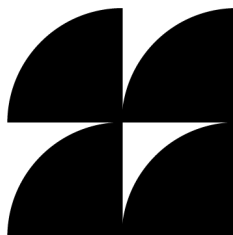


17 JANUARY 2025

Ofcom Call for Evidence: Researchers' access to information from regulated online services

Knight-Georgetown Institute (KGI) Response





About the Knight-Georgetown Institute

The Knight-Georgetown Institute (KGI) is dedicated to connecting independent research with technology policy and design. KGI serves as a central hub for the growing network of scholarship that seeks to shape how technology is used to produce, disseminate, and access information. KGI is designed to provide practical resources that policymakers, journalists, and private and public sector leaders can use to tackle information and technology issues in real time. Georgetown University and the Knight Foundation came together to launch the institute in 2024. Learn more about KGI at <https://kgi.georgetown.edu>.

Introduction

The Knight-Georgetown Institute (KGI) welcomes the opportunity to provide input to Ofcom’s call for evidence on researchers’ access to information from regulated online services. Ensuring that a broad range of researchers, journalists, and civil society organizations have sufficient access to digital platform data is essential for the development of effective, evidence-based digital platform policies and oversight.¹

Ofcom’s call for evidence comes at a critical time. Ofcom and the United Kingdom (UK) government have the opportunity to play a catalytic and essential role in expanding access to platform data.

While the European Union (EU) has recently taken steps to guarantee independent access to platform data, more work remains. Article 40 of the EU’s Digital Services Act (DSA) requires large platforms and search engines to provide researchers access to platform data in two key ways: the ability for researchers, journalists, and civil society to analyze publicly accessible platform data as well as vetted researcher access to private platform data. However, as these requirements have emerged we have also seen several platforms restrict access to data that was previously available.

In our response below, we consider the three sets of questions posted by Ofcom, focusing primarily on Question 3 related to steps to expand access to data. Beyond our specific answers to questions posed by Ofcom, our comments emphasize three key points:

1. Given the importance of platform data in the information ecosystem, Ofcom and the broader UK government have an important opportunity to establish and operationalize requirements for access to platform data, inclusive of researchers, journalists, and civil society organizations.
2. When considering how to implement data access requirements, two different categories of data are relevant: (i) internal or private platform data that is made available to researchers through “vetted” researcher access, and (ii) public data that is publicly accessible on the platform.
3. In regards to publicly accessible data, KGI is working with a diverse group of experts to develop a Gold Standard for Publicly Available Platform Data. This Standard will articulate the types and formats of publicly accessible data that should be available to researchers, journalists, and civil society organizations engaging in independent research.

¹ See, e.g., Open Data Institute. (2023) ‘Global Data Infrastructure.’ Available at: <https://theodi.org/insights/projects/global-data-infrastructure/> (accessed January 5, 2025); Coalition for Independent Technology Research. (2024). ‘Independent Tech Researchers’ Summit.’ Available at: <https://independenttechresearch.org/independent-tech-researchers-summit/> (accessed January 5, 2025); Alex Engler. (2021). ‘Platform data access is a lynchpin of the EU’s Digital Services Act.’ Brookings. Available at: <https://www.brookings.edu/articles/platform-data-access-is-a-lynchpin-of-the-eus-digital-services-act/> (accessed January 5, 2025); Nathaniel Persily and Joshua Tucker. (2021). ‘How to Fix social media? Start with independent research.’ Brookings. Available at: <https://www.brookings.edu/articles/how-to-fix-social-media-start-with-independent-research/> (accessed January 5, 2025).

Question 1: How, and to what extent, are independent online safety researchers able to obtain platform information?

Independent research is central to our collective understanding of online information ecosystems. However, independent research is increasingly restricted by platforms, and researchers, journalists, and civil society organizations today are able to obtain less data and information than they were just a few years ago.

From democracy and conflict to public health and education, research and journalism has been foundational for strengthening understanding of the positive and negative impacts of digital platforms.²

Platform research has typically relied on publicly accessible data (e.g. posts, comments, or impressions) or private data obtained through collaborative research partnerships with social media companies (e.g. confidential information, experiments, etc). Many large platforms have developed Application Programming Interface (API) access for parties external to the company to access platform data. This has included both commercial and public interest research API use cases.

Researchers, journalists, and civil society organizations have struggled to operationalize collaborative research with digital platforms. The Coalition for Independent Technology Research (CITR), a coalition of academics, journalists, and civil society representatives, has chronicled a range of challenges associated with accessing data from technology platforms.³ Researchers suggest that in collaborative research partnerships, platforms tend to exert significant control in “shaping the research questions, methodology, workflow, and design choices” of research.⁴ Communication around such research can be driven by the platform, and there have been disagreements over how to interpret research findings.⁵ Others have observed that such collaborations are unsustainable, calling them “independence by permission.”⁶ Leaks from platform employees have also led companies to restrict research collaboration.⁷

² See, e.g., Alex Abdo et al. (2022). ‘A Safe Harbor for Platform Research.’ Knight First Amendment Institute at Columbia University. Available at: <https://knightcolumbia.org/content/a-safe-harbor-for-platform-research> (Accessed January 5, 2025).

³ CITR. (2022). ‘Manifesto: The Coalition for Independent Technology Research.’ Available at: <https://independenttechresearch.org/manifesto-the-coalition-for-independent-technology-research/> (accessed January 5, 2025).

⁴ Prithvi Iyer. (2023-2024). ‘A Primer on the Meta 2020 US Election Research Studies.’ *Tech Policy Press*. Available at: <https://www.techpolicy.press/a-primer-on-the-meta-2020-us-election-research-studies/> (accessed January 5, 2025).

⁵ Justin Hendrix. (2023). ‘The Meta Studies: Nuanced Findings, Corporate Spin, and Media Oversimplification.’ *Tech Policy Press*. Available at: <https://www.techpolicy.press/the-meta-studies-nuanced-findings-corporate-spin-and-media-oversimplification/> (accessed January 5, 2025).

⁶ Michael W Wagner. (2023). ‘Independence by permission.’ *Science* 381, no. 6656: 388-391. Available at: <https://www.science.org/doi/abs/10.1126/science.adi2430> (accessed January 15, 2025).

⁷ The Facebook Files: A Wall Street Journal Investigation. Available at: <https://www.wsj.com/articles/the-facebook-files-11631713039> (accessed January 5, 2025).

Unfortunately, independent research that relies on company APIs is becoming more difficult.⁸ Some platforms have recently limited access to data that was previously publicly available. Recent platform data access restrictions include:

- In August 2024, Meta shutdown access to CrowdTangle, a public insights tool that enabled researchers to analyze public content on social media, which will frustrate research and hinder understanding of the online information environment.⁹
- In early 2023, Twitter/X restricted API access, which has undermined independent efforts to understand platform dynamics as they relate to a wide range of topics.¹⁰
- In 2023, Reddit updated access to its Data API, including new rate limits,¹¹ which researchers say negatively impacts their ability to conduct research on a range of topics, including computing, medicine, and the social sciences.¹² Following the API changes, Reddit is testing a new approach to research access,¹³ but there are concerns that this approach will limit public access.
- In early 2023, TikTok opened up API access to researchers for the first time. However, access to the API includes requirements that independent researchers describe as “strict,”¹⁴ and TikTok’s API has been found to be error-prone¹⁵ and unreliable.¹⁶

Researchers, journalists, and civil society organizations also independently collect and analyze publicly accessible platform data (sometimes called scraping). This approach is increasingly fraught, with legal, ethical, institutional, and scientific risks.¹⁷ Digital platforms have terminated account access for

⁸ Media and Democracy Data Cooperative. (2024). ‘The State of Digital Media Data Research 2024.’ Available at: <https://mddatacoop.org/dmd/> (accessed January 5, 2025).

⁹ Center for American Progress, Institute for Strategic Dialogue, et al. (2024). ‘CrowdTangle Letter.’ Available at: <https://www.americanprogress.org/wp-content/uploads/sites/2/2024/05/CAP-ISD-letter2.pdf> (accessed January 5, 2025) and Mozilla. (2024). ‘Open Letter to Meta.’ Available at: <https://foundation.mozilla.org/en/campaigns/open-letter-to-meta-support-crowdtangle-through-2024-and-maintain-crowdtangle-approach/> (accessed January 5, 2025).

¹⁰ CITR. (2023). ‘Imposing Fees to Access the Twitter API Threatens Public-Interest Research.’ Available at: <https://independenttechresearch.org/letter-twitter-api-access-threatens-public-interest-research/> (accessed January 5, 2025).

¹¹ See Reddit. (2023) ‘Creating a Healthy Ecosystem for Reddit Data and Reddit Data API Access Updating Our Terms and Building Native Moderator Tools.’ Available at: <https://www.redditinc.com/blog/2023apiupdates> (accessed January 5, 2025).

¹² See Brandi Geurkink and Sarah Gilbert. (2024) Why Reddit’s decision to cut off researchers is bad for its business—and humanity.’ Fast Company. Available at: <https://www.fastcompany.com/91014116/reddit-researchers-bad-for-business> (accessed January 5, 2025); Norman Adams, N. (2022). ‘Scraping’ Reddit posts for academic research? Addressing some blurred lines of consent in growing internet-based research trend during the time of Covid-19.’ International Journal of Social Research Methodology, 27(1), 47–62. Available at <https://www.tandfonline.com/doi/full/10.1080/13645579.2022.2111816#abstract> (accessed January 5, 2025).

¹³ Reddit. ‘Apply to join the Reddit for Researchers Beta [by August 23].’ r/reddit4researchers. Available at https://www.reddit.com/r/reddit4researchers/comments/1egr9wu/apply_to_join_the_reddit_for_researchers_beta_by/ (accessed January 5, 2025).

¹⁴ Aicha Counts. (2023). ‘TikTok’s Rules Deter Researchers From Crunching Data on Users, Misinformation.’ Bloomberg. Available at <https://www.bloomberg.com/news/articles/2023-09-21/tiktok-terms-of-service-strict-for-researchers> (accessed January 5, 2025).

¹⁵ Megan A. Brown. (2023). ‘The Problem with TikTok’s New Researcher API is Not TikTok.’ *Tech Policy Press*. Available at <https://www.techpolicy.press/the-problem-with-tiktoks-new-researcher-api-is-not-tiktok/> (accessed January 5, 2025).

¹⁶ Philipp Darius. (2024). ‘Researcher Data Access Under the DSA: Lessons from TikTok’s API Issues During the 2024 European Elections.’ *Tech Policy Press*. Available at <https://www.techpolicy.press/-researcher-data-access-under-the-dsa-lessons-from-tiktoks-api-issues-during-the-2024-european-elections/> (accessed January 3, 2025).

¹⁷ Megan Brown et al. (2025). ‘Web Scraping for Research: Legal, Ethical, Institutional, and Scientific Considerations’ Available at: <https://arxiv.org/abs/2410.23432> (accessed January 5, 2024).

individuals pursuing independent research on the platform.¹⁸ Other platforms have sued nonprofit organizations for scraping platform data to conduct research.¹⁹ While lawsuits have largely not been successful on their merits, legal claims and threats of legal claims have significant impacts on researchers and institutions, often raising questions whether independent research is worth the risk and potential cost.

Driven in part by the challenges associated with API access and scraping restrictions, researchers have also employed data donations. Data donations enable users to share their own data with third party researchers for analysis.²⁰ But data donation represents incomplete data and has multiple challenges. It is costly, it relies on individual users choosing to donate, it is dependent upon what platforms allow access to, and it tends to result in biased samples of data, given differential likelihood of opting into a data donation process.

The DSA is the most important regulatory regime that currently mandates independent access to platform data. The DSA Article 40 introduces requirements for researcher access to two types of data: private platform data and publicly accessible data. Research with these two different types of data are subject to different requirements. Research on publicly accessible data must comply with a more limited set of requirements²¹ whereas research with private or non-public data must be pursued through a more rigorous vetted research mechanism that is being developed by EU regulators.²² The EU has recently concluded a consultation on a Delegated Act that will establish the mechanism through which researchers can seek and maintain access to private platform information.

Research on publicly accessible data, as provided for in DSA Article 40.12, should already be available under the DSA, with requirements entering into force in early 2024. Unfortunately, current mechanisms created by platforms and search engines to enable publicly accessible data access are uneven and

¹⁸ See HateAid. (2023). 'For independent research: Landmark case against X.' Available at: <https://hateaid.org/en/for-independent-research-landmark-case-against-x/#case-travis-brown> (accessed January 9, 2025) and Rebecca Klar. (2021). 'Facebook suspends accounts of NYU researchers who've criticized platform.' Available at: <https://thehill.com/policy/technology/566332-facebook-suspends-accounts-of-nyu-researchers-whove-criticized-platform/> (accessed January 9, 2025).

¹⁹ See X Corp v. CCDH, California Northern District Court, July 31, 2023. Available at: https://www.pacermonitor.com/view/NV3GMHQ/X_Corp_a_Nevada_Corporation_v_Center_for_Counteracting_Digital_candce-23-03836__0001_0.pdf?mcid=tGE3TEOA (accessed January 3, 2025).

²⁰ See, e.g., Zeve Sanderson and Lama Mohammed. (2023). 'A Multi-Stakeholder Approach for Leveraging Data Portability to Support Research on the Digital Information Environment.' Available at: <https://tsjournal.org/index.php/jots/article/view/215> (accessed January 5, 2025); Mozilla Foundation. (Undated). 'About YouTube Regrets.' Available at: <https://foundation.mozilla.org/en/youtube/regretsreporter/about/> (accessed January 9, 2025); Josephine Lukito et al. (2023). 'Enabling Independent Research Without Unleashing Ethics Disaster.' *Tech Policy Press*. Available at: <https://www.techpolicy.press/enabling-independent-research-without-unleashing-ethics-disasters/> (accessed January 9, 2025).

²¹ For publicly accessible data under DSA Article 40.12, researchers need to demonstrate:

- independence from commercial interest (40.5(b)),
- disclosure research funding (40.5(c)),
- capability of fulfilling data security and confidentiality requirements (40.5(c)),
- how the proposed research is necessary and proportionate (40.5(d)), and
- how the research contributes to understanding systemic risks or mitigation measures (40.5(e), 40.4).

²² European Commission. (Undated). Digital Services Act (DSA) overview. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en (accessed January 5, 2025).

largely insufficient.²³ The European Commission, which is responsible for investigations under the DSA, has opened proceedings against multiple platforms for alleged shortcomings in facilitating researcher access to publicly accessible data under 40.12. This includes investigations focused on X/Twitter,²⁴ TikTok,²⁵ AliExpress,²⁶ and Facebook/Instagram.²⁷

Decentralized platforms are a new focus for research. Research practices on these platforms are mixed.²⁸ Bluesky, for example, has enabled broad API access, including for researchers, journalists, and civil society organizations.²⁹ While promising, more work remains to understand opportunities for independent research amongst and across decentralized platforms.

²³ Mozilla, Algorithmic Transparency Institute and National Conference on Citizenship. (2024). 'Public Data Access Programs: A First Look - Assessing Researcher Data Access Programs Under the Digital Services Act.' Available at https://assets.mofoprod.net/network/documents/Public_Data_Access_Programs_A_First_Look_Final_cExWfcH.pdf (accessed January 5, 2025) and Proof. (2024). 'Meta Is Getting Rid of CrowdTangle – and Its Replacement Isn't As Transparency or Accessible.' Available at: <https://www.proofnews.org/meta-is-getting-rid-of-crowdtangle-and-its-replacement-isnt-as-transparent-or-accessible/> (accessed January 5, 2025).

²⁴ European Commission. (2024). 'Commission sends preliminary findings to X for breach of the Digital Services Act.' Available at https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761 (accessed January 5, 2025).

²⁵ European Commission. (2024). 'Commission opens formal proceedings against TikTok under the Digital Services Act.' Available at https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926 (accessed January 5, 2025).

²⁶ European Commission. (2024). 'Commission opens formal proceedings against AliExpress under the Digital Services Act.' Available at https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1485 (accessed January 5, 2025).

²⁷ European Commission. (2024). 'Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act.' Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2373 (accessed January 5, 2025).

²⁸ Jakob Guhl, Oliver Marsh, and Henry Tuck. (2022). 'Researching the Evolving Online Ecosystem: Barriers, Methods and Future Challenges.' Institute for Strategic Dialogue. Available at: <https://www.isdglobal.org/isd-publications/researching-the-evolving-online-ecosystem-barriers-methods-and-future-challenges/> (accessed January 15, 2025).

²⁹ See Bluesky Developer APIs, Available at: <https://docs.bsky.app/> (accessed January 5, 2025).

Question 2: What challenges currently constrain the sharing of information for the purpose of research into online safety related issues?

Independent research focused on the online information ecosystem faces numerous challenges and is significantly constrained. Research that relies on access to data from online platforms – particularly large platforms – face a range of technical, legal, and operational barriers.

There are three primary challenges that currently constrain independent online safety related research: (1) the lack of access to data from digital platforms, including both publicly available platform data as well as private platform data, (2) the lack of data systematization and standardization across platforms and research efforts, and (3) privacy and ethical challenges related to researching online safety.

1. Data access

As described in our response to Question 1, researchers face significant barriers to access platform data.³⁰ Companies restrict independent access to data in a multitude of ways that hinder meaningful research. First, avenues for platform data access are restricted for private as well as publicly available data. For research that relies on internal company data, systems, assessments, and practices, researchers necessarily must collaborate with the company in defining the research questions and conducting research. As mentioned above, companies tend to exert significant controls over this research done in collaboration with the company, including whether research findings can be made public or shared with regulators.³¹

When researchers seek to conduct research on publicly available data, they still run into challenges. They face legal risks, as the legal threats to researchers described in our answer to Question 1 demonstrate. There are also significant technical challenges. Companies may choose to amend API access, timeliness, and data formats at any time. This disrupts, and ultimately may undermine, any attempt at longitudinal research. For example, Meta's choice to replace CrowdTangle with a less comprehensive Meta Content Library upended ongoing research.³² Media reporting suggests this was

³⁰ Coalition for Independent Technology Research. (2022). 'Coalition for Independent Technology Research Founding Document.' Available at: <https://independenttechresearch.org/coalition-for-independent-technology-research-founding-document/> (accessed January 5, 2025).

³¹ See, e.g., Caroline Mimbs Nyce. (2024). 'A New Development in the Debate About Instagram and Teens Meta, infamous for kicking researchers off its platform, flirts with slightly more transparency.' The Atlantic. Available at <https://www.theatlantic.com/technology/archive/2024/07/instagram-teen-well-being-studies/679048/> (accessed January 5, 2025) and Justin Hendrix. (2023). 'Examining the Meta 2020 Election Research Partnership.' *Tech Policy Press*. Available at: <https://www.techpolicy.press/examining-the-meta-2020-us-election-research-partnership/> (accessed January 5, 2025).

³² See Sarah Grevy Gofredsen and Kaitlyn Dowling. (2024). 'Meta Is Getting Rid of CrowdTangle — and Its Replacement Isn't As Transparent or Accessible.' Available at: <https://www.proofnews.org/meta-is-getting-rid-of-crowdtangle-and-its-replacement-isnt-as-transparent-or-accessible/> (accessed January 5, 2025).

in part the intent, with key executives concerned that credible and independent research presented reputational risks to Meta.³³

2. Data standardization

The lack of data standardization, combined with the lack of data access points, undermines the ability of independent researchers to draw comprehensive conclusions about the online information environment. Looking at a single platform, rather than facilitating comparison across platforms, “fails to recognize the diverse, fragmented, and complex modern media environment, of which a single media platform is only a small part.”³⁴

Due to a lack of data standardization, most research that is conducted on online safety issues is concentrated on a single platform. Piecemeal access to data means that it is difficult to compare the same risks or harms across multiple platforms. For instance, researchers interested in understanding how much low quality information is circulating on platforms are subject to what they can measure on any given platform. So one study might consider the nature of the URLs circulating on Twitter/X,³⁵ and another might use different metrics to consider the sharing of low quality information on Facebook.³⁶ While each is useful in isolation, it would be more powerful to be able to see the relative prevalence of such content across different platforms. Knowing that a platform has a certain level of a particular type of content is difficult to make sense of without comparative context. Without standardization, such comparisons are challenging.

Lack of standardization also means that studying how content travels across platforms is a very difficult undertaking, meaning understanding the information ecosystem as a whole is basically unachievable. For example, platforms are increasingly disclosing information about advertising in compliance with Article 39 of the DSA.³⁷ But because the disclosure requirements do not include specific standard requirements, it still remains difficult to compare types of ads and targeting across platforms.³⁸

Standardizing data access would mean that the same types of data are available in the same formats, at the same level of aggregation, across many platforms. Such standardization would allow for more

³³ See Kevin Roose. (2021). ‘Inside Facebook’s Data Wars.’ N.Y. Times. Available at <https://www.nytimes.com/2021/07/14/technology/facebook-data.html> (accessed January 5, 2025).

³⁴ Leticia Bode and Emily K. Vraga. (2018). “Studying politics across media.” *Political Communication* 35, no. 1: 1-7. Available at: https://www.researchgate.net/publication/319412118_Studying_Politics_Across_Media (accessed January 16, 2025).

³⁵ Lisa Singh et al. (2020). “Understanding high-and low-quality URL Sharing on COVID-19 Twitter streams.” *Journal of computational social science* 3: 343-366.

³⁶ Andrew Guess et al. (2019). “Less than you think: Prevalence and predictors of fake news dissemination on Facebook.” *Science advances* 5, no. 1.

³⁷ European Commission. (2022). “The final text of the Digital Services Act (DSA): Article 39, Additional online advertising transparency - the Digital Services Act (DSA).” Available at https://www.eu-digital-services-act.com/Digital_Services_Act_Article_39.html (accessed January 10, 2025).

³⁸ Pooja Iyer. (2024). “The Impact of the Digital Services Act on the World of Advertising Technology.” *Tech Policy Press*. Available at <https://www.techpolicy.press/the-impact-of-the-digital-services-act-on-the-world-of-advertising-technology/> (accessed January 10, 2025).

thorough research, and would also allow individuals to make clearer decisions about which platforms they wish to use, depending on the relative benefits and harms documented for each one.

3. Privacy and Ethics

A final challenge when it comes to researching online safety relates to ethical research practices. Securing informed consent is a particular challenge when conducting research focused on the online information ecosystem, given that most internet users do not explicitly give consent for their data or online activities to be researched. Just how public or private users perceive their content, posts, behaviors, or other activities to be is often unknown,³⁹ and rarely measured directly.⁴⁰ This makes applying traditional research ethics expectations around informed consent to online safety research challenging.

A second dimension of ethical challenge comes from ensuring user privacy when working with online data. This entails a tradeoff between full transparency (for example, providing the full text of an online post) and privacy protection (for example, doing so would make it easy to find out who posted it). Online safety researchers must grapple with questions of when, whether, and how to anonymize data and maintain privacy protections.^{41, 42}

When research includes hostile communities, such as those involved in coordinated hate speech or terrorism, the safety of researchers also may become a risk.⁴³ Many researchers, civil society organizations, newsrooms, and universities are only beginning to grapple with these safety and targeting tactics.⁴⁴

³⁹ Lisa Sugiura, Rosemary Wiles, and Catherine Pope. (2017). 'Ethical challenges in online research: Public/private perceptions.' *Research Ethics* 13, no. 3-4: 184-199.

⁴⁰ Matthew L. Williams et al. (2017). 'Users' views of ethics in social media research: Informed consent, anonymity, and harm.' In *The ethics of online research*, pp. 27-52. Emerald Publishing Limited.

⁴¹ Lynne D. Roberts. (2015). 'Ethical issues in conducting qualitative research in online communities.' *Qualitative Research in Psychology* 12, no. 3: 314-325.

⁴² Dag Elgesem. (2002). 'What is special about the ethical issues in online research?.' *Ethics and information technology* 4, no. 3: 195-203.

⁴³ Maura Conway. 2021. 'Online extremism and terrorism research ethics: Researcher safety, informed consent, and the need for tailored guidelines.' *Terrorism and political violence* 33, no. 2: 367-380.

⁴⁴ Researcher Support Consortium. "Researchers Increasingly Face Campaigns of Intimidation and Harassment." Available at <https://researchersupport.org/> (accessed January 14, 2025).

Question 3. How might greater access to information for the purpose of research into online safety issues be achieved?

There are concrete ways to strengthen access to platform data. Ofcom and the UK government could enable greater access to two key categories of digital platform data: (1) publicly accessible platform data and (2) private platform data. This could include enabling research on large online platforms, as required in the EU, as well as higher risk, smaller platforms.⁴⁵

1. Publicly accessible platform data

Given ongoing and increasing restrictions, there is a need to ensure that independent researchers, civil society, and journalists are able to access and analyze publicly available platform data. Such data is vital for advancing informed public debate and effective external research of social media platforms. This requires researcher access to real-time or near real-time data of public content available on a wide range of platforms. This should include data related to user engagement (likes, comments, shares, etc), user reports, and other agreed categories.

The absence of core definitions and expectations related to publicly available platform data practically means that companies alone decide what research questions can be explored.⁴⁶ To clarify expectations related to publicly accessible platform data, KGI has convened an Expert Working Group to develop a [Gold Standard for Publicly Available Platform Data](#).

The Gold Standard will articulate a uniform, cross-industry framework for what kind of platform data should be made publicly available, under what circumstances, and in what formats. This would allow for greater understanding of the online information ecosystem as a whole, including the study of online safety issues in more holistic, multi-platform ways. Greater access to data would allow and encourage research into online safety issues, not just by vetted academic researchers but by independent researchers from academia and civil society, including think tanks and journalists. Indeed, the Gold Standard group is composed of experts from academia, journalism, and civil society, representing the variety of needs and use cases for platform data access.

We expect to release the Gold Standard in mid 2025 and would be eager to engage with Ofcom about it.

⁴⁵ Henry Tuck. (2023). 'Policy Approaches to Addressing Data Access Challenges in the Evolving Online Ecosystem.' Institute for Strategic Dialogue (ISD). Available at: <https://www.isdglobal.org/wp-content/uploads/2023/09/Policy-Approaches-to-Addressing-Data-Access-Challenges-in-the-Evolving-Online-Ecosystem.pdf> (accessed January 15, 2025).

⁴⁶ Mozilla. (2024). 'Open Letter to Meta.' Available at: <https://foundation.mozilla.org/en/campaigns/open-letter-to-meta-support-crowdtangle-through-2024-and-maintain-crowdtangle-approach/> (accessed January 5, 2025).

2. Private platform data

In addition to the need to clarify expectations for independent access to publicly accessible platform data, research that relies on internal platform data is also critical. Ofcom and the UK Government can learn from the foundational work done to develop and operationalize vetted researcher access under the DSA. Ofcom should engage with key academic, civil society, and journalism stakeholders to facilitate meaningful access to platform data that is efficient, effective, and privacy protecting.⁴⁷

There are specific categories of relevant private platform data for independent research related to online safety. Key categories of data include:

- Real-time or near real-time data of public content available on the platform, its distribution and reach. This should include metrics like the definition and prevalence of harmful exposures, the reach of illegal and harmful content over specific time periods, the distribution of frequency of exposures for users, among other categories.⁴⁸
- Insights into a company's data structures⁴⁹ as well as the platform's product strategy and inventory of product experimentation. This could include data detailing product experimentation oversight structures as well as relevant team objectives and key results (OKRs) related to product features that likely impact online safety risks.
- Data related to recommendations and recommender system design, including inputs, outputs, weights, and metrics used to determine which content, accounts, ads, and other items get surfaced to users.
- Data related to the testing of new features prior to their deployment, including specific metrics and results from tests platforms use to assess new features prior to deployment. This could also include data related to holdout experiments to understand broader impacts of specific product changes. This could include data for features that were never fully deployed, where available.
- Platforms regularly conduct internal research to understand user experiences and outcomes. This often includes, for example, surveys like the Meta Bad Experiences and Encounters Framework (BEEF) Survey,⁵⁰ other experience surveys, focus groups, and qualitative interviews. The results of internal platform user experience research should be accessible to researchers

⁴⁷ For example, the Coalition for Independent Technology Research, the European Digital Media Observatory, and others. See the Commission's 2023 consultation on the Delegated Regulation on data access provided for in the Digital Services Act. Available at:

https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/feedback_en?p_id=32045757 (accessed January 5, 2025) and Status Report on Mechanisms for Research Access to Online Platform Data. (2024). Available at:

<https://digital-strategy.ec.europa.eu/en/library/status-report-mechanisms-researcher-access-online-platform-data> (accessed January 5, 2025).

⁴⁸ See Integrity Institute. (2022). Metrics & Transparency: Data and Datasets to Track Harms, Design, and Process on Social Media Platforms.' Available at:

<https://static1.squarespace.com/static/614cbb3258c5c87026497577/t/617834d31bcf2c5ac4c07494/1635267795944/Metrics+and+Transparency+-+Summary+%28EXTERNAL%29.pdf> (accessed January 9, 2025).

⁴⁹ Called a 'Data Inventory' under the draft EU Article 40 Delegated Act.

⁵⁰ See Instagram. (2021) Bad Experiences and Encounters Framework (BEEF) Survey. Available at:

<https://storage.courtlistener.com/recap/gov.uscourts.nmd.496039/gov.uscourts.nmd.496039.36.2.pdf> (accessed January 5, 2025).

studying online safety. Sharing such aggregated survey findings could help to mitigate user privacy risks.

Barriers to consistent, independent, and credible research on online safety fundamentally frustrates Ofcom's efforts to advance requirements of the Online Safety Act in the UK. We commend Ofcom for considering how to expand access to data and information to enable independent research into online safety issues by a range of researchers, journalists, and civil society organizations.

Conclusion

We appreciate the opportunity to contribute to this consultation. Please contact us should you have any questions.

Contact:

Leticia Bode
KGI Research Director

Peter Chapman
KGI Associate Director

knightgeorgetown@georgetown.edu