**  
Vetting process



German DSC  
produces initial op

Data access request

**Platform:**  
Data provision

V

Image with the simplified process to apply for platform data. A research team can apply in their home EU country or in the country where the platform they want data from is based.

The practicalities of the process still remain open because the delegated act on data access is not done and because national DSCs are not in place yet. For some background information on the Article 40 rules, see:

 [Guide to the data access rules](#) by John Albert (AlgorithmWatch blog post, December 2022)

## Isn't the "trade secrets" exemption a huge loophole?

It's not as big as it could have been, but it's still an issue.

Platforms can ask for amendments to the data access requests if they have concerns about breaches of confidentiality, including trade secrets. During the legislative process for the DSA, researchers and advocates voiced concerns that platforms could just invoke the trade secrets clause and not share data.

This risk was somewhat minimized: An overly broad exemption did not make it into the final text and, crucially, even with the clause in place now, there is hope it doesn't mean platforms can just get out of sharing data. If they say they can't share certain data with researchers, **platforms can't ask to simply stop the data access request altogether**. Instead, platforms can ask for amendments to the request and must make a suggestion for at least one alternative means to fulfill the request. So, in practice, it could happen that researchers don't get the data they asked for. Yet, alternative options must at least be discussed which could hopefully still allow them to carry out their research.

In any case, the amendment process constitutes a possible delay that researchers should plan for.


 The trade secrets clause: [Article 40\(5\) and 40\(6\) DSA](#)


# Very large online platforms and search engines (VLOPs)

## What platforms have to provide data?

Only "very large online platforms and search engines" (VLOPs) are covered by the data access rules. These are **platforms with at least 45 million monthly users** in the EU. Platforms have to report their user numbers to the regulators and the Commission ultimately designates the VLOPs.



















 [Overview of VLOPs](#) by the European Commission (continuously updated)

 [Context and analysis](#) by Martin Husovec (journal article, February 2023)

 VLOP definition and designation process: [Article 33 DSA](#)

Here is the list of VLOPs:

No.	Company	Company's VLOP(s)	Company's EU headquarter = "DSC of establishment"
1	Alphabet	Google Maps	 Ireland
2		Google Play	
3		Google Search	
4		Google Shopping	
5		YouTube	
6	Apple	App Store	
7	ByteDance	TikTok	
8	Infinite Styles Services	Shein	
9	Meta	Facebook	
10		Instagram	
11	Microsoft	Bing	
12		LinkedIn	
13	PDD Holdings	Temu	
14	Pinterest	Pinterest	
15	X	X	
16	Alibaba	AliExpress	 Netherlands

17	Booking	<a href="https://www.booking.com">Booking.com</a>	
18	Snap	Snapchat	
19	Aylo	Pornhub	 Cyprus
20	Technius	Stripchat	
21	WebGroup Czech Republic	XNXX	 Czechia
22		XVideos	
23	Zalando*	Zalando	 Germany
24	Amazon**	Amazon Marketplace	 Luxembourg
25	Wikimedia	Wikipedia	 Netherlands

- [Zalando sued](#) the Commission regarding its VLOP designation, claiming the methodology is arbitrary.

\*\* [Amazon](#) sued the Commission regarding its VLOP designation, claiming it doesn't fit in this category.

Other VLOPs might be designated later.

 [Context and commentary on the VLOP designation](#) by Eliška Pírková and Julie Fuchs (Access Now blog post, May 2023)

As the VLOP overview shows, the **Irish DSC will play an especially important role for data access matters** because many platforms of high interest to researchers have the EU headquarters there. That's why DSCs, and particularly the Irish one, need to have strong capabilities to vet researchers and understand systemic risks (see the [questions on the DSCs below](#)).

## Is cross-platform research enabled?

Yes.

Getting data from multiple VLOPs to study cross-platform systemic risks is possible.

It is unclear, however, whether separate data access requests would have to be filed or whether it's allowed to combine data access requests for multiple VLOPs into one.

## **Does the data access regime cover smaller platforms at all?**

No.

The DSA does not require “non-VLOPs”, that is, platforms with less than 45 million users in the EU per month, to follow its data access rules. Those platforms could share data on their own terms and/or based on the vetting process, but this would be on a voluntary basis.

The lack of coverage for smaller platforms is a downside of the DSA, since “systemic risks” might also stem from smaller, niche platforms. It is likely the result of the need to balance researcher access with high burdens for smaller platforms (small and micro enterprises are exempt from many DSA rules altogether).


## **Vetted researchers**

### **What are the conditions to become a vetted researcher?**

A data access request can only be filed by “vetted researchers”.

In order to become a vetted researcher, several criteria have to be met. Researchers need to...

- ... come from a research institution,
- ... be independent of commercial interests,
- ... disclose their funding,
- ... have measures for data protection in place and
- ... make their research results publicly available for free

 Who can become vetted researchers: [Article 40\(8\) DSA](#)

There are still a lot of open questions around these basic criteria, which will hopefully be addressed in the delegated act on data access. For instance, the specifics of how to prove independence of commercial interests and how to disclose funding have to be determined.

Crucially, any data access request needs to be **linked to the “systemic risks”** mentioned in the DSA (see [details below](#)).

## **Is the data access request per researcher or per research proposal?**

That’s not entirely clear from the DSA text, so both options seem possible as of now.

The delegated act on data access should hopefully clarify how the vetting process works for researchers/research teams and research proposals.

## **Can research teams/consortia apply for data access? Is data sharing among researchers allowed?**

Yes, research teams can turn in joint data access requests.

The DSA does not explicitly rule this out, so collaboration between universities and/or civil society organizations should be possible.

However, it is still open how exactly such an application would go forward and to what extent data sharing within the consortium (or even outside of it) is allowed. The delegated act on data access should hopefully clarify this, as this was a topic for the European Commission’s consultation.

## **Can researchers from outside the EU request data from platforms?**

Yes.

The data access request has to relate to “systemic risks” for the EU (see below). Beyond that, there is **nothing in the DSA that excludes non-EU researchers from requesting data**, as long as they fulfill the vetting criteria (see [above](#)).

The Commission asked for feedback on its delegated act on data access. In many contributions to the Commission, authors reiterated the need for and benefits of having non-EU researchers covered under the DSA’s provisions. Therefore, it’ll be interesting to see if the delegated act provides some more specific details about when and how non-EU researchers can file data access requests.

## Can non-academic researchers request data from platforms?

Yes.

The definition of “research institution” doesn’t only cover universities, but also other research-heavy organizations.

🇪🇺 Who can become vetted researchers: [Article 40\(4\) and 40\(8\) DSA](#)

The DSA refers to another piece of EU legislation when defining research organizations. Here’s the [full text](#):

‘research organisation’ means a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research:

(a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State;

in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation;

This definition potentially covers a wide range of non-university organizations. As such, it is a **big opportunity for civil society organizations** to become vetted researchers and gain access to platform data. However, it’s still not clear how civil society organizations can prove their research focus, so this is a big point the delegated act will have to provide some clarity on. For instance, how will it be determined whether an organization’s “primary focus” is scientific research?

Joint data access requests from academic and civil society experts seem feasible, too, if the overall criteria are met.

## Can researchers lose their “vetted researchers” status?

Yes.

If the DSC that granted the “vetted” status determines that the researchers don’t fulfill the vetting criteria anymore (see [above](#)), it needs to inform the researchers of this finding and that they will be stripped of their status. The researchers can react to that finding before their status is terminated.

🇪🇺 Rules on losing the “vetted researcher” status: [Article 40\(10\) DSA](#)

## Research topics and data

### Can researchers request data for any topic they want? What are some study examples?

On paper, no.

In practice, maybe “*almost* any topic”.

The data access request has to unequivocally relate to what the DSA calls “systemic risks” in the EU. VLOPs have to report on systemic risks on their services and have to mitigate them.

🇪🇺 Rules on systemic risks: [Article 34 DSA](#)

🇪🇺 Rules on risk mitigation measures: [Article 35 DSA](#)


Yet, these **risks are defined in relatively broad terms**, so a lot of research questions can potentially be grouped under one or more of them. The risks can stem from the “functioning” and “use” of a platform, which is also a rather wide definition that specifically includes algorithmic systems at the platforms.

Here’s a summary of the systemic risks:

1. Dissemination of illegal content
  - Important: Illegal as defined by existing EU and national law, since the DSA does not include a set list of what is “illegal”
2. 🔍 Example: Studying the spread of posts inciting violence or containing libel
3. Negative effects for the exercise of fundamental rights (e.g., right to privacy, non-discrimination, freedom of expression)
  - 🔍 Examples: Studying the uptake and consequences of privacy settings; studying biases in content moderation practices and their effects on representation/non-discrimination
4. Negative effects on civic discourse, electoral processes and public security
  - 🔍 Examples: Studying whether viewing more or less content from political parties affects the likelihood to vote; studying the spread of false information regarding an

election date; studying algorithmic recommendations of extremist content

5. Negative effects in relation to gender-based violence, the protection of public health, the protection of minors and relating to someone's physical and mental well-being

 Examples: Studying to what extent exposure to certain types of content affect individuals or groups at risk of depression

Looking at these categories, it's pretty clear that many crucial research questions can be related to "systemic risks". On the plus side, this allows researchers a lot of leeway to align their interests and capabilities with the DSA's requirements. On the downside, there is a risk that studying "systemic risks" will become interchangeable. Also, researchers are restricted in their focus on systemic **risks in the EU context**.

There are still a lot of open questions regarding what a risk assessment and what risk mitigation measures will look like. What is clear, though: Aligning data access requests strongly with one or more of these risks is essential for a successful vetting application.

 [Analysis of and proposals for risk assessments of recommender systems](#) by Anna-Katharina Meßmer and Martin Degeling (SNV policy paper, February 2023)

## What data (formats) can researchers receive?

A wide variety, it seems, but it's not clear yet.

The DSA does not include a list of data formats that VLOPs have to work with. Instead, **researchers need to specify what type of data they need to study systemic risks**. Then, the DSA requires platforms to provide "appropriate interfaces specified in the request, including online databases or application programming interfaces". This rather broad provision seems to cover various types of data and data formats. More clarifications (or limitations) could be introduced in the delegated act the Commission is working on.

 Rules on data formats: [Article 40\(7\) DSA](#)

It is theoretically possible that the Commission or DSCs might request VLOPs to publish (or at least make available to them) a list or type of codebook. It's unclear whether this is planned, though.

## Do the data access rules cover experiments and A/B testing?

In certain cases, maybe.


The data access rules as such are not meant to run social experiments on platforms. It seems possible, though, that through data access requests, A/B tests and experiments could be facilitated. This topic brings up more questions that the delegated act on data access would ideally have to clarify. For instance, can researchers set up “sandboxes” for their studies that employ privacy-by-design measures? Can real-time APIs help with running experiments?

## Digital Services Coordinators (DSCs)


### What is a DSC and why is important for the data access rules?

The Digital Services Coordinator (DSC) is a new oversight body that the DSA requires every EU member to set up. In the data access context, they are the **single-most important body that researchers will interact with**.

Each DSC coordinates various agencies at the domestic level, collaborates with other DSCs and the Commission at the EU level and enforces the DSA regarding all “not very large” platforms in their respective country. As such, they fulfill important “coordinating” tasks - as their name suggests - but also play an important role as a regulatory enforcement agency.

 [Overview of DSCs' responsibilities and ideas for strong DSCs](#) by Julian Jaurisch (Verfassungsblog post, October 2022)

The DSA includes some requirements for the DSC: It needs to be completely independent from commercial and governmental influence, it needs to have sufficient resources and be impartial and transparent. Member states have until February 17, 2024 to nominate their DSCs. The vast majority of countries is expected to house the DSC at an existing agency such as their telecommunication, consumer protection or media regulator.

 Rules on the DSCs: [Articles 49 to 51 DSA](#)

The DSCs themselves can request data from VLOPs and they are the ones “vetting” data access requests from researchers. Especially this latter point makes them such vital points of exchange for researchers.

### Do researchers need to file their requests to the DSCs in their country?

They can, but they don't have to.



The application to become a vetted researcher (see [above](#)) can be turned in to either the DSC where the researcher/research team is based or to the DSC where the platform that is supposed to provide data is based. Researchers can pick and choose based on their language preferences, established connections to a particular DSC or other considerations.

In any case, the **final decision about the request rests with the DSC where the platform is based**. That DSC would receive a first opinion from the DSC where the researchers are based, but then have the final say. It might take longer for researchers to go through “their” local DSC first, but it could also help because that DSC might be more familiar with the language and research landscape in the country. This is a trade-off researchers have to consider when applying.

Here’s the fictional example again (see [above](#)):

## Simplified process of a data access request under

Imaginary sample case: A German research team with data from a very large online platform (VLOP)

**Researchers:**  
Preparation

Re  
detailing propos  
as well as

Option A  
Sends application to become  
“vetted researchers” to G  
(because the team is based in

**Digital Services  
Coordinator (DSC):**  
Vetting process



German DSC  
produces initial op

Data access request

**Platform:**  
Data provision

V

Image with the simplified process to apply for platform data. A research team can apply in their home EU country or in the country where the platform they want data from is based.

## Isn't the Irish DSC a bottleneck for researchers?

Potentially.

Ultimately, most requests will likely be decided by the Irish DSC because those platforms that have so far been of highest interest to researchers are based in Ireland. That includes Meta's VLOPs, YouTube, TikTok and Twitter ([see the table with the all the VLOPs and corresponding DSCs above](#)).

The Irish DSC has publicly signaled - for instance, at the [Commission's "DSA stakeholder event"](#) - that they are well aware of their crucial role for data access and are already in the process of building a team and developing networks to outside experts. Ireland was one of first countries to officially designate their DSC, housing it at the new [Coimisiún na Meán](#) (Media Commission). The goal is to have a staff of 120 people in 2023, with potentially more later. These are **promising signs that Ireland will not be a bottleneck.**

However, **whether these efforts will work in practice remains to be seen.** The Media Commission has only been around since March 2023 and is in the process of building capacities. Depending on the scale of researcher requests, there is a risk the Irish DSC could be overwhelmed. Other risks apply not only to the Irish DSC, but all of them. Other DSCs also need to step up. For instance, if researchers apply in their countries of residence first, a well-done initial opinion by the local DSC can greatly aide the Irish colleagues. That means that DSCs need enough staff, resources and expertise as well as a fitting organizational structure to handle the vetting process. Without a deep understanding of systemic risks and tried networks to academia and civil society, DSCs will be at a loss to support EU platform research.

[i Overview of DSCs' responsibilities and ideas for strong DSCs](#) by Julian Jaurisch (Verfassungsblog post, October 2022)

[i Analysis on why DSCs need strong data and research units](#) by Julian Jaurisch (DSA Observatory blog post, March 2023)

[i Overview of all DSCs across Europe](#) by Julian Jaurisch (SNV, February 2024)

In addition to motivated, capable DSCs, some researchers have suggested creating institutionalized support structures for DSCs to vet researchers: an independent intermediary body.

[i Analysis and ideas on platform-to-researcher data access](#) by a working group of the European Digital Media Observatory (EDMO report, May 2022)

Reference to “independent advisory mechanisms” that might enable the set-up of such an intermediary body: [Article 40\(13\) DSA](#)

Such an intermediary body could help DSCs with a peer-reviewed, independent vetting process and take some of the load off of the DSCs, which have other regulatory tasks to fulfill after all. Several responses to the Commission’s consultation on data access referenced EDMO’s idea. It is unclear if/how the Commission will take up the suggestion for an intermediary body.

## **Can researchers receive any resources for their applications or research from regulators?**

No - at least it’s not mandatory for regulators.

The DSA does not explicitly require the DSCs or the Commission to financially support data access requests.

The delegated act by the Commission could be viewed as an indirect support measure, if it clearly spells out the rules for data access, provides templates and guidance, and as such facilitates the application process for researchers. This is especially the case for smaller universities, NGOs and early-career researchers, who would benefit the most from an easy-to-navigate vetting process.

Across all organizations, though, funding and resources will be a huge determining factor whether researchers can truly make use of the data access provisions. This is one more reason why coordination and support structures within the research company are crucial, for instance, to avoid certain pitfalls in the application or to team up for data access requests.

Moreover, the resource question also pertains to DSCs: If they were well-staffed, have a dedicated scientific outreach unit and maybe even a budget to fund studies, this would be a positive development for researchers.

[i Analysis on why DSCs need strong data and research units](#) by Julian Jaurisch (DSA Observatory blog post, March 2023)

There is also an idea to create a foundational infrastructure to be shared across researchers/research organizations, not just for the DSA but generally to study the “information environment”. The basic outline of this came in a white paper:

[i Suggestion for a CERN model for studying the information environment](#) by Alicia Wanless and Jake Shapiro (Carnegie Endowment policy paper, November 2022)

## **How do researchers even know what DSCs look for in a “good” data access request?**

More details on this are likely to emerge in 2024.

As of now, there is no way to say what a good or even just a suitable data access request looks like in detail. The Commission is still working on its delegated act, most DSCs have yet to be set up and there is little precedence that researchers can turn to. Nonetheless, until these steps have been taken by summer/fall 2024, researchers can still prepare by familiarizing themselves with the general structure of the data access regime and particularly by considering how potential research questions relate to “systemic risks” (see [above](#)).

## What remedies are there if platforms don't comply or act nefariously?

It cannot be ruled out that data access requests fail because of platforms' refusals or inaction. There are some possible avenues for researchers and regulators to act, though.

If platforms invoke the trade secrets exemption, there is still a chance that researchers can conduct their research (see [above](#)).

Yet, VLOPs might still make the process difficult or outright refuse to share data, even if all conditions are met. This would violate the DSA and as such could ultimately result in fines for the company. The DSA foresees sanctions not only for breaches of data access rules but also for other violations. However, fines are one of the last resorts. Beforehand, the Commission must try to work out an “action plan” with VLOPs to stop non-compliance.

 Rules for non-compliance: [Articles 73 to 75 DSA](#)

Even with financial sanctions in play, platforms could decide that delaying tactics, overwhelming researchers with too much or useless data, ensuing reputational damage or even paying the fines are more suitable options than cooperation. These risks can likely only be reduced by strong, independent regulators as well as a vocal and well-coordinated researcher community.

To formalize complaints about platforms, researchers could use the DSA's complaints mechanism. A complaint about potential DSA violations can be filed with a DSC. This system is rather vague, which again underscores the need for active, well-staffed regulators. Without that, a complaint might not lead to much. It would at least be in the official records, though.

 Rules for complaints: [Articles 53 DSA](#)

## Other opportunities for researchers in the DSA

## Does the DSA allow/legalize scraping?

Most likely.

The DSA contains a provision about “publicly accessible” data. VLOPs need to provide such data to vetted researchers. This rather vague **sentence is read to mandate tools such as CrowdTangle as well as to allow scraping activities.**

Researchers have long called for legal clarity regarding scraping of social media sites and other platforms. Many reiterated this call during the Commission’s consultation on Article 40. Dozens of civil society organizations emphasized the need for access to public data in an [open letter to the Commission](#). That is why there is hope that the delegated act will clarify the rules on scraping.

In any case, this would still only apply to researchers who fulfill the vetting criteria and who want to study systemic risks in the EU. Contrary to the other rules in Article 40, this type of access would not involve the DSC, it would be a straight platform-to-researcher approach.

🇪🇺 The “scraping paragraph”: [Article 40\(12\) DSA](#)

## What other rules in the DSA might be useful for researchers?

The most promising avenues for researchers in the DSA are Article 40(4) and (8) - the vetted researcher rules described at length on this page - and Article 40(12), the public data rules described [above](#). Yet, there are more opportunities for researchers to enhance their studies with data provided or facilitated by DSA rules.

### Databases on content moderation decisions and on platforms’ terms and conditions

Platforms - and not just the very large ones - are required to offer easy-to-understand terms and conditions. They also need to provide information on content removal decisions that the European Commission will put in a machine-readable database.

🇪🇺 Rules on terms and conditions: [Article 14 DSA](#)

🇪🇺 Rules for the transparency database: [Article 24\(5\) DSA](#)

The Commission set up two transparency databases on these topics:

- One lists all platform contracts. Researchers could, for example, use this to understand how and when platforms change their terms and conditions

 Commission's [Platforms Contracts Database](#)

- Another database is constantly being populated with “statements of reasons” for why content was blocked or removed. This could be a useful source for researchers interested in how content moderation on platforms works, how companies report to users and authorities and what scale the content moderation efforts have.

 Commission's [DSA Transparency Database](#)

### Database on online advertising

The DSA calls on very large online platforms to create “ad repositories”, collecting all online advertising along with some metadata on their source of funding and display time, among other things.

 Rules on online ad repository: [Article 39 DSA](#)

This ad repository/library is meant to encourage scrutiny not only by consumers and journalists, but also by researchers studying paid communications online. However, some key metrics regarding reach and impressions are missing.

The Commission may (not must) issue some guidelines on the repositories. If researchers notice serious weaknesses with the repositories - as has been the case with the [existing, voluntary](#) ones - some clarifications might be necessary in the future.


**An overview of what reports, guidelines and databases are mentioned in the DSA can be found in this [spreadsheet](#).**


---

### Next steps for researchers

#### Immediate steps:

- Bookmark this page to keep up to date
- Learn more about ideas to coordinate researchers and develop a common infrastructure

 A working group for the European Digital Media Observatory (EDMO) led by Rebekah Tromble proposed an [independent intermediary body](#) to help DSCs vet researchers.

 In a [white paper](#), Alicia Wanless and Jake Shapiro suggest building a CERN-like structure to facilitate studies on the information environment.

**i** Help [keep track of data access requests by using this form](#) created by the Weizenbaum Institute and the European New School.

- Check in with your university/institution to see if they are planning to coordinate on data access requests
- Check in with your scientific organization and/or national academy of sciences to see if they are planning to coordinate on data access requests
- Collect good and bad experiences with platform data access to feed into scientific and policy communities

### **Medium-term steps for 2024:**

- Keep up to date on DSA developments (e.g., via this page)
  - Especially crucial: Developments on the delegated act on data access

University-wide and/or national coordination efforts can feed into:

- Supporting existing efforts such as those from EDMO (see above)
- Building complementary national/university/discipline-related coordination networks
- Approaching your national DSC, once it's clear which national authority will house it
- Approaching the European Commission during the development of the delegated act
- Preparing data access requests

### **Longer-term steps for 2025 and beyond:**

Once the delegated act on data access is done (expected for 2025):

- Review and understand delegated act with specific guidelines on vetting process
  - Particularly: How does the vetting process work? Is there an intermediary body to handle vetting?

Once networks and infrastructure, both for researchers and for the DSCs, has begun to be set up, researchers can collaborate to:

- Develop, coordinate on and turn in data access requests
- Consulting with DSCs on questions related to data access

---

### **About this page and acknowledgements**

The questions collected here stem from our own work as well as a wide variety of researchers, who posed them during workshops, webinars, in publications, consultation responses and informal talks. The answers rely, very basically, on the legal text of the DSA itself, but, more



crucially, are shaped by the conversations we had with experts from academia, civil society and administration. We gratefully acknowledge their contributions and efforts.

The following events, networks, publications and people greatly supported the information gathered on this page:

- [Workshop](#) on DSA data access organized by Paddy Leerssen at the University of Amsterdam
- [Responses](#) to the EU's consultation on data access rules
- Webinars and talks given by Mathias Vermeulen and Rebekah Tromble, who lead an EDMO working group proposing an independent intermediary body for the data access regime
- German researcher network coordinated by Prof. Ulrike Klinger and Jakob Ohme
- International researcher network coordinated by Alicia Wanless based on her and Jake Shapiro's white paper "[A CERN Model for Studying the Information Environment](#)"
- Feedback from former SNV experts Dr. Anna-Katharina Meßmer, Dr. Martin Degeling and Josefine Bahro

### About the authors

This page was first created for a [tutorial](#) at [IC2S2'23 in Copenhagen](#), conducted by [Dr. Philipp Lorenz-Spreen](#) and Dr. Julian Jaursch. Philipp is a research scientist at the Max Planck Institute for Human Development, studying the self-organized online discourse and how to empower democratic and autonomous decision-making through platform design and boosting. Julian worked at interface (formerly SNV), a not-for-profit think tank focused on tech policy issues, where he analyzed and developed policy recommendations on the EU's Digital Services Act.

Reach out to the authors and follow them on Mastodon:

#### Philipp Lorenz-Spreen (MPI):

 [https://social.dev-wiki.de/@lorenz\\_spreen](https://social.dev-wiki.de/@lorenz_spreen)

 [lorenz-spreen@mpib-berlin.mpg.de](mailto:lorenz-spreen@mpib-berlin.mpg.de)

#### Julian Jaursch:

 <https://mastodon.social/@jjaursch>

The tutorial was part of the project "[Reclaiming individual autonomy and democratic discourse online](#)", which is a joint research endeavor by the Max Planck Institute for Human Development, the University of Bristol and Northeastern University, funded by the initiative "Artificial Intelligence and the Society of the Future" from the Volkswagen Foundation.