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?	**Note on submission response** This submission is informed by research conducted by the Ada Lovelace Institute under the project *Private-sector data for public good: modelling data access mandates* which aims to develop understanding of the policy, regulatory and practical conditions necessary to mandate access to private data and to test their viability as a legal and governance mechanism for disintermediating value creation from data and re-channelling it – through the powers of regulators – towards wider public benefit. As part of this project, the Ada Lovelace Institute developed a series of cases studies that explore what illustrations of social benefit are available when access to data is enabled, and what challenges, friction points and power dynamics are involved. These case studies offer insights into the different mechanisms, purposes, incentives and challenges around access to data more broadly which translate across domains and provide lessons learned for data to access for research into online safety matters. Answers in the submission form are informed by findings from our own research and also surface work from organisations and researchers working with platform data in the day-to-day, and reference the work of those working closely on Art 40 and EU Digital Service Act (DSA) implementation. **Models and current extent of access** There is a wide variety of ways in which researchers can access data from platforms, either directly or indirectly. Mechanisms for direct access from platforms include partnerships[1], application programming interfaces (APIs)[2], dashboards[3], or research environments[4] such as 'clean rooms' set up virtually or physically

[1] LinkedIn partnership with Industry Data for Society Partnership. Available at https://economicgraph.linkedin.com/data-for-impact# and Meta Centre for Open Science. Available at https://www.cos.io/about/news/meta-partners-with-cos-to-share-data-to-study-well-being-topics and Social Science One. Available at https://socialscience.one/

[2] TikTok Research API. Available at https://developers.tiktok.com/products/research-api/

[3] Data Knowledge Hub for Researching Online Discourse. Available at https://data-knowledge-hub.com/

[4] Meta's Researcher platform. Available at https://developers.facebook.com/docs/researcher-platform/overview/

Question	Your response
	by platforms[5] or by third party services offering secure environments. Indirect methods include data donations through tools such as browser extensions (for example Who Targets Me[6] and the NYU Ad Observatory[7]) or access through third party services and databases (such as BrightData[8] and Internet Archive/Wayback Machine[9]). For a typology of access to data models, please see the Open Data Institute's register of initiatives that enable access to social media data.[10] Each mechanism presents its own set of challenges and opportunities and can be appropriate in certain circumstances and less so in others. For example, research environments can provide a secure environment for handling sensitive data, but might be impractical and excessive if the nature of the data is different. Researchers have pointed out challenges around replicability of research where certain clean rooms such as Meta's Researcher Platform[11] regularly deletes data.[12] While replicability of data is not possible in every scenario, there should be flexibility for data storage in circumstances when certain data needs to be stored. In one of our case studies we looked at **industry-academia collaborations** as pathway for accessing platform data for independent research. More specifically, we explored the Social Science One initiative, as one of the early examples of partnerships between academia and platforms. Although researchers and academics have long called for access to Facebook data, the context that ultimately led to establishing Social Science One in 2018 as a formal partnership for collaboration was the growing media pressure in the aftermath of the Cambridge Analytica.[13] The Cambridge Analytica scandal not only put public pressure on the company and incentivised the platform to share data with researchers, but also brought to the public attention access to data

[5] CDT Europe (2023). CDT Europe Comments on Delegated Regulation on data access provided for in the Digital Services Act. Available at https://cdt.org/wp-content/uploads/2023/06/CDT-Europe-Submission-to-European-Commission-Delegated-Act-on-DSA-Article-40_-Data-Access.pdf

[6] For information about the browser extension, but also about the limitations to do research based on this tool see Who Targets Me (2024). Why do we keep running our browser extension? Available at https://whotargets.me/en/why-do-we-keep-running-our-browser-extension/

[7] NYU Ad Observatory by Cybersecurity for Democracy. Available at https://adobservatory.org/

[8] See https://brightdata.com/

[9] See https://web.archive.org/

[10] Open Data Institute (2024). ODI register of initiatives that enable access to social media data. Available at https://airtable.com/appYiJlYO9XJWFQLi/shrxgNuNJ4EMoNbiI/tblQdjhh4jiIknRBD/viwaXZUDoMoGyCHrL

[11] See https://developers.facebook.com/docs/researcher-platform/overview/

[12] Seiling, L., Klinger, U., Ohme, J. (2024) Non-Public Data Access for Researchers: Challenges in the Draft Delegated Act of the Digital Services Act. Available at https://www.techpolicy.press/non-public-data-access-for-researchers-challenges-in-the-draft-delegated-act-of-the-digital-services-act/

[13] King, G. and Persily, N. (2019). A New Model for Industry-Academic Partnerships. PS: Political Science and Politics, 53, 4, Pp. 703-709. Available at https://gking.harvard.edu/partnerships

Question	Your response
	as an important mechanism for transparency and independent scrutiny. However, the initiative faced various challenges and was amply criticized because of its omissions and limitations[14], ultimately resulting in leading academics to spearhead draft federal legislation in the US as the only way to address challenges and lack of incentives for platforms to share research data.[15] Further collaboration initiatives, such as the Meta initiative to provide access to data relating the 2020 US elections, while facilitating several studies, did not meet the needs of independent research. In the words of the independent rapporteur, 'the project is not a model for future industry-academy collaborations. The collaboration resulted in independent research, but it was independence by permission from Meta', with key features of independent research, such as prioritizing workflow, not possible under the collaboration. [16] The **main lessons learned** from the Social Science One example are that: 1) While access to platform data is instrumental in understanding the social impacts and designing effective policy solutions, despite positive improvements, the current understanding is still limited by **insufficient data access**. Leading academics behind the US draft legislation argue that when looking at the potential harms of social media platforms only internal data (i.e. non-public data) can provide policy makers with a sufficient foundation for effective interventions.[17] 2) In cases where access is provided, there is often missing or misleading data, and a lack of ability to validate the **quality of data**. This forces researchers to conduct independent data collection and adversarial audits in order to validate data from platforms. [18] Such initiatives have been undermined or blocked by platforms. For example, the tool developed by ProPublica often picked up on adverts which were missing in Facebook's archives, however the

[14] Timberg, (2021), 'Facebook made big mistake in data it provided to researchers, undermining academic work', Available at: https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/ and Galley by CJR, (2022), Available at: https://galley.cjr.org/public/conversations/-Msa7j9xmogdYQ9-Yx6f

[15] Persily, N., Tucker, J. A. (2021) How to fix social media? Start with independent research. Available at https://www.brookings.edu/articles/how-to-fix-social-media-start-with-independent-research/

[16] Wagner, M. (2023). Independence by permission. Available at https://www.science.org/doi/10.1126/science.adi2430

[17] Persily, N., Tucker, J. A. (2021) How to fix social media? Start with independent research. Available at https://www.brookings.edu/articles/how-to-fix-social-media-start-with-independent-research/

[18] See AI Forensics (2023). Response to the call for feedback on a planned Delegated Regulation on data access provided for in the Digital Services Act (DSA). 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/F3423644_en

Question	Your response
	browser extension was blocked by Facebook, which claimed it was enforcing its terms of service and preventing its users' data from being collected in an unexpected way.[19] Researchers at New York University had a similar experience with their accounts being disabled whilst collecting data to research how Facebook can be used to spread misinformation.[20] Considerations for data quality, validity of data and researcher liability will be further discussed in Question 3. While the focus of our case study was primarily to draw lessons about power dynamics from industry-academic collaboration under Social Science One, it is important to note that Meta explored other mechanisms for access, each facing different sets of challenges. In 2016, Meta acquired CrowdTangle, a social media monitoring tool used by researchers and journalists to analyse Facebook and Instagram public data and perform investigations that have led to exposing misinformation and hate speech.[21] However, Meta gradually reduced support for the tool and ultimately closed it down in August 2024, replacing it with Meta Content Library.[22] Researchers and civil society organisations do not deem the new tool to provide the same level of functionality or an improved one.[23] The CrowdTangle example serves as reflection around the risk for platforms to unilaterally and unexpectedly remove access. At the same time, several platforms have limited access when deciding to switch from a free access model to their API to a paid one that exceeds regular academic budgets.[24] In other cases, platforms have tampered with the research by

[19] Merrill, J-B, Tobin, A. (2019). Facebook Moves to Block Ad Transparency Tools — Including Ours. Available at https://www.propublica.org/article/facebook-blocks-ad-transparency-tools

[20] New York Times (2021). We research misinformation on Facebook, it just disabled our accounts. Available at https://engineering.nyu.edu/news/we-research-misinformation-facebook-it-just-disabled-our-accounts

[21] Jackson, J., Heal, A. (2021). Misinformation market: The money-making tools Facebook hands to Covid cranks. Available at https://www.thebureauinvestigates.com/stories/2021-01-31/misinformation-market-the-money-making-tools-facebook-hands-to-covid-cranks/

[22] Meta Content Library and API. Available at: https://transparency.fb.com/en-gb/researchtools/meta-content-library

[23] Multiple civil society open letters: Human Rights Watch (2024). Letter Urging Meta to Maintain CrowdTangle Tool Through Upcoming Elections. Available at https://www.hrw.org/news/2024/05/15/letter-urging-meta-maintain-crowdtangle-tool-through-upcoming-elections and Mozilla Foundation (2024). Open Letter To Meta. Available at https://foundation.mozilla.org/en/campaigns/open-letter-to-meta-support-crowdtangle-through-2024-and-maintain-crowdtangle-approach/

[24] Calma, J. (2023). Twitter just closed the book on academic research. Available at https://www.theverge.com/2023/5/31/23739084/twitter-elon-musk-api-policy-chilling-academic-research and Shakir, U. (2023). Reddit's upcoming API changes will make AI companies pony up. Available at https://www.theverge.com/2023/4/18/23688463/reddit-developer-api-terms-change-monetization-ai

Question	Your response
	changing their algorithms without notice, affecting the research results.[25] Limitations from industry-academic partnerships and other mechanisms mentioned reveal important questions that an effective access to data frameworks need to consider: • How can privacy protections be balanced with the access to sensitive information for research? • What approaches can be explored in order to ensure the validity of data for reliable research? • Can data quality audits performed by a third party be a mechanism for reducing the burden put on researchers to engage in alternative data collection methods, risking liability from potential violations of terms of service? • What are ways to prevent ad hoc decisions from companies to block or limit access?
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?	Our expert interviews and research from external organisations align on the fact that a key impediment in the current access to data landscape is a lack of knowledge about the types and volumes of data platforms have. This makes it more difficult to formulate research questions in the first place, and have confidence in the results. For example, in order to carry out statistical sampling, researchers need data on the volume of content (knowing that a certain amount of posts contain hate speech is insufficient for studying the phenomenon if the total number of posts is unknown). [26] Examples of data that have been widely used by researchers include engagement data and social graph data which have been used to uncover how disinformation spreads within and across groups.[27] However, there are also particular areas, where data is insufficient or inaccessible, specifically data on advertising (e.g. the content of ads, detailed ad targeting data, information about who purchased

[25] Guess, M. *et al* (2024). 'How do social media feed algorithms affect attitudes and behavior in an election campaign?' available at https://www.science.org/doi/10.1126/science.abp9364#elettersSection, referenced in Seiling, L., Klinger, U., Ohme, J. (2024) Non-Public Data Access for Researchers: Challenges in the Draft Delegated Act of the Digital Services Act. Available at https://www.techpolicy.press/non-public-data-access-for-researchers-challenges-in-the-draft-delegated-act-of-the-digital-services-act/

[26] Vogus, C. (2022). Improving Researcher Access to Digital Data. CDT workshop report. Available at https://cdt.org/wp-content/uploads/2022/08/2022-08-15-FX-RAtD-workshop-report-final-int.pdf

[27] CDT Europe (2023). CDT Europe Comments on Delegated Regulation on data access provided for in the Digital Services Act. Available at https://cdt.org/wp-content/uploads/2023/06/CDT-Europe-Submission-to-European-Commission-Delegated-Act-on-DSA-Article-40_-Data-Access.pdf

Question	Your response
	or viewed an ad, data on ad spending), recommendation algorithms (e.g. algorithmic ranking, input points for recommendations), and automated content moderation systems (e.g. details on specific enforcement decisions which can allow examination of internal policies and fairness of decisions).[28] There is a broad range of actors using data to support public interest research such as civil society organisations, non-academic researchers and investigative journalists which have been using platform data to study issues such as misinformation,[29] disinformation,[30] hate speech,[31] election interference,[32] platform loopholes and blindspots,[33] to name just a few. Insights from our case studies show that there is greater potential for social benefit when access is enabled across a plurality of actors, who can approach problems in the area with unique perspectives and tools. Alongside academia, civil society organisations and researchers working with communities, all bring different skills and perspectives and can study complementary aspects of harms. Important lessons drawn from our research are: 1) Whether the potential of data is realised is determined by **who uses it**. In our case study looking at environment data, water companies are mandated to capture data on the length and frequency of sewage spills, and must share this data with the public upon request under the Environmental Information Regulations 2004 (EIR). Researchers at an environmental advocacy group used data from EIR to illustrate that a significant number of sewage spills were in breach of their permits, potentially causing significant environmental damage.

[28] Vogus, C. (2022). Improving Researcher Access to Digital Data. CDT workshop report. Available at https://cdt.org/wp-content/uploads/2022/08/2022-08-15-FX-RAtD-workshop-report-final-int.pdf
[29] https://www.thebureauinvestigates.com/stories/2021-01-31/misinformation-market-the-money-making-tools-facebook-hands-to-covid-cranks/
[30] https://www.thebureauinvestigates.com/stories/2024-06-05/facebook-failed-to-block-thousands-of-political-ads-peddling-false-information/
[31] https://www.thebureauinvestigates.com/stories/2022-02-20/facebook-accused-of-letting-activists-incite-ethnic-massacres-with-hate-and-misinformation-by-survivors-in-ethiopia/
[32] Tracking Exposed (2022). Shadow-promotion: TikTok's algorithmic recommendation of banned content in Russia. Available at https://tracking.exposed/pdf/tiktok-russia-ShadowPromotion.pdf
[33] Alvarado Rincón, D., Meyer-Resende, M. (2024). The big loophole (and how to close it): How TikTok's policy and practice invites murky political accounts, Democracy Reporting International. Available at https://democracy-reporting.org/en/office/global/publications/the-big-loophole-and-how-to-close-it-how-tiktoks-policy-and-practice-invites-murky-political-accounts

Question	Your response
	The case study illustrates how in certain situations public benefit and the incentives of the data holder contradict each other, making it advantageous to the public to make the data more accessible to a broad range of actors. 2) Access should be appropriate for the purpose of **effective scrutiny**. In the same case study, datasets collected and published by the Environment Agency contained yearly aggregations which did not allow for the identification of certain qualifications of sewage spills. These qualifications represented important nuances which were missed out. While these are specificities in the environmental sector, other sectors such as online safety will have their unique data points which need to be analysed for effectively understand online harms and enable public scrutiny. Compounding factors such as providing access to data whilst making the application process difficult or providing the data at a resolution which does not allow for the extraction of necessary insights, might risk creating the appearance of transparency, without effective mechanisms for scrutiny.
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?	There is a distinction between access to public platform data which anyone can access via the internet, and access to non-public data which can only be accessed if platforms share it. Both types of data are relevant for online safety research, but non-public data offers higher degree of insight which can provide policy makers with a sufficient foundation for effective interventions against online harms. However, the discussion about relevant access to non-public data will largely remain opaque until there is more knowledge about what data platforms are holding.
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	Please refer to Question 1.

Question	Your response
the specific use cases for the information. • Please provide relevant examples of these governance models used in the online services industry.	
Question 1d: What technologies are typically used by providers of online services to facilitate existing information access?	Please refer to Question 1.
Question 1e: Have services and/or researchers made use of privacy-enhancing technologies to enable access?	

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?	Insights from our case study series show the discussion around challenges is two-fold: **When access is enabled** As pointed out in Questions 1a, a key impediment is **opacity**. There is a lack of knowledge about what type and volume of data platforms are collecting. Even when access is enabled, our expert interviews highlighted that a major barrier is that researchers don't know what data to ask for because they don't have information about what type of data is collected and generated by companies. Another key challenges is **friction**. When access is enabled there may be friction generated from the data not being provided in an easily usable format; the dataset having large gaps; and inconsistent formatting combined with occasional missing values means some of the data can be of a poor quality. The utility of access is mediated by the accessibility of the process, timeliness, and the quality of the data when it is received. Though access to data can be legally mandated, factors such as variations in how applications for data need to be sent, the process being extended through several appeals, companies failing to respond to requests, and the data arriving in formats which make it difficult to use, can inhibit research. Inconsistencies in data provision also means scrutiny is not applied equally across the sector, potentially

Question	Your response
	resulting in companies which are more compliant with data access requests facing a disproportionate proportion of the consequences if they open up access. **When access is not enabled** An important challenge for researchers is that they often need to rely on methods to independently collect data such as web scraping or custom web extensions. Beyond liability risks for researchers from potential violations of platform terms of service, this also entails barriers such as having a certain level of technical experience (or to have the budget to work with skilled data scientists) for data collection. Another barrier is the risk that platforms will shut down these data collection operations or introduce significant friction points (for example by constantly changing the algorithms so that researchers need to constantly update their systems for the collection tools to work). However, even if alternative data collections methods are used, a significant limitation is the inaccessibility of certain data types generated on the side of the platform (for example inferences, categories, profiles) because independent data collection tools will only capture what's 'on the outside' of the platform. As it currently stands, researchers are insufficiently equipped for the effective provision of evidence. In some cases, data is not available, whilst in others attempts at independent data collection are curtailed. These limitations restrict the kinds of questions researchers can ask, and the confidence with which they can provide answers.
Question 2a: What are the legal challenges/risks to sharing information from online services with independent researchers?	Some of the legal challenges that have to be navigated are around: • Balancing individual privacy with access to data for research. Researchers need to understand and comply with data protection obligations, while platforms need not restrict access to data based on unjustified privacy grounds. The intermediary body we discuss in Question 3 could address some of these tensions by standardizing the application process, mitigating divergences between platforms and researchers, and supporting researchers with capacity building on data protection and security requirements. • There needs to be a clarification on the definition of 'sensitive' data. Is it the definition of special category data in the UK GDPR or does it have a broader scope to include information which can become sensitive in certain circumstances? With the nature of research being

Question	Your response
	international, how can these definitions be aligned to ensure that researchers don't miss out based on jurisdiction? • Users have certain expectations around the use of data for research and rules around further use need to be clarified to include explicit consent especially in cases involving sensitive information. • With the proposed changes to the definition of scientific research in the draft Data Use and Access Bill, there needs to be a clarification around the purpose of research for online safety and the purpose of research for scientific research in the public interest or for statistical and archival purposes. Is online safety research expected to follow the research regime provisions in the UK GDPR or is it meant to be developed as a sub-regime of research specifically tailored to online safety matters?
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?	
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?	
Question 2d: What are the information quality challenges relating to online	Please refer to Question 1.

Question	Your response
services sharing information with independent researchers?	
Question 2e: What are the financial costs to online services relating to online services sharing information with independent researchers?	
Question 2f: What are the financial costs to researcher trying to make use of information shared by online services?	Before scandals around data misuse such as Cambridge Analytica and technological developments such as generative AI (where vast amounts of publicly available data are used to feed and train AI models), large platforms allowed API access free of charge. Since then, social media platforms have started to limit API access and move towards a subscription based model. For example, X (Twitter) has shifted from a free model, announcing in March 2024 that access packages will start at the cost of $42.000 per month, well outside underfunded research budgets.[34] Reddit also started to put restrictions on their API, charging for more extensive access.[35] Aside costs from APIs, another category of high costs is transcription costs for audio and video data which are increasingly being the dominant type of media content on platforms.[36] It is important to develop additional funding options and support researchers with tools and data analytics.

Question	Your response
Question 3: How might greater access to information for the purpose of research into online safety issues be achieved?	Greater access to information from platforms will allow more understanding of the risks of harms from online activity and of the impact of social media platforms. However, there is a need to facilitate effective use of platform data, beyond mandating more open access. Important nuances around the main scoping

[34] Stokel-Walker, C. (2023). Twitter's $42,000-per-Month API Prices Out Nearly Everyone. Available at https://www.wired.com/story/twitter-data-api-prices-out-nearly-everyone/

[35] Shakir, (2023), Reddit's upcoming API changes will make AI companies pony up, Available at: https://www.theverge.com/2023/4/18/23688463/reddit-developer-api-terms-change-monetization-ai

[36] CDT Europe (2023). CDT Europe Comments on Delegated Regulation on data access provided for in the Digital Services Act. Available at https://cdt.org/wp-content/uploads/2023/06/CDT-Europe-Submission-to-European-Commission-Delegated-Act-on-DSA-Article-40_-Data-Access.pdf

Question	Your response
	questions will become important indicators for the effectiveness of a new access to data framework. **1. What data should be available?** As mentioned in Question 1b, there is a distinction between access to public platform data which anyone can access via the internet, and access to non-public data which can only be accessed if platforms share it. Both types of data are highly relevant for online safety research, but non-public data offers higher degree of insight which can provide policy makers with a sufficient foundation for effective interventions against online harms. **Real-time access to data** Civil society organisations have emphasised the need for access to public data, but have also raised challenges from platforms sharing incomplete or inaccurate data[37], and from limiting access to public data.[38] To address these challenges, public data should ideally be accessible **real-time,** in full, free and open for anyone to research (this means a similar provision to Article 40(12) of the DSA which is yet to be complied with in full by platforms). **Access to non-public data** Alongside access to real-time public data, researchers should be granted access to study risks of harms from online activity. The discussion around non-public data is more complex due to privacy concerns, but will largely remain opaque until there is more knowledge about what data platforms are holding. **Data quality audits and metrics** Establishing what data is available is important in order for requests to be more precise and effective. For this reason, and to avoid the risk for platforms to unnecessarily restrict research or introduce friction, there is a need for data quality audits and metrics for social media data. This role can be performed by a third party body that investigates and scrutinises what data is collected

[37] Pershan, C. (2023). The DSA must ensure public data for public interest research. Mozilla Foundation. Available at https://foundation.mozilla.org/en/blog/the-digital-services-act-must-ensure-public-data-for-public-interest-research/

[38] Kuchta, R., Almeida Saab, B., Böswald, L-M. (2023). The Data Access Problem: Limitations on Access to Public Data on Very Large Online Platforms. Democracy Reporting International. Available at https://democracy-reporting.org/en/office/global/publications/the-data-access-problem-limitations-on-access-to-public-data-on-very-large-online-platforms

Question	Your response
	and generated by platforms and matches that with the quality of data that is being accessed.[39] The meaning of data access should not be limited to what platforms *define* as data. With quality of data being a key component for fulfilling the objectives of the Online Safety Act 2023, establishing a process for validating the quality of the data from platforms becomes an essential building block for facilitating reliable research. Another benefit to establishing this process is to reduce the pressure on researchers to use grey methods such as web scraping in order to make sure they are able to validate their results. **Researcher sandboxes** Organisations such as DSA40 Data Access Collaboratory argue that data access should not be limited to what data platforms *create* themselves. [40] Innovative models such as researcher sandboxes (where researchers can combine platforms and survey data[41] or perform A/B testing into the effectiveness of the recommendation algorithms[42]) are considered to be more aligned with the role of the researchers which actively collect and process data, instead of passive analysers of data shared by platforms. **Understanding how data was compiled** Alongside access to data itself, researchers will also need to have a comprehensive understanding of how the data they are working with was generated or compiled, as this information can significantly influence the way they interpret and analyse the data. For example, civil society organisations have pointed out that companies should also provide detailed explanations of how they define data types (e.g. does the number of posts include deleted posts? Are different data types bundled under one category such

[39] Darius, P. (2024). Researcher Data Access Under the DSA: Lessons from TikTok's API Issues During the 2024 European Elections. Available at https://www.techpolicy.press/-researcher-data-access-under-the-dsa-lessons-from-tiktoks-api-issues-during-the-2024-european-elections/

[40] Seiling, L., Klinger, U., Ohme, J. (2024) Non-Public Data Access for Researchers: Challenges in the Draft Delegated Act of the Digital Services Act. Available at https://www.techpolicy.press/non-public-data-access-for-researchers-challenges-in-the-draft-delegated-act-of-the-digital-services-act/

[41] See Meta pilot project: Instagram Data Pilot Project for Well-being Research. Available at https://www.cos.io/meta

[42] Proposed by the General Partnership on Artificial Intelligence (2024). Social Media Governance Project: How the DSA can enable a public science of digital platform social impact. Available at https://gpai.ai/projects/responsible-ai/social-media-governance/how-the-DSA-can-enable-a-public-science-of-digital-platform-social-impacts.pdf

Question	Your response
	as 'engagements' which might include reaction emojis, shares, saved content?).[43] Further considerations should be around the obligation for platforms to disclose any potential changes made to data or algorithms in response to access requests they receive. **Potential limitations on data types** While certain data types might be too sensitive to be shared (unless there is explicit consent from the data subject), as a general principle, access to data should be flexible enough to allow researchers to request a wide range of data types on a case-by-case basis. This links to the legal challenges section and the definition of sensitive data, asking questions such as whether location data should be considered sensitive or not. Location data is not covered under special category data under the UK GDPR, but is personal information and could be used for profiling (noting that the interdiction to automated decisions based on profiling might be overturned by the draft Data Use and Access Bill). The challenge is that depending on context, some information can become sensitive (for example, location logs or posts from a gay bar or log in information that might reveal sexual orientation or behaviour). **2. Who should it be available to?** In Question 1a, we've underlined the value in having a broad interpretation for 'researchers' to include civil society organisations and non-academic researchers that meet certain criteria such as independence. At the same time, there should be no geographical restrictions on who applies for access to data (whether from UK or elsewhere), which is in line with Section 162 (2) of the Online Safety Act 2023 stating that 'independent research' means the person carrying out research on behalf of a person other than a provider of a regulated service.[44] As with data protection and with other areas of the law where there are provisions on scope of access for government authorities

[43] CDT Europe (2023). CDT Europe Comments on Delegated Regulation on data access provided for in the Digital Services Act. Available at https://cdt.org/wp-content/uploads/2023/06/CDT-Europe-Submission-to-European-Commission-Delegated-Act-on-DSA-Article-40_-Data-Access.pdf

[44] One important aspect to be noted is around aligning access to data frameworks. While the EU Delegated acts confirm organisations from outside the EU can apply for access, they specify that the research must primarily study 'systemic risks in the European Union' which potentially introduces challenges around data transfers to non-EU affiliated researchers. See Vermeulen, M. (2024). Reading the European Commission's Proposed Implementation of DSA Article 40: Six Initial Observations on a New Framework for Research Data Access. Available at https://www.techpolicy.press/reading-the-european-commissions-proposed-implementation-of-dsa-article-40-six-initial-observations-on-a-new-framework-for-research-data-access/

Question	Your response
	such as police and law enforcement bodies, there needs to be a separate discussion on the necessity and legitimacy of extending the access under the new framework to government bodies with law enforcement attributions.[45] **3. What processes and safeguards need to be put in place?** For this section, we look at the developments introduced by the draft EU delegated regulation.[46] The role of the delegated acts is to clarify and harmonise procedures under the DSA and include innovations such as establishing a 'data access portal' and introducing mediation mechanisms. The aim of the portal to facilitate the application process and create a public overview of data access applications. Although the DSA governance framework includes various roles and responsibility for the regulator and national coordinators,[47] there was a need to create a 'one-stop-shop' for all actors involved and a central point managing information and mediation procedures.[48] Learning from these developments, there seems to be the need to create a third-party body that can fill in some governance and processes gaps for the UK context. **Independent intermediary body** This can take the shape of an independent intermediary under the form of a new body or an existing one already supporting research. Some of the functions and responsibilities for the intermediary that can fill in process gaps observed in the DSA are: • to oversee the application process and establish an appeal mechanism for rejected requests. Organisations submitting evidence to the EU delegated acts have pointed out

[45] US civil society organisations such as the Centre for Democracy and Technology (CDT) have highlighted that law enforcement agencies can use social media data for illegitimate purposes (such as monitoring protestors and racial minorities) and recommend explicitly excluding law enforcement agencies from the meaning of 'independent, vetted researchers'. CDT Europe (2023). CDT Europe Comments on Delegated Regulation on data access provided for in the Digital Services Act. Available at https://cdt.org/wp-content/uploads/2023/06/CDT-Europe-Submission-to-European-Commission-Delegated-Act-on-DSA-Article-40_-Data-Access.pdf

[46] European Commission Draft EU Delegated regulation on data access. 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

[47] Jaursch, J., Lorenz-Spreen, P. (updated 2024). Researcher access to platform data under the DSA: Questions and answers. Available at https://reclaimingautonomyonline.notion.site/Researcher-access-to-platform-data-under-the-DSA-Questions-and-answers-8f7390f3ae6b4aa7ad53d53158ed257c#1c81b5210ddc4f83ad9826a407629477

[48] More detailed discussion about 'data access portal' considerations in Seiling, L., Klinger, U., Ohme, J. (2024). Response to the Consultation on the Delegated Regulation on Data Access provided for in the Digital Services Act. Weizenbaum Policy Papers. Available at https://www.weizenbaum-institut.de/media/Publikationen/Weizenbaum_Policy_Paper/Weizenbaum_Policy_Paper_11.pdf

Question	Your response
	important details for the vetting procedure[49] and appeal mechanism which deserve a close look.[50] • to clarify the role of sub-grantees, consortium partners and other forms of partnership agreements in establishing the relationship between the applicant researcher and any potential requirements for establishing affiliation [51] • role of intermediary to provide research infrastructure such as secure processing environments, data analytics capability and compute power for research, together with capacity building and skills development for researchers • to perform data quality assessments (as discussed above) for data validation • ensure that the application process is standardised to reduce friction [52] • collect and analyse information about applications and produce statistics (e.g. the number of requests, number of accepted or rejected applications, grounds for rejection and lengths of processing are valuable statistics into the effectiveness of the data access framework) • the body should keep a public record tracking the number of applications, who is making the applications, the subject matter of the applications, and the independent body analysis on the applications • importantly, the independent body should keep a record of research studies published and summarize recommendations stemming from studies for platforms to act upon and for the regulator to investigate

[49] Seiling, L., Klinger, U., Ohme, J. (2024). Response to the Consultation on the Delegated Regulation on Data Access provided for in the Digital Services Act. Weizenbaum Policy Papers. Available at https://www.weizenbaum-institut.de/media/Publikationen/Weizenbaum_Policy_Paper/Weizenbaum_Policy_Paper_11.pdf and Democracy Reporting International (2024). Feedback to the Delegated Regulation on Data Access. 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/F3498223_en

[50] Democracy Reporting International (2024). Feedback to the Delegated Regulation on Data Access. 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/F3498223_en and CDT Europe (2023). CDT Europe Comments on Delegated Regulation on data access provided for in the Digital Services Act. Available at https://cdt.org/wp-content/uploads/2023/06/CDT-Europe-Submission-to-European-Commission-Delegated-Act-on-DSA-Article-40_-Data-Access.pdf

[51] Democracy Reporting International, Das Nettz, Fundación Maldita.es, and Science Feedback submission to the delegated acts. 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/F3498223_en

[52] Democracy Reporting International (2024). Feedback to the Delegated Regulation on Data Access. 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/F3498223_en

Question	Your response
	Tiered process In terms of safeguards that need to be considered, one way to address privacy challenges is potentially via introducing a tiered approach for access. Depending on the sensitivity of data, there will be a need for different safeguards. A higher level of security and confidentiality of data should be required for sensitive data such as private messages, information revealing religious affiliation or beliefs, sexual orientation etc.[53] Where researchers are not sufficiently funded or equipped to ensure a high level of security, this might lead to excluding researchers with more modest skills and resources. Where possible, additional funding options should be considered. For example, there should be ways for researchers to access high security infrastructure and skilled data analysts and security engineers, for example through a pool of resources maintained by the independent body and subsidised via portions of the fines allocated to maintain this as critical resources for access to data.
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?	OpenMined offers an innovative solution which involves querying platform data to receive answers to research questions, instead of access to raw data. The tool offers a promising approach towards addressing a range of concerns such as privacy and liability for researchers, yet more understanding is needed to assess potential limitations for research (for example, what type of research can or cannot be performed via using this system). More details are available in OpenMined's submission to the EU delegated acts.[54]
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?	Please see reference to research sandbox models in Question3.

[53] CDT Europe (2023). CDT Europe Comments on Delegated Regulation on data access provided for in the Digital Services Act. Available at https://cdt.org/wp-content/uploads/2023/06/CDT-Europe-Submission-to-European-Commission-Delegated-Act-on-DSA-Article-40_-Data-Access.pdf

[54] OpenMined (2023). Response to Call for Evidence on Contemplated Delegated Regulation on Data Access provided for in the Digital Services Act (DSA) Article 40. 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/F3423932_en

Question	Your response
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?	OpenMind's approach, mentioned in Question 3a might provide an innovative way forward.
Question 3d: What are the advantages and disadvantages of this approach? • These may include elements pertaining to financial, legal, security, technical or feasibility issues	
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?	Please refer to Question 3. There is a need for providing research infrastructure such as secure research environments, data analytics, compute power and investment in capacity building and skills for researchers.
Question 3f: What could these third-party models look like, and what are some of the benefits and challenges associated with this approach?	
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?	Please refer to Question 3.