

Social Platforms Data Access Taskforce

Final Report, September 2026

A year of evidence: insights and recommendations on social media data access for research into online harms.

Dr Alexandra Boutopoulou, Professor Katharine Dommert,
Professor Amy Orben, Mark Scott, Dr David Zendle



Social Platforms
Data Access
Taskforce



About this report

The Social Platforms Data Access Taskforce was established by Smart Data Research UK (SDR UK) to examine how researcher access to social media data could be made more responsible, ethical and secure. Bringing together independent experts from academia, regulation and civil society, the Taskforce has spent the past year gathering evidence on the barriers researchers face and the options available to address them.

This is the Taskforce's final report, following its earlier interim publication, *Social Media Data Access to Reduce Online Harms: A Call for Action* (Orben et al., 2026). It draws on extensive engagement with the

research community, analysis of the international data access landscape and close engagement with the evolving UK legislative and regulatory framework.

The views, conclusions and recommendations expressed in this report are those of the Social Platforms Data Access Taskforce and do not represent the positions of Smart Data Research UK (SDR UK) or UK Research and Innovation (UKRI).

This report represents the view of Taskforce members only and not those of any external advisors or third parties.

Acknowledgements

The Taskforce is grateful to the following individuals for their comments, feedback and expert input during the preparation of this report:

Leticia Bode, Andreu Casas, Peter Chapman, Ioanna Fokas, Brandi Geurkink, Veronica Guzman Quilaqueo, Katherine Holden, Vineet John Samuel, Alex Lawrence-Archer, LK Seiling, Tania Smith, Henry Tuck, Maeve Walsh, Megan Wood, members of the Taskforce's Policy Advisory Group and attendees at a British Academy Event on Social Media Data Access.

Table of contents

Executive Summary

About the Taskforce

Chapter 1. Introduction and context

- 1.1 Purpose and scope
- 1.2 Why the UK Government needs to take action now
- 1.3 The current state of data access for online research
- 1.4 Structure of the report

Chapter 2. Toward better data access: the UK opportunity

- 2.1 Introduction
- 2.2 Policy making action on social media data access in the UK
- 2.3 What enhanced Data Access could achieve

Chapter 3. Pathways to social media data access in the UK

- 3.1 Introduction
- 3.2 Web-Scraping
 - 3.2.1 What is public interest web-scraping?
 - 3.2.2 Why is web-scraping needed?
 - 3.2.3 Current legal foundations, implications and recommendations

3.3 Data Donation

3.3.1 What is data donation?

3.3.2 Why is data donation needed?

3.3.3 Technical Approaches and Infrastructure

3.3.4 Current legal foundations, implications and recommendations

3.4 Data Intermediaries: a new hybrid-repository social media data intermediary

3.4.1 What is a data intermediary?

3.4.2 Why is hybrid-repository social media data intermediary needed in the UK?

3.4.3 Existing intermediary models in the UK

3.4.4 Types of social media data access intermediaries possible in the UK

3.4.5 International examples of intermediary models

3.4.6 Lessons from international intermediaries

3.5 Summarising policy options for web-scraping, data donation and an intermediary body

Chapter 4: Research Community Actions for Better Data Access

4.1 Skills and capacity gaps

4.2 Institutional fragmentation and support

4.3 Legal action and collective responses to legal ambiguity

4.4 International coordination

Chapter 5 - Conclusions and Recommendations

5.1 Report summary

5.2 The opportunity to act

5.3 Recommendations: a phased approach

5.3.1 Recommendations for the UK Government and Regulators

5.3.2 Recommendations for the research community

Chapter 6 - References

Appendix:

I - Key terms and definitions

II - Key actors in the UK's Data Access landscape

Executive Summary

Access to social media data is essential for research on issues of significant public interest from children's mental health to national security. Yet researchers in the UK are currently unable to obtain this data reliably, lawfully or at scale. Without it, the harms the Online Safety Act 2023 was designed to address cannot be independently measured or verified and the effects of platforms' efforts to address them cannot be assessed. Nor can the effectiveness of the UK's own regulatory regime be fully tested.

This report, produced by the Social Platforms Data Access Taskforce, sets out a concrete and achievable pathway for change. It builds on Ofcom's 2025 report on researchers' access to information from regulated online services and is designed to inform the Government's forthcoming consultation on researcher access to social platforms, expected at the end of 2026.

The problem

Over recent years, the routes through which researchers access platform data have narrowed significantly. The withdrawal of free research APIs by X, the closure of CrowdTangle by Meta and restrictions across other platforms have created what researchers describe as a "post-API age", in which systematic, independent access has become

substantially harder to sustain. Researchers face barriers in the form of prohibitive costs, unstable and partial technical access, prolonged delays in response to formal access requests, inconsistent or incomplete data documentation, legal and institutional uncertainty, and inconsistent skills and capacity across the research community. These constraints inhibit researchers' ability to investigate and mitigate harm in the online world.

Why now?

The UK has a distinctive and time-limited window to act. The Data (Use and Access) Act 2025 (DUAA) has given the Secretary of State the power to create new statutory mechanisms for researcher data access and Ofcom's Section 162 report has provided a recent diagnosis of the infrastructure needed in relation to social media data access. The UK also possesses relevant expertise in secure data access infrastructure for research. This provides a strong foundation on which the Government can build recommendations to bring about change.

Internationally, the UK can learn from recent efforts to enable data access under the EU Digital Services Act, both adopting what has worked and avoiding the implementation difficulties encountered. Acting now would also ensure the UK keeps pace with international partners in its capacity to understand

online services and harms, rather than falling behind.

The urgency is heightened by a growing awareness of the need for better evidence about online activities, arising from live policy debates that depend on robust evidence, including policy plans around age-based restrictions on young people's social media use. As governments in the UK and internationally move to legislate in this area, independent research will be essential to ensuring such measures are informed by evidence rather than assumption.

Three pathways to better access

This report identifies three complementary pathways for improved data access, consistent with Ofcom's conclusion that no single mechanism can meet the full range of researcher needs and that a layered approach offers the best prospect of success.

- 1. Public interest web-scraping.** Researchers can already collect publicly accessible data through web scraping, but legal uncertainty currently discourages them or their host institutions from doing so. Clearer guidance and a durable legal foundation would remove that barrier.
- 2. Data donation.** Individuals are already empowered to voluntarily share their own platform data with researchers through data donation, but inconsistent platform compliance and the absence of standardised mechanisms limit its potential. Extending the Smart Data Framework to social media platforms would place this route on a reliable footing.
- 3. A statutory social media data intermediary.** An independent body that would centralise researcher accreditation, manage data security and provide platforms with a single, governed counterpart is the most significant structural reform this report proposes, reducing platform burden while expanding access. This intervention would lessen the burden on individual researchers and platforms to advance data access. An intermediary could serve as a clearing house through which platforms discharge their obligation to provide information, as outlined in Section 154A of the Online Safety Act, should the Secretary of State exercise this power.

Alongside these proposed reforms it remains important that platforms maintain and improve

direct API access, particularly real-time access to public data, which is essential for time-sensitive research such as elections, crises and public safety incidents.



Recommendations

The report makes a series of recommendations directed at two audiences. The first set of recommendations is provided for the government and regulators, while the second is for the research community. Each set is sequenced by actions that can be taken now, in the medium, and longer-term

Recommendations for the UK Government and Regulators

In the near term:

1. **Web-Scraping:** Consolidate existing legal and regulatory guidance for web scraping into domain-specific guidance for social media research.
2. **Data Donation:** Clarify data donation rules in relation to:
 - a. The right to third-party transfer of donated data under the UK GDPR
 - b. The legal basis for enriching donated data
 - c. Data donation portability standards
3. **Data Intermediary:** Identify the legal foundation and governance arrangements for a hybrid-repository social media data intermediary body.

In the medium term:

4. **Web-Scraping:** Establish a safe harbour to enable public interest web-scraping.
5. **Data Donation:** Establish a researcher complaints mechanism to address the failure of data donation processes and grant Information Commissioner's Office (ICO) powers to investigate and resolve complaints.
6. **Data Donation:** Outline standardised processes for the identification and delegation of data donation requests.
7. **Data Donation:** Enable continuous, as opposed to only momentary, data donation.
8. **Data Donation:** Require platforms to provide data transparency and change-log documentation around data donation processes.
9. **Data Intermediary:** Learn lessons from existing data infrastructure and from international social media data access efforts to inform hybrid-repository intermediary body design.
10. **Data Intermediary:** Engage the research community to inform design of a new intermediary body.

In the longer term:

11. **Data Donation:** Expand and normalise Smart Data schemes for social media platforms to enable data donation, building on work already underway through the Smart Data Donation Service.
12. **Data Intermediary:** Establish a statutory hybrid-repository social media data intermediary.

Recommendations for the research community

In the near term:

13. As a funder of research, UK Research and Innovation (UKRI) should clarify its position on web-scraping for public interest research.
14. Research institutions and researchers should privately share best practice on navigating data access challenges and look to develop collective responses to key areas of legal ambiguity.
15. UKRI should audit existing skills initiatives designed to promote legal, ethical and technical understanding of data access mechanisms.

In the medium term:

16. Research institutions and researchers should develop practical data protection impact assessments (DPIAs) templates and worked examples tailored specifically to social media scraping research, data donation and other data access mechanisms.
17. UKRI should fund the development of accessible, practical training and resources for researchers and research administrators covering the legal and technical dimensions of social media data research.
18. Research institutions and funders should introduce dedicated skills fellowships for academics and non-academic researchers to enable the development of data access skills.
19. Research institutions and researchers should work collaboratively to push for changes in platforms' terms of service to allow the use of web-scraping for research purposes.

In the longer term:

20. Research institutions should develop formal partnership agreements with key international research infrastructure bodies, including GESIS, the Digital Services Act (DSA) Observatory and equivalent organisations
21. Research institutions, including philanthropic funders and university consortia, should establish a legal defence fund and coordination mechanism to support researchers and civil society organisations facing platform litigation or cease-and-desist action for public interest scraping research.

The UK's moment to lead

These reforms are proportionate, achievable and cost-effective. A well-designed access regime should minimise new burdens on platforms but replaces the current fragmented and unpredictable patchwork with a single, governed framework that platforms can engage with consistently and at scale. Many of the steps set out in this report can be taken now, building momentum through early, practical action before progressing to the fuller statutory model over time.

The UK is exceptionally well placed to lead. The legislative foundations are in place, the regulatory diagnosis is clear, and there is broad consensus across regulators, parliamentary committees, civil society and the research community on the need for reform. Acting now would position the UK as a world leader in enabling secure, lawful and scalable access to platform data in the public interest. It would also strengthen the evidence base on which the success of the Online Safety Act ultimately depends, particularly in protecting children online. The conditions for meaningful reform are aligned and the opportunity to act is now.

About the Taskforce

Smart Data Research UK established the Social Platforms Data Access Taskforce to champion responsible, ethical, and secure access to data from a wide range of online social platforms. The Taskforce brings together independent experts to provide evidence-based analysis of challenges and opportunities in this area. Co-chaired by the University of Sheffield and the University of Cambridge, the Taskforce brings together researchers, regulators, funders and civil society actors across four integrated workstreams: data access models and methodologies; legal and policy frameworks; international landscape and best practices; and stakeholder engagement and solutions. The Taskforce convenes a Policy Advisory Group on which Ofcom, the ICO and Governmental representatives sit in an observer capacity.

The views, conclusions and recommendations expressed in this report are those of the Social Platforms Data Access Taskforce and do not represent the positions of Smart Data Research UK (SDR UK) or UK Research and Innovation (UKRI).

This report represents the view of Taskforce members only, and not those of any external advisors or third parties.

Taskforce members

Professor Katharine Dommett, Professor of Digital Politics, University of Sheffield (Co-Chair). An expert in digital technologies, platform regulation, political communication, and the societal implications of technology.

Professor Amy Orben, UKRI Future Leaders Fellow, University of Cambridge (Co-Chair). Her work focuses on digital technologies, mental health, and youth wellbeing, with a special lens on evidence-based approaches to understanding the societal impacts of online platforms.

Mark Scott, Senior Resident Fellow, Digital Forensic Research Lab's Democracy + Tech Initiative, Atlantic Council. Global technology policy specialist, EU comparative digital policy, regulation and governance.

Dr. David Zendle, University of York. Director of the Smart Data Donation Service (SDDS). His research focuses on the societal impacts of digital technologies and the development of data infrastructures that enable responsible access to digital behavioural data for public interest research and evidence-based policymaking.

Policy research team

Dr Alexandra Boutopoulou

UKRI Research Fellow for Social Media Data Access, University of Sheffield

Ioanna Fokas

Research Assistant, University of Cambridge

Independence

The Taskforce operates independently of any social media platform and has no commercial interest in the data access mechanisms it recommends. It also operates independently of its funder, UKRI and SDRUK. Its recommendations are based on evidence gathered from across the research community and are designed solely to serve the public interest.

Conflict of Interest

Conflicts of Interest for core Taskforce members are outlined below.

Prof Dommett has received funding from UK Research and Innovation. She is a paid reviewer for a range of grant givers and was paid to perform the role of Chair of the Independent Intermediary Body Pilot under the Digital Services Act. She is an unpaid member of the UK Department for Science, Innovation and Technology, the UK Department for Culture, Media and Sport College of Experts and the Research Advisory Group of An Coimisiún Toghcháin. She was previously an unpaid member of the British Academy Public Policy Committee.

In the past 36 months, AO has received funding or contracts from the Wellcome Trust, Nathoo Family Trust, Huo Family Foundation, Jacobs Foundation, UK Research and Innovation (incl. Medical Research Council, Economic and Social Research Council and Engineering and Physical Sciences Research Council), the UK Department for Science, Innovation and Technology, National Institute of Health, University of Cambridge, Emmanuel College and St John's College of the University of Cambridge, the Prudence Trust and the Livelihood Impact Fund. She has given paid talks at SWFGL, Apple University and various schools, and is a paid editor of Psychological Science, as well as a paid reviewer for a range of grant givers. She was an unpaid member of the ESRC Smart Data Research UK Programme Board, British Academy Public Policy Committee, UK Department for Education Science Advisory Council, UK Department for Science, Innovation and Technology and UK Department for Culture, Media and Sport College of Experts, Australian eSafety Commissioner Social Media Minimum Age Evaluation Academic Advisory Group, and a paid member of the Digital Futures for Children Centre Advisory Board.

Mark Scott was a member of Ofcom's independent Online Information Advisory Committee from April, 2025 to March, 2026.



Chapter 1

Introduction and context

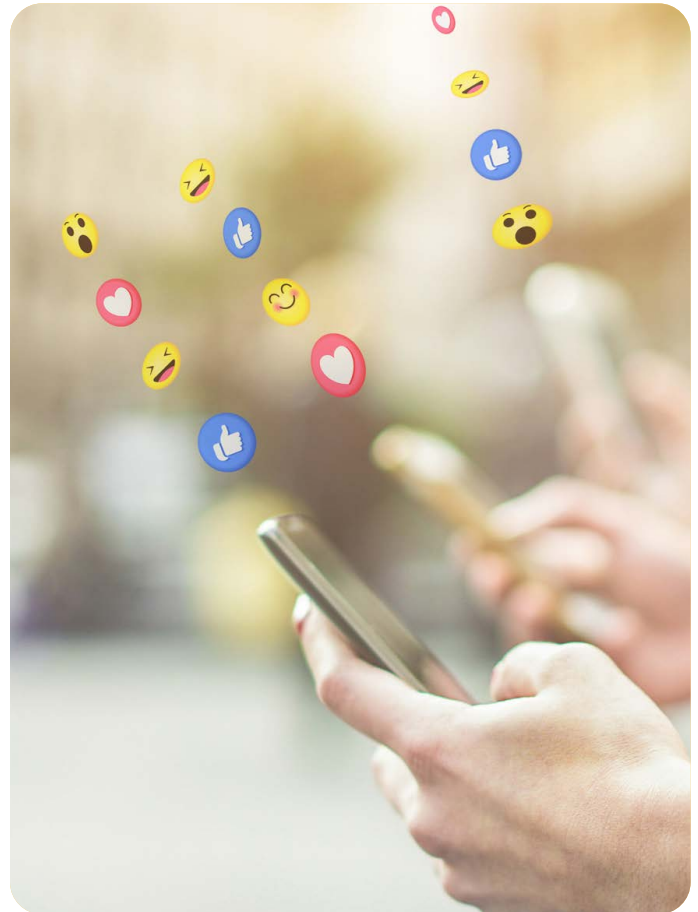
1.1 Purpose and scope

Access to social media data is essential for timely and high-quality research on matters of significant public interest, which range from adolescents' mental health to national security. Understanding the nuances of online harms depends on researchers being able to analyse platform data at scale. With such access, risks to online safety can be systematically measured, effectively mitigated and subjected to the kind of independent scrutiny that builds public trust in the digital environment.

Online harms are a growing concern for governments, regulators and civil society across the world. Social media platforms have become central to how people communicate, access information and form opinions, but they have also become environments in which harmful content, manipulative behaviour and illegal activity can spread rapidly and at scale. Research commissioned to support the Online Safety Act found that around one in 12 children in the UK has encountered priority content risks including pornography, content encouraging self-harm or suicide, bullying and hateful content, with the most vulnerable children bearing a disproportionate share of these risks (Bryce et al., 2023). Content promoting dangerous stunts and challenges represents another category of priority content under the Act, with recent UK survey data suggesting that around one in four children has encountered this type of content online (Internet Matters, 2025). Additionally, organised social media manipulation targeting elections and democratic processes has been identified in 81 countries (Bradshaw, Bailey & Howard, 2021). Addressing these harms requires not only regulation but evidence. Without the ability to study platform systems and their effects independently and at scale, neither regulators nor policymakers can be confident that their interventions are working.

The UK Government's National Data Strategy identifies five opportunities for data to transform the UK: boosting productivity and trade, supporting new businesses and jobs, increasing the speed and scope of scientific research, improving the delivery of public services and creating a fairer society (Department for Digital, Culture, Media and Sport, 2020). The Government has itself acknowledged that the vast potential of public and privately held data in the UK has yet to be realised. This potential is particularly unrealised when it comes to social media data, necessitating government action.

At the moment, the conditions for meaningful reform of social media data access are falling into place. There is growing consensus across regulators, parliamentary committees, civil society and the



research community that the current data access landscape is insufficient. There is also growing clarity about what needs to change. Inquiries by the UK Parliament's Science, Innovation and Technology Select Committee, including its 2019 report on social media and young people's health (House of Commons Science and Technology Committee, 2019), its 2025 report on social media, misinformation and harmful algorithms (House of Commons Science, Innovation and Technology Committee, 2025) and the Joint Committee on the Draft Online Safety Bill (points 414–419, Joint Committee on the Draft Online Safety Bill, 2021) have consistently highlighted the structural

imbalance between social media platforms and independent scrutiny. The push for better researcher access has a longstanding history in these debates and there have been consistent efforts to advance better access for researchers debated in Parliament. This growing cross-sector consensus now provides a strong foundation for action. This report focuses on the most recent mechanism, the Online Safety Act 2023 and Ofcom's subsequent report, Researchers' access to information from regulated online services (July 2025), published under section 162 of the Act, as the basis for its recommendations.

This report sets out a concrete pathway for the UK to act. Building on Ofcom's 2025 report on researchers' access to information from regulated online services (Ofcom, 2025), it identifies opportunities to improve social media data access and makes specific recommendations designed to inform the Department for Science, Innovation and Technology's forthcoming consultation towards the end of 2026. The goal is to position the UK as a world leader in enabling secure, lawful and scalable access to platform data in the public interest and to demonstrate that this is achievable within the existing legislative and regulatory landscape.

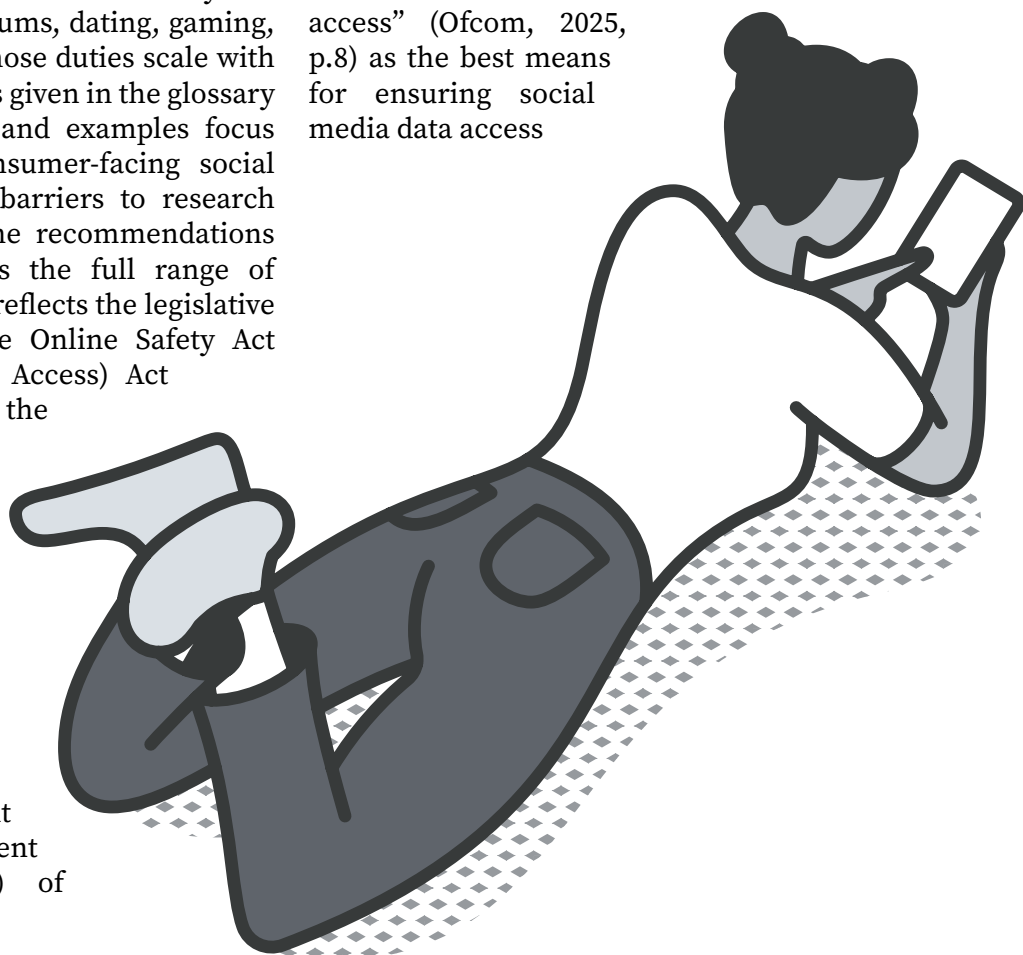
Within this report, data access refers specifically to access to social media data for the purposes of independent research into online harms. We use 'social media' as shorthand for the full range of services regulated under the Online Safety Act 2023, including messaging, forums, dating, gaming, search and adult platforms, whose duties scale with size and risk (a fuller account is given in the glossary - Appendix I). Our evidence and examples focus predominantly on major consumer-facing social media platforms, where the barriers to research access are most acute, but the recommendations are intended to apply across the full range of regulated services. This scope reflects the legislative framework established by the Online Safety Act 2023 and the Data (Use and Access) Act 2025, which together provide the primary statutory basis for any future researcher access regime in the UK. Access to other forms of online data, whilst an important concern for researchers, falls outside the scope of this report.

For the purposes of this report, "researcher" is understood broadly, consistent with the definition of independent research in section 162(2) of

the Online Safety Act 2023, to include academic researchers, civil society organisations, journalists and other parties conducting research independently of the platforms whose data is being studied.

The obligation to conduct research in accordance with applicable legal and regulatory requirements is not new. UKRI's standard terms and conditions of grant already require all grant holders to ensure their projects comply with all relevant ethical, legal and regulatory frameworks, including the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018 (UK Research and Innovation, 2025). The Economic and Social Research Council (ESRC)'s framework for research ethics and its guidance on internet-mediated research similarly recognise that social media research raises specific legal and ethical obligations, and that researchers may need to navigate questions of data identifiability, platform regulations and legal compliance as part of their work (Economic and Social Research Council, n.d.). The challenges this report addresses therefore do not reflect a gap in formal obligations but a gap in practical support that these recommendations are designed to close.

In making its recommendations, the Taskforce endorses Ofcom's judgement of the need for "[a] layered, flexible approach – combining legal clarity, technical safeguards, and independent oversight – offers the best chance of enabling responsible, timely and useful information access" (Ofcom, 2025, p.8) as the best means for ensuring social media data access



1.2 Why the UK Government needs to take action now

Governments around the world are recognising the importance of social media data access and acting on it, creating a moment of genuine opportunity for the UK to follow suit and avoid the risk of falling behind international developments. In the EU, the Digital Services Act (DSA) has established a framework mandating a safer and more transparent online environment by regulating the responsibilities of social media platforms. Under Article 40, the DSA provides the means for vetted researchers to obtain publicly accessible platform data and non-public data for public-interest research. Early evidence from implementation suggests that whilst this mechanism has promise, challenges around scope, timeliness and platform compliance remain (Ohme & Seiling, 2026). Recent research on the DSA's data access mechanism has identified what has been described as an inherent "standoff" problem: researchers must specify the data they require, but cannot know what data platforms hold, while platforms in turn use that vagueness to resist or narrow requests (Goanta et al., 2026).

This design problem has played out against a slow and uneven implementation process. Each EU member state designates a Digital Services Coordinator (DSC), the national regulator responsible for supervising compliance with the DSA and processing Article 40 access applications. Empirical research tracking data access requests across all 27 member states found that, as of late 2024, no DSC had yet processed a single Article 40 data access request, with delays attributed to the pending Delegated Act and inconsistent national implementation (Halil et al., 2025).

At a national level within Europe, countries such as Germany have invested in shared research infrastructure, including through institutions such as GESIS (the Leibniz Institute for the Social Sciences), demonstrating that collective compliance capacity can be built effectively and at scale (GESIS, Leibniz Institute for the Social Sciences, n.d.). In the United States, the bipartisan Platform Accountability and Transparency Act has been recently re-introduced in the Senate (Coons & Cassidy, 2025), reflecting growing legislative recognition of the need for structured researcher access to platform data.

Data access questions are also likely to become more important over time, as generative AI becomes more integrated into the platform environment. AI-generated content is an increasing presence on social platforms: the Reuters Institute's Generative AI and News Report 2025 reports that 21% of people

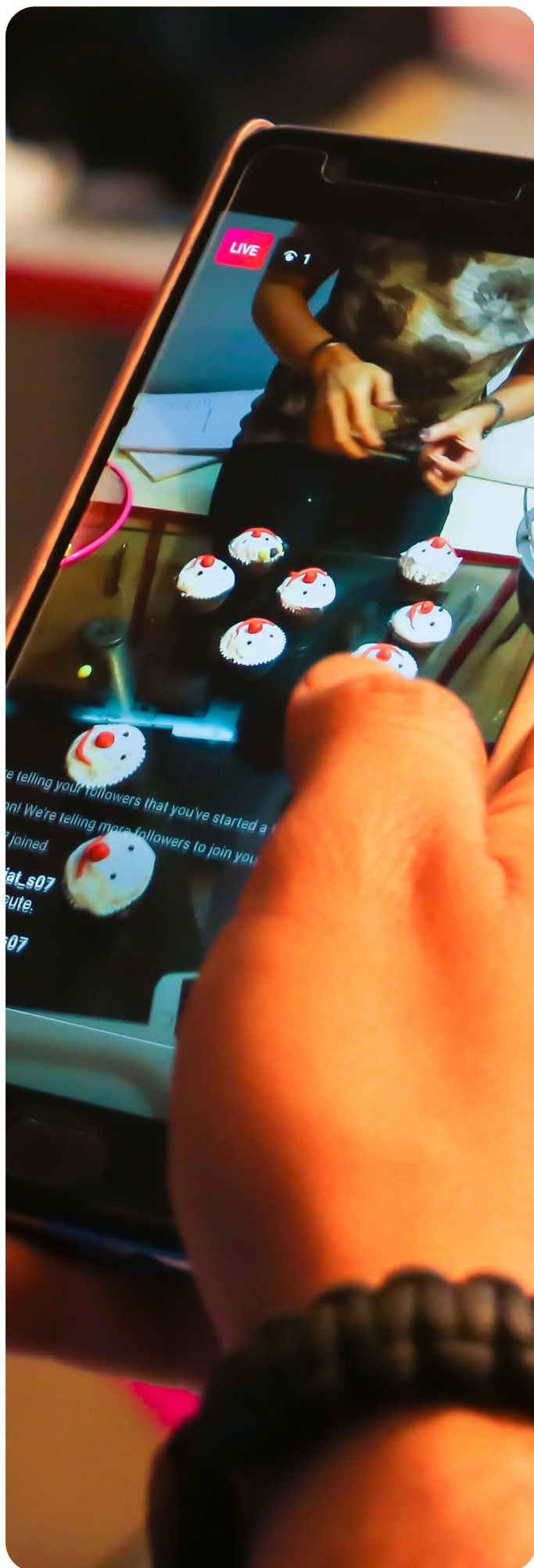
now use generative AI to create media including text, images and video and that a majority of respondents globally report some concern about distinguishing real from synthetic content online (Reuters Institute for the Study of Journalism, 2025). As platforms increasingly host and generate such content, the questions researchers ask, and the data they need to answer them, may gradually shift. The Taskforce does not make specific recommendations on access to AI systems here, considering this beyond the current legislative window, but notes it as an area that a durable data access regime may need to accommodate in time.

At a time when international momentum is building, the UK risks being left behind. In the wake of the DUAA and Ofcom's recommendations on data access for researchers, there is an opportunity to position the UK as a world leader on data access, pioneering a new response to data access that navigates challenges encountered by pre-existing initiatives. Better researcher access could also support policymakers in assessing the efficacy of measures intended to mitigate online harms. The UK already has a strong foundation on which to build, with a successful Smart Data Strategy, the Office for National Statistics (ONS) Secure Research Service and Administrative Data Research UK providing a wealth of technical expertise. This report makes the case that the UK is well placed to act decisively and sets out the specific steps needed.

1.3 The current state of data access for online research

UK researchers currently rely on a range of mechanisms to access data about the online world, including platform APIs, public archives, transparency reports, controlled research environments, web-scraping and data donation approaches (see Table 1 below). While these mechanisms provide valuable and context-specific access, they do not constitute a stable or comprehensive system. Across all routes, access remains fragmented, platform-mediated, prohibitively costly in many cases and often uneven in scope.

In recent years, major social media platforms have significantly reduced the data access tools and transparency mechanisms available to independent researchers. X discontinued its free Academic Research API in February 2023, replacing it with a tiered pricing model that has proved prohibitively expensive for most research institutions. Meta discontinued CrowdTangle in August 2024, ending the primary tool through which researchers had



accessed public Facebook and Instagram data and replacing it with Meta Content Library. However, while the Content Library represents a formal replacement, access is restricted to accredited researchers from approved institutions, it operates through a controlled environment instead of an open API and the volume and granularity of data available is significantly reduced compared to its predecessor. Reddit implemented similar restrictions on its free API in the same period (Brown et al., 2024; Davidson et al., 2023; Shiffman & Silverman, 2025). These changes have not been offset by equivalent new provision and the reduction in internal platform staffing devoted to transparency and researcher relations has compounded the problem (Ofcom, 2025). Social media data is therefore increasingly challenging to access. Taken together, these developments have resulted in what researchers have described as a “post-API age”, in which systematic, independent access to social media data has become substantially more difficult to sustain (Freelon, 2018; Perriam et al., 2020; Tromble, 2021; Breuer et al., 2022; Freelon et al., 2024; Mimizuka et al., 2025).

This assessment is consistently reflected in recent evidence from across the UK research and policy landscape. Ofcom’s statutory report under Section 162 of the Online Safety Act identifies limitations in the availability, consistency and usability of social media data (Ofcom, 2025). Submissions to Ofcom’s call for evidence (which ran from October 2024 to January 2025, including the Ada Lovelace Institute, the Open Data Institute, the Institute for Strategic Dialogue, the 5Rights Foundation and the Knight-Georgetown Institute) point to structural barriers that constrain independent scrutiny and limit the effectiveness of public-interest research (Ada Lovelace Institute, 2025; Open Data Institute, 2025; Institute for Strategic Dialogue, 2025; 5Rights Foundation, 2025; Knight-Georgetown Institute, 2025b).

These barriers are not unique to the UK. A systematic assessment of data access conditions across 15 major platforms in the EU, UK and Brazil found that meaningful transparency remains the exception rather than the norm in all three jurisdictions, regardless of the regulatory frameworks in place and that platforms consistently restrict access in ways that cannot be explained by technical or commercial constraints alone (Santini et al., 2026). What makes the UK’s position distinctive is the legislative moment. The DUAA and Ofcom’s Section 162 report have together created a specific window for reform that does not yet exist in comparable form in the other jurisdictions studied.

These constraints to data access can be understood

across four interrelated categories, consistent with the analytical framework set out in Ofcom's 2025 report on researcher access (Ofcom, 2025):

Financial barriers

In recent years there has been a sharp escalation in the cost of accessing platform data, most notably following changes to X's API pricing. Numerous civil society organisations and academics reported that access to X's API became prohibitively expensive, preventing planned research. Such pricing structures effectively exclude all but the most well-resourced actors and undermine the diversity of the research ecosystem.

These direct costs are compounded by less visible financial burdens arising from the other barriers discussed below, including the cost of alternative tools and infrastructure, legal advice sought to manage regulatory uncertainty, investment in data protection and IT security compliance, and the training or hiring of specialist staff. Financial barriers should therefore be understood not only as the direct cost of accessing platform data, but as a constraint that runs through every other barrier identified in this report.

Technical barriers

Researchers consistently report that even where access is nominally available, it is often partial, unstable, or delivered in formats that limit usability. Platform-defined APIs constrain the types of research questions that can be asked, as researchers are limited to the data structures and endpoints platforms choose to expose (Dommett, Orben & Zendle, 2025; Ofcom, 2025). The result is a fragmented research landscape in which findings are difficult to validate, replicate or compare over time.

Legal and institutional barriers

Legal uncertainty is repeatedly identified by both platforms and researchers as a central constraint on data sharing and access. Meta has highlighted concerns about joint controllership under UK GDPR and argued that unclear liability boundaries discourage broader data sharing with independent researchers. Meanwhile researchers have described how ambiguity surrounding platform terms of service, potential legal liability and uneven institutional capacity to manage risk discourages many researchers from undertaking platform-based research (Dommett, Orben & Zendle, 2025).

Capacity and skills barriers

There is a growing capacity and skills barrier associated with the diversity and volatility of data access mechanisms. Researchers are increasingly required to navigate a complex and rapidly changing landscape of platform-specific APIs, bespoke data formats, multimodal content types and evolving access conditions. This places significant demands on technical expertise, infrastructure and institutional support that many researchers do not possess. Where such capacity is lacking, research is delayed, scaled back, or abandoned altogether. These challenges disproportionately affect smaller research teams, civil society organisations and disciplines outside computer science, reinforcing existing inequalities in who is able to conduct platform research and whose questions are prioritised.

Evidence from regulators, academic institutions and civil society organisations consistently identifies the same structural problems: inadequate access, legal uncertainty, fragmented infrastructure and unequal distribution of capacity. The legislative and regulatory context now provides a credible basis for reform. The recommendations in this report are designed to translate that opportunity into concrete action.

Table 1: Current Data Access Mechanisms and their Challenges

Platform Research APIs		
What This Provides	Examples	Challenges
Application Programming Interfaces (APIs) are structured, machine-readable interfaces that allow approved researchers to programmatically query and retrieve data directly from platform systems. They enable scalable and automated data collection within predefined parameters.	• X (formerly Twitter) API • TikTok Research API • Meta Content Library API	• Not available across all platforms • Access is selective and often restricted to approved researchers • Data scope is defined by the platform rather than research needs • Significant variation in structure, fields and coverage across platforms • Subject to pricing, rate limits, or withdrawal at platform discretion • Limited transparency over sampling, filtering and missing data
Platform Archives and Public Databases		
What This Provides	Examples	Challenges
Platform-maintained, publicly accessible repositories that enable analysis of specific categories of content, typically for transparency or regulatory purposes.	• Meta Ad Library • Google Ads Transparency Center	• Limited to specific content categories (e.g. advertising) • Not available across all platforms • Data structure and access functionality vary significantly • Often incomplete or lacking key metadata • Limited historical depth; archives may not be maintained longitudinally • Not designed for systematic or cross-platform research
Controlled or Secure Research Environments		
What This Provides	Examples	Challenges
Access to sensitive or large-scale datasets through secure, access-controlled infrastructures (e.g. Trusted Research Environments), where data is analysed within governed systems and cannot be directly extracted.	• UK Data Archive • Office for National Statistics Secure Research Service	• Require formal accreditation and governance approval • Access processes can be slow and administratively intensive • Often require predefined research designs • Limit flexibility for exploratory or iterative analysis • Integration with platform-origin data remains limited restricting external analytical tools (for example, trained ML classifiers)

Transparency Reports and Disclosures		
What This Provides	Examples	Challenges
Regular platform-published reports containing aggregated information on content moderation, enforcement actions and compliance with policies or regulation.	• Meta Transparency Reports • Google Transparency Report • TikTok Transparency Centre	• Aggregated data limits analytical granularity • Lack of standardisation across platforms • Methodologies are platform-defined and not independently verifiable • Designed for accountability, not research • Limited capacity to support causal or behavioural analysis • Inconsistent IDs across releases prevent matching or hydration, even where account- or post-level data is provided (e.g. DSA transparency reports on deleted content) • Often published as non-machine-readable documents (e.g. PDFs) rather than structured datasets
Platform-Curated or Released Datasets		
What This Provides	Examples	Challenges
Datasets selectively released by platforms for research purposes, often linked to specific events, partnerships, or regulatory pressures.	• Meta datasets released via the Social Science One consortium (e.g. 2016 US election data) • Twitter/X COVID-19 research dataset	• Typically one-off releases • Narrow scope and limited temporal coverage • Limited transparency over dataset construction • Often restricted to a small group of researchers • Lack of continuity and reproducibility • Risk of undetected errors affecting completeness or accuracy, with limited external means to verify (Timberg, 2021)
Web-Scraping and Web Crawling		
What This Provides	Examples	Challenges
Automated extraction of publicly accessible data from platform interfaces (e.g. websites), enabling independent and flexible data collection outside platform-provided access routes.	• Scraping public posts from X or Reddit using automated scripts • Collection of publicly visible TikTok video metadata via browser-based tools	• Legal uncertainty (e.g. interaction with terms of service and data protection law) • Potential contractual and regulatory risks • Technical instability due to platform countermeasures • Ethical considerations where personal data is involved • Limited standardisation and reproducibility

Consent-Based Data Donation		
What This Provides	Examples	Challenges
A regulatory access pathway under Article 40 of the EU Digital Services Act, enabling vetted researchers to request platform data to study systemic risks (e.g. disinformation, online harms).	- Smart Data Donation Service (University of York / Smart Data Research UK) - PORT (open-source privacy-preserving data donation tool)	- Limited scale and representativeness - Susceptible to selection bias - Technical complexity in processing - heterogeneous data exports - Dependent on platform data portability quality and formats - Highly unstable over time, as data coverage and formats change frequently, particularly across languages for some platforms - Requires significant ethical, legal and technical infrastructure
Access via the Digital Services Act (DSA)		
What This Provides	Examples	Challenges
A regulatory access pathway under Article 40 of the EU Digital Services Act, enabling vetted researchers to request platform data to study systemic risks (e.g. disinformation, online harms).	- Article 40 DSA data access requests - Platform research tools expanded or formalised under DSA obligations (e.g. Meta Content Library, TikTok Research API)	- Does not apply within the UK, but UK-based researchers are not excluded: eligibility depends on the researcher's affiliation and research purpose, not their location. As the UK has no Digital Services Coordinator, applications route via the platform's country of establishment instead - Early-stage implementation; processes still evolving - Access limited to vetted researchers meeting strict criteria - Research related to systemic risks in the EU - Operational procedures remain unclear or inconsistent - Dependent on platform compliance and technical readiness - Does not yet guarantee timely, complete, or standardised access

1.4 Structure of the Report

The remainder of this report is organised as follows:

Chapter 2 sets out the case for action, describing the current policy moment in the UK and the opportunities that enhanced data access would create for research, regulation and the wider public interest.

Chapter 3 forms the analytical core of the report. It examines three complementary pathways to improved social media data access: public interest web-scraping (3.2), data donation (3.3) and the establishment of a social media data intermediary (3.4). For each pathway, it sets out what the mechanism involves, why it matters, the current legal and technical landscape, and the specific reforms the Taskforce recommends. The chapter also reviews existing UK and international intermediary models and the lessons they offer, before summarising the policy options across all three pathways (3.5).

Chapter 4 turns to the research community itself, identifying the actions researchers, institutions and funders can take to build skills and capacity (4.1), address institutional fragmentation (4.2), respond collectively to legal uncertainty (4.3) and strengthen international coordination (4.4).

Chapter 5 draws the report together, summarising the key findings (5.1), setting out the consequences of inaction (5.2) and presenting the Taskforce's recommendations as a phased package for the UK Government, regulators and the research community (5.3).

Chapter 6 concludes with references and an appendix containing a glossary of key terms and an overview of key actors in the UK's data access landscape.



Chapter 2

Toward better data access: The UK opportunity

public interest (UK Parliament, 2025).

It is worth noting that UK data access efforts are specifically focused on research into online safety matters. Underpinned by the Online Safety Act 2023, there is particular interest in enabling research that helps to protect children and adults online from a defined set of harms including child sexual abuse, fraud, terrorism and illegal content more broadly (Department for Science, Innovation and Technology, 2025).

To improve action against online harm, section 162 of the Online Safety Act sets out provisions relating to data access for researchers into online safety matters, calling for Ofcom to publish a report describing how researchers are currently able to obtain information from providers of regulated services, exploring the legal and other issues constraining data sharing and assessing the extent to which greater access might be achieved (UK Parliament, 2023).

Ofcom published its report entitled Researchers' access to information from regulated online services in July 2025. The report found that researchers remain concerned about "the lack of timeliness, consistency and specificity of information access" (Ofcom, 2025, p. 7). Voluntary access arrangements have also tended to favour senior researchers at prestigious institutions, reinforcing existing inequalities within the research community (Ofcom, 2025, p. 38).

Ofcom (2025, p.41) identified three policy pathways through which researcher access to platform data could be improved. These pathways are not mutually exclusive and are best understood as complementary parts of a future access regime. Ofcom explicitly noted that "no single model is likely to meet the full range of researcher needs. A layered, flexible approach – combining legal clarity, technical safeguards and independent oversight – offers the best chance of enabling responsible, timely and useful information access" (Ofcom, 2025, p.8).

1. Clarifying existing rules

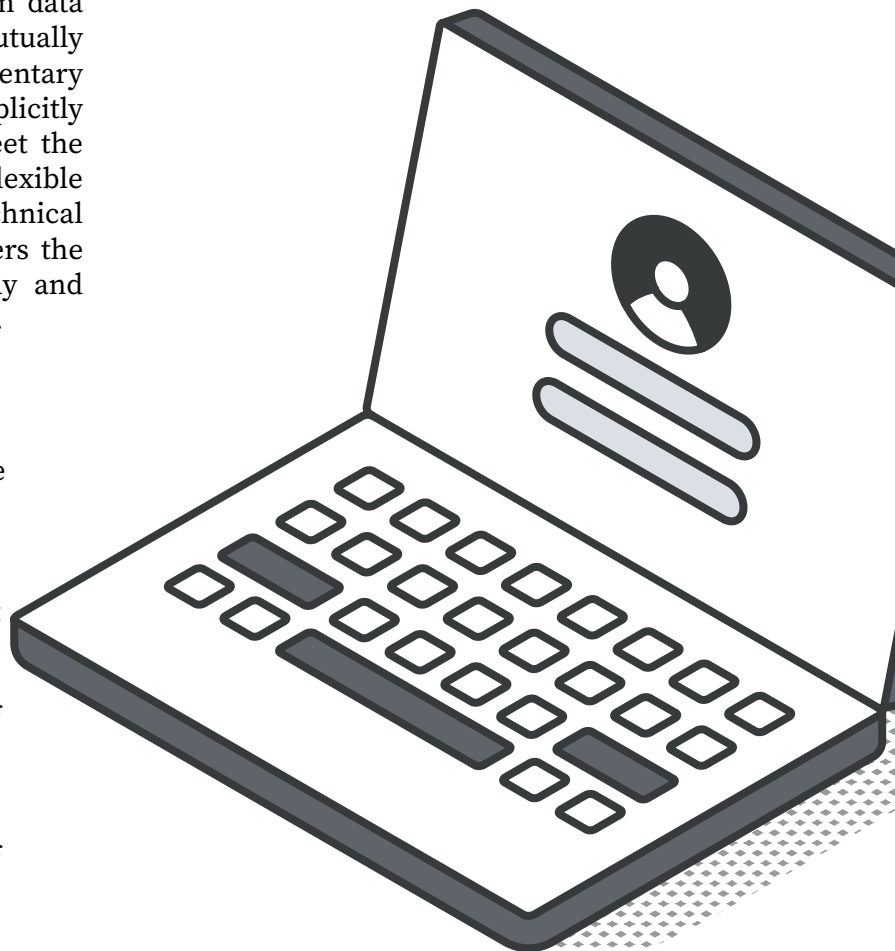
Ofcom concluded that clearer, authoritative guidance on lawful research data access, particularly where data are publicly available or processed for public-interest research purposes, would enable more consistent and responsible access within existing legal boundaries. Ofcom highlighted widespread uncertainty regarding the permissibility of common research practices, including the collection of publicly accessible data and the use of consent-based data donation models. This uncertainty arises from the interaction of data protection law, platform terms of service

and perceived liability risks and has contributed to institutional risk aversion and a chilling effect on public-interest research. Specifically they recommended that "Relevant authorities could provide additional guidance on what is already legally permitted for researcher access on important issues, such as data donations and research-related scraping" (Ofcom, 2025, p.9).

2. Create new duties for services

Ofcom proposed the creation of new duties on regulated services. Under this approach, platforms would be required to establish systems and processes through which accredited researchers could request access to defined categories of data for specified public-interest purposes. Such duties would shift access away from discretionary, platform-controlled arrangements toward enforceable expectations grounded in public law, reducing reliance on voluntary APIs or ad hoc partnerships.

Here they recommended policy makers consider that "[s]ervices could be required to put in place systems and processes to operationalise data access. This could take the form of a direct access model, which could include standardised service-led procedures for researcher accreditation and data handling, with the backstop regulator responsible for enforcement.



Services themselves would be responsible for the tasks of accrediting researchers, providing researchers with data directly or providing the interface through which they can access it and offering appeal and redress mechanisms. A new regulatory body would need to be established, or an existing organisation would need to be given new powers” (Ofcom, 2025, p. 9).

3. Establish and manage access via independent intermediary

Ofcom recommended the use of an independent trusted intermediary to manage researcher access to platform data. Rather than mandating direct data sharing between platforms and researchers, this approach would place responsibility for access governance with a third party operating under defined statutory or regulatory oversight. Ofcom outlines three possible intermediary models: a) a direct access intermediary model, in which the intermediary provides the access interface while services retain data hosting, b) a notice to service intermediary model, in which researchers submit proposals to an intermediary which liaises with services and both parties assess the request and c) a repository intermediary model, in which the intermediary itself hosts or manages access to data provided by services.

Specifically they recommended that policy makers: “[e]nable and manage access via independent intermediary: New legal powers could be granted to a trusted third party which would facilitate and manage researchers’ access to data. This could be done through facilitators who vet researchers and

provide secure access (for example through “clean rooms” or application programming interfaces (APIs)), through researchers submitting requests directly to an intermediary which then liaises with services, or through intermediaries which host or manage data access services (either locally or virtually)” (Ofcom, 2025, p. 9).

Since the publication of Ofcom’s report, the framework for action on researcher data access is taking shape. The Data (Use and Access) Act 2025 grants the Secretary of State the power to make regulations requiring regulated services to provide information for the purposes of independent research into online safety matters, and the Government is expected to bring forward a consultation on proposals for such regulations towards the end of 2026.

2.3 What enhanced Data Access could achieve

The UK has a world-leading reputation for its research, but currently academics, civil society, journalists and policymakers face significant barriers to conducting cutting-edge research into the online world. Securing greater access to social media data for public interest research would not only position UK researchers to lead new programmes of work on a range of vital social issues but would also strengthen the UK’s capacity to be a world leader in identifying, understanding and addressing online harms. This includes research into:



Platform design, algorithms, moderation and systemic risk

Platform design, recommender systems and moderation practices have a pivotal influence on our experience of the online world and can reveal problematic practices and causes of online harm. Currently, researchers are fundamentally constrained by the absence of independent access to system-level data, limiting their ability to assess how these systems operate in practice or generate risk over time. Action to provide social media data access would position UK researchers at the forefront of debates around algorithmic amplification and enforcement consistency and would ensure that UK policymakers are able to understand the effect of policy interventions on the prevalence and distribution of online harms.

Real-time behaviour, networks and influence

Effective media reporting and accountability processes require real-time research of online behaviour. Whether studying extremism, foreign influence operations, or the exposure of children and young people to harmful content, there is often a need for rapid, time-sensitive analysis of social media content. This is particularly acute at moments such as elections, crises, or periods of heightened online harm or risk. Action to enable real-time social media data access would ensure that UK researchers are able to identify, report and help remove problematic content, mitigating online harm in a timely and effective manner. Enabling this kind of independent research would also strengthen Ofcom's own capacity to fulfil its statutory functions, supplementing its evidence base with research it does not have the resources to generate alone. Ofcom itself has acknowledged this connection, noting in its crisis response guidance that its researcher access report and the powers created under the Data (Use and Access) Act 2025 are relevant to strengthening the evidence base available in moments of crisis (Ofcom, 2026).

Impacts on individuals over time

The evolution of digital technology has been rapid, affecting our lives across generational boundaries. Our understanding of the impact of this change on individual wellbeing, mental health, information consumption, educational attainment and behavioural outcomes is significantly underdeveloped due to a lack of longitudinal, comparable data about the online world. Action to enable longitudinal access to consistent platform data would enable researchers to understand individual-level harm and the real-

world impact of technologies. Such research could help to mitigate specific harms to children and other vulnerable groups by facilitating evidence-based policymaking. Whilst comparable data about the offline world underpins several crown jewel research investments, including national cohort studies, the current state of data access is prohibiting equivalent analysis of digital trends, making government action more necessary. Adding social media data to UK longitudinal data access would provide a step change for understanding the impact of digitalisation on individuals and society.

Revealing illegal practices

Social media platforms have become a key environment in which violations of civil and election law take place and where fraud, terrorism and abuse is played out. Understanding how these practices manifest and how they bypass existing safeguards and protocols is key to holding platforms to account, preventing individuals from experiencing harm and mitigating negative outcomes for society and democratic processes. Action to enable real-time data access would ensure that UK researchers are able to identify and document potential violations, provide independent evidence to electoral authorities and regulators and support the enforcement of existing law.

Platform governance, interventions and accountability

Independent access to platform data also matters for another significant reason: there is no independent verification of how platforms moderate their content, how they enforce their own rules and report back on what they remove. Data access would let researchers test these claims directly, using the same content moderation decisions, enforcement actions, transparency reports and removed content platforms already hold. This evidence would allow researchers and policymakers to hold platforms to account for gaps between their stated commitments and what actually happens in practice.

Chapter 3

Policy pathways to Social Media Data Access in the UK



3.1 Introduction

This chapter is designed to consider the policy pathways through which data access can be improved. It builds on Ofcom's 2025 [report](#) to identify tangible opportunities for improved social media data access in the UK to support platform transparency and civil engagement. The Government's consultation on data access, scheduled for the end of 2026 provides a route to enacting these proposals.

Drawing on extensive engagement with the research community undertaken by the Social Platforms Data Access Taskforce, the chapter makes specific recommendations across three areas:

1. Public-interest web-scraping,
2. Data donations; and
3. Creation of a social media data intermediary body.

3.2 Web-scraping

3.2.1 What is public-interest web-scraping?

Web-scraping refers to the use of software to automatically browse, parse and extract information from websites or platform interfaces and storing that information in a structured format suitable for analysis (Venturini & Rogers, 2025). It is technically distinct from the use of official APIs, which are structured data pipelines provided and controlled by platform operators. Researchers may scrape through several routes: parsing the HTML source code of public-facing pages; using undocumented internal API endpoints not intended for third-party access; or deploying browser plugins through which participants share their own platform activity with the research team (Brown et al., 2024). Each method carries different legal, ethical and technical implications. Platforms often update their technical architecture specifically to obstruct automated extraction via web scraping.

Web-scraping can be used to collect data from a range



of online spaces, including websites, video streaming channels or online communities. For services such as social media platforms, scraping can be conducted with the user either logged-out or logged-in. Logged-out scraping collects only data publicly visible without authentication. It carries lower legal risk but enables the collection of more limited datasets. Logged-in scraping provides richer data but requires accepting platform terms of service, which almost universally prohibit automated collection. As social media platforms have progressively moved content behind login walls (partly to prevent AI companies from training on their data), logged-out data scraping has become less insightful for efforts to conduct public interest research (Social Platforms Data Access Taskforce, 2025a).

The difference between public-interest scraping, in which researchers access platform data for transparency and civic goals, and commercial-focused scraping, in which firms extract publicly-available information often to train AI large language models is an important distinction. The former is conducted in the public-interest with specific research and society-wide objectives. The latter is typically dedicated to commercial activities that carry separate legal challenges. Due to ambiguity around the legal foundation for public interest scraping, some researchers have become reliant on commercially gathered data offered by companies such as NewsWhip and Bright Data. Distinguishing these two types of data collection is critical, as it is the foundation of public interest scraping which our report seeks to clarify and ensure.

3.2.2 Why is web-scraping needed?

Web-scraping has become an increasingly important means of data access in the wake of restrictions to platform APIs or changes in access rules. Yet it can be technically difficult, costly and time consuming for researchers without requisite computer science skills, particularly given the need to continually update systems in response to changes in platforms' interfaces and protections. Scraping is also essential as not all platforms offer data access mechanisms such as APIs, with many smaller platforms hosting high-risk content such as political extremism online often lacking access mechanisms. This makes it important to ensure routes for data access in cases where other mechanisms do not exist.

In addition, data provided via APIs can often be insufficient for research purposes. To provide an example, to study the impact of TikTok videos on viewers' mental health, TikTok's API can provide valuable data for video links including audience

metrics, total number of likes, total number of comments, voice-to-text, subtitles, the time of creation and video length. It also provides comment text, comment likes, replies and the time each comment was posted. However, it does not provide the actual video content a viewer is seeing. If looking to conduct research into viewers' response to the imagery of a specific TikTok post, that data can only be accessed via scraping (TikTok for Developers, n.d.). This gap illustrates a wider pattern: the data withheld from researchers is often accessible to commercial actors through paid partnerships and advertising APIs, creating a structural asymmetry in which those with commercial purposes can access far richer platform data than those conducting independent public interest research (Shiffman & Silverman, 2025).

In the case of TikTok, the platform also differentiates between academic researchers who can directly access its researcher API and those from civil society who are instead provided data access via a virtual clean room, in which the platform first reviews API-based data before displaying it to these researchers within the secure environment.

Scraping is also needed to assess the robustness of the data companies provide. A recent empirical audit of the TikTok Research API and the Meta Content Library found systematic gaps between the data these APIs provided and what was actually visible in the public information environment, with platform-imposed filters excluding up to 50 per cent of content and stripping up to 83 per cent of contextual metadata (Bekavac & Mayer, 2026). Scraping therefore provides a vital mechanism for ensuring the robustness, completeness, and independence of data available via other means. Where official APIs are incomplete, unrepresentative, or subject to unannounced change, scraping of publicly accessible data remains not just a workaround but a methodological necessity for independent and verifiable research.

Many civil society and academic research organisations, as well as universities, are inherently cautious about disclosing their use of public-interest scraping techniques, given that almost all social media platforms prohibit such techniques within their terms of service, meaning any use of public-interest scraping resides in a potentially legal grey area. These organisations are understandably concerned about legal action. Risks vary between platforms. One civil society organisation contacted by the Taskforce outlined how they regularly used public-interest scraping techniques to gather images from a social media company's ad library, as the platform's existing API did not provide such granular information.

3.2.3 Current legal foundations, implications and recommendations

Scraping for research in the UK does not take place within a single and comprehensive legal framework. Instead, it sits at the intersection of several overlapping bodies of law that do not specifically focus on public-interest social media research. The principal legal instruments relevant to UK-based researchers are the UK GDPR, the Data Protection Act 2018 and the Data (Use and Access) Act 2025 (which amends but does not replace the UK GDPR framework). Our goal in this section is to identify areas of ambiguity, outline the implications for researchers and make recommendations for reform to improve data access.

UK GDPR and the research exemption

Under the UK GDPR, web scraping that involves the collection of personal data (including usernames, profile information and posts that can be linked to identifiable individuals) requires a lawful basis for processing. The ICO confirmed through its 2024 consultation series on generative AI and data scraping that "legitimate interests" is the most plausible available lawful basis for processing web-scraped personal data, including for research purposes. This requires a documented three-part assessment demonstrating that the processing purpose is legitimate, that processing is necessary for that purpose and that it does not override data subjects' rights and freedoms (ICO, 2024).

The UK GDPR contains a research exemption under Article 89, which permits relaxed requirements, including exemptions from certain transparency obligations and data subject rights, where processing serves genuine scientific research purposes and is accompanied by appropriate technical and organisational safeguards. In principle, this exemption provides a workable basis for much public interest social media research.

The Data (Use and Access) Act 2025, which received Royal Assent in June 2025, introduced a new "recognised legitimate interest" basis and grants the UK Government powers to create a researcher access framework for online safety matters under the Online Safety Act 2023 (Data (Use and Access) Act 2025). The ICO opened consultation on corresponding guidance in August 2025 (ICO, 2025). The implications of these provisions for research scraping remain unsettled. But the Data (Use and Access) Act currently represents the most significant legislative development for research data access, although there are few details provided by

the UK government on how these provisions will be implemented.

Where special category data may be collected (including data revealing political opinions, health information, sexual orientation, or other sensitive attributes), the legal threshold for web scraping is higher. Article 9 of the UK GDPR governs the processing of special category data and requires researchers to satisfy not only a lawful basis but also a separate, specific condition permitting such processing. In practice, this means demonstrating the research serves a substantial public interest, is carried out with appropriate safeguards and cannot reasonably be achieved by processing non-special category data. Research into online harms, extremism, radicalisation, or mental health routinely involves such data, compounding the compliance burden for researchers working in precisely those areas where evidence of potential online harms is most urgently needed by policymakers (Ofcom, 2025). There is currently no domain-specific guidance setting out what a compliant Article 9 condition looks like in the context of social media scraping research. That gap leaves researchers without a clear evidential standard to work to.

Implications for researchers

The data protection research exemption is consistently identified as a grey area by researchers. There have been calls for domain-specific guidance to clearly establish what is permissible in the context of social media scraping, particularly related to sensitive data (Social Platforms Data Access Taskforce, 2026a). In practice, most research institutions currently judge the use of web scraping to be too legally risky, this is especially within the university context, where grant/funding conditions (such as UKRI terms and conditions) note that researchers need to remain within legal bounds when conducting research.

This decision has a systemic chilling effect on societally-relevant research (Social Platforms Data Access Taskforce, 2025a). Within universities there are examples of blanket restrictions on scraping tools or policies that place compliance burden on individual researchers - requiring often junior researchers to navigate data protection requirements with limited institutional support and legal uncertainty. Whilst within civil society organisations some are avoiding scraping altogether, whilst others have imposed criteria to assess the risk of conducting research on different topics. These responses limit research on online safety, and decrease the UK's competitiveness in this research space internationally.

In response, some researchers are seeking case-specific legal advice on research proposals to attempt to gain legal cover for the use of scraping. This creates financial pressures and is not a route open to all, with many individuals and research institutions unable to fund bespoke legal advice. The Taskforce heard numerous instances of researchers abandoning legitimate projects because the combination of legal uncertainty, institutional caution and the time cost of compliance had become unsustainable (Social Platforms Data Access Taskforce, 2025a). What researchers need is twofold: confidence that their own research practices are lawful and practical tools, such as data protection impact assessments (DPIAs), through which they can demonstrate this to institutions, funders and platforms.

Recommendations for the UK Government

Consolidate existing legal and regulatory guidance for web scraping into domain-specific guidance for social media research.

The government should direct the ICO to develop practical, domain-specific guidance for researchers conducting social media research that involves public-interest web scraping. Existing ICO guidance on web scraping is predominantly focused on generative AI and commercial data brokering. It does not adequately address the specific circumstances of public interest research.

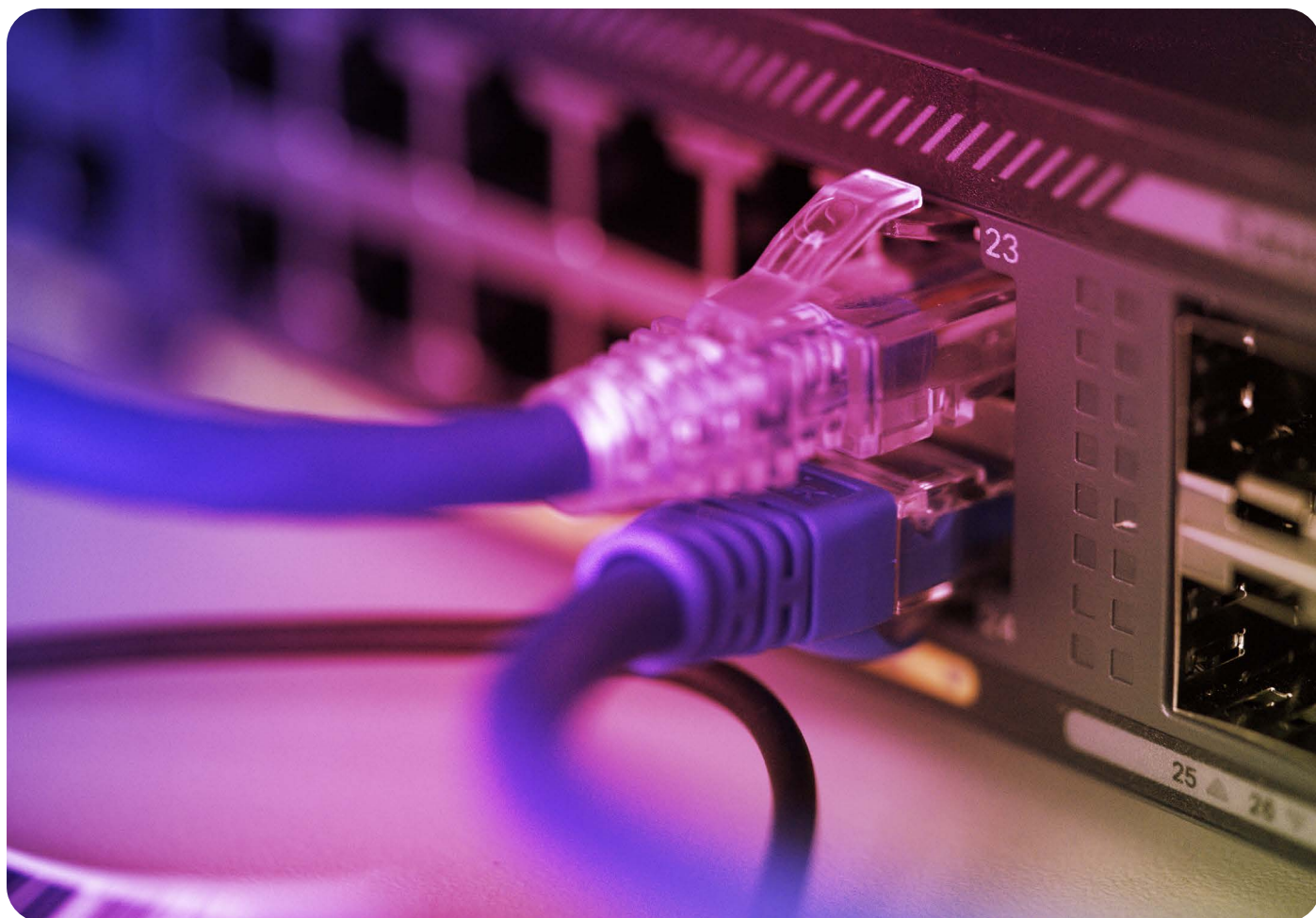
Guidance should address:

1. The application of the research exemption to social media scraping.
2. The circumstances under which logged-out versus logged-in scraping can be justified under UK GDPR.
3. The conditions under which special category data collected during scraping can be processed under Article 9, including the safeguards and documentation required.

Such guidance would not require primary legislation. It could be developed within the existing ICO mandate, informed by the evidence set out in Ofcom's 2025 report.

Platform Terms of Service and Contract Law

Web scraping is also affected by contract law. Terms of service can be imposed in two ways, a) either as a browse-wrap agreement, where terms are implied merely by visiting a website rather than accepted at login or b) as a click-wrap agreement, where acceptance is recorded at login.



The terms of service that platforms impose on users as a condition of creating an account (click-wrap agreements) almost universally prohibit automated data collection. Researchers who create accounts and then scrape (logged-in scraping) technically breach these contracts, exposing themselves to breach-of-contract claims. The enforceability of such terms varies and several US court decisions have questioned whether anti-scraping provisions are always enforceable, particularly where a platform cannot demonstrate material harm from public interest research. In *X Corp. v. Center for Countering Digital Hate, Inc.* (2024), a US federal court ruled that a civil society organisation's scraping of publicly available data for accountability purposes did not constitute actionable harm. In *Meta Platforms, Inc. v. Bright Data Ltd.* (2024), a US court held that anti-scraping provisions did not extend to data publicly visible to non-users, a ruling specifically concerned with logged-out scraping, where no account is created and no terms of service are formally accepted.

There is no directly analogous UK case law testing these questions. The enforceability of browse-wrap agreements in particular remains legally uncertain in the UK, as no court has been asked to determine whether such terms can bind a researcher scraping publicly visible data for public interest purposes.

It should be noted that some platforms, such as X, have announced intentions to amend their Terms of Service to explicitly state that eligible researchers are not contractually prohibited from scraping publicly available data (European Commission, 2026b). Such moves remain, however, uncommon, necessitating action to enable scraping without Platform action.

Implications for researchers

Multiple organisations consulted by the Taskforce described the threat of legal action by platforms, commonly referred to as ‘lawfare’, as a powerful deterrent to public interest web scraping. Even where the likelihood of legal action being successful is slim, research organisations voiced caution about the potential for lawsuits to consume organisational time and resources and the possibility of being banned from platforms. To date, there have been few researchers that have been legally targeted for public-interest web scraping, however organisations are routinely risk averse. Legal advisors consulted by the Taskforce noted that platform Terms of Service penalty clauses, including one major platform's reported rate of £10,000 per day for scraping violations, are likely unenforceable under UK contract law as ‘penalty clauses’. However, this principle has not been tested and doing so would

be cost-prohibitive for most research organisations (Social Platforms Data Access Taskforce, 2026a).

Beyond the enforceability of specific contractual provisions, researchers in the UK face a broader strategic risk from platform litigation. The threat of strategic lawsuits against public participation (SLAPPs), or litigation brought mainly to discourage research organisations rather than to vindicate genuine legal rights, is particularly acute in the UK, which currently lacks the anti-SLAPP legislative protections available in several US jurisdictions. Many researchers and civil society organisations called consistently for the introduction of UK anti-SLAPP provisions to protect public interest scraping research. It is worth clarifying, however, that few of these SLAPPs have targeted UK researchers, even if the threat of such litigation remains significant.

Recommendations for the UK Government

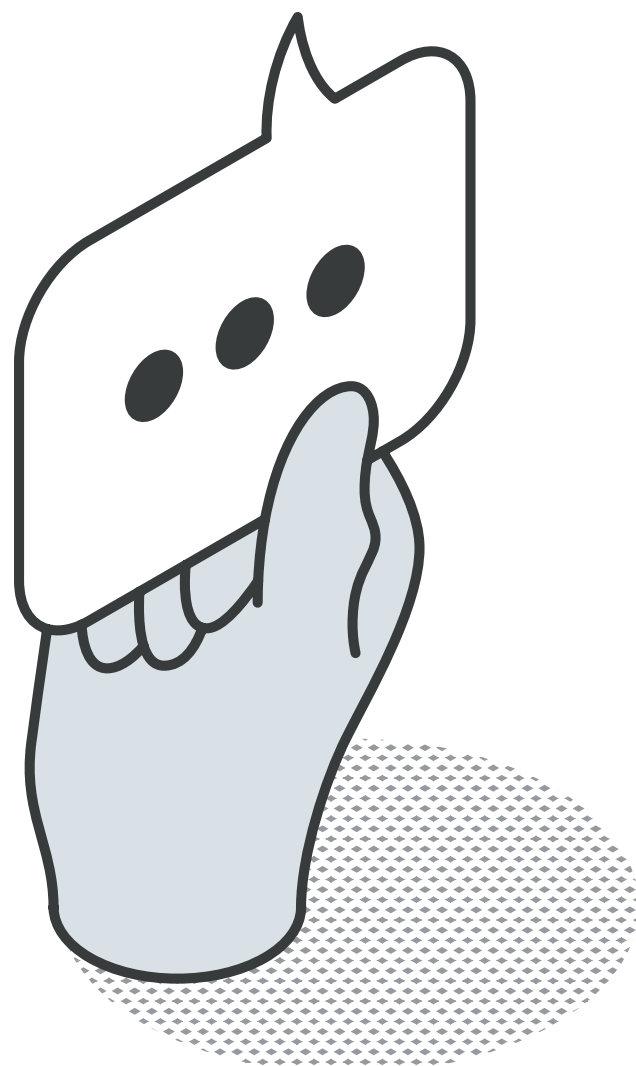
Establish a safe harbour to enable public interest web-scraping.

The most durable solution to the current ambiguity would be the establishment of a statutory legal safe harbour to protect researchers who conduct public interest scraping of publicly accessible social media data from adverse platform action. Such a provision would need to define clearly the classes of researcher and institution covered, the conditions under which scraping qualifies as “public interest” (drawing on existing UK GDPR research exemption criteria) and accompanying safeguards, including data minimisation requirements, Data Protection Impact Assessment obligations and prohibited uses of scraped data. The Data (Use and Access) Act 2025 could be amended to provide for a safe harbour from contractual liability under platform terms of service for public interest research which uses scraping. The Online Safety Act provides an alternative legislative vehicle, but this would likely only provide that protection in relation to research into ‘online safety matters’, which is only a subset of public interest research generally.

Other relevant legal provisions

A further legal instrument is relevant to the UK scraping landscape, though it falls outside the primary focus of the Taskforce’s recommendations. The Computer Misuse Act 1990 criminalises unauthorised access to computer systems and remains technically applicable to scraping in cases where a platform could characterise access as unauthorised (for example, where a researcher has bypassed technical controls or continues to scrape

after receiving a cease-and-desist notice). While the UK Government announced plans in December 2025 to reform the Act to introduce a statutory defence for cybersecurity researchers, no equivalent protection for social media researchers conducting public interest web scraping has been proposed.



3.3 Data Donation

3.3.1 What is data donation?

Data donation is broadly defined as any user-mediated pathway through which personal platform data is made available for research. One prominent definition was offered by Skatova, Ng and Goulding (2014), whose work outlined that individuals could voluntarily share their personal data with researchers for purposes they perceived as meaningful or beneficial to the public good. Building on this, Boeschoten et al. (2022) developed a practical framework for privacy-preserving digital trace data collection through data donation, providing methodological grounding for how such donation can be implemented while protecting participant privacy.

The most familiar route for data donation is a self-service export, where a user requests data from a platform and then transfers some or all of it to a research service. For example, individuals may request their personal data from a platform (typically through a “download your data” mechanism), download that data package to their own device and then choose to share some or all of it with a research team. However, this is not the only possible route. Donation may also occur through direct data transfers, portability APIs, or other mechanisms, providing the participant understands and controls the act of donation. This process is grounded in data portability rights under Article 20 of the UK GDPR. It entitles individuals to receive certain personal data in a structured, commonly used and machine-readable format and in some circumstances, to request its transmission to another controller (SDDS, 2025). Subsequent methodological work has built on this foundation, developing frameworks for integrating donated digital trace data with other individual-level sources such as surveys and experimental measures (Ohme et al., 2024).

As noted above, data donation can occur in many forms. As well as the common self-service model, it is technically feasible for data to be donated via: (a) direct platform-to-researcher transfers, in which data does not pass through participants’ devices; (b) continuous or longitudinal donation, where participants consent to ongoing data access over a defined period; and (c) selective donation, where individuals choose which data categories to share rather than donating entire packages (SDDS, 2025). At present, however, these more advanced forms of data donation are routinely held back by legal ambiguity that legislation has yet to resolve. The 2025

report prepared by Ofcom under Section 162 of the Online Safety Act recognises data donation as one possible avenue for generating user-level evidence where other access routes are limited, positioning it as one component of a wider, mixed-methods approach to platform research access (Ofcom, 2025).

3.3.2 Why is data donation needed?

Data donation is an important mechanism for understanding how individuals interact with and use social media, including what they encounter online, which can help inform research on whether and how it can cause harm. While platform-facing research APIs can provide access to aggregated, population-level data about content and behaviour, and scraping can capture publicly visible online content, data donation can reveal the unique configurations of information that individual users actually encounter, or behaviours they engage in, in their digital environments. This provides insight into not only what individuals see, but also how algorithms tailor content to individual preferences and behaviours.

Data donation offers an important counterpoint to many of the fragmented, incomplete and macro-level insights available from other data access techniques. It is also underpinned by personal data rights, meaning that when individuals exercise their right to data portability and choose to share that data with researchers, the legal basis for processing is clear and well-established. This makes data donation considerably more straightforward to pursue than web scraping or direct API access from a governance perspective, and contributes to its attractiveness as a research methodology for institutions with limited legal capacity.

The strengths of data donation have been widely recognised. Its key methodological appeal lies in the depth and granularity of digital trace data it can yield for a single user and the capacity for data linkage with other individual-level sources such as survey responses, cohort datasets, or experimental measures (Ohme et al., 2024). This individual-level mix is especially powerful when donated digital trace data is linked to longitudinal cohort studies and could enable rich, multi-layered analysis of digital behaviour and its consequences over time.

Data donation also has the capacity to help researchers look back in time. Unlike other data access mechanisms that offer snapshots of time, data donation typically provides access to historical behavioural records, enabling retrospective evaluation of policy changes or past exposure to online content (Ohme et al., 2024).

3.3.3 Technical approaches and infrastructure

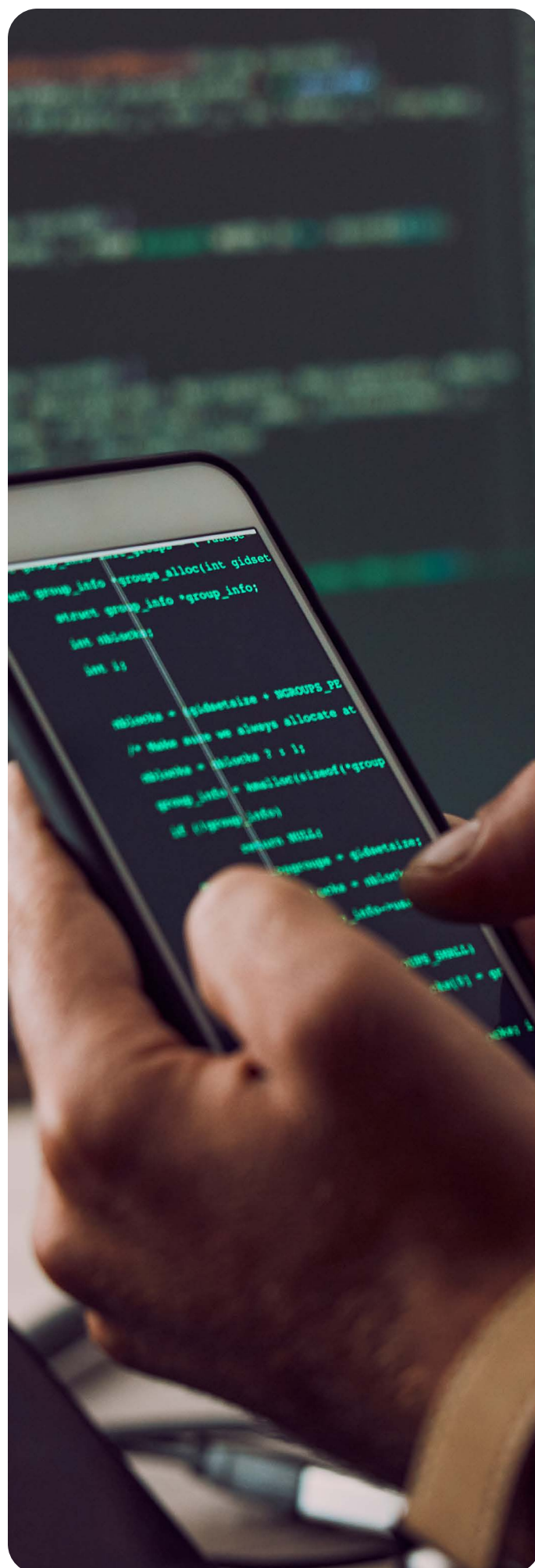
Data donation studies typically involve several technical stages: participant recruitment, data request and download from the platform, data transfer to the research team, processing and parsing of the data package, analysis and secure disposal. The architecture of the donation workflow has significant implications for privacy, data quality and participant burden and different technical approaches reflect different philosophical positions on where and how those risks should be managed.

Centralised infrastructure with controlled access

One approach retains data in a potentially identifiable state within a secure centralised environment, applying privacy and governance protections at multiple points throughout the research lifecycle. These safeguards cover the data itself, the researchers who access it, the settings in which analysis takes place, the projects approved and the outputs released. This model, exemplified by initiatives such as the Smart Data Donation Service (SDDS), enables richer and more complete datasets to be held and analysed under controlled conditions (SDDS, 2025). Rather than always minimising data at the point of collection, it manages risk through governance frameworks applied throughout the research lifecycle. This approach is particularly well suited to studies requiring longitudinal analysis, data linkage across multiple sources, or research questions that cannot be answered from minimised datasets alone.

Decentralised, client-side approaches

Another approach, associated with the PORT software tool developed by Boeschoten et al. (2023), decentralises data collection by processing participants' data on their own device. In the PORT workflow, only the specific data features of interest to the researcher are extracted and the participant reviews these features before deciding whether to consent to share them. This client-side approach supports data minimisation under UK GDPR and enables the researcher to collect exactly the digital traces needed to answer their research question. PORT is open-source and configurable for different platforms and research designs. This approach is particularly well suited to studies where minimising the volume of personal data transferred to researchers is a priority. However, where exports are held on a participant's own device, this route



may create safeguarding risks, particularly for children, vulnerable participants, or people whose devices may be monitored by others. There are also challenges to this approach as researchers are often not able to access information from platforms about the range and structure of available data, making this approach difficult in practice.

The choice between decentralised and centralised models depends on the research question, the target population, the sensitivity of the data involved, the technical capacity of the research team and the governance infrastructure available. A well-designed UK data access regime should support both models and provide researchers with clear guidance on when each is appropriate.

API-based donation

One mechanism for implementing both centralised and decentralised approaches involves API-based donation, in effect a technical implementation of the data portability right discussed in section 3.3.4, where platforms provide an API enabling direct or near-direct data transfer to a researcher-controlled system. Where API access is available, the manual download-and-upload process is eliminated, reducing participant burden and data integrity risks. Verified examples include TikTok and Google (responsible for YouTube data access), though even where such APIs exist, access and usability may remain constrained in practice due to complex approval and implementation processes (Google, n.d.; TikTok for Developers, n.d.). Taskforce evidence found that API access requests to TikTok have previously failed to be actioned without explanation or recourse (SDDS, 2025). Whilst Google does not provide API access for accounts held by under 18s, inhibiting the study of adolescents.

Data enrichment

A further significant technical challenge concerns data enrichment, or “hydration” (see Key Terms and Definitions section in the Appendix). Donated datasets frequently contain only minimal traces, such as lists of URLs visited or content identifiers rather than full content, meaning that enrichment is often a necessary step before analysis can begin. The legal implications of this process are discussed in section 3.3.4.

3.3.4 Current legal foundations, implications and recommendations

The legal foundation for data donation in the UK sits primarily at the intersection of data protection law and online safety legislation. Three legal instruments shape this environment, each playing a different role:

- a. The right to data portability under the UK GDPR provides the existing legal foundation on which donation rests, but it was not designed with research in mind and leaves several ambiguities unresolved.
- b. The Data (Use and Access) Act 2025 offers the most direct vehicle for strengthening that foundation, through its Smart Data framework (which allows the government to require data holders to share customer data in a standardised format) and its digital verification provisions.
- c. The Online Safety Act 2023 gives the government the power to create a statutory researcher access system, but that power has not yet been used, so no such system currently exists.

These instruments are not alternatives but potentially complementary parts of a single architecture and data donation infrastructure should ultimately be designed to connect with whatever researcher access system emerges under the OSA. For example, sharing the same accreditation and governance arrangements, instead of developing them separately. The sections below examine each legal foundation in turn, identifying ambiguities and opportunities, the implications for researchers and the specific reforms through which it could support data donation research.

UK GDPR and Data Portability

Data donation in the UK is grounded in the right to data portability under Article 20 of the UK GDPR. This right entitles individuals to receive their personal data from a controller relying on either the ‘consent’ or ‘contract’ legal bases in a structured, machine-readable format and to request direct transmission to another controller (SDDS, 2025). In principle, a user who wishes to donate their social media data to a researcher can exercise this right to obtain a data package and authorise its transfer to the research team. Under Article 12(3) of the UK GDPR, controllers must respond to data portability requests without undue delay and within one month of receipt, extendable by a further two months for



complex or numerous requests. However, the time taken to receive data packages varies considerably and, in the case of TikTok, data provision when provided lasts just four days before it is deleted, making this method logistically complex to deliver within large-scale settings (e.g., cohort studies).

In practice, Article 20 creates several ambiguities for donation use cases. Controllers face a lack of legislative clarity regarding what constitutes sufficient proof of identity and how to formally accept a third party acting on a data subject's behalf; no single low-friction, standardised identification or delegation flow is currently recognised in a cross-sector manner (SDDS, 2025). The ICO's guidance confirms that portability outputs must be machine-readable and interoperable and that they are distinct from Subject Access Request (SAR) outputs, yet persistent confusion between the two limits the use-value of data exports for donation pipelines (ICO, n.d.-a).

A further legal ambiguity concerns data enrichment. Donated datasets frequently contain only minimal traces, for example lists of URLs visited or content identifiers rather than full content. To conduct meaningful analysis, researchers typically need to retrieve additional contextual information about these items from the original platform, a process known as enrichment or hydration. This enrichment process typically involves scraping additional content from platform systems (see section 3.2 for a full discussion of the legal landscape surrounding web scraping), which may expose researchers to legal risk under platform terms of service, even where the underlying donation was entirely lawful (SDDS, 2025). Data portability rights as currently implemented do not guarantee that donated data will be analytically

usable without further enrichment (the additional step of retrieving contextual information from the original platform) and the legal basis for that process remains uncertain.

Implications for researchers

These ambiguities have concrete effects in practice. The absence of a standardised identification and delegation mechanism means that platforms can block or reject the technical tools researchers use to receive donated data, such as local processing software, without explanation or right of appeal (SDDS, 2025).

Article 20 of the UK GDPR gives individuals the right to receive personal data in a machine-readable format. It also gives individuals the right to request direct transmission of this data to another controller when this is technically feasible. In practice, this latter part of the right may often be overlooked. This leaves many studies dependent on manual download and upload workflows, which increase participant burden, weaken data integrity, and may compromise data protection by placing sensitive data on personal devices.

The quality and completeness of portability packages creates another constraint: packages are frequently incomplete, poorly structured or conflated with Subject Access Requests (SAR) outputs and the legal basis for enrichment remains uncertain (SDDS, 2025; Boeschoten et al., 2022). Platforms can also alter the donation process or the contents of a package without notice, disrupting in this way, ongoing studies and the research infrastructure that is built around them.

Recommendations for Government

Outline standardised processes for the identification and delegation of data donation requests.

Under the DUAA's digital verification services framework, the Government should act to mandate specific data donation consent flows. This should specify approved identity assurance provisions and introduce a recognised mechanism for delegated authorisation, enabling researchers to act formally as authorised recipients of a participant's data portability request.

Clarify data donation rules in relation to:

- a. The right to third-party transfer of donated data under the UK GDPR

The government should clarify individuals' ability to request their platform data is transferred directly to a named third party for public interest research. The government should direct the ICO to issue joint guidance on how Article 20 direct transmission should operate in public interest donation studies, or direct them to issue a GDPR code of conduct under Article 40. In this context, guidance should confirm how a participant can identify an approved research service as a recipient of a transfer, how authorisation should be recorded and how data could be transferred. A participant should be able to initiate requests themselves; select trusted third parties as recipients; and complete any authorisation step needed to complete a transfer. It is critical that this process is seamless and straightforward for participants to use. Where possible, API-based data donation should be enabled for large online platforms, so that transfers can operate reliably and at scale.

- b. The legal basis for enriching donated data

Direct the ICO to issue specific guidance on whether and under what conditions data enrichment constitutes a lawful continuation of the original donation and what safeguards are required. In the absence of such guidance, researchers are exposed to legal risk for activity that is essential to making donated data analytically usable.

- c. Data donation portability standards

Direct the ICO to issue authoritative guidance specifying minimum standards for data portability outputs, clearly distinguishing them from subject access request (SAR) outputs. Guidance should require: standardised documentation of the data provided, including how it was generated; a standardised, machine-readable data format; and at

least two months' notice to researchers of any change to the donation process or the data provided. This requires no new legislation and falls within existing ICO powers (ICO, n.d.-a; SDDS, 2025).

The Data (Use and Access) Act 2025 (DUAA)

The DUAA provided a data portability framework for "Smart Data schemes". These schemes are sector-specific, government-backed arrangements under which designated data holders may be required to share customer data with authorised third-party providers on a customer's request, through standardised APIs (Data (Use and Access) Act 2025). The Smart Data framework follows the Open Banking model established by the Competition and Markets Authority in 2016 and is intended to extend it across sectors.

In March 2026 the government published its Smart Data Strategy (Smart Data 2035), setting a target of five or more active schemes by 2030 and twenty or more by 2035, and naming an initial set of priority sectors including banking, financial services, energy, property, transport, telecommunications and digital markets (Department for Business and Trade, 2026). No scheme is yet operational in any of these sectors, with the first secondary legislation expected in late 2026. Social media is not currently among the named sectors. The Act also introduces a framework for digital verification services, which could help address the delegated authorisation problem by providing standardised identity assurance mechanisms.

The DUAA could be relevant to data donation if it were to lead a Smart Data scheme for digital platforms. If applied to social media, this potentially provides a route for individuals to share data that does not involve manual download/upload steps, and concomitant safeguarding and privacy risks: The Act's legal architecture could, in principle, be extended to include social media platforms as data holders subject to Smart Data obligations.

Implications for researchers

The absence of a standardised third party transfer route is a major constraint on data donation-based research. Researchers depend on platform tools that may be inconsistent, or liable to change without notice. Participants may be asked to manage large data files and hold sensitive records themselves. This may impact participation, sampling bias, data quality, and make donation harder to use in studies that involve children or vulnerable individuals. Smart Data schemes represent the most structurally significant reform available to donation research

(mandatory, standardised, API-based portability).

Recommendations for Government

Expand and normalise Smart Data schemes for social media platforms to enable data donation, building on work already underway through the Smart Data Donation Service.

Use the DUAA 2025 framework to designate major social media platforms as data holders subject to Smart Data obligations, requiring them to provide standardised, API-based portability mechanisms for public interest research. This is the most structurally significant reform available within the existing legislative architecture. While it can be delivered through secondary legislation under powers already granted by the DUAA, it represents a major structural undertaking requiring significant technical and institutional development from Government. For this reason, it is best understood as a longer-term reform, despite being deliverable through powers already provided by the DUAA.

This integration should be designed to leverage the role of a hybrid-repository social media data intermediary. A Smart Data scheme would impose obligations on platforms as data holders, but it would also create implementation choices regarding the management of access. The proposed intermediary would be the natural body to take on these functions, providing governance around access including researcher accreditation, project review, data stewardship, enrichment facilitation and audit.

Enable continuous, as opposed to only momentary, data donation

Introduce secondary legislation under the DUAA to allow data donation to operate not only as a single, momentary transfer but as a continuous mechanism, so that participants can consent to ongoing access for a defined future period and to selective sharing of specified data categories. A comparable approach exists under Article 6(9) of the EU's Digital Markets Act, which requires large gatekeeper platforms to provide continuous, real-time data portability (SDDS, 2025).

The Online Safety Act 2023

The OSA brought researcher access to platform data within the scope of UK online safety policy. Under section 162, Ofcom was required to produce a report on researchers' access to information from regulated online services, which it published in July 2025, fulfilling its duty under that section. The report explicitly discusses data donation as one possible avenue for generating user-level evidence, whilst acknowledging that it faces significant practical

barriers (Ofcom, 2025). The power to establish a research access framework rests with the Secretary of State under the Data (Use and Access) Act 2025, which amended the OSA to allow regulations requiring regulated services to provide information for independent research, although this framework has yet to be brought forward. The relationship between any future research access framework and data donation infrastructure is an important area for development: ideally, the two should be designed as complementary rather than parallel systems.

Implications for researchers

There is currently no enforceable right of access for researchers under the OSA, either for donation-based or other research methods. The recognition of data donation as a valid avenue in the section 162 report is significant in policy terms, but it does not resolve the practical and legal barriers researchers face. Where platforms refuse portability requests or reject API applications without explanation, there is currently no dedicated route for escalation. Nor is there any clear obligation on platforms to justify the refusal of a portability or API access request, or to provide a route of appeal. This is a structural legislative gap rather than a matter of platform discretion: no existing law (UK GDPR, OSA, DUAA) creates an enforceable right of access or a duty to give reasons for refusal. Researchers and participants therefore have no formal recourse when requests are rejected (SDDS, 2025; Dommett et al., 2025; Ofcom, 2025).

Recommendations for Government

Establish a researcher complaints mechanism to address the failure of data donation processes and grant ICO powers to investigate and resolve complaints.

Create a clear and accessible route by which researchers and participants can escalate portability failures and non-compliance to the ICO, with defined response timelines. The UK GDPR already gives individuals a route to complain in relation to their portability requests, though this offers limited practical redress in most cases. A new statutory duty requiring organisations to operate a formal complaints-handling process took effect in June 2026, but this addresses complaints from data subjects about data handling generally, not the specific problem of researchers being unable to challenge a rejected donation or a data quality issue (ICO, 2026).

To resolve these issues, the ICO should be given powers to investigate and resolve complaints from

researchers as well as data subjects. Achieving this would require primary legislation amending the Data Protection Act 2018, since only data subjects, not researchers, currently have standing to complain under section 165 of that Act. Secondary legislation under the DUAA cannot extend standing in this way, as this would go beyond a technical adjustment to an existing right (SDDS, 2025).

Require platforms to provide data transparency and change-log documentation.

Platforms should be obliged to publish and maintain documentation of changes to data export packages, including version logs and advance notice to researchers, to prevent silent disruption to ongoing studies. This obligation could be introduced through secondary legislation under the DUAA.



3.4 Data Intermediaries: A new hybrid-repository social media data intermediary

The UK currently has no independent body governing researcher access to social media platform data. This report uses the term “social media data intermediary” throughout to refer to such a body, which would be specifically designed and mandated to govern researcher access. Section 154A of the OSA 2023, inserted by the Data (Use and Access) Act 2025, outlines powers for the Secretary of State to enable access to information for research about online safety matters and could be used to establish such an organisation. An intermediary could enable access to both publicly accessible and non-public social media platform data relating to online safety matters, operating under statutory oversight and without a commercial interest in the data it handles. Any future social media intermediary should be designed to serve all public-interest researchers, including academic researchers, civil society organisations, journalists and other parties conducting research independently of the platforms whose data is being studied.

The case for a social media data intermediary in the UK builds on Ofcom’s 2025 report on researcher access which recommended an independent intermediary. This identifies three possible models for how such a body could operate in the UK: a repository model, in which platforms provide data to a secure environment managed by the intermediary; a direct access intermediary model, in which the intermediary manages accreditation and requests while data remains with platforms; and a notice to service intermediary model, in which researchers submit proposals to the intermediary which then liaises with services and both parties assess the request (Ofcom, 2025). As noted in section 2.3, Ofcom concludes that no single model is likely to meet the full range of researcher needs and that a layered approach combining elements of these models is likely to be the most effective.

3.4.1 What is a Data Intermediary?

A data intermediary is an independent body that mediates access to data on behalf of a defined group of users, operating under statutory or regulatory oversight and without itself having a commercial interest in the data it handles. The concept is well established in data governance literature and has been formalised in legislation: the EU's Data Governance Act (2022) explicitly regulates data intermediary services as a distinct category of infrastructure, defining them as providers that establish relationships between data holders and data users through technical, legal and organisational means.

The Centre for Data Ethics and Innovation (CDEI) 2021 report on data intermediaries, commissioned by the Department for Culture, Media and Sport (DCMS) identifies seven distinct types of intermediary operating across the UK economy: data trusts, data exchanges, personal information management systems, industrial data platforms, data custodians, data cooperatives and trusted third parties (CDEI, 2021). While these types differ in governance structure and technical design, they share a common function: absorbing the legal, technical and organisational costs of data sharing that would otherwise fall on individual data holders and users. The same report identifies the barriers that intermediaries are best placed to address, including lack of incentives to share data, legal and regulatory risk, the costs of data access and missed opportunities to use data in the public interest. These map directly onto the barriers facing social media researchers documented in Chapters 2 and 3 of this report. The CDEI's case studies are drawn from health, finance and transport showing intermediary models to be successfully applied across sectors where data is sensitive, commercially valuable and legally complex. These cases provide precedents for a social media data intermediary.

In June 2026 the Government opened a consultation, Empowering people through data intermediaries, looking at potential measures that could be taken to drive growth in the intermediaries market within the UK, giving individuals true power to control their data while ensuring that safety, trust and consent are all baked into the process (DSIT, 2026). This work demonstrates the well established logic for data intermediaries and the recurring challenges that can face such bodies. A social media data intermediary is well placed to draw lessons from this inquiry and the numerous data intermediaries that already exist in the UK.

3.4.2 Why is a hybrid repository social media data intermediary body needed in the UK?

As documented through this report, the current state of access to social media data is fragmented and rapidly evolving, creating challenges for researchers and research organisations. The creation of a social media data intermediary would act as a strategic hub, forging new means of data access for public interest research, curating expertise, and providing technical infrastructure to facilitate world-leading research.

Forging new means of data access

Social media platforms hold data of high public value that is essential for understanding and tackling online harms. At present, researchers lack reliable, lawful and accessible ways to access this data. An intermediary body can play a critical role in building stronger, effective partnerships with platforms, lessening the burden on platforms by consolidating what is currently a one-to-many relationship between platforms and researchers into a single institutional interlocutor. Platforms can have legitimate data protection and security concerns in providing external researchers access to often sensitive and private data. A centralised intermediary can mitigate these concerns in the most cost-effective way for the wider research community, providing a clearer and predictable interface for managing research access.

An intermediary is also able to intervene when voluntary initiatives fail to deliver required data in a way impossible for individual researchers. Should the Secretary of State exercise the power under section 154A of the OSA 2023, inserted by the Data (Use and Access) Act 2025, to create a researcher access framework, an intermediary body, underpinned by regulatory authority, could be designated to compel data access on this statutory basis.

The limitations of the current patchwork of voluntary and platform-controlled arrangements are well documented (Knight-Georgetown Institute, 2025a), and a UK intermediary would offer a more stable and equitable basis for access. An intermediary is also positioned to counter technical (see below) and non-technical barriers to data access, serving as a single source of expertise able to identify and surface issues that can prevent effective data access, even where technical infrastructure is available.

Curating expertise

As acknowledged by Ofcom, data access can not be secured by one mechanism alone, creating a need for a central source of expertise that can help researchers to navigate alternative data access avenues to enable public interest research. The Taskforce has found individual researchers and research institutions in a range of sectors to face data access challenges, creating barriers to research. An intermediary body would provide UK researchers with a roadmap for navigating the legal, financial and technical barriers documented in previous chapters. A centralised, independently governed intermediary would be able to offer training and guidance, working to promote the highest standards of trustworthy, rigorous and legally compliant research. It can also offer expertise on long-term data stewardship, preservation, and ongoing reuse/re-use protocols, ensuring that reliable and reproducible data access is secured.

Technical infrastructure for social media data

The UK has an established track record for enabling access to a range of sensitive data forms. Initiatives such as the Smart Data Strategy, the ONS Secure Research Service National Data Library and Administrative Data Research UK (and those detailed below) reveal how a variety of data can be curated and processed to enable research. Social media data, however, poses a range of particular challenges.

Social media data is heterogeneous, dynamic and drawn from an increasingly API-depleted access landscape: datasets collected through different routes arrive in different formats, with inconsistent metadata, variable completeness and no standardised structure. Preparing such data for research use requires significant data engineering capacity, including wrangling and cleaning pipelines, enrichment and hydration workflows and quality assurance processes capable of detecting gaps and biases at scale. This is not a task that can be absorbed by individual research teams: it requires sustained, dedicated technical infrastructure that goes well beyond what any single institution or project can maintain.

An intermediary could develop the expertise to navigate such complexities and, importantly - if underpinned by clear, prescriptive powers held by the regulator - would have the power required to mandate engagement and compliance from reluctant social media providers. Lessons from the EU and the implementation of the Digital Services Act have shown the need for clarity around the requirements platforms need to fulfil.

Complementing this technical infrastructure with compliance procedures such as researcher accreditation, research proposal evaluation and data protection assessments, an intermediary would ensure trustworthy, legally compliant and reproducible data access while minimising the burden on both researchers and platforms.

The practical urgency of these technical requirements is illustrated by the experience of researchers currently working on issues of the greatest public concern. Studies of adolescent mental health and social media, including work by the Born in Bradford programme and the Digital Mental Health Group at the University of Cambridge, have identified individual-level, longitudinal platform data as essential to understanding what young people are actually seeing and engaging with online, and which children are most at risk from online harms (Orben et al., 2026). Gathering data from different platforms in consistent formats is currently exceedingly challenging, and requires technical infrastructure and expertise.

As discussed in the next section, existing intermediaries within the UK show a range of tried and tested infrastructure models for delivering data access, demonstrating the capacity of these bodies to solve technical and non-technical barriers to data access.

3.4.3 Existing intermediary models in the UK

The UK Government has recognised since at least 2021 that data intermediaries are a necessary part of the national data infrastructure. The CDEI's report for DCMS concluded that intermediaries are already helping to overcome barriers to data sharing by providing governance models and technical infrastructure (CDEI, 2021). These findings show the value of using an intermediary model to secure social media data access. Table 2 below summarises the principal existing intermediary bodies in the UK research landscape. Intermediaries can be designed to perform a range of different functions including enabling access (via accreditation), serving as a data custodian by curating and securely storing data, and as a centre of expertise, ensuring compliance with legal and technical standards.

Table 2: Examples of data intermediary bodies in the UK

Existing UK data intermediary bodies	Type of intermediary	Type of data held
<u>UK Data Archive</u>	Data custodian / trusted research environment	Social science, humanities and arts research data, including survey data, qualitative data and longitudinal datasets.
<u>ONS Secure Research Service</u>	Data custodian / trusted research environment	Government administrative and statistical data, including census, health and economic data.
<u>Smart Data Research UK (SDRUK)</u>	Federated data infrastructure / building a network of Trusted Research Environments (TREs)	Smart data generated through everyday digital interactions, including financial transactions, geospatial data and digital behavioural data, accessed through secure environments following the Five Safes model . Delivered through a strategic hub embedded in ESRC and six federated data services based at leading UK universities, with a £49.3 million UKRI investment running to 2029.
<u>SAIL Databank</u>	Data custodian / trusted research environment	Health and administrative data relating to the population of Wales, including NHS records and social care data.
<u>Administrative Data Research UK (ADR UK)</u>	Trusted third party / data custodian	Public sector administrative data across all four UK nations, including data from government departments, the NHS and local authorities.
<u>Health Data Research UK (HDR UK)</u>	Trusted third party / data custodian	Health and biomedical research data, accessed via the Health Data Research Innovation Gateway.
<u>Smart Data Donation Service (SDDS)</u>	Data donation service / trusted research infrastructure	Citizen-donated digital trace data held in secure research environments at the University of York, initially focusing on social media and video game domains, with an initial target of 90,000 video game data donors and planned expansion to social media. The first national research infrastructure built around the right to data portability, aiming to facilitate research into online safety, mental health, media literacy and disinformation. Funded at £7.6 million as part of SDR UK.

Note: These intermediary type classifications drawn from Centre for Data Ethics and Innovation (2021). **Unlocking the value of data: Exploring the role of data intermediaries.** DCMS.SDR UK and SDDS type classifications reflect their own published descriptions as they postdate the CDEI (2021) taxonomy.

The UK's experience in establishing intermediary bodies provides a range of expertise that could inform the creation of a social media data intermediary. For example, through the Office of National Statistics, the UK has developed the 'Five Safes' framework that serves as best practice in data protection whilst fulfilling the demands of open science and transparency. This model has been adopted by Health Data Research-UK (HDR-UK) and the National Institute for Health Research Design Service (NIHR).

Previous examples also offer lessons for establishing an effective intermediary. For example, the SAIL Databank, established at Swansea University in 2007, demonstrates that a privacy-by-design model can operate effectively over the long term, combining technical safeguards, independent project oversight and active public engagement across nearly two decades of operation. Its rapid response during the COVID-19 pandemic illustrated the value of having mature infrastructure in place before a crisis emerges (SAIL Databank, n.d.).

Similarly, Administrative Data Research UK (ADR UK), established in 2018 and funded by ESRC as part of UKRI, demonstrates that a statutorily grounded intermediary can build sustained trust between data holders and researchers, reduce per-project compliance burden and deliver policy-relevant research at scale. The Digital Economy Act 2017 provides its statutory foundation, and a £168 million UKRI investment confirmed in July 2025 reflects the long-term commitment required to sustain infrastructure of this kind (Administrative Data Research UK, n.d.).

3.4.4 Types of social media data access intermediary possible in the UK

The Ofcom report outlined three forms of data access:

- **Direct access intermediary:** Researchers could request data with an intermediary facilitating secure access. In this model, services could retain responsibility for hosting and providing data while the intermediary maintains the interface by which researchers request access.
- **Notice to service intermediary:** Researchers could apply for accreditation and request access to specific datasets via the intermediary. This could include data that would not be accessible in direct access

models. The intermediary would review and refuse or approve access. Services would then be required to provide access to the approved data.

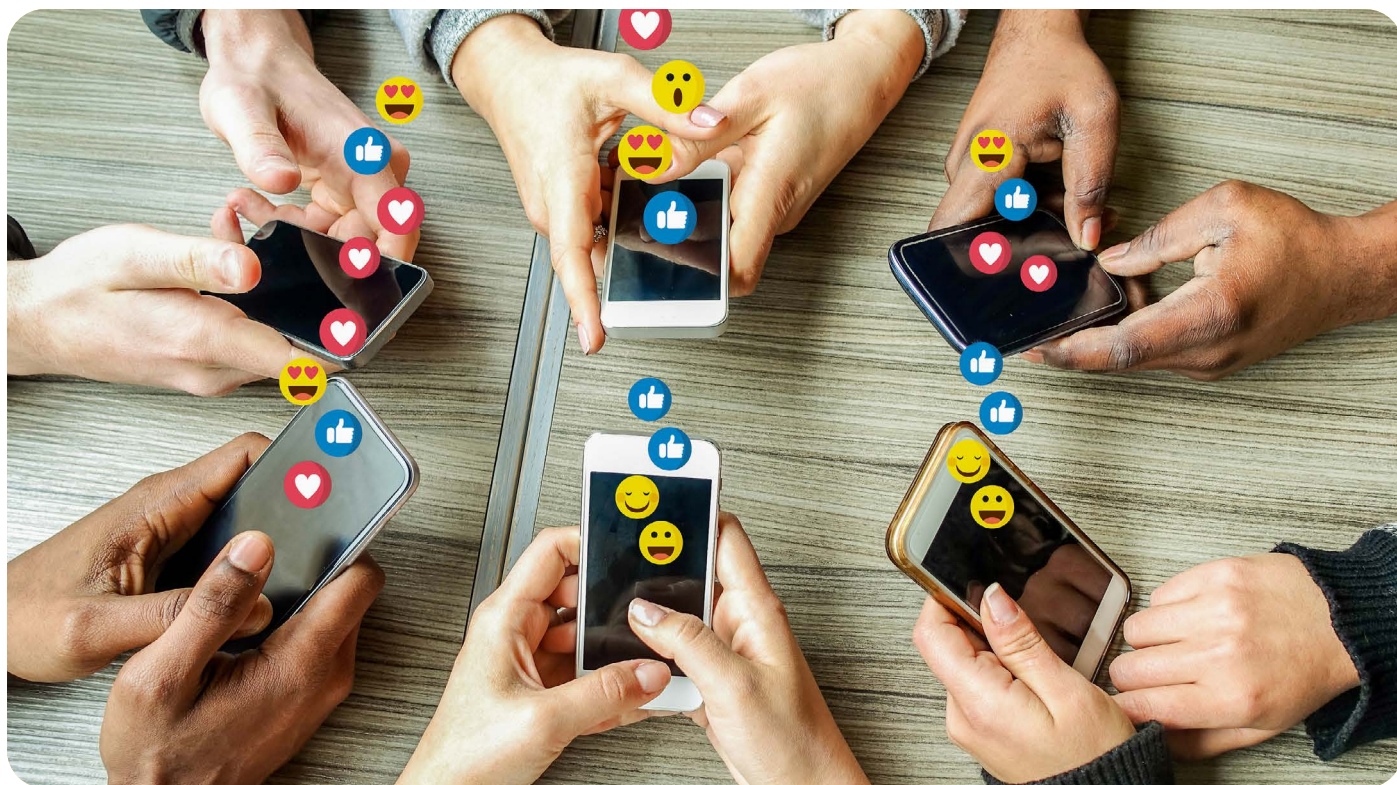
- **Repository intermediary:** The intermediary could itself provide direct access to data, by providing an interface for data access and/or hosting the data itself and taking responsibility for data governance. This could also include data that would not be accessible in direct access models.

Whilst each form of intermediary has different advantages, previous challenges experienced with platform-provided data around accuracy and completeness lead us to focus on the second two models as the most promising route for data access via an intermediary.

Notice to service and repository intermediary models provide different advantages for researchers seeking data access. The notice to service model mirrors the approach taken under the Digital Services Act and enables researchers to request particular data sets. In theory, this approach provides researchers with access to a wide range of data relating to online harm and would align the UK with regulatory practice in the EU. However, as outlined above, implementation of this model has been challenging and researchers have reported barriers in accessing data via this mechanism. This experience shows the potential for procedural and legal barriers to hinder access and indicates the value of a single body able to support researchers and broker productive relationships with platforms to secure effective data access.

In contrast, repository intermediary bodies provide access to datasets held and curated by the intermediary body, providing infrastructure for researchers to access data through secure systems. Such a model tackles inequities in researchers' access to suitable secure infrastructure and can ensure high standards of data cleaning and management. By their nature, repository intermediaries focus on specific datasets and hence do not provide the means for researchers to request their own datasets, requiring robust processes for identifying datasets of primary interest and need.

The Taskforce supports the creation of a hybrid-repository model that combines notice to service and repository components. This would more closely align the UK with practice in the EU and facilitate a process for researcher-led data requests, whilst also ensuring a base level of data access by providing high quality curated datasets on topics of the most pressing social interest via a repository model. Datasets held by a repository could include core



datasets (i.e. those requested by the intermediary body itself) and datasets requested via the Notice to Service model. One key advantage of a hybrid-repository model is the potential to link core datasets with other national data infrastructure, helping to facilitate cross-cutting research on, for example, the relationship between social media and mental health.

3.4.5 International examples of intermediary models

The UK is not the only country globally that tackles the challenge of building sustainable, scalable infrastructure for social media data research. Several international models demonstrate what shared, independently governed research infrastructure can achieve when properly resourced and institutionally grounded. These examples are instructive not only as precedents but as illustrations of the different forms an intermediary-type structure can take.

GESIS (Germany)

In Germany research infrastructure organisations such as GESIS (the Leibniz Institute for the Social Sciences) have invested in shared technical and legal resources that allow researchers across multiple institutions to conduct data collection under a common compliance framework. That includes the creation of virtual research environments in which sensitive datasets held by researchers can be

uploaded into GESIS' systems to then be accessed and analysed securely, without data being downloaded to researchers' own computers. It includes the development of research tools such as data donation facilities to allow researchers to upload data provided directly by citizens in structured and privacy-preserving modalities. Computational research capabilities and other technical infrastructure are also provided, as this would be uneconomic if produced by an individual researcher or research body. By pooling research capabilities into a centralised institution, GESIS is aiming to offer a "baselayer" of social media data infrastructure that is secure, privacy-enabling and scalable in ways that would not be possible for one-off research projects.

Social Media Archive (SOMAR)

The University of Michigan's Social Media Archive (SOMAR) provides an example of scaled data access infrastructure. It contains a variety of datasets from social media platforms including X (formerly Twitter), Facebook, Instagram and Reddit. These datasets come from partnerships with platforms, researcher deposits and SOMAR's independent active collection of data via its own infrastructure. This represents a hybrid data access infrastructure model in which researchers provide existing datasets for community use via SOMAR, active collection of social media data takes place, and access to platform-specific data is provided through partnerships with social media platforms.

In 2023, the University of Michigan's Inter-university

Consortium for Political and Social Research (ICPSR), through SOMAR, entered into a partnership with Meta to allow researchers to analyse data from the tech company, facilitated through SOMAR's Virtual Data Enclave (Meta, 2023). The university serves as an independent gatekeeper for researcher data access, as all applications for Meta's Content Library (an online repository of publicly available data) are reviewed by SOMAR. In late 2025, Meta took over the application process for this social media data access. Researchers can now decide whether to use SOMAR's secure virtual data enclave to access this data or Meta's own virtual research environment. As of early 2026, SOMAR also instituted a nominally monthly fee of less than £370 per research team to cover infrastructure costs.

Social Science One

Social Science One was a voluntary industry-academic partnership established in the United States in 2018, described and analysed in King and Persily's (2020) paper "A New Model for Industry-Academic Partnerships." Launched in the wake of the Cambridge Analytica scandal, Social Science One was intended to broker researcher access to Facebook data on elections and democratic processes, operating through an independent academic commission. Despite significant investment from philanthropic foundations and sustained effort from the academic community, the partnership encountered persistent delays, with data releases consistently falling short of what had been promised. Persily subsequently resigned from the project and the experience informed his conclusion that legal compulsion and legal immunity for data sharing were necessary conditions for any reliable researcher access regime (King & Persily, 2020). This directly informed subsequent legislative proposals including the Platform Accountability and Transparency Act in the United States.

Digital Services Act (Europe)

The EU's DSA, which applied to very large online platforms from August 2023, represents the most ambitious legislative attempt to date to mandate researcher access to social media platform data. Its provisions are instructive for the UK as a case study illustrating what statutory backing can achieve and where efforts can fall short without centralised institutional support. Article 40 of the DSA creates two distinct access routes for researchers:

- Under Article 40(12), researchers meeting independence and data security requirements can obtain access to publicly accessible

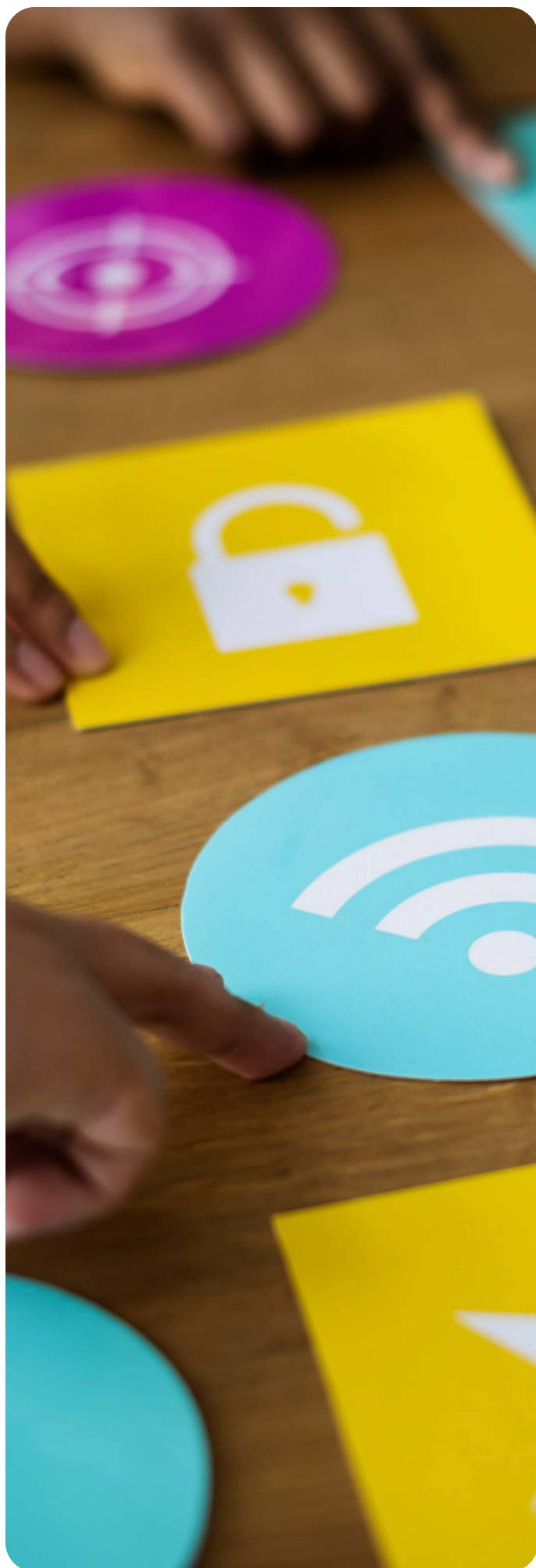
platform data by applying directly to the relevant platform or using public-interest scraping techniques, provided the research contributes to the detection, identification and understanding of systemic risks in the EU. Recent Commission enforcement action against X and AliExpress illustrates how these obligations are being applied in practice.

- Under Article 40(4), vetted researchers can apply for access to non-public data, submitted via the relevant national Digital Services Coordinator, for the same systemic risks purpose, and additionally for assessing the adequacy, efficiency and impact of platforms' risk mitigation measures (European Commission, 2022).

Both routes are deliberately focused on systemic risk as defined under EU law, a concept that does not map directly onto the categories of online safety research that a UK regime grounded in the Online Safety Act 2023 would need to serve.

The procedural architecture took several years to develop. A Delegated Act for Article 40(4) established a centralised DSA Data Access Portal as a single point of entry for researchers, platforms and national regulators. It was adopted in July 2025 (European Commission, 2025a), with the portal going live in October 2025. The portal allows researchers to make applications via a single point of entry. These applications are then initially checked by national regulators where each researcher is based or by the regulator in whose jurisdiction the company in question is headquartered. The application is then passed onto the regulator where each company is legally established for final approval. In most cases, Ireland's Digital Services Coordinator will be the approving regulator.

Early evidence from the first wave of applications highlights a number of implementation challenges. Publicly documented Article 40(4) requests suggest that the mechanism has proven more procedurally complex and administratively demanding than many researchers anticipated. For example, one documented application relating to TikTok content and electoral risks was rejected on the majority of grounds assessed, including procedural requirements (DSA Observatory, 2026). More broadly, emerging evidence suggests that significant challenges remain not only in platform compliance but also in the institutional infrastructure established to facilitate access, with researchers identifying bottlenecks, procedural complexity and uncertainty regarding implementation across the DSA governance architecture (Goanta et al., 2026). Public statements by Ireland's Digital Services Coordinator (DSC) have



emphasised data protection and security standards comparable to those used in sensitive public-sector research environments, suggesting a relatively demanding threshold for approval. This threshold is substantially higher than the governance and security requirements currently associated with most social media research projects.

An independent intermediary body bringing together researchers and platforms was envisioned under the EU Code of Practice on Disinformation. This intermediary was designed with the intention of creating robust and accessible processes for managing data access requests. Major platforms withdrew their commitment to supporting an intermediary body (DSA Observatory, 2025). In its absence, responsibility for processing applications rests with both national Digital Services Coordinators of varying capacity, as well as regulators in which each platform is legally established.

Access to public social media data under Article 40(12) has faced similar challenges. Although the EU's Digital Services Act did not lay out a procedure for accessing such data, the European Commission, which holds the legal oversight for this data access mechanism, has relied on platform-led applications and procedures to permit researchers to access publicly-available platform data.

So far, at least 50 research groups have voluntarily reported applying for Article 40(12) data access to the DSA 40 Collaboratory's tracker, of which 20 have been accepted (DSA 40 Collaboratory, n.d.). Platforms' own biannual transparency reporting under the Code of Conduct on Disinformation indicates request numbers have been higher, with YouTube and TikTok reporting acceptance rates of above 50 per cent in the second half of 2025 (DSA 40 Collaboratory, 2026). Platforms have nonetheless imposed varying degrees of gatekeeping to publicly-available data, including complex registration forms, differentiation between data access for academic researchers and civil society groups, and a series of data protection and security questions that go beyond what was outlined within Article 40(12).

The Commission's actions against AliExpress and X have been interpreted as signalling that independent public-interest access to publicly available data, including through scraping techniques, can be compatible with the objectives of Article 40(12). In AliExpress's case, this took the form of a binding commitment. In X's case, a non-compliance decision imposing a €120 million fine, which X is now challenging before the General Court of the EU in *X Corp. v European Commission* (Case T-114/26, pending). As neither the Commission's decision nor any future court judgment is yet final,

the extent to which independent, scraping-based access is protected under Article 40(12) remains to be conclusively settled and legal uncertainty and technical complexity have so far limited researchers' willingness to use such public-interest scraping techniques.

These examples illustrate that shared research infrastructure for social media data can take a range of institutional forms, from legal and regulatory coordination bodies to technical data archives to hybrid platform partnership models. These mechanisms could not have emerged from individual research projects alone: all required deliberate, sustained institutional investment and, in most cases, a degree of regulatory or legislative backing.

3.4.6 Lessons from international intermediaries

International examples of data intermediary designs offer several lessons for a new UK intermediary.

Legislative Underpinning

Pre-existing initiatives attest the importance of creating intermediary data access mechanisms underpinned by a legislative foundation. Social Science One is a useful reference point for the UK: it demonstrates that even well-resourced, well-intentioned voluntary intermediary arrangements are not a substitute for a statutory framework, and that without independent legal backing an intermediary remains dependent on the very platforms it is meant to hold to account. It is also important to note that the Social Science One initiative occurred at a point when platforms were more willing to engage with external researchers, and that such initiatives are less likely to occur in the current climate.

Effective implementation

The European experience of striving to improve data access provides a further important lesson. Legislative provision alone does not guarantee access in practice, and the UK has an opportunity to learn from early DSA implementation to design a more effective regime. One consideration worth building in from the start is flexibility for iterative and longitudinal research. The DSA requires researchers to justify their data needs upfront with precision, but social media research does not always work that way: questions evolve, unexpected patterns emerge and data needs change as a study develops. This tension has already been identified in

the literature as a potential barrier to certain types of research under the DSA framework (Botero Arcila et al., 2026). A well-designed UK regime should take this into account from the outset.

The need for coordinating infrastructure

A separate lesson from the EU experience of efforts to improve data access is how the lack of a central intermediary body has slowed public-interest social media access. Under the DSA's Article 40(4) data access provision, related to private and sensitive data, individual researchers must submit complicated, legalistic and technical applications to gain access to platform data. An early empirical study by Halil et al. (2025), conducted before the Delegated Act specifying these procedures was adopted in July 2025, found that none of the 27 Digital Services Coordinators had been able to process a request, citing legal uncertainty, missing national measures and the absence of the Delegated Act as the principal barriers. The picture has since developed, though meaningful access remains elusive: by May 2026, Digital Services Coordinators had received 49 vetted researcher applications, none of which had yet resulted in a platform receiving a request. Applications remained held at the DSC assessment stage or, in at least one documented case, were rejected before reaching the platform (European Commission, 2026a; DSA Observatory, 2026). National guidance is also now emerging: for example Ireland's Coimisiún na Meán published detailed procedural guidance on Article 40(4) applications in May 2026. The process it describes remains lengthy and demanding, requiring extensive documentation on security, confidentiality and proportionality before a request is even issued to the relevant platform. A centralised intermediary, which could meet the legitimately significant technical, privacy and security requirements for such social media research projects, could reduce the technical lift required in accessing such social media data.

Under the EU's Article 40.12 data access provisions, which oversee publicly-accessible social media data, the DSA assigns the vetting and approval of research applications to the platforms themselves, rather than to the Commission or national regulators. This structure entails researchers submitting research project outlines to individual social media platforms before accessing such data. As of May 2026, platforms' vetting decisions have been inconsistent, with outcomes varying widely between platforms rather than reflecting any clear or predictable standard (Social Platforms Data Access Taskforce, 2026c). Rejections sometimes reflect legitimate privacy and security concerns, or research questions

falling outside the DSA's scope, but also appear to reflect platform-side inconsistency rather than the merits of individual applications. There have also been significant delays in researchers receiving responses and platforms have had to be repeatedly chased, having a particular impact on time sensitive research. An intermediary body could mitigate such variability by providing a one-stop-shop for privacy and security protocols to facilitate the fast-tracking of legitimate data access. It could also build long-term relationships with platforms to enable the appropriate level of access without overwhelming these companies with ad hoc research requests that can take months to respond to.

Definitional precision

The EU case also shows the need for clarity around the meaning of key terminology integral to data access. Challenges around the interpretation of what is captured by the term 'systemic risk', for example indicate the need for clear guidance on interpretation by intermediaries and their underpinning regulatory partners to ensure that platforms are unable to utilise a narrow definition to reject valid research.

There is also some evidence that the UK is already benefiting indirectly from EU regulation through what has been described as a Brussels Effect: platform practices developed to comply with the DSA being extended to the UK market without any domestic legal requirement to do so (Santini et al., 2026). This is a fragile and uneven benefit. It applies to only a small number of platforms, it is not guaranteed and it can be withdrawn at any point. The UK cannot rely on regulatory spillover as a substitute for its own framework.

Scope of regulation

The DSA is focused on access to data from the very largest online platforms. Whilst enabling scrutiny of practices on these platforms it means that data is not available from smaller, potentially high risk services. The Online Safety Act contains provision to capture medium or smaller online services, particularly where risk is high, offering a route for more expansive insight into the study of online harm.

3.4.7 Recommendations for Government

Establish a statutory hybrid-repository social media data intermediary

A statutory intermediary is the most significant structural reform available. It would replace the current fragmented, platform-controlled access landscape with a single governed institution that centralises accreditation, manages data security and provides platforms with a consistent and predictable counterpart, reducing their compliance burden while expanding researcher access. The international evidence reviewed in section 3.4.3 confirms that voluntary arrangements are not a reliable substitute, and that statutory backing is a necessary condition for any durable access regime. For child online safety research in particular, an intermediary with access to longitudinal and individual-level data would provide the evidence base the OSA's effectiveness ultimately depends upon. In recommending the creation of an intermediary, we do not propose this is the only data access mechanism. Rather, mirroring Ofcom's recommendation for a layered and flexible approach, we recommend an intermediary body would, over time, gather and share expertise on different routes for access, but the mechanisms themselves



would be separate. This approach is essential for ensuring access across academia and civil society.

Learn lessons from existing data infrastructure and from international social media data access efforts to inform hybrid-repository intermediary body design

A social media data intermediary presents a complex prospect, but there are clear foundations on which to build such infrastructure. Existing data archives and intermediaries offer a range of technical and legal expertise, whilst precedent set by the DSA and by organisations such as GESIS, provide a range of domain specific guidance that could be adapted for the UK context. As such, urgent efforts should be directed to learning lessons from existing initiatives to ensure the UK benefits from best practice and avoids previous pitfalls. Specific attention should be paid to insights regarding application management and access assessments, platform relationships and data access notices, and data infrastructure and stewardship.

Identify the legal foundation and governance arrangements for a hybrid-repository social media data intermediary body.

A range of actors could take responsibility for leading a social media intermediary, including UKRI, Ofcom or a newly created body. As a matter of urgency, the Government should seek to identify a responsible authority and clarify the legal and technical governance responsibilities of a new intermediary body.

Engage the research community to inform design of a new intermediary body

Access to social media data for the study of online harms is essential for researchers across academia and civil society. To ensure that an intermediary body is effective for researchers from different sectors it will be vital to involve stakeholders from a range of sectors to ensure the practical demands of securing ethical, legal and methodological review are reflected in the design of a new intermediary.

3.5 Summarising policy options for web-scraping, data donation and an intermediary body

The case for improving researcher access to social media data rests ultimately on public safety. The harms the OSA was designed to address, including the exposure of children to harmful content, the

targeting of vulnerable users and the spread of illegal extremist material, cannot be independently measured or verified and the effects of platforms' efforts to address them cannot be assessed without systematic access to platform data. The recommendations in this chapter seek to ensure that the UK has the research infrastructure needed to protect online users and indeed, that case is most urgent when it comes to children. Research into the impacts of social media on children and young people, on their mental health, their exposure to harmful content and their susceptibility to predatory targeting, depends on exactly the kinds of individual-level, longitudinal and real-time data access that a strengthened UK data access regime could provide. The reforms needed are achievable, proportionate and cost-effective.

Our three policy routes for policy intervention present a route for reform that can be implemented in the immediate, medium and longer term. Each element is critical to securing a diverse and resilient landscape for data access and there are opportunities for the strands to interact. For example, an intermediary body could curate expertise on web-scraping and data donation as well providing secure infrastructure. However, these responses do not need to be progressed together in one combined framework, they can each be advanced independently to ensure diverse routes to data access. Investment is, however, required in each area to ensure a robust system of data access.

Chapter 4

Research Community Actions for Better Data Access

Alongside the legislative and regulatory reforms recommended in sections 3.2, 3.3 and 3.4, the Taskforce identified a set of challenges requiring action from within the research community - by which we mean research funders, both philanthropic and state, professional research bodies or associations, such as the Russell Group or national academies, or individual research organisations and collectives in civil society and academia. These reflect structural gaps in capacity, training and coordination that currently prevent many researchers, particularly those in smaller institutions, civil society organisations and academic disciplines outside computer science, from making effective use of the limited, rapidly evolving data access routes that exist.

The four categories of barriers identified in section 1.3 (financial, technical, legal & institutional and capacity & skills) require responses at both legislative and community level. The legislative and regulatory reforms recommended in Chapter 3 address the financial, technical and legal dimensions. This chapter focuses on the capacity, skills and institutional barriers that legislation alone cannot resolve and that require action from within the research community itself.

4.1 Skills and capacity gaps

The data access landscape has changed profoundly in the last decade. Previously existing data access avenues have closed, new routes have opened and new requirements have been put in place. These changes require researchers to adapt to newly regulated and structured forms of access, but many lack the knowledge or competencies required to navigate these systems.

Within academia, training to build data management competencies have recently been incorporated into PhD training, and services such as Smart Data Research UK are making efforts to offer training and development opportunities around data science, but a substantial proportion of more senior academics have not received such training. These challenges are also evident within civil society, journalism and government, where many researchers lack the combination of legal, technical and subject expertise required to navigate them. As a result, researchers, and particularly those without legal training, often lack the knowledge needed to navigate data protection law in the context of their research, while those without a computer science background face equivalent barriers navigating scraping, data donation and platform API workflows. The combined technical and legal knowledge needed

to navigate these processes confidently is far from commonplace.

Some capacity-building efforts exist in specific areas. The DSA Observatory at the University of Amsterdam convenes researchers to discuss and monitor DSA implementation, while the DSA 40 Collaboratory (Weizenbaum Institute) specifically supports researchers in navigating and applying for data access under Article 40(4) and 40(12). These efforts are useful but focused on a specific regulatory framework. Other work has been progressed by networks such as the Coalition for Independent Technology Research (CITR), an international body that works to advance, defend, and sustain the right to study the impact of technology on society. The Coalition convenes researchers, technologists and civil society organisations in defence of research that is ethical, transparent and privacy-preserving, and has been a leading voice in debates around platform data access. It convenes a data access task force whose members work across tooling, advocacy, litigation and capacity building to expand and defend researchers' access to platform data.

Whilst some expertise exists, there is often a lack of relevant knowledge, particularly legal knowledge, within the research institutions of which researchers are part. Administrative staff and review processes also commonly lack the legal and technical expertise required to successfully navigate data access infrastructure. When it comes to web scraping, data donation or access via platform infrastructure, there is a lack of expertise required to manage complex donation pipelines, parse heterogeneous data formats, navigate legal constraints or conduct the forms of analysis that digital trace data demands (SDDS, 2025).

Recommendation for the Research Community

Given the current lack of an intermediary body, there is a need for the research community to gather expertise and develop infrastructure to support researchers in navigating a wide range of data access systems in a way that is legally compliant and ethically robust.

Whilst a range of research organisations could play a role, within academia UKRI - the UK's umbrella organisation responsible for the distribution of research funding - provides one body capable of undertaking this work. Specifically Smart Data Research UK, a programme funded by the UKRI is the UK's national programme for improving access to digital data. It has developed successful data infrastructure that provides national access to a range of new and emerging forms of data, and would

bewell placed to undertake this work. However, there are also a range of civil society organisations and professional associations who can play a key role in this work. The Coalition for Independent Technology Research, in particular, has been a leading force in convening researchers to address data access challenges.

Specifically we identify the following steps to enhancing knowledge and expertise:

UKRI should audit existing skills initiatives designed to promote legal, ethical and technical understanding of data access mechanisms.

The research community should seek to identify available training sources within and beyond academia and assess the sufficiency of existing mechanisms. This exercise will allow signposting of existing resources and identification of gaps for future investment or targeted interventions by existing investments (such as those recently funded via UKRI infrastructure funded via UKRI Digital Research Technical Professional Skills NetworkPlus; Responsible AI UK: skills programme; Digital infrastructure: new approaches to skills or software; UKRI DRI: digital research infrastructure skills hubs for accelerated compute).

Research institutions and funders should introduce dedicated skills fellowships for academics and non-academic researchers to enable the development of data access skills.

This would help to address current skill gaps, building specialist capacity in data protection law, computational methods and platform research.

UKRI should fund the development of accessible, practical training and resources for researchers and research administrators covering the legal and technical dimensions of social media data research.

Such training should cover rights and obligations under the UK GDPR research exemption, the circumstances under which scraping of publicly accessible data is defensible, how platform terms of service are typically enforced in practice as distinct from their formal wording, how to conduct a proportionate Data Protection Impact Assessment, how to respond to platform cease-and-desist communications and the management of data donation pipelines and the computational analysis of digital trace data. Training resources should be available to researchers across all institution types, including social enterprises and community interest companies that currently fall outside university-focused support structures.

Research institutions and researchers should develop practical DPIA templates and worked

examples tailored specifically to social media scraping research, data donation and other data access mechanisms.

Examples of compliant DPIA documentation are essential for building awareness amongst researchers of the requirements of trustworthy data access. In the future there is potential to utilise the capacities of generative AI to support the production of DPIA documentation, but identifying high quality examples will be key to sharing best practice.

4.2 Institutional fragmentation and support

A persistent structural problem identified through the Taskforce's evidence is the fragmentation of the research landscape in the UK. Both within and across sectors, research organisations lack a common approach to navigating data access requirements. Some institutions have adopted blanket restrictions on scraping tools, others are more permissive (either on the whole, or on a case-by-case basis) or lack clear policies. The result is a fragmented landscape in which the ability to conduct social media data research depends in large part on where a researcher is based.

Within academia, there is a further disconnect between funder expectations and institutional sign-off. UKRI requires that all research is legally compliant, and yet many applications have been funded which outline plans to use currently legally ambiguous methods such as web scraping. UKRI terms and conditions can therefore lead to institutional caution, meaning research can be prevented from going ahead or incur significant costs as bespoke legal advice is sought. Smaller organisations, social enterprises and community interest companies are particularly disadvantaged by the lack of clear guidance on data access as they often lack in-house legal capacity (Social Platforms Data Access Taskforce, 2026a).

Some of the most advanced practice in navigating these challenges has emerged from well-resourced civil society organisations that have invested in building internal legal expertise and standardised risk assessment frameworks. Several organisations consulted by the Taskforce have developed internal protocols that systematically evaluate data collection projects against structured criteria, including whether the research purpose is genuinely public interest, whether data collected is limited to what is publicly accessible, whether platform terms of service constitute an enforceable contract in the specific circumstances, and whether any material harm

could be attributed to the research (Social Platforms Data Access Taskforce, 2025a). These approaches are not yet widely shared or systematically available to smaller or less well-resourced organisations.

Recommendation for the Research Community

Research institutions and researchers should privately share best practice on navigating data access challenges and look to develop collective responses to key areas of legal ambiguity.

This may manifest as shared guidance on navigating legal risk in social media data research, coordinating approaches across institutions to replace the current patchwork of institution specific procedures. Shared approaches will have particular value in relation to social media data scraping, where approaches are currently disparate. Sharing privately (as opposed to publicly) will help to address any risks incurred through sharing strategies publicly.

As a funder of research, UKRI should clarify its position on web scraping for public interest research.

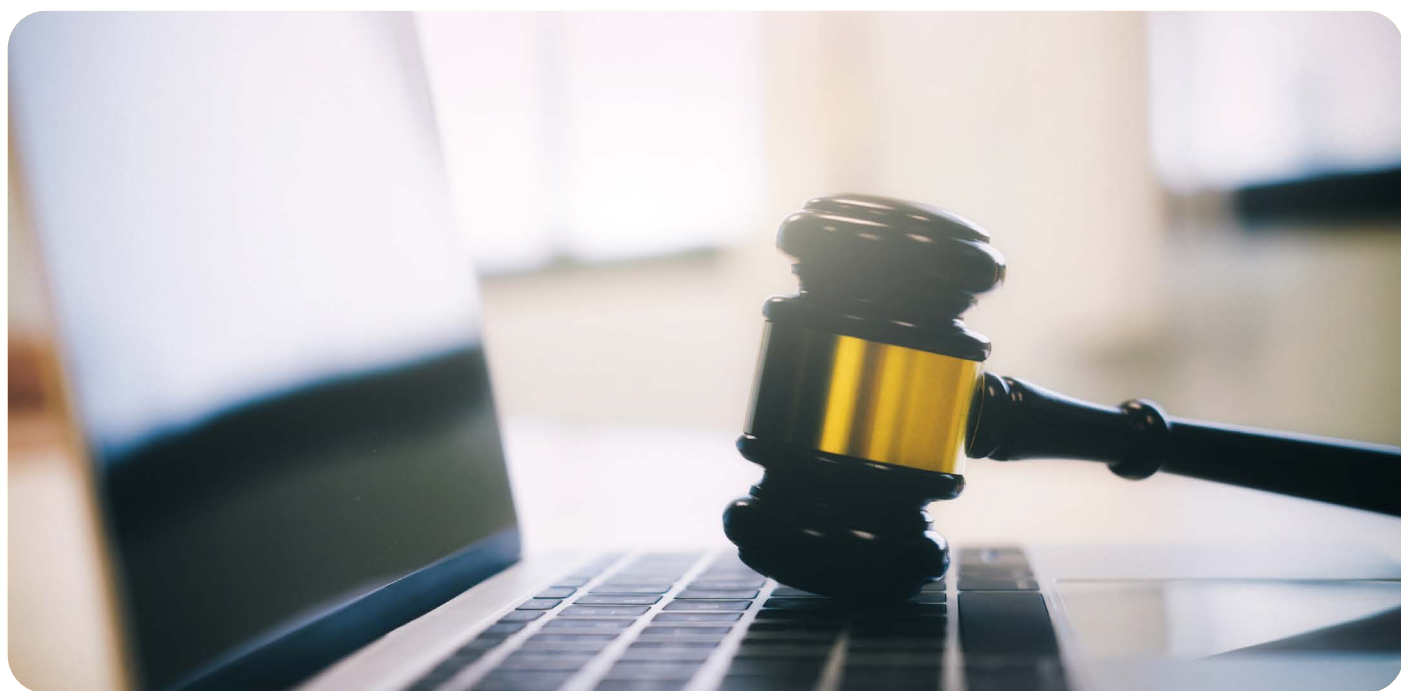
UKRI requires that all research is legally compliant, and yet many applications have been funded which outline plans to use currently legally ambiguous methods such as web scraping. UKRI should issue guidance on social media research and permissible methodologies. This guidance should be incorporated into grant assessment and institutional approval processes.

4.3 Legal action and collective responses to legal ambiguity

One emerging avenue for challenging platform authority over research access is through strategic litigation designed to test the enforceability of anti-scraping provisions. Several significant US court cases have begun to erode the ability of platforms to use their terms of service to block public interest research. In *X Corp. v. Center for Countering Digital Hate, Inc.* (2024), a US federal court ruled against X's lawsuit against a civil society research organisation, finding that scraping publicly available information for accountability purposes did not constitute the kind of harm required to sustain a breach-of-contract claim. Similarly, *Meta Platforms, Inc. v. Bright Data Ltd.* (2024) held that anti-scraping provisions do not automatically apply to data publicly visible to non-users.

These cases have not been tested in UK courts and there is no directly analogous domestic case law. They nevertheless provide persuasive evidence of how similar principles might apply under UK contract law and some civil society organisations have drawn on them explicitly in their internal risk assessments. There is a growing argument within the research and civil society community that strategic litigation in the UK, testing whether platforms can enforce anti-scraping provisions against public interest researchers who cause no material harm, could help establish clearer domestic precedents. The risk of SLAPPs by platforms against researchers has been widely identified as a structural barrier to public interest research, including by Ofcom (2025).

The need for collective legal support for researchers



facing platform litigation is not unique to the UK. In the United States, CITR was established specifically to provide crisis response support, including legal referrals and solidarity mechanisms, for researchers facing legal and political challenges to their work (CITR, n.d.). The Knight First Amendment Institute at Columbia University has similarly litigated cases on behalf of researchers whose methods platforms have sought to restrict through legal action (Knight First Amendment Institute, 2022). These models demonstrate that collective legal infrastructure for public interest research is both feasible and necessary. In the meantime, the absence of equivalent UK infrastructure leaves researchers to rely on individual institutional risk assessments, which can themselves be sources of uncertainty and inequity.

Recommendation for the Research Community

Research institutions, including philanthropic funders and university consortia, should establish a legal defence fund and coordination mechanism to support researchers and civil society organisations facing platform litigation or cease-and-desist action for public interest scraping research.

This fund should be able to support strategic litigation designed to test the enforceability of platform terms of service against public interest researchers in UK courts, with the aim of establishing domestic precedents that reduce legal uncertainty across the research community. This would ensure that individual researchers and smaller organisations do not bear the risks of potential legal action.

4.4 International coordination

The international landscape for social media data research infrastructure is developing rapidly. New shared infrastructure, research consortia and legal frameworks are emerging across the EU and the United States, as illustrated by the examples examined in section 3.4.3. The UK is already engaged in this space through a number of bilateral research collaborations and UKRI-funded researchers participate in several international consortia. However, these arrangements are largely informal and depend on individual initiative rather than a coherent national strategy.

This presents an opportunity to identify and learn from international best practice, helping to share and advance trustworthy data access approaches. However, the infrastructure to facilitate such learning is presently lacking.

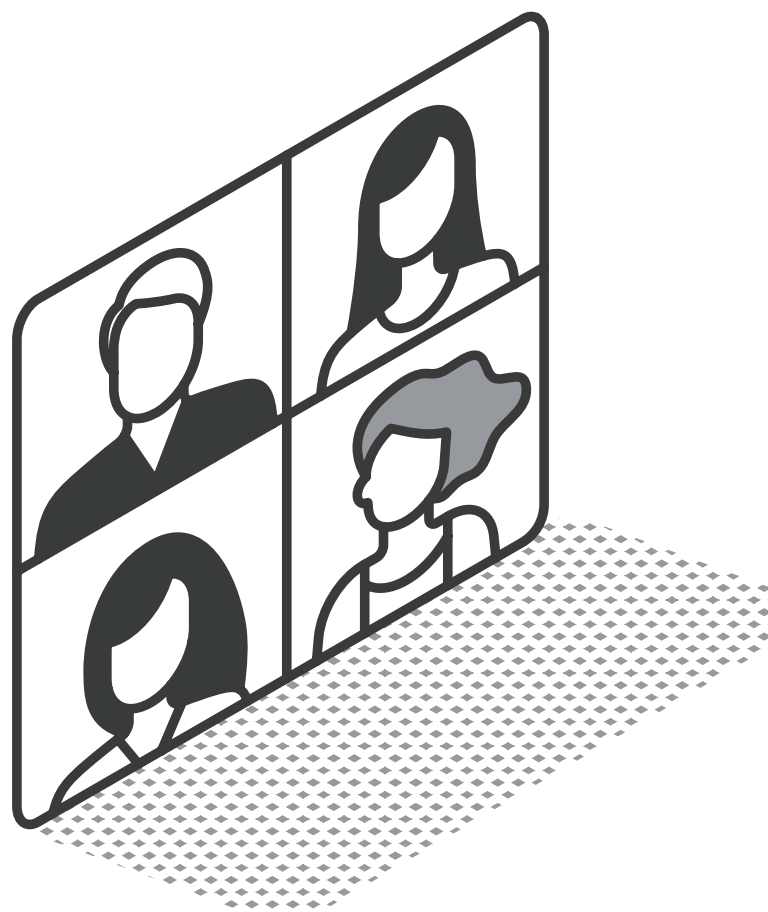
Recommendation for the Research Community

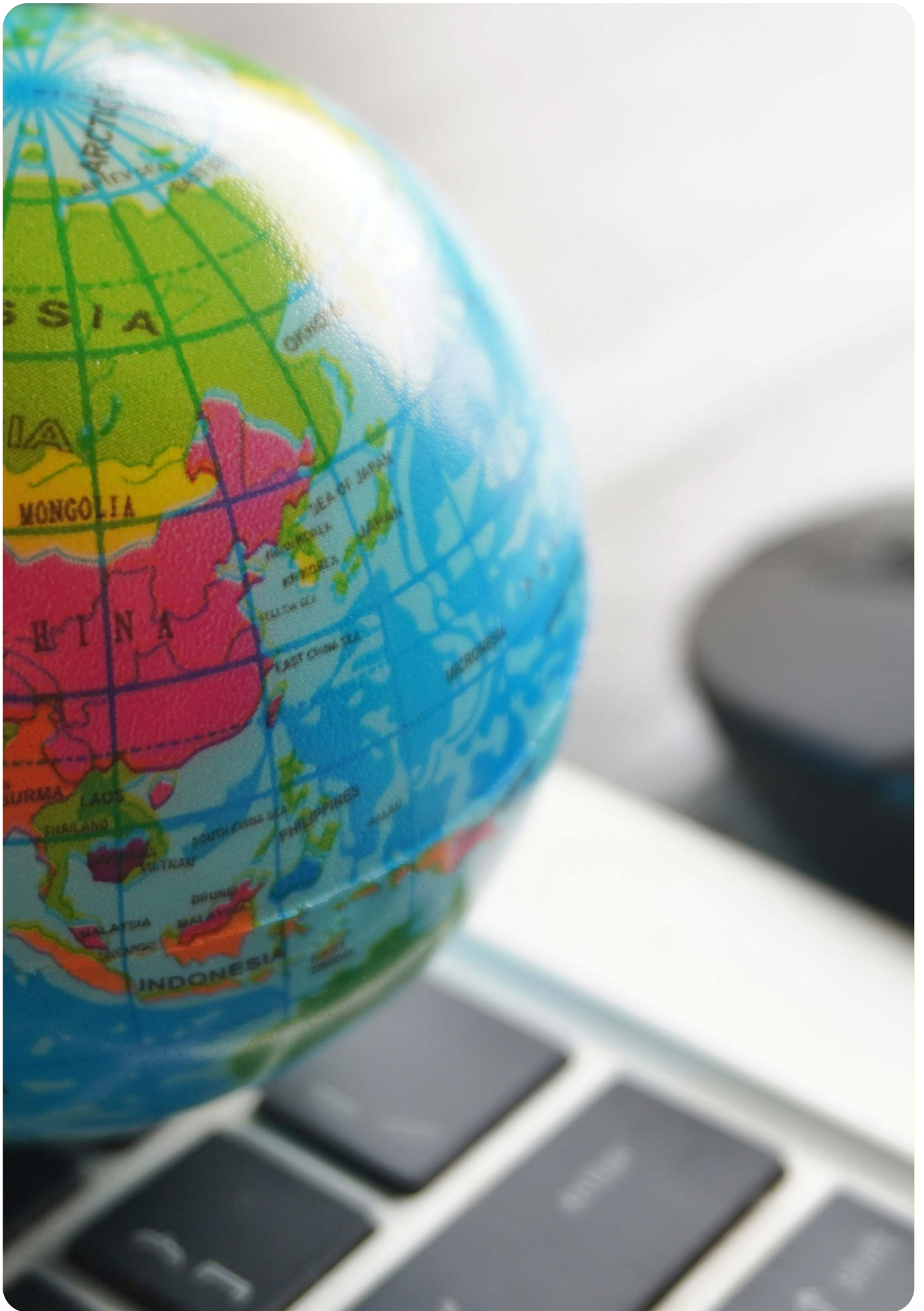
Research institutions should develop formal partnership agreements with key international research infrastructure bodies, including GESIS, the DSA Observatory and equivalent organisations.

This would ensure that UK researchers have clear and consistent routes to participate in international data access initiatives and that UK interests are represented in the governance of those initiatives.

Research institutions and researchers should work collaboratively to push for changes in platforms' terms of service to allow the use of web-scraping for research purposes.

Some platforms have shown they will make such changes under regulatory pressure: AliExpress, for example, committed to a research exception in its terms of service under the DSA, allowing researchers meeting Article 40(12) conditions to access publicly accessible data for systemic risk research, via a dedicated application page and form for data not otherwise readily available. Researchers should work collaboratively across national boundaries to call for other platforms to adopt similar provisions.





Chapter 5

Conclusions and recommendations

5.1 Report Summary

This report set out to make the case for access to social media data for the purposes of online safety research. Within this report we have outlined numerous barriers to social media data access and their implications for our understanding of the online world. What is clear is that, as Ofcom argues, “A layered, flexible approach – combining legal clarity, technical safeguards, and independent oversight – offers the best chance of enabling responsible, timely and useful information access” (Ofcom, 2025, p.8).

In line with this diagnosis, we conclude that a range of actions are needed to advance data access. This includes enacting change across the three areas discussed in this report: web scraping, data donation and the creation of an intermediary body. In addition, we have argued that the research community can play an important role in helping researchers to navigate an increasingly complex data access landscape.

The barriers documented in this report are significant, but the conditions for reform are more favourable than at any previous point: the Data (Use and Access) Act 2025 has created new powers that can be extended to social media platforms, Ofcom has published a clear diagnosis of the access problem and the Government has committed to a consultation on researcher access towards the end of 2026. The recommendations in this report are designed to ensure that consultation produces durable, proportionate and enforceable outcomes.

5.2 The opportunity to act

The UK currently has an opportunity to establish itself as a world-leader on social media data access, positioning its researchers to produce the world-leading research needed to understand and combat online harms. The Government’s consultation on data access, scheduled for the end of 2026, is the key mechanism through which many of our proposals can be actioned to produce tangible change.

By acting now, the UK can take full advantage of the range of benefits identified in the UK Government’s National Data Strategy and keep pace with international developments in this field. More profoundly, decisive action would allow the public benefit of research into online harms to be realised at pace. The harms the OSA was designed to address (children’s exposure to harmful content, the targeting of vulnerable users) can be independently measured and mitigated with systematic access to platform data. Seizing this opportunity will allow the

effectiveness of existing legislation to be tested and strengthened.

5.3 Recommendations: a phased approach

Bringing together the recommendations outlined across previous chapters, this section presents them as a sequenced package. Rather than grouping them by data access route, the recommendations are organised here by the time horizon over which they can realistically be delivered, reflecting what each requires to put into practice. The phasing matters as many of the recommendations can be advanced now, without primary legislation, to support the research community. Others can be delivered through secondary legislation under the Data (Use and Access) Act 2025. A smaller number require primary legislation or the establishment of new infrastructure, and represent the longer-term structural reforms towards which the earlier actions build. This sequencing is deliberate. Clarifying the rules that already apply to researchers is a necessary first step that can be achieved without new infrastructure and is distinct from the more complex task of establishing a mandated access regime or a statutory intermediary. Treating these as separate phases, rather than components of a single reform, allows researchers to benefit from clearer guidance in the near term while the longer-term structural changes are developed.

Presenting the recommendations in this way is intended to support a realistic and achievable implementation pathway, beginning with the actions that can deliver value soonest. The policy recommendations thus, are set out below under three time horizons: near-term, medium-term and longer-term. Recommendations for the research community, which can largely be progressed now, follow separately.

Recommendations for the UK Government and Regulators

Near-term (no primary legislation required)

WEB-SCRAPING:

1. Consolidate existing legal and regulatory guidance for web scraping into domain-specific guidance for social media research.

The government should direct the ICO to develop practical, domain-specific guidance for researchers conducting social media research that involves public-interest web scraping. Existing ICO guidance on web scraping is predominantly focused on

generative AI and commercial data brokering. It does not adequately address the specific circumstances of public interest research.

Guidance should address:

1. The application of the research exemption to social media scraping
2. The circumstances under which logged-out versus logged-in scraping can be justified under UK GDPR
3. The conditions under which special category data collected during scraping can be processed under Article 9, including the safeguards and documentation required

Such guidance would not require primary legislation. It could be developed within the existing ICO mandate, informed by the evidence set out in Ofcom's 2025 report.

DATA DONATION:

2. Clarify data donation rules in relation to:
 - a. The right to third-party transfer of donated data under the UK GDPR

The government should clarify the ability of individuals to request that their platform data is transferred directly to a named third party for public interest research. The government should direct the ICO to issue joint guidance on how Article 20 direct transmission should operate in public interest donation studies, or direct them to issue a UK GDPR code of conduct under Article 40. In this context, guidance should confirm how a participant can identify an approved research service as a recipient of a transfer, how authorisation should be recorded, and how data could be transferred. A participant should be able to initiate requests themselves; select trusted third parties as recipients; and complete any authorisation step needed to complete a transfer. It is critical that this process is seamless and straightforward for participants to use. Where possible, API-based data donation should be enabled for large online platforms, so that transfers can operate reliably and at scale.

- b. The legal basis for enriching donated data

Direct the ICO to issue specific guidance on whether and under what conditions data enrichment constitutes a lawful continuation of the original donation and what safeguards are required. In the absence of such guidance, researchers are exposed to legal risk for activity that is essential to making donated data analytically usable.

c. Data donation portability standards

Direct the ICO to issue authoritative guidance specifying minimum standards for data portability outputs, clearly distinguishing them from subject access request (SAR) outputs. Guidance should require: standardised documentation of the data provided, including how it was generated; a standardised, machine-readable data format; and at least two months' notice to researchers of any change to the donation process or the data provided. This requires no new legislation and falls within existing ICO powers (ICO, n.d.-a; SDDS, 2025).

DATA INTERMEDIARY:

3. Identify the legal foundation and governance arrangements for a hybrid-repository social media data intermediary body.

A range of actors could take responsibility for leading a social media intermediary, including UKRI, Ofcom or a newly created body. As a matter of urgency, the Government should seek to identify a responsible authority and clarify the legal and technical governance responsibilities of a new intermediary body.

Medium-term (secondary legislation under the DUAA 2025

WEB-SCRAPING:

4. Establish a safe harbour to enable public interest web-scraping

The most durable solution to the current ambiguity would be the establishment of a statutory legal safe harbour to protect researchers who conduct public interest scraping of publicly accessible social media data from adverse platform action. Such a provision would need to define clearly the classes of researcher and institution covered, the conditions under which scraping qualifies as "public interest" (drawing on existing UK GDPR research exemption criteria) and accompanying safeguards, including data minimisation requirements, Data Protection Impact Assessment obligations and prohibited uses of scraped data. The Data (Use and Access) Act 2025 could be amended to provide for a safe harbour from contractual liability under platform terms of service for public interest research which uses scraping. The Online Safety Act provides an alternative legislative vehicle, but this would likely only provide that protection in relation to research into 'online safety matters', which is only a subset of public interest research generally.



DATA DONATION:

5. Establish a researcher complaints mechanism to address the failure of data donation processes and grant ICO powers to investigate and resolve complaints

Create a clear and accessible route by which researchers and participants can escalate portability failures and non-compliance to the ICO, with defined response timelines. The UK GDPR already gives individuals a route to complain in relation to their portability requests, though this offers limited practical redress in most cases. A new statutory duty requiring organisations to operate a formal complaints-handling process took effect in June 2026, but this addresses complaints from data subjects about data handling generally, not the specific problem of researchers being unable to challenge a rejected donation or a data quality issue (ICO, 2026).

To resolve these issues, the ICO should be given powers to investigate and resolve complaints from researchers as well as data subjects. Achieving this would require primary legislation amending the Data Protection Act 2018, since only data subjects, not researchers, currently have standing to complain under section 165 of that Act. Secondary legislation under the DUAA cannot extend standing in this way, as this would go beyond a technical adjustment to an existing right (SDDS, 2025).

6. Outline standardised processes for the identification and delegation of data donation requests.

Under the DUAA's digital verification services framework, the Government should act to mandate

specific data donation consent flows. This should specify approved identity assurance provisions and introduce a recognised mechanism for delegated authorisation, enabling researchers to act formally as authorised recipients of a participant's data portability request.

7. Enable continuous, as opposed to only momentary, data donation.

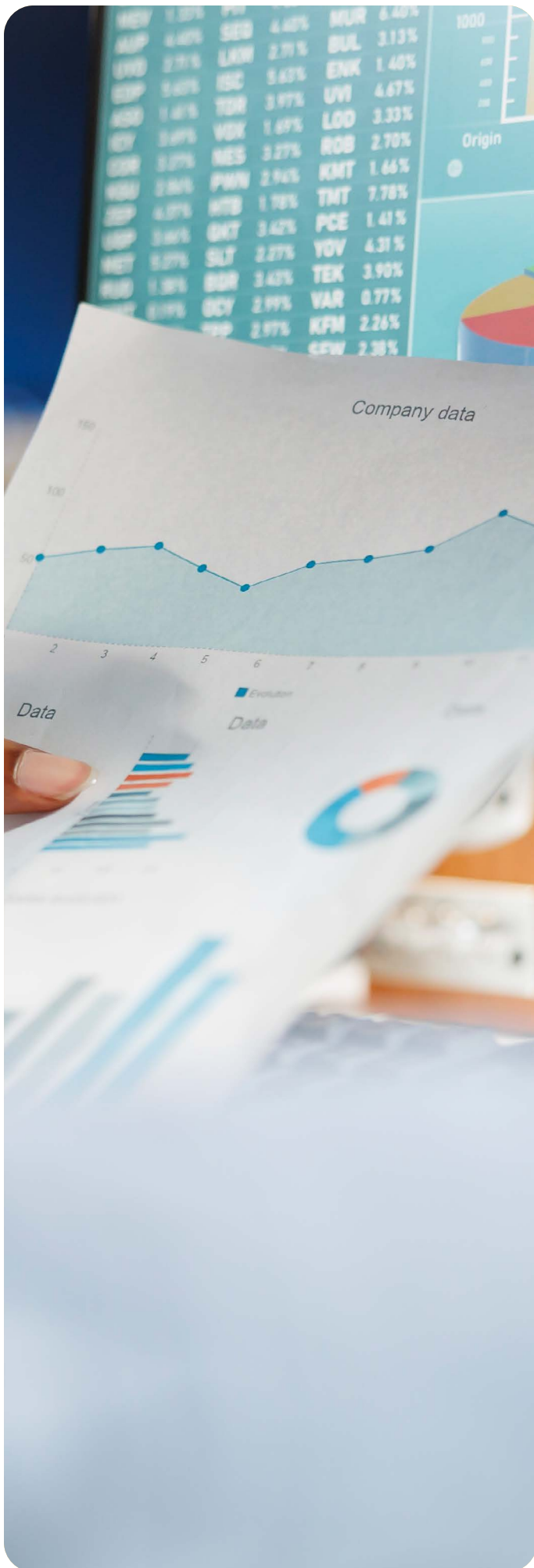
Introduce secondary legislation under the DUAA to allow data donation to operate not only as a single, momentary transfer but as a continuous mechanism, so that participants can consent to ongoing access for a defined future period and to selective sharing of specified data categories. A comparable approach exists under Article 6(9) of the EU's Digital Markets Act, which requires large gatekeeper platforms to provide continuous, real-time data portability (SDDS, 2025).

8. Require platforms to provide data transparency and change-log documentation around data donation processes.

Platforms should be obliged to publish and maintain documentation of changes to data export packages, including version logs and advance notice to researchers, to prevent silent disruption to ongoing studies. This obligation could be introduced through secondary legislation under the DUAA.

DATA INTERMEDIARY:

9. Learn lessons from existing data infrastructure



and from international social media data access efforts to inform hybrid-repository intermediary body design.

A social media data intermediary presents a complex prospect, but there are clear foundations on which to build such infrastructure. Existing data archives and intermediaries offer a range of technical and legal expertise, whilst precedent set by the DSA and by organisations such as GESIS, provide a range of domain specific guidance that could be adapted for the UK context. As such, urgent efforts should be directed to learning lessons from existing initiatives to ensure the UK benefits from best practice and avoids previous pitfalls. Specific attention should be paid to insights regarding application management and access assessments, platform relationships and data access notices, and data infrastructure and stewardship.

10. Engage the research community to inform design of a new intermediary body.

Access to social media data for the study of online harms is essential for researchers across academia and civil society. To ensure that an intermediary body is effective for researchers from different sectors it will be vital to involve stakeholders from a range of sectors to ensure the practical demands of securing ethical, legal and methodological review are reflected in the design of a new intermediary.

Longer-term (primary legislation or major structural build.)

DATA DONATION:

11. Expand and normalise Smart Data schemes for social media platforms to enable data donation, building on work already underway through the Smart Data Donation Service.

Use the DUAA 2025 framework to designate major social media platforms as data holders subject to Smart Data obligations, requiring them to provide standardised, API-based portability mechanisms for public interest research. This is the most structurally significant reform available within the existing legislative architecture. While it can be delivered through secondary legislation under powers already granted by the DUAA, it represents a major structural undertaking requiring significant technical and institutional development. For this reason, it is best understood as a longer-term reform, despite being deliverable through powers already provided by the DUAA.

DATA INTERMEDIARY:

12. Establish a statutory hybrid-repository social media data intermediary.

A statutory intermediary is the most significant structural reform available. It would replace the current fragmented, platform-controlled access landscape with a single governed institution that centralises accreditation, manages data security and provides platforms with a consistent and predictable counterpart, reducing their compliance burden while expanding researcher access. The international evidence reviewed in section 3.4.3 confirms that voluntary arrangements are not a reliable substitute, and that statutory backing is a necessary condition for any durable access regime. For child online safety research in particular, an intermediary with access to longitudinal and individual-level data would provide the evidence base the OSA's effectiveness ultimately depends upon.

In recommending the creation of an intermediary, we do not propose this is the only data access mechanism. Rather, mirroring Ofcom's recommendation for a layered and flexible approach, we recommend an intermediary body would, over time, gather and share expertise on different routes for access, but the mechanisms themselves would be separate. This approach is essential for ensuring access across academia and civil society.

Recommendations for the research community

In the Near-Term

13. As a funder of research, UKRI should clarify its position on web scraping for public interest research.

UKRI requires that all research is legally compliant, and yet many applications have been funded which outline plans to use currently legally ambiguous methods such as web scraping. UKRI should issue guidance on social media research and permissible methodologies. This guidance should be incorporated into grant assessment and institutional approval processes.

14. Research institutions and researchers should privately share best practice on navigating data access challenges and look to develop collective responses to key areas of legal ambiguity.

This may manifest as shared guidance on navigating legal risk in social media data research, coordinating approaches across institutions to replace the current patchwork of institution specific procedures. Shared approaches will have particular value in relation to

social media data scraping, where approaches are currently disparate. Sharing privately (as opposed to publicly) will help to address any risks incurred through sharing strategies publicly.

15. UKRI should audit existing skills initiatives designed to promote legal, ethical and technical understanding of data access mechanisms.

The research community should seek to identify available training sources within and beyond academia and assess the sufficiency of existing mechanisms. This exercise will allow signposting of existing resources and identification of gaps for future investment or targeted interventions by existing investments (such as those recently funded via UKRI infrastructure funded via UKRI Digital Research Technical Professional Skills NetworkPlus; Responsible AI UK: skills programme; Digital infrastructure: new approaches to skills or software; UKRI DRI: digital research infrastructure skills hubs for accelerated compute).

In the Medium-Term

16. Research institutions and researchers should develop practical DPIA templates and worked examples tailored specifically to social media scraping research, data donation and other data access mechanisms.

Examples of compliant DPIA documentation are essential for building awareness amongst researchers of the requirements of trustworthy data access. In the future there is potential to utilise the capacities of generative AI to support the production of DPIA documentation, but identifying high quality examples will be key to sharing best practice.

17. UKRI should fund the development of accessible, practical training and resources for researchers and research administrators covering the legal and technical dimensions of social media data research.

Such training should cover rights and obligations under the UK GDPR research exemption, the circumstances under which scraping of publicly accessible data is defensible, how platform terms of service are typically enforced in practice as distinct from their formal wording, how to conduct a proportionate Data Protection Impact Assessment, how to respond to platform cease-and-desist communications and the management of data donation pipelines and the computational analysis of digital trace data. Training resources should be

available to researchers across all institution types, including social enterprises and community interest companies that currently fall outside university-focused support structures.

18. Research institutions and funders should introduce dedicated skills fellowships for academics and non-academic researchers to enable the development of data access skills.

This would help to address current skill gaps, building specialist capacity in data protection law, computational methods and platform research.

19. Research institutions and researchers should work collaboratively to push for changes in platforms' terms of service to allow the use of web-scraping for research purposes.

Platforms such as AliExpress have recently committed to modifying their terms of service under the DSA. Researchers should work collaboratively across national boundaries to call for other platforms to adopt similar provisions.

In the Long-Term

20. Research institutions should develop formal partnership agreements with key international research infrastructure bodies, including GESIS, the DSA Observatory and equivalent organisations

This would ensure that UK researchers have clear and consistent routes to participate in international data access initiatives and that UK interests are represented in the governance of those initiatives.

21. Research institutions, including philanthropic funders and university consortia, should establish a legal defence fund and coordination mechanism to support researchers and civil society organisations facing platform litigation or cease-and-desist action for public interest scraping research.

This fund should be able to support strategic litigation designed to test the enforceability of platform terms of service against public interest researchers in UK courts, with the aim of establishing domestic precedents that reduce legal uncertainty across the research community. This would ensure that individual researchers and smaller organisations do not bear the risks of potential legal action.



Summary Table A: Recommendations for the UK Government and Regulators (by theme)

#	Timeframe	Recommendation
Web-Scraping		
1	Near-Term	Consolidate existing legal and regulatory guidance for web scraping into domain-specific guidance for social media research
2	Medium-term	Establish a safe harbour to enable public interest web-scraping
Data Donation		
3	Near-Term	Clarify data donation rules on third-party transfer, the legal basis for enrichment, and portability standards
4	Medium-Term	Establish a researcher complaints mechanism to address the failure of data donation processes and grant ICO powers to investigate and resolve complaints
5	Medium-Term	Outline standardised processes for the identification and delegation of data donation requests
6	Medium-Term	Enable continuous, as opposed to only momentary, data donation
7	Medium-Term	Require platforms to provide data transparency and change-log documentation around data donation processes
8	Long-Term	Expand and normalise Smart Data schemes to social media platforms to enable data donation
Data Intermediary		
9	Near-Term	Identify the legal foundation and governance arrangements for a hybrid-repository social media data intermediary body
10	Medium-Term	Learn lessons from existing data infrastructure and from international social media data access efforts to inform hybrid-repository intermediary body design
11	Medium-Term	Engage the research community to inform design of a new intermediary body
12	Long-Term	Establish a statutory hybrid-repository social media data intermediary

Summary Table B: Recommendations for the Research Community

#	Timeframe	Recommendation
13	Near-Term	As a funder of research, UKRI should clarify its position on web-scraping for public interest research
14	Near-Term	Research institutions and researchers should privately share best practice on navigating data access challenges and look to develop collective responses to key areas of legal ambiguity
15	Near-Term	UKRI should audit existing skills initiatives designed to promote legal, ethical and technical understanding of data access mechanisms
16	Medium-Term	Research institutions and researchers should develop practical DPIA templates and worked examples tailored specifically to social media scraping research, data donation and other data access mechanisms
17	Medium-Term	UKRI should fund the development of accessible, practical training and resources for researchers and research administrators covering the legal and technical dimensions of social media data research
18	Medium-Term	Research institutions and funders should introduce dedicated skills fellowships for academics and non-academic researchers to enable the development of data access skills
19	Medium-Term	Research institutions and researchers should work collaboratively to push for changes in platforms' terms of service to allow the use of web-scraping for research purposes
20	Long-Term	Research institutions should develop formal partnership agreements with key international research infrastructure bodies, including GESIS, the DSA Observatory and equivalent organisations
21	Long-Term	Research institutions, including philanthropic funders and university consortia, should establish a legal defence fund and coordination mechanism to support researchers and civil society organisations facing platform litigation or cease-and-desist action for public interest scraping research

References

- Ada Lovelace Institute. (2025). Response to Ofcom's call for evidence on researcher access to information from regulated online services. Ofcom. <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/call-for-evidence-researchers-access-to-information-from-regulated-online-services>
- Administrative Data Research UK. (n.d.). About ADR UK. <https://www.adruk.org/about-us/about-adr-uk/>
- Bekavac, L., & Mayer, S. (2026). Auditing Meta and TikTok Research API data access under Article 40(12) of the Digital Services Act. arXiv. <https://arxiv.org/abs/2601.12390>
- Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research*, 4(2), 388–423. <https://doi.org/10.5117/CCR2022.2.002.BOES>
- Boeschoten, L., de Schipper, N. C., Mendrik, A. M., van der Veen, E., Struminskaya, B., Janssen, H., & Araujo, T. (2023). Port: A software tool for digital data donation. *Journal of Open Source Software*, 8(90), Article 5596. <https://doi.org/10.21105/joss.05596>
- Botero Arcila, B., Ramaciotti, P., & Cabale, E. (2026). Seeing in the dark: Towards a broad construction of the access to data provisions of the DSA. *Internet Policy Review*, 15(1). <https://doi.org/10.14763/2026.1.2085>
- Bradshaw, S., Bailey, H., & Howard, P. N. (2021). Industrialized disinformation: 2020 global inventory of organized social media manipulation. Programme on Democracy & Technology, Oxford Internet Institute, University of Oxford. <https://demotech.oii.ox.ac.uk/wp-content/uploads/sites/12/2021/01/CyberTroop-Report-2020-v.2.pdf>
- Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2022). User-centric approaches for collecting Facebook data in the 'post-API age': Experiences from two studies and recommendations for future research. *Information, Communication & Society*, 26(14), 2649–2668. <https://doi.org/10.1080/1369118X.2022.2097015>
- Brown, M. A., Gruen, A., Maldoff, G., Messing, S., Sanderson, Z., & Zimmer, M. (2024). Web scraping for research: Legal, ethical, institutional, and scientific considerations. arXiv. <https://doi.org/10.48550/arXiv.2410.23432>
- Bryce, J., Livingstone, S., Davidson, J., Hall, B., & Smith, J. (2023). Evidence review on online risks to children. NSPCC. <https://learning.nspcc.org.uk/research-resources/2023/online-risks-to-children-evidence-review>
- Case T-114/26, X Corp. v European Commission, OJ C, 2026/2392 (pending before the General Court of the European Union).
- Centre for Data Ethics and Innovation. (2021). Unlocking the value of data: Exploring the role of data intermediaries. Department for Digital, Culture, Media and Sport. <https://www.gov.uk/government/publications/unlocking-the-value-of-data-exploring-the-role-of-data-intermediaries/unlocking-the-value-of-data-exploring-the-role-of-data-intermediaries>
- Coalition for Independent Technology Research. (n.d.). About us. Retrieved 10 June 2026, from <https://independenttechresearch.org/about-us/>
- Coimisiún na Meán. (2026, May 25). Guidance on Article 40(4-11) DSA (Version 2). https://www.cnam.ie/app/uploads/2026/05/DSA-Article-40-4-11-Data-Access-Guidance_V2-.pdf
- Competition and Markets Authority. (2017). The Retail Banking Market Investigation Order 2017. GOV.UK. <https://www.gov.uk/government/publications/retail-banking-market-investigation-order-2017/the-retail-banking-market-investigation-order-2017>
- Competition and Markets Authority. (2026, January 28). CMA proposes package of measures to improve Google search services in UK. GOV.UK. <https://www.gov.uk/government/news/cma-proposes-package-of-measures-to-improve-google-search-services-in-uk>
- Coons, C., & Cassidy, B. (2025). Platform Accountability and Transparency Act, S. 3292, 119th Congress. United States Senate. <https://www.congress.gov/bill/119th-congress/senate-bill/3292/text>
- Data (Use and Access) Act 2025, c. 18. <https://www.legislation.gov.uk/ukpga/2025/18>
- Davidson, B. I., Wischerath, D., Racek, D., Parry, D. A., Godwin, E., Hinds, J., & Cork, A. G. (2023). Platform-controlled social media APIs threaten open science. *Nature Human Behaviour*, 7(12), 2054–2057. <https://doi.org/10.1038/s41562-023-01750-2>

Department for Business and Trade. (2026). Smart Data 2035: The UK's Smart Data Strategy (CP 1551). <https://www.gov.uk/government/publications/smart-data-strategy>

Department for Digital, Culture, Media and Sport. (2020). National data strategy. <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy>

Department for Science, Innovation and Technology. (2026). Empowering people through data intermediaries [Consultation]. GOV.UK. <https://www.gov.uk/government/consultations/empowering-people-through-data-intermediaries/empowering-people-through-data-intermediaries>

Department for Science, Innovation and Technology. (2025, April 24). Online Safety Act: Explainer. GOV.UK. <https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer#types-of-content-the-act-tackles>

Desai, T., Ritchie, F., & Welpton, R. (2016). Five Safes: Designing data access for research (Working Paper No. 1601). Bristol Business School, University of the West of England. <https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf>

Digital Markets, Competition and Consumers Act 2024, c. 13. <https://www.legislation.gov.uk/ukpga/2024/13>

Dommet, K., Orben, A., & Zendle, D. (2025). Submission to Ofcom call for evidence: Researchers' access to information from regulated online services. Ofcom. <https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/call-for-evidence-researchers-access-to-information-from-regulated-online-services/responses/dommet-dr-k-orben-dr-a-and-zendle-dr-d.pdf?v=403885>

DSA 40 Collaboratory. (2026, June 12). #DSA40 Data Access Newsletter: The fine print is still being written (Issue 4). Weizenbaum Institute. https://dsa40collaboratory.eu/newsletters/04_fineprint/

DSA Observatory (n.d.). About the DSA Observatory. <https://www.dsaobservatory.eu/>

DSA Observatory. (2025, May 23). Workshop report: Researchers on data access and preparing for DSA Article 40(4). <https://dsa-observatory.eu/2025/05/23/researchers-on-data-access-and-preparing-for-dsa-article-404/>

DSA Observatory. (2026, March 12). If at first you don't succeed: Reflections on a rejected Art. 40 DSA data access request. <https://dsa-observatory.eu/2026/03/12/if-at-first-you-dont-succeed-reflections-on-a-rejected-art-40-dsa-data-access-request/>

Economic and Social Research Council (UKRI) & Smart Data Research UK. (2025). Response to Ofcom's call for evidence: Researchers' access to information from regulated online services. Ofcom. <https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/call-for-evidence-researchers-access-to-information-from-regulated-online-services/responses/economic-and-social-research-council-uk-research-an-innovation-smart-data-research-uk.pdf?v=403863>

Economic and Social Research Council (UKRI). (n.d.). Internet mediated research. <https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/internet-mediated-research/>

European Commission. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act). Official Journal of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065>

European Commission. (2025a). Commission Delegated Regulation (EU) 2025/2050 of 1 July 2025 supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council by laying down the technical conditions and procedures under which providers of very large online platforms and of very large online search engines are to share data with vetted researchers. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg_del/2025/2050/oj/eng

European Commission. (2025b, June 18). Commission makes AliExpress' commitments under the Digital Services Act binding [Press release]. <https://digital-strategy.ec.europa.eu/en/news/commission-makes-aliexpress-commitments-under-digital-services-act-binding>

European Commission. (2025c, December 5). Commission fines X €120 million under the Digital Services Act [Press release]. <https://digital-strategy.ec.europa.eu/en/news/commission-fines-x-eu120-million-under-digital-services-act>.

European Commission. (2026a, May 20). Commission holds roundtable on data access for vetted researchers. <https://digital-strategy.ec.europa.eu/en/news/commission-holds-roundtable-data-access-vetted-researchers>

European Commission. (2026b, July 15). Commission accepts X's corrective measures to terminate breaches of the DSA. <https://digital-strategy.ec.europa.eu/en/news/commission-accepts-xs-corrective-measures-terminate-breaches-dsa>

5Rights Foundation. (2025). Response to Ofcom's call for evidence on researcher access to information from regulated online services. Ofcom. <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/call-for-evidence-researchers-access-to-information-from-regulated-online-services>

Freelon, D. (2018). Computational research in the post-API age. *Political Communication*, 35(4), 665–668. <https://doi.org/10.1080/10584609.2018.1477506>

Freelon, D., Monzer, C., Jeon, G., Moy, C., & Williams, N. (2024). The post-API age of social media data access: Past, present, and future. *The ANNALS of the American Academy of Political and Social Science*, 711(1). <https://doi.org/10.1177/00027162251372557>

Goanta, C., Zannettou, S., Kaushal, R., van de Kerkhof, J., Bertaglia, T., Annabell, T., Gui, H., Spanakis, G., & Iamnitchi, A. (2026). The great data standoff: Researchers vs. platforms under the Digital Services Act. *Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)*, 874–888. <https://ojs.aaai.org/index.php/ICWSM/article/view/42671/50231>

Google. (n.d.). Data Portability API. Google for Developers. <https://developers.google.com/data-portability>

Halil, D., Kollnig, K., & Tamò-Larrieux, A. (2025). Regulating pressing systemic risks – but not too soon? *Internet Policy Review*, 14(2). <https://doi.org/10.14763/2025.2.2010>

Health Data Research UK. (n.d.). About HDR UK. <https://www.hdruk.ac.uk/about/>

House of Commons Science and Technology Committee. (2019). Impact of social media and screen-use on young people's health: Fourteenth report of session 2017–19 (HC 822). UK Parliament. <https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/822/822.pdf>

House of Commons Science, Innovation and Technology Committee. (2025). Social media, misinformation and harmful algorithms (HC 441). UK Parliament. <https://publications.parliament.uk/pa/cm5901/cmselect/cmsctech/441/report.html>

Information Commissioner's Office. (n.d.-a). Right to data portability. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/right-to-data-portability>

Information Commissioner's Office. (n.d.-b). The research provisions. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/the-research-provisions>

Information Commissioner's Office. (2024). Response to the consultation series on generative AI: The lawful basis for web scraping to train generative AI models. ICO. <https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/response-to-the-consultation-series-on-generative-ai/the-lawful-basis-for-web-scraping-to-train-generative-ai-models/>

Information Commissioner's Office. (2025). ICO launches consultations for Data (Use and Access) Act 2025 amendments. <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2025/08/ico-launches-consultations-for-data-use-and-access-act-2025-amendments/>

Information Commissioner's Office. (2026, February 12). How to deal with data protection complaints. <https://ico.org.uk/for-organisations/how-to-deal-with-data-protection-complaints/>

- Institute for Strategic Dialogue. (2025). Response to Ofcom's call for evidence on researcher access to information from regulated online services. Ofcom. <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/call-for-evidence-researchers-access-to-information-from-regulated-online-services>
- Internet Matters. (2025, June 23). Rising harms, new rules: Why the Online Safety Act matters. <https://www.internetmatters.org/hub/research/rising-harms-new-rules-why-the-online-safety-act-matters>
- Joint Committee on the Draft Online Safety Bill. (2021). Draft Online Safety Bill (HC 609 / HL Paper 129). UK Parliament. <https://publications.parliament.uk/pa/jt5802/jtselect/jtonlinesafety/129/12912.htm>
- King, G., & Persily, N. (2020). A new model for industry-academic partnerships. *PS: Political Science and Politics*, 53(4), 703–709. <https://doi.org/10.1017/S1049096519001021>
- Knight First Amendment Institute. (2022). A safe harbor for platform research. Columbia University. <https://knightcolumbia.org/content/a-safe-harbor-for-platform-research>
- Knight-Georgetown Institute. (2025a). Better Access: Data for the Common Good. <https://kgi.georgetown.edu/research-and-commentary/better-access/>
- Knight-Georgetown Institute. (2025b). Response to Ofcom's call for evidence on researcher access to information from regulated online services. Ofcom. <https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/call-for-evidence-researchers-access-to-information-from-regulated-online-services/responses/knight-georgetown-institute.pdf>
- Meta. (2023, November 21). New tools to support independent research. Meta Newsroom. <https://about.fb.com/news/2023/11/new-tools-to-support-independent-research/>
- Meta. (n.d.). Data portability. Meta for Developers. <https://developers.facebook.com/docs/data-portability>
- Mimizuka, K., Brown, M. A., Yang, K. C., & Lukito, J. (2025). Post-post-API age: Studying digital platforms in scant data access times. *arXiv preprint arXiv:2505.09877*.
- Ofcom. (2025). Researchers' access to information from regulated services: Report prepared under section 162 of the Online Safety Act 2023. <https://www.gov.uk/government/publications/researchers-access-to-information-from-regulated-services>
- Ofcom. (2026). Crisis response protocol: Online Safety – Additional Safety Measures (para. A1.53). <https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/protecting-children/illegal-harms/crisis-response-protocol-online-safety-additional-safety-measures.pdf>
- Ohme, J., & Seiling, L. K. (2026). Finally, access: How Article 40 DSA changes platform research in practice. *Political Communication*. <https://doi.org/10.1080/10584609.2026.2664157>
- Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2024). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, 18(2), 124–141. <https://doi.org/10.1080/19312458.2023.2181319>
- Online Safety Act 2023, c. 50. <https://www.legislation.gov.uk/ukpga/2023/50>
- Open Data Institute. (2025). Response to Ofcom's call for evidence on researcher access to information from regulated online services. Ofcom. <https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/call-for-evidence-researchers-access-to-information-from-regulated-online-services>
- Orben, A., Dommett, K., Scott, M., Zendle, D., Fokas, I., & Boutopoulou, A. (2026). Social Media Data Access to Reduce Online Harms: A Call for Action. *Social Platforms Data Access Taskforce*. <https://doi.org/10.17863/CAM.131053>
- Perriam, J., Birkbak, A., & Freeman, A. (2020). Digital methods in a post-API environment. *International Journal of Social Research Methodology*, 23(3), 277–290. <https://doi.org/10.1080/13645579.2019.1682840>
- Reuters Institute for the Study of Journalism. (2025). Generative AI and news report 2025: How people think about AI's role in journalism and society. University of Oxford. <https://reutersinstitute.politics.ox.ac.uk/generative-ai-and-news-report-2025-how-people-think-about-ais-role-journalism-and-society>

- SAIL Databank. (n.d.). About SAIL Databank. <https://www.swansea.ac.uk/research/research-highlights/health-innovation/sail-databank/>
- Santini, R. M., Leal, H., Salles, D., Belisário, A., Mattos, B., & Pinho, D. (2026). Data Not Found: Social Media Data Transparency for Information Integrity. NetLab UFRJ & Minderoo Centre for Technology and Democracy. <https://doi.org/10.17863/CAM.128975>
- SDDS. (2025). Data donation: Barriers and legislative solutions (Internal document v1, 18 September 2025). Taskforce material.
- Shiffman, N., & Silverman, B. (2025). The case for transparency: How social media platform data access leads to real-world change. Institute for Data, Democracy and Politics, George Washington University. https://iddp.gwu.edu/sites/g/files/zaxdzs5791/files/2025-05/case_for_transparency_shiffman_silverman.pdf
- Skatova, A., Ng, E., & Goulding, J. (2014). Data donation: Sharing personal data for public good? N-Lab, Application of Digital Innovation. https://www.researchgate.net/publication/269101036_Data_Donation_Sharing_Personal_Data_for_Public_Good
- Skatova, A., & Goulding, J. (2019). Psychology of personal data donation. PLOS ONE, 14(11), e0224240. <https://doi.org/10.1371/journal.pone.0224240>
- Smart Data Research UK. (n.d.). Smart Data Donation Service. <https://www.sdruk.ukri.org/data/smart-data-donation-service/>
- Social Platforms Data Access Taskforce. (2025a). Social platforms data access webinar on scraping: Executive report for DSIT. SDRUK / University of Sheffield.
- Social Platforms Data Access Taskforce. (2026a). Interviews with legal departments and civil society: Common legal challenges detected. Internal report for Policy Advisory Group meeting, 23 January 2026.
- Social Platforms Data Access Taskforce. (2026b). Written evidence submitted to the Science, Innovation and Technology Committee inquiry into platform data access. UK Parliament, March 2026.
- Social Platforms Data Access Taskforce. (2026c). Internal evidence log: Researcher experiences of DSA Article 40(12) applications. Unpublished internal document.
- TikTok. (n.d.). Data portability API. TikTok for Developers. <https://developers.tiktok.com/products/data-portability-api>
- TikTok for Developers. (n.d.). About Research API. <https://developers.tiktok.com/doc/about-research-api/>
- Timberg, C. (2021, September 10). Facebook made big mistake in data it provided to researchers, undermining academic work. The Washington Post. <https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/>
- Tromble, R. (2021). Where have all the data gone? A critical reflection on academic digital research in the post-API age. Social Media + Society, 7(1), 1–8. <https://doi.org/10.1177/2056305121988929>
- UK General Data Protection Regulation 2018 (as retained in UK law by the European Union (Withdrawal) Act 2018, as amended). <https://www.legislation.gov.uk/eur/2016/679>
- UK Parliament. (1988). Copyright, Designs and Patents Act 1988 (c. 48).
- UK Parliament. (2023). Online Safety Act 2023 (c. 50), Section 162. <https://www.legislation.gov.uk/ukpga/2023/50/section/162>
- UK Parliament. (2025). Data (Use and Access) Act 2025. <https://www.legislation.gov.uk/ukpga/2025/18/contents>
- UK Research and Innovation. (2025). Terms and conditions of FEC grants. <https://www.ukri.org/publications/terms-and-conditions-for-research-grants/ukri-terms-and-conditions-of-fec-grants/>
- University of Michigan. (n.d.). Social Media Archive (SOMAR). <https://somar.umich.edu/>
- Venturini, T., & Rogers, R. (2025). Digital methods: A short introduction (Chapter 6: Scraping documents and interfaces). Manchester University Press.

Appendix

I. Key terms and definitions

The following definitions apply throughout this report. Where terms carry specific statutory meanings under UK legislation, the relevant provision is noted. Where terms are used in a way that differs from or narrows their common usage, this is made explicit.

Application programming interface (API)

A structured technical interface that allows software systems to communicate with a platform's servers, enabling automated, programmatic retrieval of data within parameters defined by the platform. Platforms use APIs to provide controlled access to their data for approved third parties, including researchers. API access is distinct from web scraping in that it operates through formally provided and documented data pipelines rather than automated extraction from public-facing interfaces.

Data access / researcher data access

For the purposes of this report, data access refers to the ability of independent researchers to obtain social media platform data for research into online safety matters. This encompasses access via official APIs, consent-based data donation, web scraping of publicly accessible data and access mediated through an independent intermediary body. The term does not refer to access for commercial or platform-operational purposes.

Data donation

A consent-based process through which individuals voluntarily share their personal platform data with researchers. In its most common form, this involves a user requesting a copy of their personal data from a platform under the right to data portability (Article 20, UK GDPR), downloading that data to their own device and choosing to share some or all of it with a research team. See section 3.3 for a full discussion of the legal framework and technical approaches.

Data enrichment / hydration

The process of supplementing an existing dataset with additional information retrieved from external sources to improve its completeness and analytical usability. In social media research, this typically involves retrieving contextual information about content items (such as full post text, metadata or engagement data) from the original platform using content identifiers contained in a donated or scraped dataset. The term "hydration" is used specifically in social media research contexts, originating from the practice of retrieving full content from tweet IDs on X (formerly Twitter).

Data intermediary

An independent body that mediates access to data between data holders (such as social media platforms) and data users (such as researchers), operating under statutory or regulatory oversight and without a commercial interest in the data it handles. The term is used in this report specifically to refer to a proposed social media data intermediary of the kind recommended in section 3.4. See section 3.4.1 for a full typology of intermediary models.

Data protection impact assessment (DPIA)

A structured assessment, required under Article 35 of the UK GDPR for processing likely to result in high risk to individuals' rights and freedoms, that documents the purpose, necessity and proportionality of a data processing activity, identifies risks to data subjects and sets out the measures taken to mitigate them. For researchers, a DPIA serves both as a due diligence exercise and as evidence that can be presented to institutions, funders and platforms to demonstrate that a research project, including one involving web scraping, has been assessed as lawful.

Independent research / independent researcher

See Researcher / independent researcher / public interest researcher.

Online harms / online safety matters

Online safety matters, as defined under the Online Safety Act 2023, refers to harms arising from content or activity on regulated services, including child sexual abuse material, terrorism, fraud, hate speech and content that encourages self-harm or suicide. The term online harms is used more broadly throughout this report to refer to the full range of harmful

content, conduct and contact facilitated by social media platforms. Both terms are used consistently with the regulatory framework established by the Online Safety Act 2023.

Post-API age

A term widely used in the research community to describe the current period in which major social media platforms have significantly restricted or withdrawn the application programming interfaces previously made available to independent researchers, substantially reducing the routes through which systematic, independent access to platform data can be obtained. The term was coined by Freelon (2018) and has been widely adopted in the academic literature on platform research. See section 1.3 for a full discussion of the current data access landscape.

Privacy enhancing technologies (PETs)

Technical tools and approaches designed to minimise privacy risks associated with data collection, sharing and analysis while preserving the analytical value of data. In the context of social media research, PETs include local processing frameworks such as PORT, differential privacy techniques and secure multi-party computation. They are increasingly recognised as a necessary condition for responsible data access in sensitive research contexts, particularly where special category data is involved.

Public interest research

Research conducted for the benefit of society rather than for commercial gain, typically by academic researchers, civil society organisations, journalists or think tanks. For the purposes of this report, public interest research refers specifically to independent research into online safety matters, consistent with the framework established by the Online Safety Act 2023 and the Data (Use and Access) Act 2025. See also Researcher / independent researcher / public interest researcher.

Researcher / independent researcher / public interest researcher

These terms are used interchangeably throughout this report to refer to any party conducting research independently of the platforms whose data is being studied. This includes academic researchers, civil society organisations, journalists, think tanks and other organisations engaged in research for public

benefit rather than commercial gain. This definition is consistent with section 162(2) of the Online Safety Act 2023, which defines independent research as research carried out on behalf of a person other than a provider of a regulated service. Eligibility for access should be determined by the public interest nature and governance of the research, not by institutional affiliation.

Regulated service

A service regulated under the Online Safety Act 2023. Regulated services comprise user-to-user services (any service enabling users to share content or interact, including social media, messaging apps, forums, dating and gaming services), regulated search services and, under Part 5 of the Act, providers of pornographic content. Duties apply to all in-scope services but scale with size and risk: the largest and highest-reach services are designated Category 1, 2A or 2B and carry additional obligations (for example, transparency reporting), while smaller services assessed as high-risk are subject to separate ‘small but risky’ supervisory arrangements. This report focuses on major consumer-facing social media platforms, where the barriers to research access documented here are most acute, but its recommendations are intended to apply across the full range of regulated services.

Social media platform / platform / service

These terms are used interchangeably throughout this report to refer to regulated user-to-user services as defined under section 4(4) of the Online Safety Act 2023, specifically those services that enable user-generated content to be shared publicly or semi-publicly at scale. The report focuses on major consumer-facing social media platforms including but not limited to Meta (Facebook and Instagram), X (formerly Twitter), TikTok, YouTube and Reddit.

Subject Access Request (SAR)

A request made under Article 15 of the UK GDPR through which an individual can obtain a copy of the personal data an organisation holds about them, together with information about how and why that data is processed. A SAR is distinct from a data portability request: its purpose is transparency rather than reuse, and its output is typically human-readable (for example, a PDF or collection of documents) rather than the structured, machine-readable format required for data portability. In data donation contexts, platforms sometimes respond to portability requests with SAR-style outputs, which

are less suitable for research because they cannot be readily processed at scale.

Trusted research environment (TRE)

A secure, access-controlled infrastructure in which researchers can access and analyse sensitive datasets without data being extracted or downloaded to their own systems. TREs apply governance frameworks to ensure data are used safely and only for approved purposes. The Five Safes model (covering safe people, safe projects, safe data, safe settings and safe outputs) is widely used to structure TRE governance in the UK. Examples relevant to this report include the ONS Secure Research Service and the SAIL Databank.

Web scraping / public interest web scraping

Web scraping refers to the automated extraction of data from websites or platform interfaces using software tools, outside of formally provided API pipelines. Public interest web scraping refers specifically to scraping conducted by independent researchers for purposes in the public interest, as distinct from commercial scraping conducted for profit or AI training. See section 3.2 for a full discussion of the legal framework and the Taskforce's recommendations.

II. Key Actors in the UK's Data Access Landscape

There are a number of key actors, institutions, and initiatives that currently shape debates and practices on UK data access. Particular attention is given to the role of UK Research and Innovation and Smart Data Research UK (SDRUK), as commissioners of this report and as central convenors within the UK data access ecosystem.

For analytical clarity, UK actors are grouped into three broad categories:

Regulators and legislative actors

The UK Government Responsibility for data access has lain with a range of Governmental Actors. In existence between 2023-2026, the Department for Science, Innovation and Technology (DSIT) provided strategic policy leadership across data, digital regulation and innovation. DSIT played a

coordinating role across the Online Safety Act, the DUAA and wider digital governance initiatives and acted as a key interface between regulators, research funders and industry. While DSIT did not itself operate data access mechanisms, its policy direction and sponsorship of legislative frameworks materially shape the scope, priorities and future trajectory of data access in the UK. DSIT's duties transferred to the Department for Business, Innovation, Science and Trade (DBIST) and the Department for Culture, Media and Sport (DCMS).

The Information Commissioner's Office (ICO) is a non-departmental public body sponsored by the Government to regulate data protection and information access. While the ICO does not directly mandate data sharing, its guidance on data protection, lawful bases for processing, anonymisation and research exemptions under UK GDPR substantially shapes the boundaries of what forms of access are considered permissible. The ICO's position has therefore been influential in both enabling and constraining research access, particularly in contexts involving personal or pseudonymous platform data.

Ofcom is the regulator for communications services, overseeing rules for broadcasting, the internet, telecommunications and postal industries in the UK. Ofcom's role in the UK's data access debate stems from a one-off statutory duty under section 162 of the Online Safety Act (OSA) 2023, which required it to report on the adequacy of researcher access to platform data. Ofcom discharged this duty by publishing its report in July 2025 (Ofcom, 2025) and holds no ongoing statutory duties in relation to researcher access. Nor does the ICO. The Act itself does not create a research access framework: section 154A (as amended by the Data (Use and Access) Act 2025) gives the Secretary of State the power to create one, should they choose to do so. No organisation has yet been designated a role or responsibility under any future framework, beyond a duty on the Government to consult Ofcom, the ICO and other persons as the secretary of state considers appropriate as part of the process of drafting regulations.

Key academic research infrastructure

Smart Data Research UK (SDR UK) is a flagship UK Research and Innovation (UKRI) investment designed to provide safe, secure and sustainable access to "smart data" for public-interest research. SDR UK functions as the primary national vehicle for developing shared technical infrastructure, including the expansion and standardisation of Trusted Research Environments (TREs) and for

piloting governance models that enable access to sensitive platform data while maintaining high standards of security and ethical oversight.

The UKRI Social Platforms Data Access Taskforce, convened by UKRI and led by the University of Sheffield in partnership with the University of Cambridge, operates as a coordinating body to align academic research needs with emerging technical standards, legal constraints and regulatory expectations. The Taskforce brings together researchers, regulators, funders and civil society actors to identify barriers to access and to develop practical, policy-relevant proposals for improving the UK data access landscape.

Academic and independent research organisations

The UK benefits from a dense ecosystem of academic and civil society organisations engaged in the study of digital platforms. Academic centres including the Oxford Internet Institute, the Minderoo Centre for Technology and Democracy at the University of Cambridge and HateLab at Cardiff University produce internationally recognised research on disinformation, platform governance and online harms. These efforts are complemented by civil society organisations among which, the Ada Lovelace Institute, the Institute for Strategic Dialogue, the Centre for Countering Digital Hate, Reset and Demos, which contribute applied research, real-time monitoring, and policy-facing analysis.

Social Platforms Data Access Taskforce

Final Report, September 2026

A year of evidence: insights and recommendations on social media data access for research into online harms.

Dr Alexandra Boutopoulou, Professor Katharine Dommett,
Professor Amy Orben, Mark Scott, Dr David Zendle

