16 January 2025

Attention: Ofcom

UK Online Safety – Ofcom Call for Evidence on Researcher Access

RESPONSE ON BEHALF OF META PLATFORMS INC

Meta Platforms, Inc. ("**Meta**" or "we") welcomes the opportunity to participate in this call for evidence under the UK Online Safety Act (the "**OSA**" or the "**Act**") organised by the Office of Communications (**Ofcom**). We are pleased to share our insights regarding our experience and ongoing efforts to support independent research in the field of online safety.

Meta is committed to supporting independent research that will enhance our understanding of the impact platforms, like those provided by Meta, have on society. We recognize the importance of being transparent and sharing meaningful data with researchers - data that is robust, representative and thereby can serve as a basis to contribute to understanding how our services work and their potential impact on society.

This presents specific challenges that are common across all research fields and go beyond safety-related research. We are deeply committed to protecting users' privacy and maintaining a safe and secure community, and thus strive to promote data sharing programs and processes that provide meaningful transparency while maintaining the privacy and security of users' data. We are actively working to meet researchers' objectives and to create privacy-protective solutions, and are engaged in finding solutions to these challenges. *See* Meta's Response to the European Digital Media Observatory's Call for Comments on the GDPR and Sharing Data for Independent Social Scientific Research.

The protection of users' privacy not only involves Meta, as a potential data transferor, but also the researchers, as potential transferees, and policymakers and regulators, in defining standards that both platforms and researchers should apply to achieve an appropriate level of protection and to be eligible data transferees.

We welcome the open approach adopted by Ofcom, and we would be pleased to continue this dialogue to inform pending legislative changes.

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?

- **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?**
- **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?**

Meta has multiple tools and resources available for public interest and scientific research that aim to provide valuable data while preserving user privacy. As described in more detail below, Meta provides information that can be used to study online safety matters through (1) transparency reports, (2) a publicly

accessible library of ads currently running across Meta's platforms, (3) Meta Content Library and API, which provides access to publicly accessible data across Facebook, Instagram, and Threads, and (4) specific datasets available to certain researchers and other stakeholders. In addition, Meta seeks feedback and engages in pilots to advance how best to provide useful data while balancing user privacy.

1) **Regular transparency reports concerning content shared on our platforms.** These include reports such as the Community Standards Enforcement Report, the Content Restricted Based on Local Law Report, and the Adversarial Threat Report, as well as regulatory reports, including new reports coming under the OSA.
    - The Community Standards Enforcement Report: We publish this report publicly in our Transparency Centre on a quarterly basis to more effectively track our progress and demonstrate our continued commitment to making our services safe and inclusive. The report shares metrics on how we are doing at preventing and taking action on content that goes against our Community Standards.
    - Content Restrictions Based on Local Law Report: When we restrict content based on local law, we do so only in the country (see UK-related geo blocked content) or region where it is alleged to be illegal. We report the number of pieces of content restricted in each country where our products are available. When a single piece of content is restricted in multiple countries, each restriction is counted independently.
    - The Adversarial Threat Report: Meta publicly shares findings about coordinated inauthentic behaviour (CIB) that is detected and removed from our platforms. As part of our quarterly adversarial threat reports, we publish information about the networks we take down to make it easier for people to see the progress we're making in one place.

2) **A publicly accessible Ad Library and an Ad Library API (**Ad Library tools**)** displaying and archiving all ads delivering impressions.
    - The dedicated website for the Ad Library allows users to search all of the ads currently running across Meta technologies. For all ads, this includes the ad content and the basic information, such as when the ad started running and which advertiser is running it. For ads about social issues, elections or politics that have run in the past seven years, it includes the ad content, the basic information, such as when the ad started running and which advertiser is running it, and additional transparency about spend, reach and funding entities.

3) **Meta Content Library User Interface (**MCL**) and Meta Content Library API (**MCL-API). Launched in 2023, these research tools provide the means for eligible individuals, vetted by a third party, to conduct public interest and scientific research on near real-time publicly accessible data from Meta platforms (e.g., public posts and properties from Instagram, Threads, and Facebook). Details about the content, such as the number of reactions, shares, comments and for the first time, post view counts are also included. Data from MCL can be searched, explored and filtered on a graphical user interface or through a programmatic API.
    - **MCL** is a web-based, controlled-access environment, where eligible individuals can explore publicly accessible data in a user-friendly environment. It provides a number of easy to use features across different content types, including searching, comprehensive sorting and filtering options. Photos, text, videos and reels are available for dynamic search and exploration. Users can build customisable collections of content producers to refine search results and apply custom producer lists to a search query in order to surface public content from specific content owners. In response to researcher feedback, this year we added a dashboard functionality to the MCL which enables researchers to display, monitor, and rank content on one surface.

- **MCL-API** is an application programming interface (API) that allows eligible individuals to programmatically query data from the MCL in order to perform deeper analysis on publicly accessible content. The MCL-API is designed for computational research by users who are familiar with querying data in APIs, and it facilitates quantitative, longitudinal research. The MCL-API offers stronger search capabilities than MCL, returning up to 100,000 results per query as opposed to the ranked sample of the top 1,000 results generated in MCL. Data queried through MCL-API is moved into a Virtual Data Centre ("VDC") for analysis. The VDC is hosted within the secure Virtual Data Enclave ("VDE") (sometimes called a 'clean room') built by SOMAR of the ICPSR. The VDE imposes controls against risks of unauthorised access to and use of the data. The VDE also offers researchers a wide variety of programming languages and tools (e.g., Stata, SPSS) that provide users with increased flexibility and control.
- Meta continues to iterate on and improve MCL and MCL-API. Over the course of 2024, we added new data and features to MCL and MCL-API. Specifically, we made it possible to download a subset of publicly accessible content posted by widely known individuals and entities. We also added a subset of personal Instagram accounts' and Facebook profiles' content and 'comments' as a new data type within MCL. This will help users study how people around the world receive, discuss and reinterpret content across publicly-accessible pages and posts.
- **Vetting and application reviews by a third-party independent partner:** Individuals, including journalists affiliated with qualified institutions pursuing scientific or public interest research topics, apply for access to these tools through the University of Michigan's Inter-university Consortium for Political and Social Research (ICPSR). This partnership enables individuals to analyse data from the API in the aforementioned VDE. The ICPSR reviews applications independently. By partnering with the ICPSR, Meta seeks to ensure that the application procedure is objective and independent. The relevant steps required to access the research tools, as well as the eligibility and review criteria are explained here. Detailed information on the MCL and MCL-API tools is publicly available on the MCL and MCL-API Transparency Center page.

4) **Specific datasets are also available to eligible researchers**, including:
   - The **Data for Good program** empowers partners with data to help make progress on major social issues. Data for Good shares datasets with universities, nonprofits and governments around the world conducting independent research on topics ranging from public health to wealth and poverty. These trusted organisations build maps, surveys and insights to strengthen communities and advance social issues. In order to enable a largely open access model for Data for Good, Meta incorporates many privacy preserving techniques such as protecting against re-indentification, including tools like differential privacy, which you can read more about here.
   - The **Influence Operations (IO) Research Archive** data set is hosted in MCL and provides access to previously public information associated with Groups, Pages, Facebook accounts, and Instagram accounts that Meta has removed for violating its Coordinated Inauthentic Behavior policy, for eligible individuals to conduct short-term and

long-term research.[1] This includes comparing tactics across threat actors globally and over time and producing independent reports.[2]

- The **URL Shares** dataset includes differentially private individual-level counts of the number of people who viewed, clicked, liked, commented, shared, or reacted to any URL on Facebook between January 2017 and September 2022. Counts are aggregated at the level of country, year-month, age bracket, and gender. In order to maintain the independence of individuals who use the data, vetting and application reviews to the URL Shares are performed by a third-party partner, Social Science One.
- The **Ad Targeting Dataset** includes targeting information for social issue, electoral, and political ads that have run globally since August 2020. 140+ researchers globally have access to Ads Targeting API since it launched publicly in September 2022.
- In addition, Meta engages in special projects and pilots to increase access to data related to public interest and scientific topics. Examples of these projects are available here on the Transparency Center.

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.**

As described in the **response to Questions 1(a)-(b)**, the governance model used depends on many factors (e.g., data type, purpose for sharing the data, who is accessing the data). By way of example, anyone can access certain aggregated data such as advertisements run on Meta platform and respective information about those advertisements via Ad Library, while access to other data may require additional protections, such as vetting by independent, third parties or access controls. For example, Meta provides access to publicly accessible data to eligible individuals who are vetted by third parties via Meta Content Library and API and other ad hoc datasets referenced above. The following documents describe the data governing models for a few of Meta's tools in more detail:

- Meta Content Library and API
- Researcher Platform developer documentation
  - Access to Ads Targeting dataset
  - Access to URL shares dataset
- Data for Good tools

Third-party bodies should play an important role in governance and research data sharing models. Examples of governance models in partnership with a third-party organization (more information at https://transparency.meta.com/researchtools/):

[1] *See* Community Standards, Inauthentic Behavior, Meta, https://transparency.meta.com/policies/community-standards/inauthentic-behavior/.
[2] *See* Meta's Adversarial Threat Report, Dec. 1, 2021, Meta, https://about.fb.com/news/2021/12/metas-adversarial-threat-report/

- **ICPSR**: Meta has partnered with the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. One aspect of the partnership is enabled through ICPSR's industry-leading Social Media Archive (SOMAR) initiative. SOMAR at ICPSR independently processes and reviews applications for access to the Meta Content Library. Meta Content Library API is hosted within SOMAR's secure Virtual Data Enclave. Eligible individuals use this secure environment to access and analyze the data from the MCL-API.
- **Social Science One**: Social Science One hosts applications for access to the URL Shares dataset, and similar to ICPSR, independently processes and reviews applications. Applications that are approved by Social Science One will be granted access once per quarter following completion of onboarding with Meta.
- **Center for Open Science (COS)**: Meta partners with the Center for Open Science (COS) on a pilot program sharing certain privacy-preserving social media data with select researchers to study topics related to the well-being of teens and young adults. This pilot program will use research processes that have been popularized in the open science movement, such as pre-registration and early peer review. Meta does not select the researchers or provide input on their research questions, and the pilot program with COS is based on the Registered Reports publishing process to ensure research independence, transparency, and scientific integrity.

As part of our commitment to support independent research, we also engage in pilots in partnership with the research community to consider new models for privacy-protective data sharing with researchers, such as:

- **CASD**: Meta partners with the Secure Data Access Center's (Le Centre d'Accès Sécurisé aux Données (CASD)) to provide selected pilot researchers with access to their clean room environment to analyze the respective Instagram data of their study sample. More information is available on the COS website.
- **EDMO**: One model to address these challenges is proposed by the European Digital Media Observatory's (EDMO) Working Group on Platform Data Sharing, which published a report and code of conduct on platform-to-researcher data sharing that proposes a process to vet and approve research projects and relying on a third-party intermediary as an ex ante body. Last year, Meta participated in a pilot project to test the EDMO Code of Conduct related to compliance with the GDPR while sharing individual-level data for research purposes. The pilot highlighted the value of an independent intermediary body to facilitate data access by providing necessary technical and organizational standards for access and sharing of data, including mediating between stakeholders.

1d: What technologies are typically used by providers of online services to facilitate existing information access?

For many years, Meta has empowered the global study of the political, economic, and social impact of Meta's platforms, including by building datasets, surveys, maps and APIs. Meta has developed, and continues to evolve a range of tools and processes to help individuals gain access to information and analytical capabilities to support their research. MCL-API uses a virtual data enclave (also called a "clean room") to provide access in a secure and privacy-protective environment. See **response to Question 1e** for more information.

Meta also makes available developer documentation pages that provide information and guidance to developers using a particular software or API.

Meta's own 'clean room', used for access by eligible researchers to certain data, is described in the Researcher Platform developer documentation.

1e: Have services and/or researchers made use of privacy-enhancing technologies to enable access?

Meta deploys several privacy-enhancing technologies to share data in a privacy-protective way, including:

- **Differential Privacy**: This technique adds noise to data to protect individual privacy while allowing aggregate data analysis. It is used in various applications, including market insights and ads optimization. For example, the Data for Good program has incorporated differential privacy into many of its datasets on mobility, social connections and social capital such that researchers may perform independent analyses without needing to incorporate PETs into their analyses.
- **Reporting:** Reporting can be a more privacy-protective way of providing information than sharing underlying data. Many of our (and our partner's) aggregate reporting uses simpler, vetted techniques such as minimum reporting thresholds and random rounding of aggregate metrics to protect user confidentiality.
- **Clean rooms for research:** Virtual clean rooms are a standard industry practice that include data security protections, enforce purpose limitations, and protect user confidentiality through strict import and export controls. Clean rooms may allow Meta to expand the range of data accessible to researchers, enabling streamlined and consistent access.
- **Redaction:** When possible, Meta attempts to redact user identifiers prior to sharing with individuals conducting research. However, redaction is typically limited to structured fields.
- **Best practices for data protection, such as purpose limitation, data minimization, data access and management, data retention and user transparency and controls:** Many of these practices were incorporated into the Registered Reports program. In addition, Meta utilized privacy enhancing technology for MCL, including a clean room environment for API access, redaction of usernames for certain accounts, etc.

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

- **2a: What are the legal challenges/risks to sharing information from online services with independent researchers?**
- **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?**

- **User Expectations & Privacy:** Honoring user privacy choices and rights and being transparent with users is a legal limitation to the ability to share information with researchers. For example, users decide the audience of their posts (and, as per the UK GDPR, the default is not that the posts are available to the public). Facebook posts could have been deleted by users after creation, and, at the same time, researchers request stable datasets to study data.
- **Compliance with Global Privacy laws:** Meta has various obligations such as honoring user choices like the changing of an audience setting or requesting deletion of their data. Examples of such obligations include requirements in most global privacy laws to provide for a right to delete or "right to be forgotten" (e.g. UK GDPR, LGPD, CCPA, etc.). A lack of harmonization between research data sharing laws can also pose a challenge where data sharing laws conflict with privacy laws.
- **Data Misuse:** Meta does not sell access to our APIs or platform data, and we only provide limited

data to third-parties. We attempt to prevent data misuse and unauthorized data distribution to protect our users and business from malicious actors (e.g. stalking apps, foreign intelligence agencies).
    - We take the privacy of our users seriously, and preventing emerging risks through a system of layered protections (e.g., vetting of researchers, requirements for downstream security as a condition for onboarding, and limiting data that may pose privacy and security for users if misused) is a key part of our view of what we should do as responsible parties in this space.
- **Liability**: Without a clear definition of acceptable access and security practices, including immunity from liability for downstream third-party misuse (which includes that platforms and researchers will not be considered joint controllers under UKGDPR), platforms are left in a difficult position of trading security for accessibility. It is important that these downstream third parties are liable via contractual terms or legal obligations prior to their accessing data in the case of misuse or mishandling. Joint controllership is not appropriate for research data sharing and instead entails a joint and several liability regime that is consistent with the scope of duties and control of the respective parties.
- **Clarity around key legal concepts would facilitate research data sharing, including:**
    - A legal provision that facilitates the use of the UK GDPR scientific research exemption.
    - A legal definition of scientific research for UK GDPR purposes, including the duty to have peer review, professional accreditations and ethical rules.
    - A legal definition of the criteria to define a researcher that guarantees that the output has the relevant quality.
    - Legal certainty regarding anonymization, which could be addressed by either a UK GDPR amendment or an ICO anonymization guidance amendment.

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?

2d: What are the information quality challenges relating to online services sharing information with independent researchers?

- **Data is low accuracy or not reliable accuracy**: When data is low or not of a reliable accuracy, the relevance and use of the data may not directly support the intended research analysis. Meta invests in data accuracy measurements to help ensure that data shared is fit for purpose. High accuracy data generally means that data should a) be complete, such that researchers do not draw conclusions based on a small sample of available data; and b) inspire confidence in its accuracy, in order to prevent uncertainties from compounding in downstream analysis. It bears emphasizing that “existence” and “accuracy” are distinct criteria; data artifacts constructed in the service of low-level analyses should not be misconstrued as valid, or useful, for researchers’ needs.
- **Metric vetting**: Any data points created need to be reviewed and vetted to meet the appropriate standards for externalization. This validation process is not only to ensure proper internal disclosure process but also to validate that the shared information is as accurate, reliable and meaningful as possible.
- **Data may not exist, and challenges to capture new data**: Requested data which is not already tracked by a provider will have significant impacts to the provider’s capacity to meet transparency notice requests. These changes require adequate time to understand the request, strategise how to collect the desired information, build any logging or tracking mechanisms required, and to execute the information gathering. While capabilities vary depending on various factors, as a

rough guideline building new logging or tracking systems can take a minimum of 6 months and up to 12 or more for highly complex requests and so the availability of data must be considered when determining how much time services may need to respond to transparency notices and produce reports.
    - Requests should be proportionate, and priority should be given to existing data, rather than requiring the creation of new data.
    - Data must be reasonably accessible: We understand that researchers may want to look into a wide set of research topics and thus seek to request multiple types of data that can serve as the basis for their findings. However, there is an essential distinction to be made between data that exist and data that do not exist as well as data that are not legally available for data sharing. As such, not only should requests remain **proportional,** but providers should only be required to provide access to data that exists - and not be required to create new data - and that it is reasonably accessible to them and legally shareable. Note that due to privacy guidelines in various jurisdictions, Meta does not retain certain kinds of data beyond a certain number of days.
- **Challenges to presenting large volumes of data in near real-time**: It is challenging to present significant volumes of data (and associated metrics) in near real-time. Challenges include the technical delivery, the accuracy of the data and how often it can be updated, and user privacy challenges (like count, view count metrics; or if a user deletes their post, should the researcher still have access to it?)
- **Data includes personally identifiable information or information that could be used to directly or indirectly identify users, so we need to apply privacy enhancing technologies**: Depending on the enhancements applied and the research taking place, there could be a compromise of data usability for a given project.

2e: What are the financial costs to online services relating to online services sharing information with independent researchers?

Meta and its third-party partner have significant teams and resources devoted to the onboarding, maintenance, documentation, hosting, development, and user support for Meta Content Library alone. Additional teams and third party partners continue to support a broader set of researcher tools and datasets available to independent researchers. There are also significant hosting and data processing costs associated with providing resources like Meta Content Library.

2f: What are the financial costs to researchers trying to make use of information shared by online services?

There are no fees associated with access or computation for Meta Content Library and MCL-API, Ad Library Tools, Data for Good datasets, and/or other research tools and datasets currently provided by Meta to share information for purposes of public interest and scientific research. More information about the other research tools is available here.

There are also no fees associated with user support or services related to these research tools. Meta and its third party partners provide application and onboarding support, product user support, and product education for these research tools and datasets at no cost.

While providers may reasonably bear some costs associated with data access, it is not reasonable that they should bear computational costs associated with subsequent processing and analysis of data, especially given the potential scale of such actions. For example, sensitive or complex datasets requiring data access in a special environment may require cost sharing.

The privacy and security programme of the platform sharing the data also relies on the existence of the researchers' own privacy and security programme. Researchers, as much as the platforms, are legally accountable regarding their data protection and security duties. Researchers' financial costs include the UK GDPR compliance and cybersecurity costs, in particular, regarding the preparation of the data sharing request and accreditation of the public interest, researchers' qualifications and methodology to ensure the existence and quality of the scientific research, the proportionality of the requested datasets, the security measures in place while storing and analysing the data and sharing them for peer reviews, and those to prevent unauthorised further uses.

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

1) **Address privacy and liability issues with sharing data**: Regulation can facilitate data sharing by focusing on ways to protect privacy while unlocking the benefits of data. For example, regulation could help address the difficult privacy and liability issues (described in **response to Questions 2a and 2c**) that arise with research data sharing -- e.g., by incentivizing the development and use of privacy-enhancing technologies, developing standardized language for data-sharing agreements, providing safe harbors for good actors and penalties for bad ones, and modifying the UK GDPR or its interpretation regarding anonymization, the scientific research regime, joint and several controllership's regime, and international data transfers. Further, regulators should take into account platforms' responsibility to protect confidential information, such as source code or trade secrets, and allow platforms to push back against unreasonable data requests. Finally, regulators could also address researchers' UK GDPR accountability in data sharing as data transferees and potential further uses (in particular, inappropriate uses) as well as guarantees to ensure that the researchers' work meets standards for scientific research and serves identified public interests.
2) **Clarify with whom data should be shared**: While platforms are often best placed to determine how third parties receive the data, policymakers can help clarify who should receive data and under what circumstances.
3) **Partner with third-party bodies**: Third-party organizations can play a key role in facilitating the relationship between data transferor and transferee and in promoting privacy-protective data sharing. *See* **response to Questions 1c and 3b-3f.** In addition, technical standards organizations can play a role partnering with policymakers to standardize data definitions and privacy protections. This can also help researchers conduct comparative research that crosses multiple online services.
4) **Harmonize data sharing processes:** Streamlined processes and regulation could help remove barriers to sharing data with researchers. This could include addressing conflict of law issues that arise by working toward a standardized framework that enables broader access to data for researchers in a privacy protective manner.

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?

The question does not define what "sensitive information" means. The UK GDPR provides for a specific regime regarding "special categories of data," which does not facilitate their data sharing, in particular, if this concept is construed by the privacy regulator in an expansive manner or there is no law that clarifies how to use the scientific research exemption. In addition, **privacy preserving technologies,** including but not limited to data access via a clean room and confidentiality rules for reporting, should be used. For

this, there should be legal clarity regarding the concept and regime of anonymized data, and **peer review** of research proposals that can evaluate scientific rigor (vetted research that is sufficient quality to advance knowledge and thus respect user data and time). In addition, privacy principles like **purpose limitation** (use the right data for the right questions) and **data minimization** (use only the data you need) should be prioritized. The open science principles are a valuable framework to consider when engaging in collaborative research.

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?

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?
3d: What are the advantages and disadvantages of this approach?

- **These may include elements pertaining to financial, legal, security, technical or feasibility issues**

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?

3f: What could these third-party models look like, and what are some of the benefits and challenges associated with this approach?

1. **Independent third-party bodies are valuable in facilitating both vetting and data sharing access for research purposes**.

   A third party independent body could serve multiple purposes, including overseeing the vetting process for researchers, their requests, the scope of the research objectives as well as the privacy and security conditions once the data are under the researchers' control. At the same time, such a body could facilitate access to data and act as an intermediary between researchers, a third-party clean room and other actors, including by providing the necessary approvals of technical and organizational safeguards to ensure that data is shared in an appropriate manner.

2. **Third-party organizations, regulators, and platforms can work together to shape and facilitate data sharing mechanisms and to develop standards to protect the privacy, security and confidentiality of the data**.

   Platforms, researchers, regulators, and other third parties can work together to establish mechanisms and standards for sharing data with researchers, including proportionate technical and organizational measures necessary to preserve the confidentiality of the data. The standards should build off the UK GDPR's requirements, including that authorities and researchers' personal data access requests and subsequent processing should comply with the principles of "lawfulness of processing", "purpose limitation," "data minimization," "storage limitation," and "integrity and confidentiality".

   Building a dataset from scratch requires necessary steps to ensure the accuracy of the dataset and that appropriate privacy and security safeguards are in place. In our experience, the safest mechanism for sharing data is often via a clean room that meets technical and security standards. In addition, requirements should be proportionate. For example, where there are requirements to share bespoke datasets, it is important to find a proportional path to identify what

data is appropriate for sharing, under what data protection processing, and to adopt realistic expectations about timelines for delivering the data.

3. **Third-party bodies or regulators can provide safeguards to address violating behavior**.

   There is a lack of clarity on how to define responsibilities in the case of violating behavior or misuse of data on the part of the researcher. Taking into account that there is no reason for the the data owner or the data sharer to be held responsible for actions that are not under its control (i.e., the researcher's actions), the lack of legal certainty on this point makes it difficult for providers to assume risks they have not created and cannot control. In addition, standards should reaffirm that researchers and authorities, acting as independent "controllers" under the UK GDPR (meaning that they are not "joint controllers"), must comply with all obligations under the UK GDPR, including with all applicable users' rights, implementing appropriate technical and organizational measures at the design stage of the processing to ensure data minimization, designing and implementing appropriate security features, notifying platforms of personal data breaches, conducting a data protection impact assessment, etc.

4. Specific examples
   - *See* partnership with ICPSR described in **response to Question 1.**
   - Another potential model is data sharing consortiums. Data sharing consortiums may not conduct vetting, but can coordinate access across data providers and teams requesting data from public or research institutions. Data sharing consortiums can greatly facilitate relationships and onboarding between technology companies and external partners.
     - The Development Data Partnership, which is managed by the World Bank and includes several private sector data providers and multilateral institutions requesting data.
     - The UK Data Archive, which is the UK's largest collection of social, economic and population research data and the lead partner of the UK Data Service.
     - NORC, based out of Chicago, IL, hosts large-scale data in secure data enclaves for many government and private entities. Examples are available here.
     - Le Centre d'Accès Sécurisé aux Données, CASD, a third-party provider of secure data hosting services in France, hosts and provides researchers access to data from French statistical and government agencies.

In terms of financing these data sharing arrangements, the businesses and organizations listed above implement a variety of models depending on the agreement with a data depositor. In some cases, the data depositor pays the service provider to host the data, the application forms, and IT infrastructure. However, in nearly all cases, the data depositor is the entity to review applications, vet researchers, and make the final decision around access, even when application forms are hosted with a third party. Data depositors may decide to charge researchers for access to certain datasets via the third party, to cover some or all costs of using the third party service, or to subsidize the data access and usage to fund the costs of using a third party service. In other cases, data depositors may pass costs onto researchers/institutions who use grants to pay for access to the data and processing via the third-party (oftentimes if the institution or research team is under-resourced they can apply for exceptions or special grants to help offset costs). And, in some cases, the costs are shared between the data depositors and researchers/institutions.

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?

The categories that should be provided include statistics available through reporting, ads, and publicly accessible information. In addition, where the request is proportional and feasible (***see* response to Question 2b**), data that exists and has high accuracy could be made available for independent, public interest research where users have provided consent to use the data for research or the data can be anonymized and protected with privacy-enhancing technologies, and where that data does not expose trade secrets or business confidential information. Additionally, it is important that certain categories of commercially sensitive information (for example, nuances of content moderation processes), notably from a security perspective, are protected where disclosed or accessed for the purposes of research, where public disclosure of that information could defeat Meta's goal to prevent abuse by bad actors.