

Overview of Platform Data Access Mechanisms (2025)

17 December 2025

Introduction

On 5 December, we published our latest [Data Access Resource](#) update, in which we assess the data access programmes of selected VLOPs and non-VLOPs across four dimensions: *eligibility*, *access modalities*, *vetting processes*, and *data availability*. This analysis supports researchers in identifying where platforms are making progress, and where significant barriers continue to impede effective compliance with the DSA's data access obligations.

Our review takes a comprehensive approach: it examines all initiatives that allow researchers to retrieve data from **Meta (Facebook, Instagram and Threads), TikTok, YouTube, X, LinkedIn, Telegram, and Bluesky** -not only the tools explicitly developed under Article 40(12). By looking beyond formal compliance mechanisms, the analysis highlights the current options available to researchers and exposes where meaningful access remains limited or inconsistent.

This brief highlights the most relevant findings for VLOPs, which may be relevant to the Commission's ongoing enforcement efforts. It also includes actionable recommendations on issues that could enhance both the utility of data access tools for research purposes and their alignment with Article 40(12) requirements.

Eligibility



Key findings

- **TikTok restricts civil society data access in ways that may be inconsistent with Article 40(12).** TikTok provides different data access tools to academic researchers and civil society organizations. Academic researchers can access an API, while civil society organizations are restricted to a Virtual Compute Environment (VCE) -a clean room that significantly constrains the scope and quality of research. As we detailed in our June [opinion piece](#), the VCE's restrictions are preventing CSOs from conducting meaningful analysis.
Article 40(12) makes no distinction between types of researchers. All vetted researchers meeting the conditions in Article 40(8)(b–e) should have access to the same publicly available data under the same conditions. TikTok is the only platform in our analysis that implements such differentiated access based on organizational affiliation.
- **Most VLOP data access programs may effectively exclude journalists as eligible researchers.** While application forms do not explicitly exclude journalists, their design creates significant barriers. Programs emphasize project-based research with defined timelines and funding sources, and request detailed methodological frameworks that may not align with investigative journalism practices, which typically involve responsive reporting.
By contrast, more accessible options such as YouTube's Data API v3 and the APIs offered by Telegram and Bluesky, (which require only user account registration) provide lower barriers to

entry that better accommodate journalism workflows and other non-academic research approaches, not to mention that it also allows non-EU researchers to access data. This more open model appears more aligned with the spirit of Article 40(12).



Recommendations

- **Establish clear guidance that Article 40(12) requires equal treatment of all vetted researchers, regardless of whether they are affiliated with academic or civil society organizations.** Platforms must provide equivalent access modalities and tools to all researchers meeting Article 40(8)(b-e) criteria.
- **Require TikTok to eliminate its two-tier access system.** In ongoing enforcement proceedings, the Commission should require TikTok to provide API access to civil society organizations on the same terms currently offered to academic researchers.
- **Explicitly recognise journalists conducting public interest investigations into online systemic risks as eligible researchers under Article 40(12).** Platforms should avoid imposing requirements incompatible with journalism practices, including predetermined project timelines, or rigidly defined research questions that preclude iterative investigation.

Access modalities



Key findings

- **Platforms employ varied access modalities with significantly different levels of usability.** These include APIs, user interfaces, secure research environments, and permitted scraping. The modalities differ substantially in data accessibility and technical complexity.
- **Meta VLOPs (Facebook and Instagram) and TikTok have the most restrictive access modalities among platforms analysed.** TikTok restricts civil society researchers to the VCE, a data clean room that audits every script by researchers, prevents data downloads and limits processing to aggregated outputs only. Meta's Content Library API similarly restricts data exports to aggregated results and requires VPN connections, raising concerns about platform control over research processes.
- **These restrictions appear disproportionate for publicly available data.** The Delegated Act on Data Access (Recital 9) requires that "access modalities should be proportionate to the sensitivity of the data." The [EDMO report](#) identifies clean rooms as appropriate only for high-risk processing data, for example non-public data under Article 40(4). Even then, clean rooms should permit data import, analysis, export, peer review, and replication (pages 90-91). For medium-risk processing, which would better characterise publicly available data under Article 40(12), the EDMO report recommends API access allowing researchers to export and store data in their own secure systems (page 88). This approach is particularly justified given that Article 40(12) already requires researchers to demonstrate adequate data protection measures to be eligible for access.

- **Other VLOPs demonstrate that less restrictive access to public data is viable.** YouTube permits data scraping under the [Google Researcher Programme](#), showing that researcher accountability mechanisms can effectively substitute for restrictive technical environments when handling publicly available data.
- **LinkedIn provides minimal information about its data access modalities.** Information about LinkedIn's Beta API remains extremely limited. The [application form](#) briefly references the modality of access: *"Selected researchers will be asked to enter into terms governing their access to the public data (and/or use of APIs, as applicable) for the purpose of the approved research (...)."* However, this text does not specify which LinkedIn API product researchers would access. The only LinkedIn API product that is public and is related to regulatory compliance, rather than commercial purposes, is the [Pages Data Portability API](#). This tool was developed to meet data portability obligations under the DMA, meaning it provides users access only to their own data, not the publicly available data required under Article 40(12).



Recommendations

- **Review the proportionality of Meta's and TikTok's data access modalities.** As part of ongoing enforcement proceedings, the Commission may wish to examine whether certain technical restrictions in Meta's Content Library API and TikTok's VCE clean rooms -including prohibitions on data export, mandatory script auditing, limitations to aggregated results, and constraints on research replication- are proportionate to the sensitivity of publicly available data under Article 40(12), particularly given that eligible researchers have already demonstrated adequate data protection capabilities.
- **Highlight best practices in data access implementation.** The Commission should encourage VLOPs and VLOSEs to consider more open approaches to data access, such as the Google Research Program model, which for YouTube permits limited scraping for eligible researchers. Such models are consistent with the Commission's established position in the [AliExpress commitments](#) and the recent [X non-compliance decision](#), that researchers should have access to public data via automated means, including scraping.
- **Require LinkedIn to clarify and document its data access mechanism under Art. 40 (12).** The Commission should request that LinkedIn provide additional clarity in its application materials regarding which API product researchers will access upon approval, along with documentation describing available data endpoints and data points.

Vetting Processes



Key findings

- **Inconsistent interpretation of "without undue delay" requirement.** Platforms vary significantly in their timelines for reviewing data access applications, ranging from 15–30 business days, to 4-

6 weeks. Transparency is also inconsistent: some platforms, like [YouTube](#)¹ and [LinkedIn](#)², only disclose timeline information in the audit reports under Article 37 DSA, rather than in public-facing materials. Meanwhile, X provides no timeline details in either source. Based on DRI's experience, X's review process can extend beyond eight months.

- **Lack of independent oversight in data access vetting processes.** Platforms, with the exception of Meta (which has worked with SOMAR), rely on internal teams to review data access applications. Researchers have no independent mechanism to challenge denied applications or discuss technical aspects of data access modalities with an independent body. As a result, platform decisions are final, unless contested through judicial proceedings.

While the 2022 Code of Practice on Disinformation envisioned, in Commitment 27, an independent third-party body to support researcher vetting and enhance transparency, [all VLOP and VLOSE signatories withdrew](#) from this commitment during the Code's transition to a Code of Conduct in January 2025.

- **Problematic Provisions in Terms of Use and Data Agreements.** VLOPs and VLOSEs condition access to data for eligible researchers on their acceptance of Terms of Use and Data Agreements. Our analysis highlights provisions that risk exposing researchers to legal liabilities, potentially hindering their willingness to apply to data access. Key issues include:

- *Data retention, refresh intervals, or mandatory deletion of accessed datasets.* Meta³, TikTok⁴, LinkedIn⁵, and X⁶ impose data retention and deletion policies, including requiring researchers to delete data upon platform request. X demands deletion within 24 hours of a written request, while LinkedIn requires deletion within 10 days. TikTok also mandates data refreshment every 30 days. TikTok, LinkedIn, and X also require researchers to provide written certification confirming data deletion.

These requirements create administrative burdens for researchers and restrict their ability to retain data for replication or follow-up analyses. The Delegated Act on Data Access (Article 15(3)) highlights this concern and specifies that *“data providers shall not impose on vetted researchers’ data management requirements such as archiving, storage, refresh and deletion requirements, or limitations to the use of standard analytical tools, that may hinder the performance of the relevant research”*. If this applies for non-public data, it should also apply to data accessed under Art. 40 (12).

- *Pre-publication review requirements.* Meta, TikTok, and LinkedIn impose pre-publication review requirements, with differences in scope and impact on research independence.

¹ According to [YouTube’s 2025 Audit Report](#): *“the audited service usually provides access within 15-30 days unless special circumstances require an in-depth analysis”*.

² According to [LinkedIn’s 2025 Audit Report](#): *“Undue delay: significant unreasonable delay under the circumstances as determined by management, including provision of data within 45 business days of execution of all associated agreements with the researcher”*.

³ According to [Product Terms for Meta Research Tools, Article 6 \(d\)](#) (Updated version from 8 December 2025).

⁴ According to [TikTok Research Tools Terms of Service](#) (updated version from 4 November 2025), III Rules and Restrictions, (e).

⁵ According to [Additional Terms for the LinkedIn Research Tools Program](#) (updated version from 21 August 2023), Article 4.2.

⁶ According to [X Developer Agreement](#) (updated version from 17 October 2025), IV. Updates and Removals, B.

Meta⁷ requires researchers to notify them upon publication and submit drafts for review, whereas TikTok⁸ mandates a seven-day pre-publication review to assess potential privacy risks. LinkedIn⁹ requests that researchers provide a courtesy copy of their work at least five business days before sharing it with third parties. Although these requirements aim to protect privacy and ensure proper attribution, they restrict researchers' independence.

- Scraping prohibitions. LinkedIn explicitly prohibits scraping or any automated means of data collection under all circumstances in its Terms of Use. Meta, TikTok, YouTube, and X only allow scraping when explicitly approved by the platform. These restrictive policies contradict the European Commission's established position in the AliExpress commitments and the recent X non-compliance decision, both of which affirm that researchers should be able to access publicly available data through automated methods, including scraping.
- **No reporting on data access applications.** VLOPs in our study no longer report how many data access applications they receive, approve, or reject. Previously, such information was provided in transparency reports under the Code of Practice on Disinformation, but platforms have since opted out of these reporting obligations.



Recommendations

- **Clear guidance on what constitutes "without undue delay"** under Article 40(12), including maximum review timeframes.
- **Require all platforms to publish expected processing times** in their public data access materials.
- **Support the development of an independent oversight mechanism.** The Commission could encourage the reintroduction of an independent third-party body for researcher vetting, whether through industry commitments, regulatory guidance, or other mechanisms. Such body could provide researchers with appeal channels, reduce inconsistencies in decision-making across platforms, and potentially lower litigation risks.
- **Require transparency reporting on data access applications.** The Commission should require platforms to publish regular reports on data access applications, including the number of applications received, approval rates, rejection reasons, and average processing times. This could be asked in their Risk Assessment reports, for example.
- **Require standardised, DAS-compliant Terms of Use.** The Commission could develop model terms of use for data access programs that balance legitimate platform interests with research needs, addressing, among others, (i) reasonable data retention periods aligned with research timelines and replication needs, (ii) clear permissions for automated data collection methods, (iii) liability

⁷ Product Terms for Meta Research Tools, Article 5 (c).

⁸ TikTok Research Tools Terms of Service, III Rules and Restrictions, (f).

⁹ Additional Terms for the LinkedIn Research Tools Program, Article 5.1.

protections for good-faith research activities. Platforms should be encouraged to adopt these model terms or justify deviations.

Data Availability



Key findings

- **Significant gaps remain in the data points and data endpoints that VLOPs make available,** limiting researchers' ability to study key aspects of online discourse:
 - Content format gaps: Popular formats including Shorts, Reels, and Stories remain inaccessible to researchers across platforms, despite representing substantial portions of user engagement.
 - AI-generated content labels: Among reviewed VLOPs, only YouTube provides data on AI-generated content labels, despite all major platforms hosting such content and claiming to label it.
 - Sponsored content: Only Facebook, Instagram and YouTube provide data on sponsored content labels.
 - Account verification status: TikTok and YouTube do not provide data on verified account status.
- **We could not find detailed codebooks or API documentation for LinkedIn's data access tools,** preventing researchers from assessing what data would be available before applying for access.
- **We identified available data by reviewing platform API codebooks (technical documentation describing data offerings).** This approach reflects what platforms claim to provide but cannot confirm the actual quality or consistency of data delivered to researchers.



Recommendations

- **Clarify the scope of publicly available data.** Issue guidance establishing clear principles for what constitutes "publicly available data" under Article 40(12).
- **Address data gaps in enforcement proceedings.** The Commission should require platforms to provide meaningful access to all publicly available data. Priority gaps include high-engagement content formats (Shorts, Reels, Stories), integrity labels (AI-generated content, sponsored content, verified accounts), and verification tools (post/comment URLs, download capabilities).
- **Mandate publication of API documentation.** Require all VLOPs to publish detailed API codebooks that clearly describe available data endpoints, data points, and data formats. LinkedIn's failure to provide such documentation should be addressed as an immediate priority.
- **Establish data quality verification mechanisms** to verify that data delivered to researchers matches platform documentation in completeness, accuracy, and consistency. This could be achieved through enhanced Article 37 audit requirements, mandatory platform reporting on data quality metrics, researcher feedback channels, or independent technical reviews.