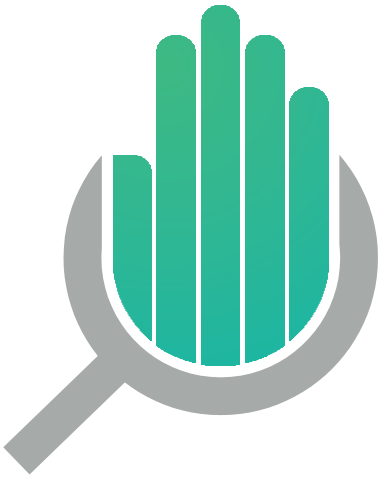
REPHRAIN

Protecting citizens online



REPHRAIN's response to Ofcom's Call for Evidence: Researchers' Access to Information from Regulated Online Services

January 2025

UK Research and Innovation · University of Bristol · UCL · The University of Edinburgh · King's College London · University of Bath

REPHRAIN
Protecting citizens online

REPHRAIN's response to Ofcom's Call for Evidence: Researchers' Access to Information from Regulated Online Services

Thank you for the opportunity to provide our response to this call for evidence. We are writing on behalf of REPHRAIN, the National Research Centre on Privacy, Harm Reduction and Adversarial Influence Online. REPHRAIN is the UK's world-leading interdisciplinary community focused on the protection of citizens online.

Led by the University of Bristol and partnered with University College London, King's College London, the University of Edinburgh, and the University of Bath, REPHRAIN unites experts across disciplines such as Computer Science, Law, Psychology, and Public Policy to explore how to keep people safe online while enabling full participation in digital technologies. Announced by UKRI in October 2020, REPHRAIN now has over 100 experts from 23 UK institutions, working across 50+ research projects to address our missions:

- Delivering privacy at scale while mitigating its misuse to inflict harms
- Minimising harms while maximising benefits from a sharing-driven digital economy
- Balancing individual agency vs. social good.

Executive summary

- **Overall, collecting data from online services is increasingly challenging for academic researchers.**
- Barriers to accessing data include:
    - **Financial costs** – X's API, which was previously free in legacy Twitter, is now unfeasibly costly for researchers, since it now starts at $42,000 per month.
    - **Administrative burdens** – Processes for ethics approval, applying for data access, and reaching a data sharing agreement, are unnecessarily burdensome and lengthy.
    - **Unclear reasons for rejection** – Online services can reject researchers' applications for data access for unclear reasons, as discovered by the Data Access Collaboratory (2024).
    - **A lack of standardised processes** – The onus is often on researchers to find a pathway to accessing data from online services, costing them valuable time and funding.
- **Online services have too much influence over the sharing of data from their platforms –** Providing access to data is often not within their interests, since they risk reputational damage if found to be responsible for adverse consequences.

Recommendations

- **Mandate online services to provide data access** – Data access for research into online safety matters must be mandated, with online services incurring financial penalties if they refuse.
- **Standardise processes** – Applying for data access should follow a standardised, seamless protocol, which would remove the administrative burden from researchers.
    - This could follow a 'traffic light' system, in which publicly available data is made much more accessible, with sensitive information carefully safeguarded.
- **Utilise third-party organisations** – To remove subjectivity, third-party organisations could serve as intermediaries between online services and researchers by:

    - Making decisions regarding data access applications
    - Providing a secure holding site for data
    - Reducing contact points between parties
    - Offering ethical and legal oversight
    - Providing guidance for platforms.
- **Recognise and align with EU regulations** – Ofcom should consider the impact of Article 40 of the Digital Services Act on UK researchers, recognise the similarities between their and the European Commission's aims, and align their proposals to this Act. This would enhance the credibility of the UK's regulations.

Please find a detailed response to your questionnaire below.

Consultation title	**Call for Evidence: Researcher Access to Regulated Online Services Information**
Full name	Dr Ignacio Castro, Lecturer in Data Analytics, Queen Mary University of London Josie Curtis, Policy Engagement Associate, REPHRAIN, University of Bristol Dr Leonie Tanczer, Associate Professor in International Security and Emerging Technologies, University College London Dr Mark Warner, Lecturer in Information Security, Dept of Computer Science, University College London Prof. Stephan Lewandowsky, Chair in Cognitive Psychology, School of Psychological Science, University of Bristol Dr Tariq Elahi, Lecturer of Security and the Internet of Things, School of Informatics, University of Edinburgh
Contact phone number	N/A
Representing (delete as appropriate)	Organisation
Organisation name	REPHRAIN
Email address	rephrain-centre@bristol.ac.uk

Confidentiality

We ask for your contact details along with your response so that we can engage with you on this consultation. For further information about how Ofcom handles your personal information and your corresponding rights, see Ofcom's General Privacy Statement.

Your details: We will keep your contact number and email address confidential. Is there anything else you want to keep confidential? Delete as appropriate.	Nothing
Your response: Please indicate how much of your response you want to keep confidential. Delete as appropriate.	None
For confidential responses, can Ofcom publish a reference to the contents of your response?	N/A

Your response

Question	Your response
Question 1: How, and to what extent, are persons carrying out independent research into online safety related issues currently able to obtain information from providers of regulated services to inform their research?	Confidential? – N **<u>Current situation – limited access to APIs</u>** Collecting data from online services is increasingly challenging for academic researchers. Previously, much of our research was carried out on X (formerly Twitter), since Twitter provided a free Application Programming Interface (API) to researchers. With this, researchers could gather tweet IDs and then "hydrate" them (use the API to retrieve detailed metadata such as timestamps, geolocation, and user information). Hydrating data means that user-deleted data is inaccessible, which provides users with the control and privacy they deserve. This enabled researchers from REPHRAIN and beyond to access information on online harms, such as the spread of mis/disinformation, hate speech, and cyberbullying and harassment. The process of accessing Twitter's API ensured compliance with their terms of service and privacy policies while enabling researchers to study real-world phenomena responsibly. REPHRAIN projects which used the free Twitter API include the *Key2Kindness* project, which aimed to uncover the effectiveness of a more proactive, rather than reactive, approach to content moderation (REPHRAIN, 2025a). Using Twitter's free API, our researchers were able to simulate a public online service, like Twitter. They then used different language models to detect different types of toxic content. However, such a project may be unfeasible nowadays. Following Elon Musk's acquisition of the platform in 2022, the cost of accessing X's API starts at $42,000 per month, rendering it virtually impossible for academic researchers to use in-depth data from X within projects (Hutchinson, 2024). There are a few ways around this cost – for example, by navigating X's data access application process to gain a more limited access to data – but as will be discussed, this comes with countless administrative barriers. This drastically hinders vital computational social science research being conducted on crucial topics, including radicalisation, military propaganda, and racial discrimination. This is an issue we raised in our September 2023 report, *Making Sense of the Twitter Takeover* (REPHRAIN, 2023a). But the decreasing accessibility of data for researchers is not unique to X. Our recent paper, '*"Edit: I'm sorry for being offensive, this is getting downvoted and I feel terrible": Implicit Social Norms as Governance in Identity-Based Communities',* explores self-governance and self-regulation across various community spaces,

Question	Your response
	focused on the social media platform Reddit (Beadle, Warner and Vasek, 2025). It utilised data obtained from Pushshift.io, an external API used for accessing data from Reddit, and developed by academics. Pushshift.io provided enhanced access to Reddit data by combining data acquired from Reddit's official API with additional datasets created by periodic scraping. This approach allowed the researchers to overcome some of the limitations of Reddit's official API, which offers restricted access to data. However, Reddit eventually revoked Pushshift.io's access, meaning the API is no longer available. While Reddit's official API still exists, it doesn't provide the same breadth of functionality or ease of use for research as Pushshift.io did. Further, Meta's discontinuation of CrowdTangle, a social media analytics tool that helped researchers, journalists, and content creators track, analyse and monitor content, identify trends, and understand the spread of information, has hindered research profoundly. This was particularly useful for understanding online harms such as misinformation and political polarisation. Meta replaced CrowdTangle with the Meta Content Library, yet this has serious limitations including incompleteness, and the inability to transfer data to a CSV file or search within an image, limiting research capabilities (Lobo, 2024). **These cases highlight the challenges researchers face when relying on APIs or external tools for data access, as platform policies can abruptly change**, limiting their ability to collect or share data for research purposes. Other barriers imposed by online services include **highly bureaucratic processes for researchers applying for data access**. Researchers often must navigate complicated, unnecessarily burdensome processes, such as having to submit a detailed description of their research project including a literature review, in the case of TikTok (Correia de Carvalho, 2024). Resultantly, researchers at REPHRAIN and elsewhere are discouraged from applying. **<u>Unclear reasons for data access request rejection</u>** When researchers do apply for data access, many of their applications are rejected for unclear reasons. The Data Access Collaboratory have started compiling a tracker for researchers' applications to access data from online services under the Digital Services Act (DSA) (Data Access Collaboratory, 2024). Of the 24 applications in the tracker that have been decided by online services so far, 15 have been accepted and 9 rejected. Interestingly, there is great disparity between online services. Of

Question	Your response
	the 10 decided applications from TikTok, 9 were accepted, whilst of the 13 decided applications from X, 8 were rejected. The most frequent reason for rejection was that the research project did not adequately convey that the data collected would be used for research contributing to the detection, identification and understanding of systemic risks in the EU under Article 34 of the DSA (Data Access Collaboratory, 2024). **Allowing platforms to determine whether a researcher's request qualifies under the DSA gives them significant flexibility in interpreting what constitutes systemic risks.** The DSA provides only broad definitions of systemic risks, such as the spread of illegal content, misinformation, or threats to democracy, without offering detailed criteria for assessing research proposals. This **lack of standardisation opens the door to subjective decision-making by platforms. Instead, an independent body, like Ofcom, should have the decision-making power over data access applications.** Moreover, the Data Access Collaboratory also found that TikTok's average decision-making wait time was 37.5 days, whereas X was 71.23 days, with an average of 116.2 days to accept a decision. This **lengthy wait time prohibits researchers from conducting timely research into pressing online harms.** **<u>Online services have too much influence</u>** To try to enable greater researcher access to data from online services, the Social Science One project, launched in 2018, aimed to create partnerships between academia, industry, and funding bodies. However, the project has been met with challenges from private companies, with Facebook handing over what was deemed to be 'incomplete data' to researchers (Murgia, Criddle, and Murphy, 2021). This demonstrates the problems inherent in platforms having full control over their data, with no accountability measures. Here, **it must be remembered that platforms have a vested interest in limiting researcher access to avoid potential risks**. This includes legitimate risks, such as data breaches, misuse of information, or legal liabilities, but also the **overarching risk of potential reputational damage if researchers find that the platforms are responsible for adverse consequences.** By making applications arduous, rejecting research proposals under the guise of not aligning with the DSA's objectives, or by lengthening the decision wait time, platforms can reduce exposure to regulatory scrutiny or public backlash. This dynamic highlights the **need for greater transparency and independent oversight of platforms to enhance researchers' access to data.**

Question	Your response
Question 1a: What kinds of online safety research does the current level of access to information enable? • What type of independent researchers are carrying out research into online safety matters? • What topics/issues they are researching?	Confidential? – Y / N
Question 1b: Are there types of information that independent researchers are currently unable to access that may be relevant to the study of online safety matters? If so, what are they and what kind of research would they facilitate?	Confidential? – Y / N
Question 1c: What data governance models are currently used to allow access to online services' information for researchers? • This might include: open-access forms of information-sharing, such as publicly-accessible information libraries or databases; information-sharing models that rely on vetting or accreditation of individuals or organisations; and/or models that rely on the accreditation of the specific use cases for the information. • Please provide relevant examples of these governance models used in the online services industry.	Confidential? – Y / N Please see our discussion regarding the Digital Services Act (page 14).
Question 1d: What technologies are typically used by providers of online services to facilitate existing information access?	Confidential? – Y / N None

Question	Your response
Question 1e: Have services and/or researchers made use of privacy-enhancing technologies to enable access?	Confidential? – Y / N

Question	Your response

Question 2: What are the challenges that currently constrain the sharing of information for the purpose of research into online safety related issues?	Confidential? – Y / N
	
	Alongside the aforementioned barriers, such as online services reducing the accessibility of APIs and increasing the administrative burden of applying for data access, further problems arise from:
	
	• Researchers struggling to find a contact point within an organisation
	• Researchers having to develop adequate rapport with this contact point to be able to set up a data sharing agreement
	• Conflicts between the legal teams of universities and online services
	• Lengthy ethics approval processes
	  ○ These processes put the onus on researchers to justify the right to scrutinise online services. They also create a huge overhead for institutions and researchers themselves
	• Limited practical guidance on how data controllers should facilitate data transfers, meaning that data portability is inconsistent amongst online services
	  ○ Whilst the right to data portability is enshrined within the General Data Protection Regulation (GPDR), our findings from interviews with academics and industry experts demonstrate that many researchers cannot make use of it when applying for data access (Turner and Tanczer, 2024).
	
	Sometimes universities prefer that data is cleaned and processed by the data provider, since they believe it is too high risk to be stored within their institution. Other times, data is obtained physically from an organisation via a USB stick, or else uploaded onto a cloud for researchers to download.
	
	Ultimately, there are no standardised processes for accessing data, and it is the **researcher's responsibility to figure out how to access data** by navigating numerous teams and processes each

Question	Your response
	time. Therefore, **online services should be mandated to provide a standardised process for data access**. However, it is important to note that the **problems do not only arise from online services; universities can also impose various obstacles to conducting research**. Often, like online services' legal teams, it appears that universities' legal teams see it as too risky, if it has the potential to lead to reputational damage.
Question 2a: What are the legal challenges/risks to sharing information from online services with independent researchers?	Confidential? – Y / N
Question 2b: What are the technical challenges relating to sharing information from online services with independent researchers? What are the challenges relating to the scale and complexity of the information involved?	Confidential? – Y / N
Question 2c: What are the security challenges relating to sharing information from online services with independent researchers? • What are the security challenges relating to the potential sensitivity of information? • What are the security protocols required to protect information from misuse? • To what extent do you view security as a governance issue compared to a technical infrastructure issue?	Confidential? – Y / N **Security challenges** REPHRAIN projects, which have focused on private or anonymous communication platforms, have struggled to access data due to security challenges. For instance, the *Key2Kindness* project had to simulate a private communication platform akin to WhatsApp, due to WhatsApp's end-to-end-encryption. We have faced similar barriers when studying anonymous communication platforms like Tor. In one project, our aim was to understand which websites users were accessing. We were able to gain partial access to data by directly participating in the system by operating machines within the Tor network, allowing us to capture traffic routed through those machines. However, we were only able to see the traffic that passed through our own machines and we did not have access to the whole network. Tor, like other platforms offering regulated services, provides only restricted, high-level statistical information through public APIs. The restrictions on access are in place to enhance safety, ensuring that detailed information about the entire network cannot be exploited for malicious purposes. Therefore, the main challenge with accessing data from this type of organisation is trying to persuade platforms that they should provide this information.

Question	Your response
	Platforms do of course have legitimate concerns about the safety of users, and legal requirements to protect personally identifiable information. In another project, our aim was to investigate users of a VPN service to find out when they used the service and what websites they accessed. Of course, this was met with blockades from the VPN service, since their purpose is to provide user anonymity. Eventually, we found a way around this and were able to access the data after a burdensome bureaucratic process. Despite this, **the VPN service did not make this a pathway for future researchers to draw upon, even though it would save both researcher and service time**. Often, **even after accessing data from a service, researchers must go through the same process from start to finish with the same organisation to access further data, wasting valuable time and resources.** This demonstrates how the **onus is on researchers to create their own pathway to data access,** rather than on organisations being required to put standardised procedures in place.
Question 2d: What are the information quality challenges relating to online services sharing information with independent researchers?	Confidential? – Y / N
Question 2e: What are the financial costs to online services relating to online services sharing information with independent researchers? (won't be able to answer this one)	Confidential? – Y / N
Question 2f: What are the financial costs to researchers trying to make use of information shared by online services?	Confidential? – Y / N Due to excessive costs for accessing data from certain online services, namely X, the **cost of many projects is now mainly in terms of labour hours** for perhaps reverse engineering a platform, or ethics approval processes.

Question	Your response
Question 3: How might greater access to information for the purpose of	Confidential? – Y / N **Standardised processes**

Question	Your response
research into online safety issues be achieved?	As mentioned, there needs to be standardised processes for researchers to access data from online services. This would save both researchers' and services' time. This should include the requirement for greater consistency across the data that is provided to researchers by platforms, with regards to formatting and naming conventions. This would enable researchers to compare and contrast data across different platforms. **'Traffic light' system** For data access, there should be a clear distinction between platforms that host public data and those that handle sensitive, private information. A **traffic light system** could be an effective approach: green for public data accessible on mainstream platforms, and red for sensitive topics, such as private communications on platforms dealing with issues like child sexual abuse. This system would help prioritise ethical considerations and safeguard privacy while still enabling responsible data collection for research purposes. 'Green data' should be easily accessible for researchers, and platforms should be mandated to provide it. **Beyond privacy risk** However, the traffic light system must not only consider sensitive data in terms of privacy risk. The implications of research must also be considered. For instance, Chung et al's (2017) study investigated what private information may be inferred from publicly available data on Event-based social network, Meetup. They found that sensitive information such as LGBT status could be predicted with 93% accuracy. Information such as this could be used to create models for misuse. This highlights a key issue: **even if users are comfortable with their anonymised data being used for research, careful thought must be given to the potential outcomes and risks of the resulting analysis or models.** Similarly, our *CSAC* project, which developed a child sexual abuse conversation dataset, aimed to advance our understanding of how perpetrators of child sex grooming engage online with young people through computer-mediated communication tools and platforms (REPHRAIN, 2025b). This project laid the foundations for developing reactive and proactive mechanisms for limiting this behaviour across platforms. Although this data is of course highly sensitive information, the risk here is not just in terms of privacy, but such data could be misused to develop a model to automate grooming.

Question	Your response
	Therefore, a nuanced approach to assessing risk from data access is essential. Beyond safeguarding privacy, it is critical to evaluate the potential applications and implications of research findings to ensure they do not inadvertently enable harmful uses of data. This requires a framework for responsible action, similar to "responsible disclosure" in vulnerability research. For example, if researchers discover that sensitive inferences can be made from publicly available data, they should disclose these findings to the relevant platforms. The platforms, in turn, should collaborate with researchers to implement mitigations that limit the potential for misuse, ensuring the findings are applied ethically and do not amplify harm. **Data (Use and Access) Bill** Due to our recommendation for standardised processes, **we support the provisions in the Data (Use and Access) Bill which may mandate that platforms provide information to researchers, and that they will be faced with penalties if they refuse** (Data [Use and Access] Bill, 2024). This could effectively hold online services to account. Developing “researcher access notices” which would set guidelines for procedures, data access protocols and security standards, is also a positive step towards standardisation across services and therefore enhanced data accessibility for researchers (Data [Use and Access] Bill, 2024; pp. 153-155). However, as noted, it is not only online services which impose barriers to data access; universities can make it difficult due to administrative burdens. Therefore, **alongside standardised processes for platforms, there also must be a standard for universities.** Platforms need assurance that universities are capable of securely handling data, and that research requests are legitimate.
Question 3a: What models, arrangements or frameworks exist for allowing researchers access to sensitive information beyond the online services industry? What are the benefits and risks of those models, and how might they apply to the online services context?	Confidential? – Y / N The importance of standardisation of procedures is evident from examples beyond the online services industry. **Examples from public services** For instance, the procedures for researchers gaining access to NHS data is standardised and it simplifies the process. This includes the new NHS England Secure Data Environment, in which approved researchers can access anonymised data from patients through a secure research portal. In this way, no identifying data ever leaves the server, greatly enhancing data protection. Although this is not

Question	Your response
	free, it is a more accessible fee for researchers at recognised institutions (NHS England, 2025). The ONS has a similar process; a researcher must become an 'accredited researcher' by undertaking the ONS Safe Researcher Training (Office for National Statistics, 2025). Similar approaches could be adopted by the online services industry. Moreover, it is essential to streamline the points of contact between research institutions and online platforms. For example, one of our REPHRAIN researchers has experience working within a telecommunications unit for law enforcement. This unit was responsible for requesting communications data from telecommunications providers, which involved numerous police officers calling providers to request information and ultimately overwhelming the provider, slowing down the process. As a result of these problems, the unit transitioned to the model of having a single point of contact between themselves and the provider. This enabled the process to be streamlined, and relationships between both parties to be built and maintained. Within the context of online services, we can learn from this example, by reducing points of contact between researchers and platforms.
Question 3b: Are there any models or arrangements that exist in the online services industry already that might provide increased access to information for research purposes if applied more generally across the industry? If so, what are these and what are the benefits and disadvantages of these models/arrangements?	**API models** The original Twitter API model provided an excellent framework that could be revisited. It allowed researchers to query the API to collect data identifiers, which could then be "hydrated" to retrieve associated metadata. This model was highly beneficial for research as it ensured reproducibility and gave users control over their data, enabling them to exercise their right to be forgotten by deleting tweets, which would then become inaccessible via the API. **Requiring platforms to provide accessible APIs could be a model for enabling greater access to data for researchers**. One of the challenges of using an API model is validating the credentials of the users who can access the API. However, in the case of legacy Twitter, it was simple: if a user had an educational email address, then access would be granted. This approach could work well for researchers in academia. **'Clean rooms'** Another pathway to data access has been exhibited by Facebook in collaboration with Social Science One. Facebook mandates that researchers physically attend their offices and sit in a 'clean room', so that the data cannot leave the organisation. However, this is very expensive and inaccessible for many researchers. Yet,

Question	Your response
	technological solutions to get around this exist – for example, there are now virtual clean rooms available. **<u>Digital Services Act</u>** Yet there does already exist an important, overarching framework that aims to enhance researcher access to data: the EU's recent draft delegated act which lays down the specific conditions under which researchers will be able to access data from large online platforms and search engines under Article 40 of the DSA (European Centre for Algorithmic Transparency, 2024). This delegated act outlines the procedures to be put in place to standardise data access, including the development of the DSA Data Access Portal (*Ibid*). It appears that the move towards trying to enable greater researcher access to data by the UK parliament and Ofcom – as exhibited through the Data (Use and Access Bill) and this consultation – aligns broadly with EU initiatives such as these. Therefore, rather than attempting to create a similar framework, **we recommend that Ofcom and the UK Government align their data access protocols with those of the EU**. This would give our regulations more credibility, which is crucial when standing up to large online platforms like X and Meta. We have previously called for greater alignment between UK and EU regulations for online services. For example, in our white paper, *The Metaverse and Web 3.0*, we called for regulatory frameworks on user generated harmful content to be harmonised across the UK and EU (REPHRAIN, 2023b). We argued that this framework should follow the proposals of the DSA, in which users are held legally accountable for illegal content that they generate. Similarly, the UK should follow the EU's DSA in the context of researcher access to data. **Alongside enhancing regulatory credibility, it would also enable greater coordination between states, which is vital when dealing with global services like social media companies.** However, there are limitations within the draft regulation which must be addressed. Such limitations are discussed at length within UCL's Gender and Tech Research Lab's response to the European Commission's consultation on the draft regulation, which our REPHRAIN researcher, Dr Leonie Tanczer, led (Gender and Tech Research Lab, 2024). Some of these limitations include the DSA's planned process which involves an independent researcher applying to their national representative body for vetting status, which would enable them to request data from a large social media platform or search engine (European Commission, 2024). Yet, since there is no

Question	Your response
	provision for UK researchers to be able to apply for vetting status – and it is unclear whether EU researchers within UK institutions qualify – the DSA's proposals place researchers in the UK at a huge disadvantage (EU Digital Services Act, 2022; 'Directive (EU) 2019/790 of the European Parliament and of the Council', 2019). This may result in an exodus from UK institutions, or an increase in second appointments at an EU university. Therefore, regardless of whether the UK chooses to align their regulations to the EU or not, the DSA and its frameworks will drastically affect UK researchers. It is thus **crucial that Ofcom and the UK Government recognise the DSA and its implications**.
Question 3c: What are some possible models for providing researchers with access to relevant information that may not exist or be widely used yet, but which might be implemented by industry?	Confidential? – Y / N
Question 3d: What are the advantages and disadvantages of this approach? • These may include elements pertaining to financial, legal, security, technical or feasibility issues	Confidential? – Y / N
Question 3e: What role could third party organisations, such as regulatory bodies, civil society or public sector organisations have in facilitating researcher access to online safety information?	Confidential? – Y / N **Data gathering, holding and processing** Third-party organisations could play a key role in facilitating researcher access to online safety information by acting as trusted intermediaries between researchers and online services. A potential model involves establishing a dedicated third-party platform responsible for gathering, storing, and managing sensitive data from online services. **This platform would ensure that data is shared securely with researchers under strict guidelines, reducing the burden on individual services and ensuring consistent standards for data handling and privacy.** The main benefit of this model is that it **centralises data management, creating a single secure entity with the resources to implement robust security measures.** It also fosters long-term trust between platforms and researchers by streamlining the data-sharing process. However, a potential drawback is that such a

Question	Your response
	platform could become a high-value target ("honeypot") for attackers, requiring significant investment to maintain its security. **However, we do not believe that it is necessary to require a secure processing environment or holding facility for less sensitive, publicly accessible data**. This would create an unnecessary administrative burden and slow down the research process, as exhibited by Meta's Content Library. As mentioned, a traffic light system could be used to classify levels of sensitivity of data. The SoMe4Dem response to the EU's DSA Article 40 draft delegated regulation, which our REPHRAIN researcher, Prof. Stephan Lewandowsky, contributed to, similarly raised this point (SoMe4Dem, 2024). Aside from holding or processing data, third parties may also serve as intermediaries by: **Making decisions regarding data access applications:** • As mentioned, online services can act as 'gatekeepers' of data, rejecting data access applications for unclear reasons. Having an external body, such as Ofcom, make these decisions would reduce subjectivity and enable a more transparent decision making process for researchers. **Providing ethical and legal oversight:** • Ensuring researchers comply with ethical guidelines and legal regulations when accessing and using platform data. • For instance, an independent body could review research proposals to confirm they align with privacy laws like the UK GDPR or ethical standards for social science research. **Offering guidance for platforms**: • Helping platforms understand their obligations for providing data access while safeguarding user privacy and platform security. **Creating incentives for data access**: • Encouraging voluntary participation from platforms by offering benefits such as reduced regulatory scrutiny or public recognition for cooperation. • For instance, if a platform does not provide a mandated API, it could instead agree to regular audits by a regulator like Ofcom. Conversely, platforms using a public API could allow Ofcom to audit them via that API, reducing the administrative burden.

Question	Your response
Question 3f: What could these third-party models look like, and what are some of the benefits and challenges associated with this approach?	Confidential? – Y / N
Question 3e: What categories of information should online service providers give researchers access for the study of online safety matters? Why would this information be valuable for the study of online safety matters?	Confidential? – Y / N

References

'Directive (EU) 2019/790 of the European Parliament and of the Council' (2019) Available at: https://www.legislation.gov.uk/eudr/2019/790/article/2#:~:text=%E2%80%98research%20organisation%E2%80%99%20means%20a%20university%2C%20including%20its%20libraries%2C,activities%20involving%20also%20the%20conduct%20of%20scientific%20research%3A (Accessed: 09/01/2025).

Beadle, K., Warner, M., & Vasek, M. (2025) '"Edit: I'm sorry for being offensive, this is getting downvoted and I feel terrible": Implicit Social Norms as Governance in Identity-Based Communities', *Proceedings of the ACM on Human-Computer Interaction*, CSCW. Association for Computing Machinery (ACM). Available at: https://discovery.ucl.ac.uk/id/eprint/10200690/1/Implicit%20Social%20Norms%20as%20Governance%20in%20Identity-Based%20Communities.pdf (Accessed: 08/01/2025).

Chung, T. *et al* (2017) 'Privacy Leakage in Event-based Social Networks: A Meetup Case Study', *Proceedings of the ACM on Human-Computer Interaction*, Volume 1, Issue 35, CSCW. Available at: https://doi.org/10.1145/313467 (Accessed: 09/01/2025) pp. 1 – 22.

Correia de Carvalho (2024) 'Researcher Access to Platform Data and the DSA: One Step Forward, Three Steps Back', *TechPolicy.Press*. Available at: https://www.techpolicy.press/researcher-access-to-platform-data-and-the-dsa-one-step-forward-three-steps-back/ (Accessed: 08/01/2025).

Data (Use and Access) Bill (2024). Parliament: House of Lords. Bill no. 57. Available at: https://bills.parliament.uk/bills/3825 (Accessed: 08/01/2025).

Data Access Collaboratory (2024) *Tracker Overview: Decision Insights*, Available at: https://dsa40collaboratory.eu/tracker-insights/ (Accessed: 08/01/2025).

EU Digital Services Act (2022) *The final text of the Digital Services Act (DSA*) Available at: https://www.eu-digital-services-act.com/Digital_Services_Act_Article_40.html (Accessed: 09/01/2025).

European Centre for Algorithmic Transparency (2024) *Delegated act on data access published for consultation - European Commission*. Available at: https://algorithmic-

transparency.ec.europa.eu/news/delegated-act-data-access-published-consultation-2024-10-31_en (Accessed: 08/01/2025).

European Commission (2024) *Delegated Regulation on data access provided for in the Digital Services Act*. Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en; (Accessed: 09/01/2025).

Gender and Tech Research Lab (2024) 'Feedback from: UCL Gender and Tech Research Lab' *Delegated Regulation on data access provided for in the Digital Services Act*. Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3498995_en (Accessed: 10/01/2025).

Hutchinson, A. (2024) *X Increases Its API Access Fees*. Available at: https://www.socialmediatoday.com/news/x-formerly-twitter-increases-api-access-fees/731151/#:~:text=X%20has%20also%20changed%20the%20pricing%20of%20its,that%20connects%20to%20X%E2%80%99s%20API%20through%20their%20app. (Accessed: 08/01/2025).

Lobo, M. (2024) *CrowdTangle Shutdown: What it Means for Misinformation Tracking*. Available at: https://www.medianama.com/2024/08/223-meta-discontinues-crowdtangle-criticism-fact-checkers-researchers/ (Accessed: 08/01/2025).

Murgia, M., Criddle C., and Murphy, H. (2021) 'Investigating Facebook: a fractious relationship with academia', *Financial Times*. Available at: https://www.ft.com/content/1f409239-9e4a-4988-b6fa-cad4dbe7c344 (Accessed: 08/01/2025).

NHS England (2025) *Secure Data Environment - NHS England Digital.* Available at: https://digital.nhs.uk/services/secure-data-environment-service (Accessed: 08/01/2025).

Office for National Statistics (2025) *Become an accredited researcher*. Available at: https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/becomeanaccreditedresearcher (Accessed: 08/01/2025).

REPHRAIN (2023) *Making Sense of the Twitter Takeover*. Available at: https://bpb-eu-w2.wpmucdn.com/blogs.bristol.ac.uk/dist/1/670/files/2023/09/REPHRAIN-Making-sense-of-the-Twitter-Takeover.pdf (Accessed: 08/01/2025).

REPHRAIN (2023b) *REPHRAIN White Paper: The Metaverse and Web 3*. Available at: https://bpb-eu-w2.wpmucdn.com/blogs.bristol.ac.uk/dist/1/670/files/2023/05/REPHRAIN-White-Paper-Metaverse-and-Web-3.pdf (Accessed: 10/01/2025).

REPHRAIN (2025a) *Key2Kindness*. Available at: https://www.rephrain.ac.uk/key2kindness/ (Accessed: 08/01/2025).

REPHRAIN (2025b) *CSAC*. Available at: https://www.rephrain.ac.uk/csac/ (Accessed: 09/01/2025).

SoMe4Dem (2024) 'Feedback from: Some4Dem', *Delegated Regulation on data access provided for in the Digital Services Act*. Available at: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3498896_en (Accessed: 09/01/2025).

Turner, S. and Tanczer, L. (2024) 'In principle vs in practice: User, expert and policymaker attitudes towards the right to data portability in the internet of things', *Computer Law & Security Review*, Volume 52, 105912. Available at: https://doi.org/10.1016/j.clsr.2023.105912. (Accessed: 10/01/2025).