British Academy response to Ofcom's call for evidence on researchers' access to information from regulated online services

15th January 2025

Background

In January 2025, the Office of Communications (Ofcom) initiated a consultation about researchers access to information from regulated online services. The purpose of the consultation was to inform a report that is to be produced by Ofcom under the Online Safety Act.

The British Academy is the UK's national academy for the humanities and social sciences, funded by the UK government and a range of charities to promote the SHAPE disciplines and the institutions that support them. In 2024, as part of the Digital Society Policy Programme, the Academy convened three workshop events that were aimed at generating dialogue around, and to provide a better understanding of the challenges and opportunities related to SHAPE researchers' (Social Sciences, Humanities and the Arts for People, the Environment and the Economy) access to privately held data from digital media companies. A summary brief was produced and used as evidence to answer the following questions included in the Ofcom consultation brief.[1]

Questions

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?**

[1] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief'.

The Academy's summary brief, mentioned above, highlighted that researchers from SHAPE disciplines carrying out research into online safety related issues may be accessing data through the following avenues (this list is non-exhaustive): [2]

- **Web Scraping/Web Crawling:** Web scraping as a collection method involves the researcher extracting data from a website. Some platforms allow researchers to scrape their data only for non-commercial purposes without a need for an application. Web Crawling involves using a type of bot to index links on the web, continuously discovering new potentially relevant links.
- **Research Application Programming Interfaces (API's):** Research API's that are tied to online platforms such as Google, TikTok, Meta, where researchers can work with large data holders to access data sets e.g., Youtube Research and TikTok Research.
- **Data Donations:** A data collection method where researchers partner with individuals who are interested in donating their digital traces for research purposes.
- **Purchasing Data Sets:** Researchers access online data by purchasing specific data sets online.
- **Data Repositories:** Data repositories that specialise in collecting sensitive data. These include UK Data Archive and ICPSR.
- **Data Sharing Programmes:** UKRI funded programmes such as Smart Data Research UK aim to streamline the process of access to data from online spaces.

2. What kinds of online safety research does the current level of access to information enable?

The British Academy's summary brief highlighted that researchers and expertise from the Social Sciences and the Arts are highly valuable in providing insights on topics related to online safety such as online misinformation, disinformation, hate speech, threatening communications, as well as understanding wider aspects of engagement with online services such as device addiction, online bullying and trolling.[3]

The British Academy commissioned a large piece of work to explore and analyse impact case studies in the SHAPE disciplines from the most recent research excellence framework (REF2021), the United Kingdom's research assessment exercise. The report – and this accompanying online dashboard – aims to serve as a starting point for the SHAPE community to explore the scope and reach of the evidence base. The full report provides more information about the project and its results. The dashboard can also be used to browse the impact of SHAPE research in other topics relevant to online safety.

For example, one case study that appears in the dashboard is research by the University of Liverpool, which has been utilised by Good Things Foundation to directly support projects, interventions, and advocacy – including in supporting the development of online safety skills and awareness as outcomes of national programmes.

3. What data governance models are currently used to allow access to online services' information for researchers?

[2] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief".
[3] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief".

See response to Question 1 above for further details. Data governance models that are currently used include:

- Research Application Programming interfaces (APIs)
- Data Donations
- Purchasing data sets
- Data repositories
- Data sharing programmes (e.g. Smart Data Research UK)

The summary brief published by the Academy also emphasised the importance of a holistic approach to online data access for research, moving away from a 'one size fits all' approach that assigns responsibility for enabling access to a single group of stakeholders, and towards a more trustworthy and open data ecosystem.[4] This would ensure that there are avenues for data access that allow a range of topics to examined (not restricted to just those related to online safety, for instance) and create a wider range of opportunities for societal benefit.

4. What technologies are typically used by providers of online services to facilitate existing information access?

As noted in the Academy's summary brief, the access and sharing of data has been central to a range of actions being led by the UK government.[5] For example, the Integrated Data Service (IDS), is a cross-sector government initiative that provides a platform for accredited researchers to access data from across government departments in a more streamlined way as well as access to analytical and visualisation tools, using a secure multi-cloud infrastructure. The Office for National Statistics (ONS) is the delivery partner for this initiative. As discussed further in the response other examples include:

- Research Application Programming Interfaces (API's)
- Data Repositories
- Data Sharing Programmes

5. What are the challenges that currently constrain the sharing of information for the purpose of research into online safety related issues?

The British Academy's summary brief on online data access for SHAPE research identified a range of challenges that constrain access and sharing of information for research purposes.[6] These include:

Legal risks

Methods such as data scraping can lead academics to violate the terms of service of some platforms, placing them at legal risk. E.g. in 2021, Meta took legal action against a US-based research group for engaging in data scraping, claiming privacy concerns.[7]

[4] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief".
[5] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief", p. 7.
[6] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief", p. 8.
[7] Edelman, G. (2021), Facebook's Reason for Banning Researcher's Doesn't Hold Up, Wired

Adhoc and unreliable processes

Current pathways for access and sharing are unreliable and subject to change. APIs provided by some online companies are not guaranteed and in recent years some companies have moved to either close APIs or introduce use charges.[8]

Outdated, partial and fragmented data sets

Datasets can be outdated, partial, fragmented, or unverifiable, e.g. in one publicised case, a dataset made available by Meta was shown to be inaccurate. [9]

Undermining research integrity, innovation and creativity

Initiatives to improve data can also run the risk of undermining broader values of research independence and can raise concerns about equity. For example, initiatives in which researchers are placed in private companies, and are therefore granted access to datasets, often require researchers to develop strong relationships with large data holders. These initiatives are certainly a positive opportunity for building cross-sectoral relationships and sharing knowledge and data, but they can also create challenges for researchers by potentially undermining the independence and legitimacy of research projects. Moreover, such partnerships are often most accessible to more senior academics, particularly those based in the USA, raising questions about the equity of access to data.

Some initiatives compel researchers to work according to strict parameters, requiring them to have a fully formulated project before accessing data rather than allowing them to develop their research questions in a more organic and open fashion. This can limit the exploratory analysis that uncovers truly novel insights and unexpected trends from online data.[10]

6. What are the legal challenges/risks to sharing information from online services with independent researchers?

Businesses that own data from online services and hold large shares of personalised and sensitive data often have concerns relating to maintaining the privacy of users; security implications; the risk of compromising their legal compliance, and the potential to undermine their commercial interests and competitiveness by disclosing information, ideas, or strategies. Large companies refer data access requests to legal and compliance teams, who tend to be conservative and risk averse, and this can result in restrictive arrangements. Some businesses may be cautious of data sharing due to the potential for reputational damage if data is misused.

7. 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?

The British Academy's summary brief highlighted that digital infrastructure, capacity, and resourcing are essential to safely and more effectively enable data access and sharing.[11] This might include hosting online platforms for data access, providing large-scale compute facilities, cloud and data storage repositories, permission-based access,

---

[8] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief', p. 8.
[9] The Washington Post (2021), Facebook made big mistake in data provided it provided to researchers, undermining academic work.
[10] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief', p. 8.
[11] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief'.

accreditation, vetting researchers, and carrying out the appropriate auditing processes to track who is accessing data. Maintaining this infrastructure so that it is implemented effectively requires significant resource, so creating ways that enable services to do this is important.[12]

8. What are the information quality challenges relating to online services sharing information with independent researchers?

The summary brief noted that quality of information can be impacted by online services providing, outdated, partial and fragmented data sets, which can lead to inaccuracies in evidence and research.[13]

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

The workshops convened by the British Academy revealed the need to approach the challenge of access to information for research in a more holistic way. The summary brief suggests moving away from a 'one size fits all' approach that assigns responsibility for enabling access to a single group of stakeholders, as workshop participants instead called for attempts to build a more trustworthy and open data ecosystem – one that can provide incentives for businesses, improve research, and generate valuable evidence for policymakers.[14] The brief points to the following mechanisms to do so:

Legislating on data access for research

To effectively strengthen sharing agreements and access protocols for researchers, making actions such as reforming data sharing and standards, improving data laws, and implementing corresponding protections for personal data and privacy more explicit in legislation such as the Data Protection and Digital Information Bill will be crucial.

Addressing conflicting messaging and shifting narratives on the value of data (e.g. for online safety)

Participants at the workshops noted that narratives around the opportunity for data to help post-Brexit Britain achieve a status as an innovation superpower contrasts with narratives around Britain taking a more community-oriented approach to data and being a leader in online safety. These contradictions in messaging can undermine trust in decision makers and those working to facilitate greater access to information for research.

Careful consideration of how to communicate the value of information is vital to help shift these narratives. Demonstrating how data can provide socially beneficial insight into human behaviour and lead to better policy analysis and decision-making can incentivise data sharing. This is of particular importance when articulating to the wider public whose data has been collected and what its value could be in research contexts.

Supporting and enhancing the role of trusted data institutions

Data institutions such as Smart Data Research UK, Health Data Research UK, Research Data Scotland, Administrative Data Research UK, subsidiary organisations such as Smart Data Foundry, and software based platforms such as OpenSAFELY have an explicit focus on ensuring that data is available in ways that do not cause harm to people, promoting

[12] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief', p. 11.
[13] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief', p. 8.
[14] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief'.

approaches that steward data in a responsible and safe manner, and often work with or on behalf of private companies. These trusted data institutions:

- Have the capacity and resources to be more proactive in setting best practice and are equipped with the knowledge on how to support safe and secure access for researchers, working directly with policymakers, industry and the public.
- Provide the necessary digital infrastructure that is required to unlock the power of data. This might include hosting online platforms for access, providing large-scale compute facilities, cloud and data storage repositories, permission-based access, accreditation, vetting researchers, and carrying out the appropriate auditing processes to track who is accessing data.
- Help to build trust in relationships between researchers and data owners alongside facilitating dialogue on challenges and solutions around data access and sharing

Enabling cross-sector engagement

Cross-sector engagement is vital to a holistic approach to online data sharing. This includes creating incentives and processes that reassure stakeholders across sectors that sharing can be done safely. This could entail establishing online data forums, networks or committees to help share information between policymakers, industry and the research community on the challenges and opportunities of online data. Such mechanisms can move the agenda on data access forward in the UK in ways that encompass a broader diversity of voices on the challenge, rather than leaving engagement to occur in a siloed way or exclusively in industry circles. It is particularly important that engagement is sustained over time rather than undertaken as a short-term activity, as addressing complex issues of access may take time and will require ongoing attention.

10. 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 British Academy's workshop series gathered insights on existing models that can provide researchers access to sensitive information. These include research based models such as Smart Data Research UK, Health Data Research UK, Research Data Scotland, Administrative Data Research UK, Consumer Data Research Centre. Subsidiary organisations such as Smart Data Foundry, and software-based platforms such as OpenSAFELY have an explicit focus on ensuring that data is available in ways that do not cause harm to people, promoting approaches that steward data in a responsible and safe manner, and often work with or on behalf of private companies. Although these institutions work with a range of datasets and across various sectors, they represent examples of best practice and have a targeted capacity to address a range of challenges that exist in this space.

11. 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?

The British Academy's summary brief noted that organisations and networks with functions at the intersection of research and policy can play a role in advancing access by engaging in discourse and facilitating the creation of policy networks.[15]

NGOs such as Data for Policy foster dialogue on the impact and potential of data, AI and related technologies in government, governance, and policy. The Open Data Institute work to cement trust in data and data sharing practices to further strengthen engagement. The UK's national academies can also play an informal role in the space by harnessing their convening power to encourage more exploratory and creative avenues of engagement, such as hosting workshops that can enable industry professionals, researchers and policymakers to work more collaboratively and identify a range of innovative solutions to data access. Research councils such as UKRI can help push this agenda forward through cross-sector engagement, creating the infrastructure for academics to be able to access data in a safe and secure way, removing risks from individual academics and helping to build processes that assuage companies' concerns while also including subjects such as Archaeology and Geography that combine SHAPE and STEM methodologies to help develop a regulatory environment that will promote the benefits of AI for all. This requires an intellectual property rights environment (legislation and enforcement) that on the one hand encourages the best research, the best education and the right kinds of competition amongst key actors and, on the other hand, ensures proper and fair rewards for creativity.

However, there are concerns that the current consultation on copyright and AI does not identify the most important issues that face the university sector, the creative industries, and the research and development environment more generally. The key issues are the following:

- Removing obstacles to improving the quality and quantity of top-quality scientific research and education for the leading-edge UK university sector, while also recognising the specific imperatives of the humanities, social sciences and cultural industries.
- Improving and clarifying the licensing environment. This must include transparency, so that access to quality content for R&D becomes easier for all engaged in scientific, educational and creative work with appropriate recognition, consistent with open access principles.
- Improving revenues of primary creators as well as, where appropriate, recognising the moral rights of creators to support the UK's world-leading arts and creative industries.

Addressing these issues is not straightforward. Doing so will necessarily engage questions of copyright, but also many wider issues. The British Academy is uniquely placed to bring together leading researchers in the SHAPE disciplines to address these questions. We call

[15] The British Academy (2024), 'Data as a Tool for Researching the Digital World: Summary Brief'.

upon the government to engage with the British Academy and similar institutions to develop this wider conversation.