DSA Open Lectures #2

DSA: Is Science Opening the Black Box?

Ana Malheiro, Deputy Head of Unit DG Connect, Joana Gonçalves-Sá PhD, João Vinagre PhD

whatnext.law

May 6th, 3 p.m. – 6 p.m. | Campus of Campolide, Room 123.

Commentary

Daniela Bruto da Costa Antão[1]

The NOVA School of Law's Knowledge Centre, whatnextlaw, hosted the second DSA Distinguished Lecture, open to the community, on May 6, at the Campolide Campus, in classroom 123.

From left to right: Daniela Bruto da Costa Antão PhD Candidate, Ana Malheiro Deputy Head of Unit DG Connect, Unit D1, João Vinagre,PhD, JRC, Joana Gonçalves-Sá, PhD, LIP and NOVA.

DISCLAIMER: The views expressed herein represent the author's best effort to faithfully summarise the lectures and the ensuing debate. Any errors or inaccuracies remain the sole responsibility of the author. The views and opinions expressed by the European Commission officials are their personal views only and do not represent, reflect, or bind the official position of the European Commission.

[1] Daniela Bruto da Costa Antão is a PhD Candidate at NOVA School of Law and a Fellow at Knowledge Centre whatnext.law, a partnership between NOVA School of Law and VdA Academy.

LECTURERS

Ana Malheiro is Deputy Head of Unit for Platform Regulation and Digital Markets at the European Commission (DG CONNECT), where she works on enforcing the Digital Services Act and regulating Very Large Online Platforms and Very Large Online Search Engines. She previously served as a case handler in the Commission's Digital Markets Act Task Force in DG COMP and as lead administrator on the DSA and DMA in the European Parliament.

João Vinagre PhD is a researcher at the European Commission's Joint Research Centre and a member of the European Centre for Algorithmic Transparency (ECAT), which provides scientific and technical expertise for the enforcement of the DSA's systemic-risk obligations. His work focuses on auditing and evaluating recommender systems and supporting the implementation of Article 40.

Joana Gonçalves de Sá PhD is a Principal Investigator at LIP and NOVA LINCS and an ERC-funded researcher whose work focuses on human and algorithmic biases, particularly in auditing recommender systems, such as search engines and LLMs, to uncover systemic risks to democracy and fundamental rights. She leads pioneering projects developing policy-aligned methodologies for implementing Article 40 of the DSA, bridging independent research and platform regulation.

DSA Open Lectures #2

DSA: Is Science Opening the Black Box?

Distinguished lecturers: Ana Malheiro, Deputy Head of Unit DG Connect; João Vinagre PhD, JRC; Joana Gonçalves de Sá PhD, LIP and NOVA.

whatnext.law

May 6th, 3 p.m. – 6 p.m. | Campus of Campolide, Room 123.

10 TAKEAWAYS

3-minute reading

1. Based on our own estimations, between December 2023 and April 2026, the Commission opened 17 formal proceedings against designated VLOPs. Some of these concern the obligations under Articles 34 and 35 (duty to identify and mitigate systemic risks) and Art 40 (data access obligations):

Platform / Service	Illegal Content/ Products	Minors Protection	Addictive Design	Ads Transpar/ Repository	Data Access Art 40	RecSys	Info Integrity Elections	Consumer Protection / Marketplaces	GBV / Well-being	Risk Management / Not &Action
Meta	x	x	x	x	x	x	x	-	-	x
X	x	-	-	x	x	x	x	-	x	-
TikTok	-	x	x	x	x	-	x	-	-	-
TikTok Lite	-	-	x	-	-	-	-	-	-	-
AliExpr.	x	x	-	x	x	x	-	x	-	x
Temu	x	-	-	-	-	-	-	x	-	-
Shein	x	-	x	-	-	x	-	x	-	-
Pornhub Stripchat XNXX Xvideos	-	x	-	-	-	-	-	-	-	-
Snapchat	x	x	-	-	-	-	-	-	-	-

2. Illegal content/products seems to be the most recurrent trigger of enforcement, reflecting the DSA's baseline safety function.

3. Although structurally different, data access (Article 40) and protection of minors (POM) appear almost as frequently, confirming their enforcement concerns: POM as an end in itself, data access as instrumental to tackling all other systemic risks.

4. Addictive design and ads transparency form a second tier of recurring risks, tightly linked to platform architecture.

5. Gender-based violence/well-being appears only once, but notably in a high-salience case (X's Grok).

6. Science is intrinsic to DSA enforcement. There are 3 types of access to VLOPSEs data:

Type of data	Entitled entity	Scope & Purpose of Access	Article 40
Reasonably necessary	Commission & DSC of establishment	- Monitoring and enforcement of DSA	Para. 1
Publicly available	Qualified researchers	- Systemic risks *ex vi* Art 34	Para. 12
Non-publicly available	Vetted researchers	- Systemic risks *ex vi* Art 34 - Mitigation measures *ex vi* 35.	Para. 4

7. Access to data under Article 40(12) DSA is independent of platforms and obtained through APIs, scraping or crawling tools, and user data donations, though the latter may be restricted by the Digital Omnibus Regulation Package. This provision is being actively enforced by the Commission notably against X, Meta, AliExpress and TikTok. On 5 December 2025, the Commission adopted the first non-compliance decision against X for, among others, violation of Article 40 (12) DSA. As a follow -up to this, X will soon publish an Action Plan setting out the measures it intends to adopt to implement this important provision.

8. Technical conditions and procedures for the vetting of researchers and submission of access applications to non-publicly available data under Article 40(4) are further specified in the Commission Delegated Regulation (EU) 2025/2050 and operationalised in the DSA Data Access Portal. The DSA of the country of establishment plays a key role in this form of data access.

9. The official portal for European data is the point of access to **European open data**. The DSA Data Access Portal contains a listing of VLOPSEs pages dedicated to data access requests (Data Catalogues), and the DSA Transparency Database contains statistics on the statements of reasons of content moderation decisions of online platforms in the last six months.

10. Voluntary data-sharing programmes (pre-DSA) coexist with DSA-mandated data access schemes. With misaligned incentives and enormous resource & information advantages, VLOPSEs engage in defensive strategies: from delays and intermittent or partial supply of data to outright refusals or even threats of millionaire damage claims, criminal accusations and defamation campaigns against researchers. Collective organisational structures of researchers and publicly funded insurance could attenuate such asymmetries. A Good Samaritan principle, akin to a leniency rule, could create incentives to amicable self-regulation.

COMMENTARY

1. INTRODUCTION

The **Digital Services Act (DSA) is a quasi-constitutional** [2]chart for the European digital society. At the heart of the safeguards supporting the *European way of life* are Articles 34 and 35. Building on the profound societal transformations brought about by the rapid adoption of social media, search engines, and other online platforms over the past two decades, these provisions impose obligations on very large online platforms and search engines (VLOPSEs) to address the systemic risks they pose.

These **systemic risks** include threats to fundamental rights, democracy and public discourse, the protection of minors, gender equality, and the physical and mental well-being of individuals. Non-compliance may lead to penalties and liability. It is at this juncture that the role of data becomes essential. To understand how VLOPSEs exert their impact – either positive or negative – it is necessary to examine the interstices of algorithmic content management, the impact of user interface features, and the practices of consumer data exploitation. To operate this framework, the DSA created a framework of access to VLOPSEs' data.

Malheiro (2) provided an overview of the **Digital Services Act (DSA)** with an emphasis on **enforcement proceedings** opened by the Commission since its entry into force and on **Article 40 data access for researchers**. She outlined the scope of obligations imposed on designated VLOPSEs, explained the differentiated enforcement roles of the European Commission, national Digital Services Coordinators (DSC), and the European Board for Digital Services (EBDS), and presented a detailed enforcement track record of formal proceedings, fines, binding commitments, and preliminary findings.

Vinagre (3) introduced the students to the European Centre for Algorithm Transparency and walked the audience through the researchers' rights of access to **publicly available data** and to **non-publicly available data**, under paragraphs 12 and 4, respectively, of Article 40 of the DSA.

Gonçalves-Sá (4) **showcased the possibilities and the limitations of publicly available data** for scientific auditing of VLOSEs, the necessity of non-publicly available data, and **her firsthand crushing experience with Article 40(4) data access applications** in the first months after the framework's entry into force.

In the end (5), we answer the question "*is science opening the black box*?".

[2] Digital constitutionalists pursue the goal of transplanting the constitutional frameworks of the rule of law and representative legitimacy to the governance of online platforms in reaction to the fragmentation of sovereignty of the traditional liberal constitutional state by the phenomenon of big tech. The DSA is considered a paradigmatic example – alongside the GDPR and the DMA – of constitutionalisation of fundamental rights in EU law, in the wake of judicial activism. Blurring the dichotomy of public and private law, may, however, miss the fundamental nature of a private contractual bond established between platform providers and their users. For an overview of the digital constitutionalism project and their most influential authors, Abreu Duarte, Francisco, De Gregorio, Giovanni; and Golia, Angelo Jr, *Perspectives on Digital Constitutionalism* (2023). in O. Kanevskaia, P. Palka, B. Brozek, Handbook on Law and Technology (Edward Elgar) Forthcoming; http://dx.doi.org/10.2139/ssrn.4508600. For a critique of the flaws of a *public law bias*, Mills, Gilad. 2024. A Contractual Approach to Social Media Governance. Yale Law and Policy Review. Vol 42, Issue 2: 522-625.

2. Exploring evidence-based oversight of very large online platforms under Article 40 and beyond, by ANA MALHEIRO

"Because the DSA exists, when you engage with online platforms as users, you have rights, and they have obligations." This was the opening statement of the Deputy Head of Unit, Ana Malheiro.

2.1. Overview of the commission DSA enforcement actions

As of this day, the Commission has **opened 17 formal proceedings**,[3] sent **over 170 requests for information**, and overseen a significant increase in corrective action by platforms, with **over 52 million content or account-related decisions reversed**, representing **35% of removals being overturned** following DSA implementation.[4]

Formal proceedings have been opened against several platforms, addressing several key areas of **systemic risks.**

- Proceedings concerning **minors' protection**, **risk management**, **illegal content**, **information integrity**, **addictive design**, and **data access** were opened against **Meta (Facebook & Instagram)**

- Proceedings related to **illegal content**, **information integrity**, **deceptive design**, **recommender systems**, and **data access for researchers**, including later extensions linked to AI functionalities (Grok) and gender-based violence were opened against **X (formerly Twitter)**

- Proceedings addressing **addictive design**, **minors' protection**, **advertising repository transparency**, **elections-related risks**, and **data access** were opened against **TikTok (including TikTok Lite).**

- Proceedings focusing on **illegal products**, **consumer protection**, **trader traceability**, **platform design**, and **systemic risk management were** opened against online marketplaces **AliExpress, Temu, Shein.**

- Proceedings concerning **age-verification failures** and **inadequate safeguards for minors** were opened against pornographic platforms **Pornhub, Stripchat, XVideos, XNXX.**

- Proceedings related to **grooming risks**, **illicit content**, and **minors' protection** were opened against **Snapchat.**

[3] The official overview of DSA enforcement actions is available here.

[4] Statistical information of content moderation decisions of around 300 platforms is available in the DSA Transparency Database (here).

Table 1 presents a matrix correlating platforms with the systemic risk category at issue in the respective proceedings, preliminary findings, commitments, or decisions for each platform(s). The legend detailing the content of each category is presented below.

Table 1 - Systemic Risk or Infringement Categories Raised in the Proceedings with Respect to Each Platform

Platform / Service	Illegal Content/ Products	Minors Protection	Addictive Design	Ads Transpar/ Repository	Data Access Art 40	RecSys	Info Integrity Elections	Consumer Protection / Marketplaces	GBV / Well-being	Risk Management / Not &Action
Meta	x	x	x	x	x	x	x	-	-	x
X	x	-	-	x	x	x	x	-	x	-
TikTok	-	x	x	x	x	-	x	-	-	-
TikTok Lite	-	-	x	-	-	-	-	-	-	-
AliExpr.	x	x	-	x	x	x	-	x	-	x
Temu	x	-	-	-	-	-	-	x	-	-
Shein	x	-	x	-	-	x	-	x	-	-
Pornhub Stripchat XNXX Xvideos	-	x	-	-	-	-	-	-	-	-
Snapchat	x	x	-	-	-	-	-	-	-	-

Systemic Risk Categories (Legend):

Illegal Content/Illegal Products — Dissemination of illegal content; illegal goods in online marketplaces.

Minors Protection — Age verification failures; exposure of minors; grooming risks.

Addictive Design — Reward-based mechanisms; design choices affecting user well-being.

Ads Transparency/Ad Repository — Completeness, accuracy, timeliness, and accessibility of advertising repositories.

Data Access (Art. 40) — Access to public data (Art. 40(12)) and non-public data for vetted researchers (Art. 40(4)).

RecSys (Recommender Systems) — Algorithmic amplification; system design and transparency.

Information Integrity / Elections — Integrity of civic discourse; electoral processes.

Consumer Protection / Marketplaces — Trader traceability; affiliate programmes; hidden links; unsafe products.

GBV / Well-being — Gender-based violence; serious negative effects on mental or physical well-being.

Risk Management/N&A — Systemic risk assessment; mitigation measures; Notice-and-Action mechanisms.

Table 2 below summarizes the Commission's enforcement cases from December 2023 through April 2026.

Table 2 - DSA Enforcement Cases by Chronological Order and Procedural Details

Platform / Service	Date	Procedure	Main issue(s) identified	Status / Outcome
X (formerly Twitter)	Dec/23	OP	Illegal Content, Information Integrity, Data Access, Deceptive Design, RecSys	Proceedings ongoing
TikTok	Feb/24	OP	Addictive design; Protection of minors; Ad repository transparency; Data access	Proceedings ongoing
AliExpress	Mar/24	OP	Illegal content	Proceedings ongoing
Meta (Facebook & Instagram)	Apr/24	OP	Illegal content, Information Integrity	Proceedings ongoing
TikTok Lite	Mar/24	OP	Addictive design (reward-based features)	Led to Comm
Meta	May/24	OP	Addictive design, Protection of Minors.	Proceedings ongoing
X	July/24	PF	Deceptive design; Data access	PF issued
TikTok Lite	Aug/24	Comm	Addictive design targeting users through reward-based mechanisms; withdrawal in the EU; non-relaunch commitment	**Binding commitments adopted**
Temu	Oct/24	OP	Illegal content	**€200M fine imposed**
TikTok	Dec/24	OP	Electoral processes and elections-related risks	Proceedings ongoing
TikTok	May/25	PF	Transparency of advertising repository	PF issued
Porn platforms (Pornhub, Stripchat, XVideos, XNXX)	June/25	OP	Protection of minors; Age-verification failures; Inadequate safeguards	Proceedings ongoing

Platform / Service	Date	Procedure	Main issue(s) identified	Status / Outcome
AliExpress	June/25	Comm + PF	Monitoring and detection of illegal content; "Hidden links" and Affiliate Programme risks; Products affecting health and minors; Notice-and-Action; Internal complaint handling; Advertising; RecSys transparency; Trader traceability; Access to public data by researchers	**Binding commitments adopted; PF issued**
Temu	July/25	PF	Systemic risks related to illegal content	PF issued
TikTok	Oct/25	Comm	Advertising repository: completeness, accuracy, update speed, search and filtering functionalities	**Binding commitments adopted**
TikTok	Oct/25	PF	Data access for researchers	PF issued
Meta (Facebook & Instagram)	Oct/25	PF	Data access for researchers; Notice-and-Action mechanism	PF issued
X	Dec/25	Fine	Transparency of ad repository; Data access for researchers; Use of dark patterns (blue-check verification)	**€120M fine imposed**
TikTok	Dec/25	Comm	Advertising repository (commitments confirmed and strengthened)	**Commitments confirmed**
X	Jan/26	OP	Grok functionalities; Illegal content; Gender-based violence	Proceedings ongoing
X	Jan/26	Ext. Inv.	RecSys	Investigation extended
Shein	Feb/26	OP	Illegal products; Consumer protection; Addictive design; Transparency; RecSys	Proceedings ongoing
TikTok	Feb/26	PF	Addictive design	PF issued
Snapchat	Mar/26	OP	Protection of minors; Grooming risks; Illicit content	Proceedings ongoing

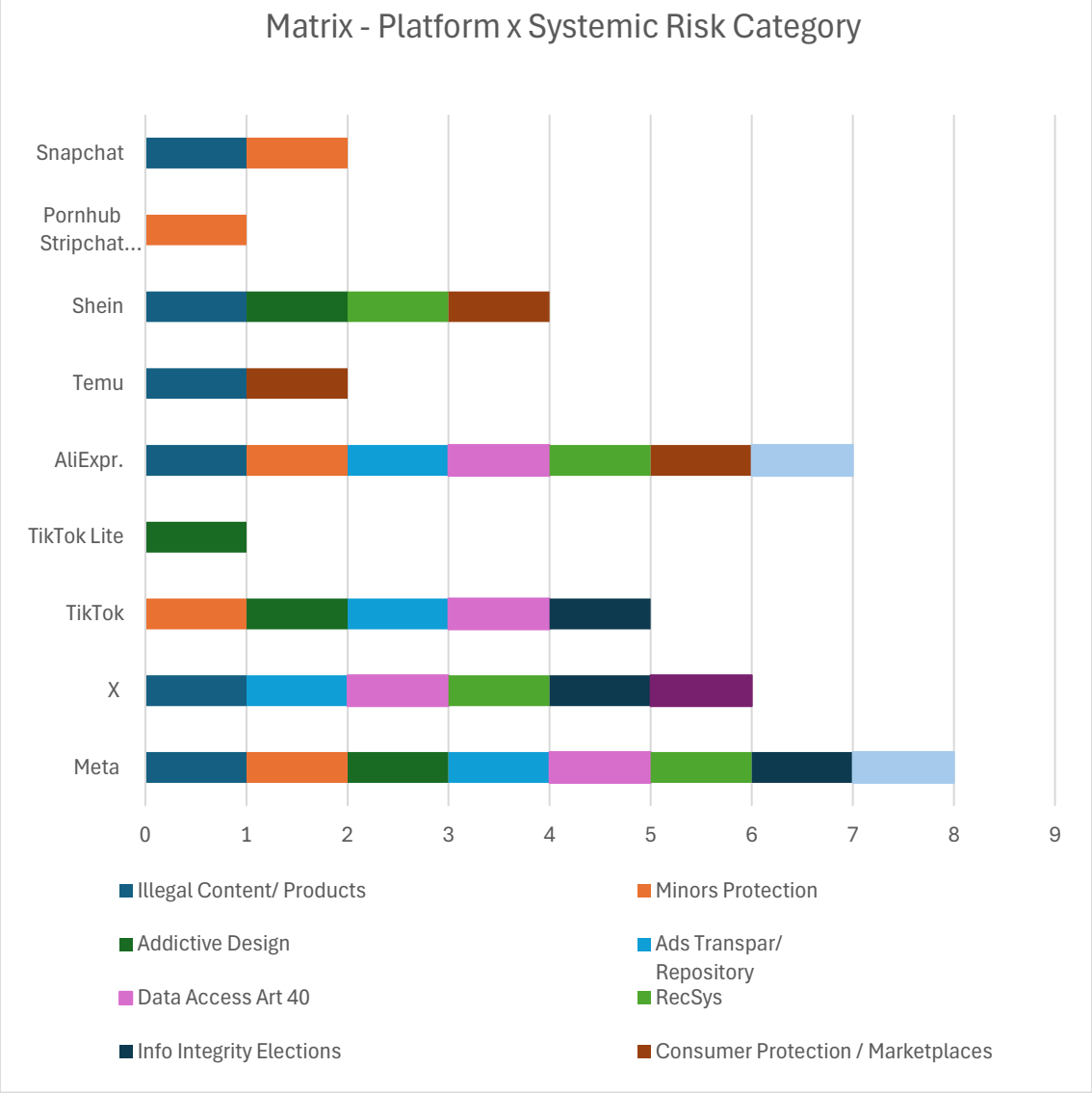
Platform / Service	Date	Procedure	Main issue(s) identified	Status / Outcome
Porn platforms	Mar/26	PF	Protection of minors	PF issued
Meta (Facebook & Instagram)	Apr/26	PF	Protection of minors; Systemic risk management	PF issued

Legend:

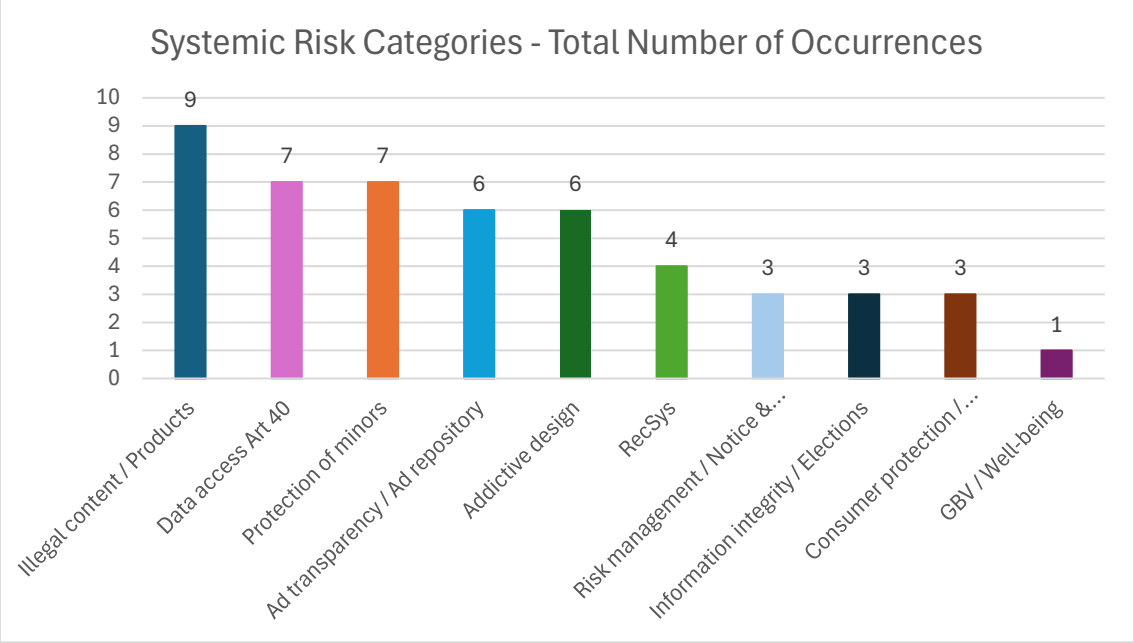
OP	Opening of proceedings
PF	Preliminary Findings
Comm	Acceptance of commitments decision (binding)
Fine	Decision of non-compliance with financial penalty
Ext. Inv.	Opening of Extended investigation
RecSys	Recommender systems

Graph 1 depicts, with colours, the systemic risks (and data access infringement) addressed against each VLOPSE and Graph 2 the total number each systemic risks (and data access infringement) occurs in the Commission's enforcement decisions.

Graph 1


Matrix - Platform x Systemic Risk Category
Snapchat
Pornhub Stripchat...
Shein
Temu
AliExpr.
TikTok Lite
TikTok
X
Meta
0
1
2
3
4
5
6
7
8
9
Illegal Content/ Products
Minors Protection
Addictive Design
Ads Transpar/ Repository
Data Access Art 40
RecSys
Info Integrity Elections
Consumer Protection / Marketplaces

Graph 2



In Graph 2, each count represents one explicit appearance of a systemic-risk (data access) issue in an Opening of Proceedings (OP), Preliminary Findings (PF), a Commitment decision (Comm), a Fine, or an Extended Investigation. The same platform is counted multiple times if the risk reappears in later stages or new proceedings. This is a raw frequency measure: it has no weighting, no platform aggregation, and no interpretation.

The key facts observed from the raw counts are (vertical Graph 2):

1. **Illegal content/products are the most recurrent trigger of enforcement, reflecting the DSA's baseline safety function.**
2. **Although structurally different, data access (Article 40) and protection of minors (POM) appear almost as frequently, confirming their enforcement concerns: POM as an end in itself, data access as instrumental to tackling all other systemic risks.**
3. **Addictive design and ads transparency form a second tier of recurring risks, tightly linked to platform architecture.**
4. **Gender-based violence/well-being appears only once, but notably in a high-salience, escalated context (new proceedings).**

Vertical patterns (Graph 2) reveal enforcement priorities (e.g. POM, addictive design), whereas horizontal breadth (Graph 1) highlights platforms with multi-dimensional systemic risk exposure.

Takeaways from the horizontal Graph 1:

i. **Meta, X, and AliExpress platforms are notable for their multi-dimensional systemic risk exposure.**
ii. **Sparse rows (such as TikTok Lite's) evidence narrow but structurally significant exposure to systemic risk.**

iii. **Coincidentally or not, only Chinese companies ended infringement proceedings with binding commitment decisions (TikTok, TikTok Lite, AliExpress).**
iv. **Data access occurrences may identify most requested platforms for Article 40-based research initiatives. In fact, data access under Article 40 (both 40(12) and 40(4)) occurs repeatedly, evidencing a compliance problem and an enforcement target.**

Malheiro stressed that the focus of enforcement extends, however, beyond major social networks, to marketplaces, adult platforms, and messaging services. Some of these areas have so far deserved less attention from researchers. Malheiro highlighted that enforcement and research attention should not focus exclusively on the most visible platforms, and attention by researchers could also be drawn to **less-discussed VLOPSEs**, such as online marketplaces, pornographic platforms, and certain social and messaging services, which are central to current enforcement actions—particularly around **minors' protection, illegal products or addictive design**.

The Deputy Head of Unit encouraged the academic and research community to submit **concrete research ideas and initiatives** that use Article 40 mechanisms to study systemic risks in the Union. She further stressed that the Commission sees independent research as essential to effective oversight and repeatedly underlined that **data access—both under Article 40(4) for *vetted* researchers and Article 40(12) for *qualified* researchers—is a priority area of enforcement**.

2.2. Data access in the DSA transparency framework

Access to VLOPSEs data is not merely a technical add-on to digital regulation; it is a **core enabler of digital safety**. The underlying premise is clear: without meaningful access to platform data, regulatory oversight remains largely formalistic.

There are three principal categories of access under Article 40 of the DSA, each reflecting a different balance among **control, transparency, and systemic risk management**.

1. Access by public authorities - European Commission and the national Digital Services Coordinators (DSCs)

This first layer **concerns data access by public authorities** — primarily the European Commission and Digital Services Coordinators- and is solely **enforcement-driven: i**t supports investigations, monitoring, and compliance assessment. Platforms are legally obligated to provide data, while authorities retain ample discretion over the scope and use of the data they request.

2. Access by Vetted Researchers

The second regime introduces access for **independent researchers**, subject to strict vetting conditions. This is arguably the most innovative — and politically sensitive — aspect of the DSA architecture. The ambition is to enable **external scrutiny of systemic risks and mitigation measures**. The model reflects a cautious compromise between the public interest and the legitimate interests of VLOPSEs. Access is conditional, mediated by platforms, and constrained by data protection and trade secrecy considerations.

This regime attempts to reconcile **epistemic openness** with **platform control**, but a key tension remains whether researchers can obtain data meaningful enough to challenge platform narratives rather than simply confirm them.

3. Public Access by Qualified Researchers and Civil Society

The third layer consists of **indirect or public-facing access**, primarily through transparency reporting obligations and data repositories. For researchers and civil society organisations, Article 40 of the DSA provides access to tools such as APIs with **public data** and **protects those who use automated access means** (e.g. scraping) to study **systemic risks in the EU.**

This helps inform the public, policymakers, and civil society, as later demonstrated by Joana Gonçalves-Sá, but its effectiveness is contingent on how platforms frame and curate disclosures, thereby introducing risks of selective transparency.

Malheiro highlighted that **access to data is a pillar of the *État de Droit* in today's digital society**. The most consequential dimension of this assertion is not whether access exists, but **how operational and meaningful it becomes in practice**. Ultimately, the effectiveness of the DSA's data access cooperative model depends less on its legal design than on its **implementation dynamics** — including institutional capacity, platform cooperation, and the willingness to confront informational asymmetries head-on. In this sense, data access is not a peripheral issue but the central lever by which the DSA's broader ambition to govern systemic risk in digital environments can succeed.

3. DATA ACCESS AND ECAT, by JOÃO VINAGRE

João Vinagre, PhD introduced the students to the European Centre for Algorithm Transparency and walked the audience through the researchers' rights of access to **publicly available data** and to **non-publicly available data**, under paragraphs 12 and 4, respectively, of Article 40 of the DSA.

3.1. ECAT

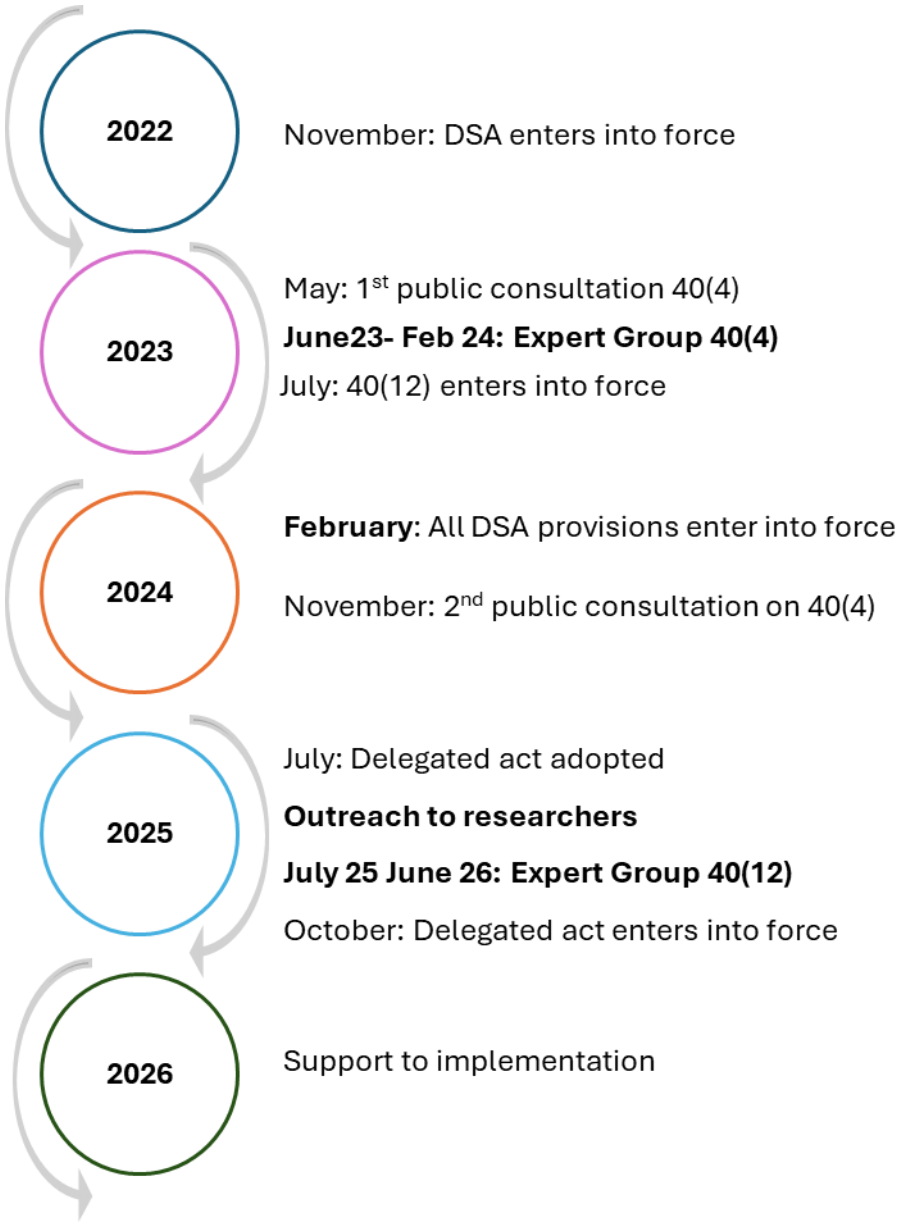
"Science is embedded in the DSA implementation." Access to data is not a purely legal entitlement. It is essential to enforcement. ECAT is at the centre of the European Commission's effort to put into practice the DSA's data access provisions. ECAT functions as a **technical and scientific support structure for enforcement**. ECAT studies **how algorithmic systems shape behaviour and create systemic risks** and develops methods to measure and audit those effects. It provides evidence, analytical input, and expertise necessary to assess compliance with DSA obligations. ECAT has 2 types of research teams: the anticipatory research team and the inspective teams working with DG Connect. The centre operates in three locations: Brussels, Seville, and Ispra.

In its task of **producing and coordinating research on algorithmic systems, ECAT's** work combines interdisciplinary approaches to the study of human behaviour and machine intelligence, with the explicit objective of informing regulatory oversight. In this respect, research is not presented as an ancillary activity, but as part of the enforcement ecosystem itself, contributing to the identification and understanding of systemic risks defined under the DSA.

This support extends to the **development and implementation of regulatory instruments** and **to assisting Digital Services Coordinators** in handling requests and procedures. Another significant dimension of ECAT's activity concerns **structuring the emerging research community around Article 40**. This community-building function is closely linked to the development of **data access infrastructure**, most notably the **DSA Data Access Portal**. The portal is a centralised interface through which researchers can access procedures, submit applications, and interact with regulators and platforms. Complementary measures, such as the publication of data catalogues and guidance materials, are designed to standardise access conditions and clarify what data is available and under what modalities.

Graph 3 describes the activities of ECAT supporting the implementation of Article 40, between 2023 and today.

Graph 3 – Timeline of ECAT's Activities Supporting Article 40 Implementation



Between 2023 and 2025 ECAT was mostly focused on providing technical and scientific support to the preparation of the Delegated Act on data access under the DSA (Commission Delegated Regulation (EU) 2025/2050 of 1 July 2025). The centre issued two public consultations, coordinating expert groups, and occasionally gathered evidence for enforcement purposes. In 2025, ECAT rolled out its outreach and community-building strategy through the mapping of research communities, tutorials, webinars, and special sessions in scientific events. Community building continues in 2026, but this year and onwards marks the implementation stage, supporting DSCs and producing and publishing the Data Access Portal and informative material (FAQ and newsletter).

3.2. Access to publicly available data – Art 40(12)

Access to publicly available data can be requested by ***qualified*** *researchers* directly from the platform (this is a **two-actor process**: the researchers and the VLOPSEs). VLOPSEs have the obligation to provide access to data "without undue delay" and "where technically possible, to real-time data". This d**ata can also be accessed independently of providers**: independent data access techniques, such as scraping, crawling, and some types of user data donations, ***cannot be prohibited for eligible researchers***.

Article 40(12) reads:

> *Providers of very large online platforms or of very large online search engines shall give access without undue delay to data, including, where technically possible, to real-time data, provided that the data is publicly accessible in their online interface by researchers, including those affiliated to not for profit bodies, organisations and associations, who comply with the conditions set out in paragraph 8, points(b),(c),(d) and(e), and who use the data solely for performing research that contributes to the detection, identification and understanding of systemic risks in the Union pursuant toArticle34(1).*

In order to qualify under paragraph 12 of Article 40, researchers must be

- *independent* from commercial interests,
- disclose their funding,
- demonstrate **capacity to protect personal data**, data security, and confidentiality requirements,
- demonstrate the ***proportionality & necessity*** of the request and its **contribution to its general objective**,
- and use the data for the sole purpose of **detecting, identifying and understanding** of EU systemic risks (mitigation is not included).

The Expert Group on Article 40(12), established by ECAT (whose mandate runs until June 2026), has assessed four categories of systemic risks – related to illegal content, gender-based violence, civic discourse and electoral processes, and the protection of minors. The experts have collected scientific and procedural evidence on:

- **6 social media**: X (including Grok), Facebook, Instagram, TikTok, YouTube, LinkedIn;
- **4 marketplaces**: Amazon, AliExpress, Shein, Temu;
- **3 pornographic platforms**: Pornhub, Xvideos, XNXX.

3.3. Access for vetted researchers

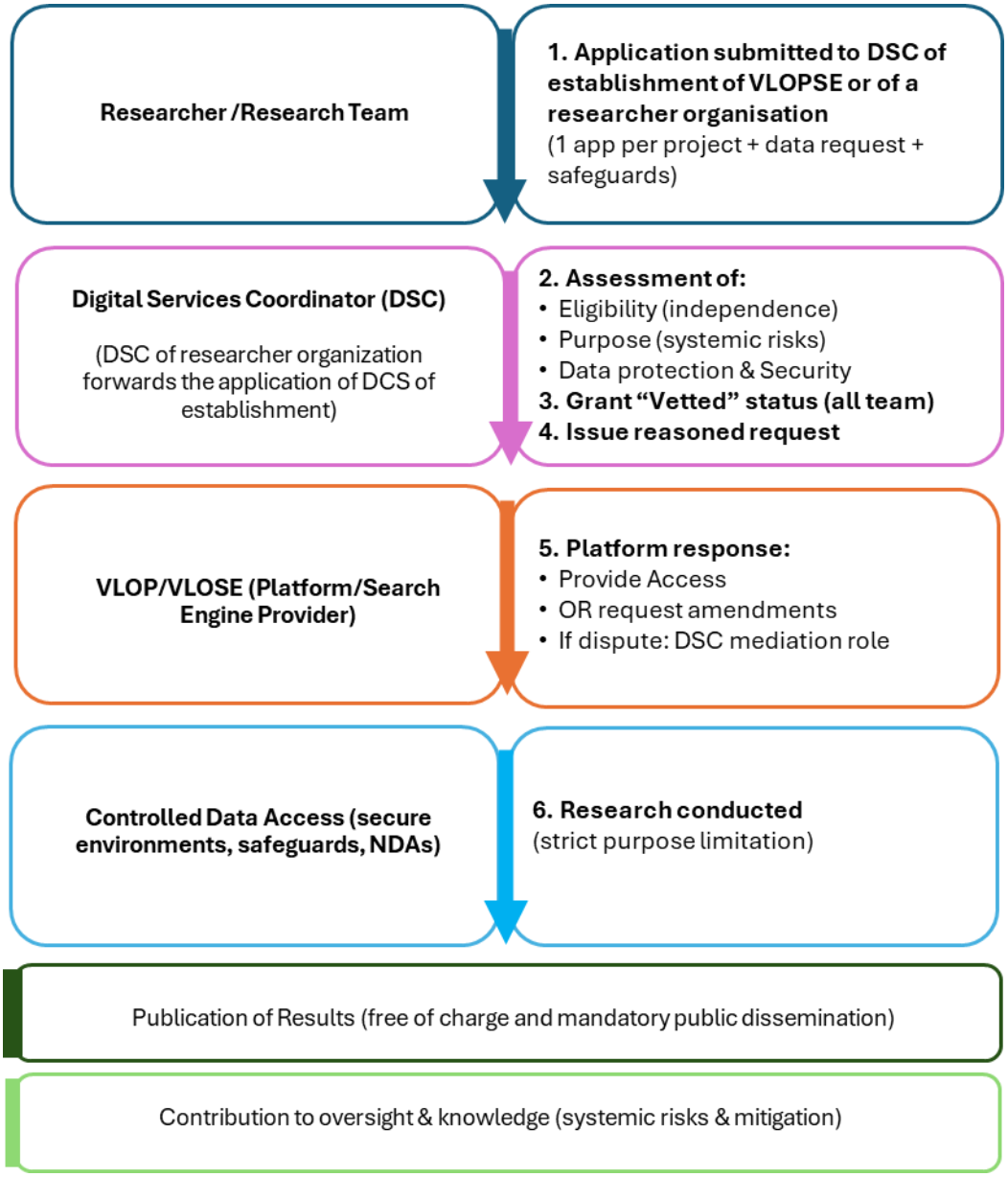
A**ccess to non-publicly available data** is understandably much more demanding for applicants and requires prior **vetting of the applicant researcher**.

> **DSA Article 40(4) reads:**
>
> *Upon a reasoned request from the Digital Services Coordinator of establishment, providers of very large online platforms or of very large online search engines **shall, within a reasonable period, as specified in the request, provide access to data to vetted researchers** who meet the requirements in paragraph 8 of this Article, for the sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the Union, as set out pursuant to Article 34(1), **and to***

the assessment of the adequacy, efficiency and impacts of the risk mitigation measures *pursuant to Article 35.*

The flowchart in Image 1 provides a very simplified visual depiction of this process.

Graph 4 – Vetting researchers and access process to non-publicly available data



Access to non-public data is framed as an **administrative process rather than a right to direct access.** It is institutionally mediated. Digital Services Coordinators are involved at every critical stage: vetting researchers, issuing requests, and mediating disputes.

The **system is multi-actor structured.** It requires coordination among researchers, national authorities, platforms, and the European Commission. This layered structure implies a degree of procedural complexity, as each actor plays a distinct role in evaluating, transmitting, and implementing access requests. Hence, adoption of **technical and organisational arrangements** in enabling access, notably, the Data Access Portal and the list of data catalogues

Access is conditional. It is not general or open-ended but limited to clearly defined research purposes linked to systemic risks. It is further constrained by requirements relating to proportionality, necessity, and the ability to ensure data protection, confidentiality, and security.

In order to obtain vetted access, researchers must fulfill the following conditions:

- Affiliated to research organisations (cf. Copyright Directive)
- Independent from commercial interests
- Disclose their funding
- Demonstrate capacity to protect personal data, confidentiality of information, including trade secrets and of not compromising the security of the data and the service. VLOPSEs demand several safeguards:
    - Technical: clean rooms,/data vaults, differential privacy, activity logging, etc.
    - Organizational: restricting access to few individuals, no right to subcontract/add people to the project, oversight by [Data Protection Officers/ethics board/other]
    - Legal: case-by-case Non-Disclosure Agreements between provider and researchers
- Demonstrate the proportionality & necessity of the request and the contribution to the general objective.
- **Commit to making the research results publicly available free of charge when the project is over**
- Use the data for the sole purpose of contributing to the understanding of EU systemic risks, **and to the assessment of the adequacy, efficiency, and impacts of the risk mitigation measures**

This reflects the sensitivity of the datasets involved and the need to balance access with other legal interests.

The Delegated Act (Recital 13) exemplifies the **wide scope of data** that can be the subject of a request by vetted researchers:

1. Data related to users of the services: such as profile information, relationship networks, individual-level content exposure, and engagement histories;
2. Interaction data: such as comments or other engagements;
3. Data related to content recommendations: including data used to personalise recommendations;
4. Data related to the targeting of advertisements and profiling, including cost per click data and other measures of advertising prices;
5. Data related to the testing of new features prior to their deployment, including the results of A/B tests;
6. Data related to content moderation and governance, such as data on algorithmic or other content moderation systems and processes, including changelogs, archives, or repositories documenting moderated content, including accounts, as well as
7. Data related to prices, quantities, and characteristics of goods or services provided or intermediated by the data provider.

All DSCs must publish their points of contact (also on the DSA Data Access Portal. **Data providers** must make easily findable on their interface four sets of information: details of their **point of contact**; a link to the DSA Data Access Portal; and **data catalogues, with** data assets, structure and metadata; their suggested access modalities, according to different levels of sensitivity.

4. *VLOSE Auditing and Data Access under Article 40.4*, by Joana Gonçalves-Sá, PhD

Gonçalves-Sá **showcased both the possibilities and the limitations of publicly available data** for scientific auditing of VLOSEs, the necessity of non-publicly available data, and **her firsthand experience with Article 40(4) data access applications** both as participant in a pilot programme

to support the penning of its Delegated Act, and in the first months after the framework's entry into force.

Gonçalves-Sá leads a multidisciplinary research group at the Social Physics and Complexity Lab (SPAC-LIP) where they study **how information access influences internet users' experience and how human and algorithmic biases reenforce each other.** SPAC investigated how users searching for politically relevant topics may encounter results that cannot be explained by previous interactions or external factors. In particular, they have uncovered uneven representation of political actors before key electoral periods. The empirical work relied on the systematic collection and analysis of **search engine results pages (**Google, Bing, DuckDuckGo, and Yahoo!) **and LLM applications' responses** (Copilot and ChatGPT), comparing the outputs across these sources.

"What is the best party/Who should I vote for in the European Elections?" The research tested the results for this query, using a system of automatized bots that purport to be humans, and identified **measurable patterns of overrepresentation and underrepresentation of political groups**, clustered in five categories: Radical Right, Mainstream Right, Greens, Mainstream Left, and Radical Left. The results showed an overwhelming overrepresentation of Radical Right appearances and underrepresentation of all other political groups, but with quite expressive values for the Mainstream Right in all analysed (five) European Countries.

The ensuing question is ***why****?* Why does over/underrepresentation occur? Gonçalves de Sá's researchers raised possible interacting factors. Could it be differences in user behavior (such as search frequency and click patterns), statistical amplification of smaller underlying bias, search engine optimization techniques, exploitation of search engines' algorithmic fragilities by third parties, or other factors?

To answer this question, the **different hypotheses need to be tested**.

> **Publicly available data (Article 40(12)** for Bing and Google, should be found in the repository of these search engines' catalogues but the available information in insufficient as it does not include the required search terms, periods or underlying algorithmic choices. So, it is basically useless to test for the systemic risk "negative effects on civic discourse and electoral processes, and public security." The researchers need to have access to non-publicly available data.
>
> In other words, **publicly accessible data provides visibility into outputs** – what platforms display to users – and thus **enables the observation of patterns and correlations across contexts. However, it does not directly reveal the internal mechanisms through which those outputs are produced, such as ranking parameters, personalisation processes, or moderation decisions.** The research therefore illustrates both the possibilities and the limits of Article 40(12): it allows for systematic observation of platform behaviour, while leaving the underlying decision-making processes only partially accessible.
>
> **What about non-public data?** Gonçalves-Sá's team initiated a data access application for vetting and access to non-public data.

The 33 fields (if our counting is correct) in the form fall into 5 main categories: funding transparency, research purpose, data request specification, risk & data protection, and legal commitments & compliance.

- *The application form in detail*

On **funding information**, the applicant researcher needs to answer 10 questions, mostly aimed at evaluating ability to conduct the research and possible conflicts of interest:

1. What is the **nature of the funding** (EU funding or private)?
2. Is the funding **from the EU**?
3. Is the funding **private**?
4. What is the **country of establishment** of the funding source?
5. What is the **name of the funding source**?
6. What is the **grant reference number**?
7. In which **year was the funding obtained**?
8. What is the **duration of the funding** (start and end dates)?
9. What is the **amount of funding (EUR)** (if applicable)?
10. What **non-financial contributions** were provided (e.g., facilities, premises)?

On **the research purpose and relevance, applicants must explain:**

11. How will the **expected research results contribute** to:
    12. Detecting systemic risks,
    13. Identifying systemic risks,
    14. Understanding systemic risks,
    15. Assessing risk mitigation measures under the DSA?
16. Which **systemic risks and mitigation measures** (Articles 34(1) and 35 DSA) is the research related to?

On **the data request description:**

17. What **type and format of data** is being requested?
18. What is the **description of the requested data** (format, scope, attributes, metadata, documentation)?
19. Why is access to the data **necessary and proportionate**?
20. What are the **timeframes of the research** for which the data is requested?

For illustration purposes, **to test for "Statistical amplification of a smaller underlying bias/personalization,**" the application details as follows:

Table 3

Data / Metadata Type	**Description**	**Format / Notes**
Suggested ads	Ads suggested by the system (including advertiser and landing page)	JSON or CSV
Data structure / schema	Schema defining SERP fields, ranking explanations, ads, news, and query context	JSON
User segmentation factors	Anonymized personalization attributes used in ranking	Fully anonymized; categorical
Provenance	Source, platform, system, and method used to collect/generate data; model versioning	JSON or CSV

Data / Metadata Type	Description	Format / Notes
Contextual signals (incl. trending topics)	External/contextual influences (news surges, local events, time-of-day patterns)	JSON or short descriptors
Session-level context	Prior queries/actions influencing results within same session	JSON; anonymized
External signals	Non-algorithmic influences (trending spikes, paid promotions, news cycles)	JSON or descriptive text

Source: Joana Gonçalves-Sá, May 6th, 2026, *VLOSE Auditing and Data Access under Article 40.4,* keynote in whatnext.law: *Open Lecture «DSA: Is Science Opening the Black Box?»*, p-26

On **Data protection and risk assessment**, applicants must detail:

21. What are the **risks related to confidentiality, data security, and personal data protection**?
22. Does the requested data **include personal data** (Yes/No)?
23. If yes, what are the **legal processing grounds under GDPR**?
24. What are the **proposed modalities for accessing the data**?
25. What **measures (technical, legal, organisational)** will be implemented to mitigate risks?

Finally, the last chapter on **documentation and compliance requests the following:**

26. Can you **provide documentation** proving capacity to ensure:
    27. confidentiality,
    28. data security,
    29. personal data protection?
30. Do all applicants **commit to using the data only for research on systemic risks**?
31. Do all applicants **commit to making results publicly available (free of charge)**?
32. What is the **estimated publication date** of the research results?
33. Do you **confirm compliance with Article 40(8) DSA requirements** and declare that the application is accurate and not misleading?

If the reader feels overwhelmed by reading this list, what can be said about applicants for vetting and data access? Applicants must submit o**ne application for each project and each VLOP/VLOSE**. Gonçalves-Sá submitted 10 incremental data requests.

Opening the floor for discussion, Gonçalves-Sá remarked, *"There is a huge asymmetry of knowledge and power between the platforms and the researchers and It is overwhelming how much we don't know."*

5. Debate and final reflections

The Commission's enforcement activity shows that it is moving quickly to exercise its powers to address systemic risks posed by VLOPSEs, even when some risk factors are challenging to substantiate through evidence-based analysis. These risks include the lack of transparency in advertising and recommender systems and their potential effects on fundamental rights,

democratic processes, and public well-being, as well as the addictive design of services and its effects on minors and young people.

The European debate on age verification – a question raised by the audience - has shifted from a binary question of whether age checks should exist to how they can be implemented without undermining fundamental rights. In fact, the EU approach seeks to reconcile child protection with privacy by promoting **privacy-preserving age assurance** through digital **identity wallets** and **zero-knowledge proofs**, which allow users to demonstrate age attributes without revealing their identities. During recent parliamentary hearings in Portugal on a legislative initiative proposing a social media ban for minors, Meta suggested that app stores should perform age verification.

Although **financial scams** are not expressly listed among the systemic risks in the DSA, the Commission's enforcement practice reflects an extensive interpretation of Article 34, under which fraud-related harms fall within broader categories such as illegal content, consumer protection, and platform manipulation. This interpretive move recasts scams from isolated unlawful acts into systemic risks tied to platform design and scale.

Article 40 of the DSA is instrumental to integrate scientific research into the governance of socio-technical systems, particularly in areas such as AI, the digital economy, and human–machine interaction.

Researchers, however, face a complex vetting process and strict requirements under **Article 40(4),** including demonstrating the proportionality and necessity of the data request, for instance, by showing that the research cannot be conducted using publicly available data alone and that the applicants have sufficient technical capacity to ensure the security of the data. The entire process puts an overwhelming burden of proof on the side of the research teams and institutions. To date, no art 40(4) data access requests have been approved by the Digital Service Coordinators. This is compounded by resource and information asymmetries and the circular dependency problem: data access is often needed to secure funding, but funding is required to pursue access in the first place.[5]

Under Article 40(12), researchers also face **generalized resistance from platforms** through delays, restrictive alternatives to the requested data, subtle forms of obstruction through the strategic, intermittent provision of requested data, and even threats of lawsuits.

Digital Services Coordinators (DSCs) are central actors, particularly in overseeing and mediating data access requests**.**

The interpretation of **Article 40(12)** is polemical. Researchers may rely on publicly available data independently from platforms, including real-time data, but **technical methods such as scraping and crawling remain difficult in practice**. CrowdTangle has historically provided access to platform data, but it has been discontinued, without an equivalent replacement. Similarly, X has removed free-of-charge access to researchers, replacing it with expensive tiered alternatives with stronger access restrictions and technical blocks to scrapping attempts.

[5] Pierri, F., Araujo, T., Kruikemeier, S., Lorenz-Spreen, P., Vanden Abeele, M. M. P., Vandenbosch, L., Gonçalves-Sá, J., & Grabowicz, P. A. (2026). *Research opportunities and challenges of the EU's Digital Services Act.* Communications of the ACM. Advance online publication. https://doi.org/10.48550/arXiv.2512.14223

Platform responses to data access requests vary widely: some redirect researchers to DSCs, while others impose **highly complex and burdensome data-sharing agreements**, or simply reject requests as insufficiently justified ("you have not sufficiently demonstrated…").

Since data is accessed independently, a recurring question **under Article 40(12) is: who will check whether the researcher is compliant?** The first step should be to verify the platform's terms of service, Malheiro suggested.

User data donations are an alternative, but their future may be affected by the proposed Digital Omnibus Regulation, which could potentially restrict GDPR-based rights of access to personal data. This points to a deeper tension between data protection law and the DSA's ambition to enable independent scrutiny.

As a final reflection on whether science is opening the black box, I would say that science is making progress, but coordinated action and reinforced collaborative regulation may accelerate its role in promoting digital welfare.

In light of the prevailing constraints on data access, two complementary avenues may be contemplated to rebalance the data access ecosystem under Article 40. First, greater **collective organisation among researchers**—including forms of professional association or unionisation, as well as shared legal support and insurance mechanisms—could help mitigate existing asymmetries in resources and exposure to legal risks, particularly in the face of well-resourced and more aggressive platforms. Such structures could also support standardisation of practices and mutual assistance in navigating access procedures. Second, efforts should be directed toward identifying **areas of convergence between researchers and platforms**, especially where both share an interest in detecting and mitigating systemic risks. In this respect, it may be worth exploring whether the Commission could develop incentives—akin to a **"Good Samaritan" principle or leniency framework**—that would encourage platforms to cooperate in good faith with vetted researchers, for instance by recognising or mitigating liability risks when platforms proactively facilitate independent scrutiny aimed at improving compliance and safety.