ISD | Powering solutions to extremism and polarisation

CASM technology

Policy Approaches to Addressing Data Access Challenges in the Evolving Online Ecosystem

Henry Tuck

About this publication

Over the past 18 months, the Institute for Strategic Dialogue (ISD) and CASM Technology have been conducting research, funded by Omidyar Network, aimed at monitoring and analysing harmful online activity on a range of small or medium-sized online platforms, and the data access challenges they present to researchers.

This policy briefing assesses the extent to which existing or proposed legislation in key jurisdictions would improve access to data for researchers from smaller or medium-sized platforms. It provides a series of recommendations for the private sector, policy-makers and the research community to address the data access challenges identified related to smaller and medium-sized platforms.

Acknowledgments

This policy briefing would not have been possible without funding support from Omidyar Network. We would like to express our gratitude to Wafa Ben-Hassine, Anamitra Deb and Emma Leiken for their vision, continued support and insightful feedback.

The author would also like thank the wider team for their contributions that have made this project possible: Jakob Guhl, Julia Smirnova and Lea Gerster at ISD, and Oliver Marsh and Nestor Prieto Chavana at CASM.

Author

Henry Tuck is the Head of Digital Policy at ISD, where he leads work on digital regulation and tech company responses to terrorism, extremism, hate and dis/misinformation online. Henry oversees ISD's Digital Policy Lab (DPL) programme and advisory work on key digital regulation proposals in Europe and Five Eyes countries, and collaborates with ISD's Digital Analysis Unit to translate research into actionable digital policy recommendations.

ISD | Powering solutions
to extremism
and polarisation

Amman | Berlin | London | Paris | Washington DC

www.isdglobal.org

Contents

Introduction

Extremist, hate and disinformation actors are increasingly adopting not just major online platforms, but also a wide range of smaller and medium-sized platforms across the online ecosystem. Many of these platforms present significant barriers to researchers attempting to understand the types of risks or harms present. These barriers to online research can have a negative impact on the quality, comparability and replicability of online research into key societal issues. Despite recent or upcoming regulatory changes in key jurisdictions, and the risks smaller and medium-sized platforms can present, many of the barriers to conducting research on them are yet to be fully or even partially addressed.

Over the past 18 months, the **Institute for Strategic Dialogue (ISD)** and **CASM Technology** have been conducting research, funded by Omidyar Network, aimed at monitoring and analysing harmful online activity on a range of these small or medium-sized platforms, and the data access challenges they present to researchers. This Policy Brief begins with a summary of the key findings from the project and feedback received during engagements with other researchers, policy-makers, regulators and the private sector. It then provides an overview of existing or proposed legislation in key jurisdictions, including the EU, US, UK, Australia, New Zealand and Canada, assessing the extent to which they would impact access to data for researchers from smaller or medium-sized platforms. It concludes with a series of recommendations for the private sector, policy-makers and the research community to address the data access challenges identified related to smaller and medium-sized platforms.

Key Project Findings

Our initial research, covered in the first project report, Researching the Evolving Online Ecosystem: Barriers, Methods and Future Challenges (July 2022)[i], led us to distinguish between three key types of barriers to researching and countering harmful activity on a wide range of online platforms:

- **Technological barriers** that can be posed for example by encryption, AI-generated content, blockchain, decentralised platform structures or different content formats (e.g. audio-visual).
- **Ethical and/ or legal barriers** that can arise from expectations of privacy, legal restrictions, platform Terms of Service (TOS) that prohibit legitimate public interest research, or difficulties in obtaining informed consent from users.
- **Fragmentation barriers** presented by platforms where online content is public and accessible but not systematically searchable, for example via an API.

Our second report, Researching the Evolving Online Ecosystem: Telegram, Discord & Odysee (April 2023), demonstrated through three platform case studies how **these barriers combine in ways that force researchers to make difficult trade-offs** between competing goods. These included the desire to understand and mitigate harmful content and behaviour online, the need to develop a systematic evidence base for online harms, and the preservation of privacy or the avoidance of legal risks.

Overall, the research we conducted illustrates that **it is possible to conduct digital research on platforms such as Telegram, Discord and Odysee, but it can be difficult to do so in a way that is simultaneously systematic, ethical and legal**. We were nevertheless able to conduct useful research on each platform, especially with more manual or ethnographic methods, which was complemented with some more systematic, larger-scale data collection and analysis.

However, there were also **significant limitations on each platform** that prevented the use of more comprehensive research methods and approaches:

- On **Telegram**, ethical considerations around the right to privacy or expectations of privacy prevented us from researching smaller, potentially more harmful communities.
- On **Discord**, legal concerns around breaking TOS stopped us from systematically collecting data from Discord servers altogether, forcing us to rely on more manual methods.
- **Odysee** presented fewer ethical concerns due to its public nature, yet technical complexity and unclear TOS regarding data collection also made more comprehensive analysis challenging.

These technological, ethical and legal, and fragmentation barriers encountered on Telegram, Discord and Odysee were also deeply interconnected and overlapping. For example, legal considerations prevented us from overcoming fragmentation barriers on Discord through systematic search methods, and technological features on Telegram created significant ethical barriers. The complications of integrating multiple tools to access Odysee data were worsened by the platforms' TOS, which are unclear as to whether different tools are providing personal data, thereby creating ethical and legal uncertainty. Ethical and legal barriers therefore often limit researchers' ability to study harmful communities on these platforms in a systematic way, even where such approaches may be technically possible.

[i] See also: ISD, Researching the Evolving Online Ecosystem: Executive Summary & Annex - Ethical, Legal & Security Risks (July 2022)

Implications for Researchers

Our research during this project has demonstrated how platforms can, whether deliberately or inadvertently, inhibit public interest research in several ways:

- **Technological choices by platforms can prevent access to data or hosted content, or make data collection unnecessarily burdensome from a technical perspective**. For example, where platforms' front-end or back-end architectures are fragmented or poorly maintained, or the tools available to access data are poorly documented. The expertise (and time) required to explore the technical possibilities across a variety of different platforms may not be available to all researchers, particularly in civil society.

- **Unclear platform legal terms can create additional burdens or risks for researchers, raising the costs and barriers to entry for online research**. We were fortunate to have access to external, pro bono legal support that enabled us to more carefully assess platforms' TOS and any resulting legal risks related to data access, but again such support may not be available to all researchers.

- **Platforms can deliberately invoke legal or ethical concerns to justify restricting researchers' access to data, including around data protection or users' rights to privacy**. This can occur even in cases where users have actively consented to the sharing and use of their data for specific, limited research purposes. Alternatively, the designation of online spaces as private can create ethical uncertainty despite the content or data being easily and widely accessible to both users and researchers in practice.

- **There is a danger that the research community's scarce resources for addressing these challenges are being employed in a disconnected and siloed way**. Consequently, without strong mechanisms and incentives for greater collaboration, opportunities to achieve greater economies of scale are potentially being missed.

- **Barriers to online research can create a problematic set of incentives for researchers when it comes to balancing or mitigating different legal and ethical risks**. For example, deception may actually serve to reduce legal risks for researchers utilising third-party research tools that contravene platform TOS. If platforms or users are not aware that data collection is taking place, then the likelihood of researchers facing legal action is reduced significantly.

- **The use of third-party tools or deceptive methods can disincentivise researchers from publishing their work and/or incentivise them to be vague about research methodologies**. This in turn has a negative impact on the quality, comparability and replicability of online research into key societal issues and can engender further distrust between platforms and the wider research community.

Overall, we argue there is a pressing need for a set of consistent baseline standards for researchers' access to platform data, both in terms of what data researchers are allowed to access and how it can be accessed. This is especially the case with research into new or emerging online platforms. Such standards would help to level the playing field and reduce the current asymmetries between platforms and researchers.

Stakeholder Engagement

Research Community Experiences

In June 2023, ISD held a roundtable with other researchers from academia and civil society, alongside policy-makers and representatives from regulators, to gather feedback on our findings and consider any additional challenges they have faced when attempting to access data from online platforms. Participants shared their experiences of researching a wide range of platforms, and overall stressed that while emerging platforms and technologies do present significant and novel challenges, many larger more established platforms also still present considerable barriers to data access. Examples cited by participants included:

- **Live-streaming social media platforms** (e.g. Twitch), **audio platforms** (e.g. Spotify) and **virtual reality spaces and video games**, particularly those with added social media elements (e.g. Fortnite), where systematic research is particularly time-consuming due to the volumes of audio-visual content available;
- Social media platforms with **ephemeral or time-limited content** (e.g. Snapchat) where relevant content is not retained and therefore out of the reach of researchers;
- **Decentralised social media platforms** (e.g. Mastodon) with complex technological architectures or dispersed 'instances' or servers that cannot be accessed or searched simultaneously;
- **Alternative social media platforms** popular with extremists (e.g. Gab, 4Chan, 8kun and Bitchute) that may be hostile to external researchers and/or have unclear TOS;
- **Messaging services** (e.g. WhatsApp, Signal or Threema) where privacy concerns present ethical barriers, and technological barriers to access mean that little data is available;
- **Donation and fundraising platforms** where little data or transparency is available on the sources of funds;
- **Search engines** that do not provide access to certain types of data required for researchers.

Beyond the technological, ethical, legal, and fragmentation barriers outlined above, participants also highlighted **financial barriers as a growing challenge** in their work. This was raised particularly in the context of X's (formerly Twitter) recent decision to significantly raise the costs of API access with reasonable rate limits, but was also flagged as a challenge when needing to store large amounts of audio-visual data from other platforms. As outlined in the Annex to our first report, several participants also noted that online research can present **safety or security risks**, for example if a platform is monitored by security agencies in authoritarian states, or if researchers are harassed and threatened by those they monitor.

Participants from academia also cited specific sectoral and institutional barriers. These included time-restricted contracts leading to knowledge about risks and their mitigations being lost as colleagues moved on, and institutional challenges to greater interdisciplinary collaboration. They also identified challenges associated with the slower pace of research in academia compared with the speed at which the online ecosystem and data access challenges can evolve, with significant time spent preparing funding applications and awaiting the outcomes of the peer-review process. Participants also noted that ethical barriers tend to be higher for academic researchers, with institutional review boards varying in terms of their flexibility and expertise regarding the specificities and challenges of online research.

Finally, while noting their importance in protecting users' privacy, both academic and civil society researchers noted **legal barriers related to data privacy regulations** that can be difficult to navigate in terms of the types of data that can be collected, stored and shared, particularly across different jurisdictions and data protection regimes. For example, sharing data between researchers, partner organisations or different legal entities within an organisation that are based in different jurisdictions can create additional barriers to cooperation and data sharing, particularly between the EU and other jurisdictions such as the US.

Looking ahead, participants cited reasons for being both optimistic and pessimistic about future prospects for data access from platforms. Participants were concerned about the recent trend of platforms raising barriers to data access, and the relatively slow process of developing and implementing data access mechanisms required in incoming regulation such as the EU's Digital Services Act (DSA). Participants also expressed concerns that platforms will continue closing themselves off to systematic external data collection to prevent their data from being used to train AI models, or to evade scrutiny regarding the nature or extent of harmful activity on their platforms. Conversely, evolving research tools, including future potential applications of AI to conduct research, as well as incoming regulation were noted as positive developments. Participants expressed hope that the DSA, and comparable legislation in other jurisdictions, should lead to positive shifts in data access, at least for larger online platforms. However, they also highlighted the risks of large platforms delaying implementation, challenging, or not fully complying with regulatory or co-regulatory obligations, citing the recent example of X pulling out of the EU Code of Practice on Disinformation in May 2023.[ii]

Insights from Policy-makers and Regulators

ISD held a second roundtable in June 2023 with policy-makers and regulators to explore possible solutions to the challenges researchers are facing when attempting to access data from online platforms. During the discussion, participants highlighted the challenges they face in assessing the risks different online platforms can pose given the breadth, complexity and diversity of the online ecosystem, and acknowledged the short-term priority of protecting existing means of data access, such as APIs, especially in the context of real-time access to data.

Participants emphasised the importance of developing and leveraging relationships with the research community to support regulators. As a result, participants stressed the need for policy-makers and regulators to engage with researchers to understand the range of research tools and methodologies they rely on, the barriers they face, and the types of data and support they require. Public consultations are one avenue for this, with recent requests for inputs on data access requirements and mechanisms by the European Commission and France's ARCOM cited as examples.[iii] ISD submitted responses to both of these consultations, which can be found here and here respectively, and also collaborated with the Mozilla Foundation and other civil society experts in crafting a joint statement to the European Commission under the same consultation.

Another example cited was the recent establishment of the Canadian Digital Media Research Network (CDMRN), which will be independently managed by The Media Ecosystem Observatory, a research initiative led by McGill University and the University of Toronto, with support from Heritage Canada's Digital Citizen Initiative.[iv] The CDMRN will conduct ongoing data collection and analysis, develop infrastructure to share data across the Canadian research community, build a community of practice across academia, civil society and the media, offer training on digital research methods, and support the dissemination of research to policy-makers and the Canadian public.[v]

Participants also stressed the importance of establishing suitable data access infrastructure, such as the intermediary body envisioned under the EU DSA currently being discussed by the multistakeholder European

[ii] TechCrunch, Elon Musk takes Twitter out of the EU's Disinformation Code of Practice (27 May 2023)

[iii] ARCOM, Access by researchers to platform data: summary of responses to the ARCOM consultation and proposals (June 2023);
European Commission, Call for evidence: Delegated Regulation on data access provided for in the Digital Services Act (May 2023)

[iv] Government of Canada, Government of Canada working with Civil Society to Strengthen Defences against Online Disinformation (June 2023)

[v] Media Ecosystem Observatory, Canadian Digital Media Research Network (CDMRN)

Digital Media Observatory (EDMO) Working Group for the Creation of an Independent Intermediary Body to Support Research on Digital Platforms. The Working Group was established in May 2023, and over the next six months will: "(a) identify appropriate governing principles for the new intermediary body, (b) lay out its core functions, (c) outline an organizational structure, staffing, and budgetary needs, (d) identify an appropriate form and place of establishment, and (e) provide a timeline for the body's initial phases of work".[vi]

In the context of this intermediary body, participants raised concerns that it could be overwhelmed by requests for data once it is established, and would therefore need both sufficient resources and a suitable approach to prioritising requests. To help mitigate this, the research community was also encouraged to ensure requests would be as streamlined and precise, and to collaborate and coordinate as widely as possible. Regulators also noted that the experience of setting up and managing new data access mechanisms for larger platforms would provide valuable learnings for potentially working with smaller platforms in future, even if they were not required to do so under the DSA. These lessons could be applied either in the context of future regulatory requirements, or approaches to encouraging smaller platforms to voluntarily providing greater access to data for research.

Privacy, data protection, confidentiality and cyber-security issues were also raised, with participants suggesting that an intermediary body would be well situated to 'host' data from platforms in a secure way, addressing platforms' concerns over potential liability if data were to be misused. The definition of 'researchers' was also discussed, with agreement that this would need to be carefully designed to ensure it would incorporate both academic and civil society or non-profit organisations conducting public interest research, but not allow for commercial, for-profit research. In this context, participants also stressed the need to develop appropriate criteria, systems and processes to vet and accredit researchers and, where possible, ensure these are aligned across jurisdictions where similar mechanisms may be introduced in future to avoid unnecessarily divergent requirements that would create additional burdens for researchers and platforms.

The Centre d'accès sécurisé aux données (CASD) was cited as another example of an existing secure infrastructure for sharing data with researchers covering areas such as health, business and finance, employment, justice, education, agriculture and the environment.[vii] The CASD was established in France in 2018 as a consortium to bring together various research-focused state bodies and French universities to establish mechanisms and infrastructure to facilitate the secure sharing of and access to confidential data for non-profit research, and promote the technology developed under the initiative. Researchers accessing data receive unique 'digital fingerprints', and cannot remove data from CASD servers. This infrastructure helps to reduce technical and financial costs for researchers, although does not allow for access to real-time data. To date, the initiative hosts 505 data sources and has facilitated 1368 research projects from 934 research institutions, resulting in over 400 publications.

As part of the Digital Policy Lab, ISD also previously organised a series of working group sessions on data access between October and November 2022. The working group consisted of representatives from national ministries and regulators from Canada, France, Ireland, Slovakia, Switzerland, the US and the UK, as well as participants from civil society, academia and the private sector. Based on these discussions, ISD produced a policy paper reviewing key lessons learnt from industry-academia partnerships and other types of existing voluntary initiatives on data access, and exploring potential future avenues for international collaboration among liberal-democratic governments, and providing targeted recommendations applicable across government initiatives to support the alignment of data access approaches.[viii]

[vi] European Digital Media Observatory (EDMO), Launch of the EDMO Working Group for the Creation of an Independent Intermediary Body to Support Research on Digital Platforms (May 2023)

[vii] Centre d'accès sécurisé aux données (CASD)

[viii] ISD, Bundtzen, Sara & Schwieter, Christian, 'Access to Social Media Data for Public Interest Research Lessons Learnt & Recommendations for Strengthening Initiatives in the EU and Beyond' (May 2023)

Overview of Legislation in Key Jurisdictions

The following section provides an overview of existing or proposed pieces of regulation in key jurisdictions, and briefly assesses whether they could have an impact on the barriers to research presented by smaller and medium-sized online platforms.

EU: Digital Services Act (DSA) and Strengthened Code of Practice on Disinformation

Under the EU's DSA, passed in July 2022, platforms will be designated as 'very large online platforms' (or VLOPs) if they average 45 million or more 'monthly active users' (MAUs) across EU countries. In April 2023, the European Commission designated seventeen platforms as VLOPs (and two as 'very large online search engines' or VLOSEs).[ix] These larger platforms will be required to provide real-time data access where technically possible, provided that the data is publicly accessible (Article 40). This could include metrics like aggregated interactions with content from public pages, public groups or accounts, including impression and engagement data, such as the number of reactions, shares and comments.

Access would need to be provided within a reasonable period upon request by the national-level regulator, the Digital Services Coordinator (DSC), in the EU country in which the company is established, or the European Commission (EC). DSCs and the EC may only use the data provided for monitoring and assessing compliance with the DSA, and they must consider the rights and interests of providers and users, including the protection of personal data, confidential information (trade secrets) and maintaining the security of the service. Data would need to be provided through appropriate interfaces specified in the request from DSCs or the EC, including online databases or APIs.

Similarly, if requested by the DSC, data can be provided to vetted researchers to support the detection, identification and understanding of the systemic risks identified in the DSA in EU contexts. Vetted researchers must be affiliated with a research organisation, be independent from commercial interests and be able to meet certain data protection and security standards. Researchers will need to demonstrate the necessity and proportionality of their data access requests, disclose their funding and make any research freely and publicly available. Platforms will need to anonymise or pseudonymise personal data unless doing so would make the intended (and legitimate) research objectives impossible. Upcoming delegated acts (secondary EU legislation) will introduce further conditions, procedures and independent advisory mechanisms to support the sharing of data with external researchers. While the EC has conducted a consultation on the delegated acts, as of August 2023 it has yet to publish a draft.

In 2022, the EU also introduced a new Strengthened Code of Practice on Disinformation, a voluntary code currently with 34 signatories, including a range of organisations with an interest in combatting disinformation and large tech companies like Meta (Facebook and Instagram), Google (YouTube), TikTok, Microsoft and Twitch, as well as a number of smaller companies such as Clubhouse and Vimeo.[x] Under 'Section VI. Empowering the Research Community', the code contains a series of data access provisions, including Commitment 26 for signatory platforms to provide non-personal data and anonymised, aggregated or public data for research purposes on disinformation.

As with the DSA, platforms should provide real-time (or near real-time) machine-readable data access through APIs or other open and accessible technical mechanisms, as well as ensure reasonable tools and processes are in place to mitigate risks of abuse (such as malicious or commercial uses of data). In order to qualify, proposed research must be compliant with ethical and methodological best practices, for example, those contained in EDMO's draft Code of Conduct on Access to Platform Data, and research teams may include civil society as well as academic organisations.[xi] The Strengthened Code of Practice on Disinformation also includes Commitment 27 to create an independent, third-party body to vet researchers and research proposals.

[ix] European Commission, 'DSA: Very large online platforms and search engines' (April 2023)
[x] European Commission, 'Signatories of the 2022 Strengthened Code of Practice on Disinformation' (June 2022)
[xi] EDMO, 'EDMO releases report on researcher access to platform data' (May 2022)

Of the platforms assessed during this project, none are likely to reach the threshold required in the DSA for the data access requirements outlined above to apply.[xii] As a result, while the DSA's requirements should go a long way to addressing current barriers to research on larger platforms, they may not help to solve many of the challenges outlined in this report related to smaller or medium-sized platforms. Similarly, none of the platforms we examined are currently signed up to the voluntary Strengthened Code of Practice on Disinformation, meaning that it will also not address barriers to research on smaller and medium-sized platforms such as Telegram, Discord or Odysee unless or until they commit to its provisions.

US: Platform Transparency and Accountability Act (PATA), Digital Services Oversight and Safety Act (DSOSA) and Kids Online Safety Act (KOSA)

Similarly to the DSA, proposed legislation in the US such as PATA[xiii], DSOSA[xiv] and to a lesser extent KOSA[xv] contain user number thresholds to determine the obligations for different sized platforms. In PATA, which was updated and reintroduced in June 2023 with additional bipartisan support, platforms hosting user-generated content would need to have at least 50 million (up from 25 million) 'unique monthly users' in the US to qualify (Section 2). The legislation would require the National Science Foundation (NSF) to identify the data and information that platforms must make available to qualified researchers. These have to be feasible for the platform to provide; proportionate to the needs of researchers to complete the research; and not create undue burdens for the platform (Sec. 2). Qualified researchers seeking access to data must be affiliated to a US university or non-profit and submit applications to the NSF for each specific research proposal, and would not be able to request access to direct or private messages, biometric or precise location data (Sec. 2).

The NSF would also establish a process to solicit research applications from researchers and define guidelines and criteria to determine how to evaluate those applications. The Federal Trade Commission (FTC) would determine appropriate privacy and security safeguards for the data that platforms would be required to provide, such as encryption or anonymisation, and researchers would need to demonstrate compliance (Sec. 3). Researchers would also be required to submit a pre-publication version of their research to the FTC for evaluation to confirm that the analysis does not expose personal information, or trade secrets (Sec. 3). If they also comply with these safeguards, platforms would receive liability immunity for any misuses of data released for research purposes, and would be able to restrict access if they determined researchers were not adhering to the safeguards (Sec. 4). The FTC would also have enforcement powers to ensure compliance from both platforms and researchers (Sec. 7), although a previous provision that would have removed platforms' Section 230 liability protections for non-compliance have been removed from the updated version of the bill.[xvi]

PATA would also provide protection to researchers via the creation of a 'safe harbour' to prevent platforms from taking legal action against researchers who obtain information consensually and with other privacy protections in place (Sec. 8). Finally, the bill would also allow the FTC to require certain transparency reporting or disclosures from platforms be made available to researchers or the public. This could include ad libraries, information about widely disseminated content or major public accounts, information about content moderation decisions or information about algorithms (Sec. 9).

[xii] At around 38.5 million, Telegram claims to have fewer users than the 45-million threshold for VLOPs in the EU. See: Telegram, 'FAQ'.
[xiii] US Congress, PATA – the text of the revised bill is available here. A one-pager on the bill is available here.
A section-by-section summary of the bill is available here.
[xiv] US Congress, DSOSA – the text of the bill is available here. A section-by-section summary of the bill is available here.
[xv] US Congress, S.3663 - Kids Online Safety Act
[xvi] Tech Policy Press, 'Platform Accountability and Transparency Act Reintroduced in Senate' (8 June 2023)

DSOSA, modelled to some extent on the EU's DSA, sets the threshold for 'large covered platforms' at 66 million MAUs, and 10 million for 'covered platforms'. Similarly to PATA, the bill would require the FTC to issue rules regarding the types of data that should be made available to certified researchers (Sec. 10(c)), and the mechanisms for data access. It would also require large covered platforms to provide a detailed ad library (Sec. 10(f)); access to key metrics for high-reach and high-engagement public content; and transparency on content that a platform amplifies (Sec. 10(g)). The FTC could choose to sponsor a Federally Funded Research and Development Center (or FFRDC) to facilitate information sharing between covered platforms and certified researchers. The FTC would also be required to ensure that data access does not infringe upon reasonable expectations of personal privacy of users (e.g. requiring platforms to anonymise any information that is not public content) and consider when privacy-preserving techniques, such as differential privacy and statistical noise, should be used (Sec. 10(c)).

The bill would also create an Office of Independent Research Facilitation at the FTC that would certify researchers from academia and civil society to study the impact of content moderation processes, product design decisions and algorithms on society, politics, the spread of hate, harassment and extremism, security, privacy, and physical and mental health (Sec. 10(a)). To qualify, research organisations would need to be either a higher education institution or a civil society organisation (501(c)(3)), which has a mission that includes developing a deeper understanding of the impacts of platforms on society; has the capacity both to comply with rules for secure researcher access and to use appropriate data science; and adopts investigative and qualitative research methods and best practices (Sec. 10(b)). The bill would also introduce safe harbour provisions for certified researchers that create accounts solely for a research project or collect information provided for research purposes by a user (e.g. via a browser extension or plug-in) where the user has provided informed consent (Sec. 10(c)).

Several other tech-related bills are also being considered in the Senate and House that include provisions on researcher access to platform data. For example, KOSA would allow academic and civil society (501(c)(3)) research organisations to access platform data to assess specific child-related harms, but only on 'covered platforms' with more than 10 million MAUs (Sec.7). 'Covered platforms' would include "a social media service, social network, video game, messaging application, video streaming service, educational service, or an online platform that connects to the internet and that is used, or is reasonably likely to be used, by a minor" (Sec. 2(3)). KOSA does also include 'safe harbor' protections for researchers violating platform terms of service when conducting public interest research into child-related harms, as long as they have taken reasonable measures to protect user privacy and security.

As with the DSA in an EU context, the thresholds set in these proposals are likely to exclude many smaller-sized platforms. While the lower 10 million thresholds for 'covered platforms' in DSOSA and KOSA may potentially capture some medium-sized platforms, such as Telegram or Discord, the most impactful data access requirements are aimed at large covered platforms. As a result, none of these proposals would comprehensively address the barriers to research on smaller platforms encountered during our work for this project. It should also be noted that the likelihood of these federal bills being passed in Congress is considered to be relatively low due to partisan disagreements and a lack of consensus over the path forward for social media-focused regulation. This suggests that the US may lag behind on regulation in comparison to new data access requirements being introduced in other jurisdictions, at least in the short to medium term.

There have however been some positive developments from the US-EU Trade and Technology Council meetings held in May 2023 that demonstrate the Biden administrations' desire to improve researchers' access to platform data, and align with the provisions for data access under the DSA.[xvii] An Annex to the joint statement provided by the White House and EC states that "it is crucially important for independent research teams to be able to investigate, analyze and report on how online platforms operate and how they affect individuals and society". It also recognises the limitations of existing

[xvii] White House, 'U.S.-EU Joint Statement of the Trade and Technology Council' (31 May 2023)

data access options provided by platforms on a voluntary basis, echoing our findings during this project, stating that these are "often narrow in scope, unclear in their limitations, unpredictable in terms of access conditions, and bespoke such that cross-platform research and replicability are difficult." The Annex calls for platforms to "share meaningful data and testing opportunities for the purpose of independent research by vetted researchers" with sufficient data protection and privacy safeguards across a similar range of risks as those included in the DSA, and that "information sharing mechanisms should build upon and be consistent with applicable legal and policy frameworks".[xviii]

UK: Online Safety Bill (OSB)

The OSB (as of July 2023) still lacks clarity over a potential future data access regime, with the UK Government having made some changes but rejected other proposed amendments during its passage through the House of Lords.[xix] Currently as the designated regulator, the Office of Communications (Ofcom) would have powers to require data from platforms (Section 101) and commission third-party 'reports by skilled persons' (Section 105) to support the regulator.

Additionally, within 18 months of the legislation being adopted, Ofcom would be required to produce a report describing how and to what extent those carrying out independent research into online safety are currently able to obtain information from regulated platforms (Section 163). The report would also explore legal and other issues that currently constrain data access, as well as assess the extent to which greater access could be achieved. Ofcom would be required to consult with various third-parties, including the Information Commissioner (the regulator responsible for data protection) and other stakeholders including platforms, the Centre for Data Ethics and Innovation (CDEI) and United Kingdom Research and Innovation (UKRI). Following the publication of the report, Ofcom would then have to produce guidance for regulated platforms and researchers on access to data, and include an assessment of its effectiveness in their own transparency reporting. However, Ofcom would not have powers to compel platforms to comply with the guidance, as data access is not included under the 'enforceable requirements' (Section 132).

However, while the OSB may not currently provide strong data access provisions, it does have a broader definition for in-scope platforms without the strict user number thresholds seen in other legislation in the EU or proposed legislation in the US. It covers "user-to-user services" (Section 3) with "a significant number of UK users", where "UK users form one of the target markets for the service", or where the service can be used in the UK or could present risks of harms to individuals in the UK (Section 4). Regulated services will then be designated as Category 1, Category 2A (search services) or Category 2B services, with Category 1 facing the strictest obligations. The precise threshold conditions for each category will be determined in secondary legislation, but will be based on: "(a) number of users of the user-to-user part of the service, (b) functionalities of that part of the service, and (c) any other characteristics of that part of the service or factors relating to that part of the service that the Secretary of State considers relevant", as well as "the level of risk of harm to individuals from illegal content and content that is harmful to children" (Schedule 11).

Overall, in comparison to the EU's DSA, the OSB contains relatively weak provisions and a lack of clarity on data access for researchers, which may not be clarified until several years into the new regulatory regime. It therefore remains unclear the extent to which, if at all, the OSB would address the barriers to data access identified during this project on either smaller or medium-sized platforms, or larger platforms. Despite this, the way in which in-scope services are defined does provide some potential flexibility in future for the data access guidance to be applied to smaller or medium-sized platforms, depending on the eventual secondary legislation that is put in place. As the OSB has not yet been finalised and passed into law, there may still be further opportunities to strengthen the data access provisions when it returns to the House of Lords and then the House of Commons in September 2023.

[xviii] European Commission, 'Transparent and accountable online platforms' (31 May 2023)
[xix] UK Parliament, 'Online Safety Bill [AS AMENDED ON REPORT]' (July 2023)

Australia: Online Safety Act

Australia passed a new Online Safety Act in June 2021, which came into force in January 2022. The Act is intended to strengthen Australia's existing online safety regulations, and provides the eSafety Commissioner with additional powers to address a range of illegal online content and harms to children.[xx] The legislation covers a wide range of online services, including social media platforms, messengers, search engines, app stores, internet service providers (ISPs) and hosting services, who are expected to develop cross-industry codes. The Act also introduced the 'Basic Online Safety Expectations' that require online services to take reasonable steps to mitigate various risks, which the Commissioner can then require them to report on to provide additional transparency, backed by enforcement powers such as civil penalties.

To date, the Commissioner has issued three rounds of notices since August 2023, the first two focusing on child sexual exploitation and abuse, and the latest in June 2023 to X regarding their efforts to combat online hate.[xxi] The first set of notices have been responded to, with a summary of responses published in December 2022.[xxii] The summary provides information that had not previously been available through companies' own transparency reporting or other voluntary initiatives. However, beyond these notices, and a responsibility for the Commissioner to "to support, encourage, conduct and evaluate research about online safety for Australians" (Art.27g), there are no other mechanisms to broaden transparency or data access for independent research in the Act.

New Zealand

The New Zealand Government launched a public consultation process in June 2023 on 'Safer Online Services and Media Platforms' as an initial step towards the introduction of new legislation to regulate both social media and traditional media platforms.[xxiii] The accompanying document provides a broad overview of the direction New Zealand is considering - a risk-based approach centred around codes of practice that include safety obligations for larger or riskier platforms, with a focus on the most severe forms of online risks, such as harms to children or online terrorist activity. However, the consultation document only briefly mentions access to platform data for third parties (in the context of the EU's DSA), but does not outline specific plans for a data access regime under the proposed legislation.

Canada

The Canadian Government has also been developing legislation to regulate online platforms, and has conducted an open consultation and convened a series of roundtables, a citizens' assembly and an expert advisory group since 2021.[xxiv] While Canada is expected to also take a broadly similar risk-based approach to the EU or UK, as a draft bill is yet to be published, it remains unclear whether it will focus predominantly on larger platforms, or whether it will include strong provisions for access to data for independent researchers.

[xx] eSafety Australia, Learn about the Online Safety Act

[xxi] eSafety Australia, Basic Online Safety Expectations

[xxii] eSafety Australia, Responses to transparency notices (December 2022)

[xxiii] New Zealand Government, Discussion Document: Safer Online Services and Media Platforms (June 2023)

[xxiv] Canadian Heritage, The Government's commitment to address online safety (January 2023)

The Emerging Policy & Regulatory Context

Outside of this project, we have also encountered similar challenges with conducting research on a range of other emerging platforms across the constantly evolving online ecosystem.[xxv] As the platform scoping exercise we conducted in the first phase of this project demonstrated, there are a wide range of other smaller and medium-sized online platforms that are also being adopted by extremists and other harmful actors.[xxvi] We suspect this trend is only likely to increase if and when larger platforms become less hospitable for these kinds of online communities and activities, especially as regulation like the EU's DSA comes into force.[xxvii] Regardless of whether the prospects for data access improve, the online ecosystem will inevitably continue to evolve at pace, increasing the number, diversity and complexity of online platforms that researchers may wish to access, and can pose significant risks or harms.

Overall, in our view the current prospects for researcher access to data from platforms across the evolving online ecosystem are mixed. Incoming regulation in key jurisdictions offers considerable promise in terms of increasing platform transparency and broadening external data access for the largest platforms. It also remains to be seen whether regulations in key jurisdictions like the EU will have spill-over benefits and whether platforms will decide to extend similar access in contexts where they are not legally required to. However, examining the data access provisions and userbase thresholds in specific pieces of existing or proposed regulation in the key jurisdictions above suggests that many of the challenges raised in this project related to smaller and medium-sized platforms will not be fully addressed. Key regulation such as the DSA relies on 'monthly active users' as a key metric to determine whether platforms meet the threshold for the strongest set of regulatory requirements; this means that many smaller platforms, which nonetheless pose significant online risks. would not be required to address existing barriers to data access.

Smaller and medium-sized platforms are not currently incentivised to voluntarily open up access to data. This status quo will likely result in a continuation of the disempowerment researchers face, with few guarantees that this situation will not further deteriorate. Some companies may attempt to restrict data access via technological, legal or financial means in order to avoid further regulatory or independent scrutiny into the risks and harmful content or behaviours present on their platforms, at least in the short-term and/or in jurisdictions without regulation requiring data access. Without strong regulatory data access requirements, platforms may have an incentive to avoid the scrutiny that independent research can provide, or attempt to profit from the provision of data access. For example, as noted above, X has restricted free access to their API, and it has been reported that Meta has plans to shut down existing access to data via the Meta-owned CrowdTangle.[xxviii]

However, some changes to the regulatory environment for online platforms may still impact smaller- and medium-sized online platforms, albeit to a lesser extent than for larger platforms. Depending on the jurisdiction, these platforms may still be required to provide more comprehensive and consistent TOS, additional transparency and/or better access to data for regulators if not third-party researchers. They may also need to have systems in place to quickly and effectively address reports of illegal content or activity on their platforms. If medium-sized platforms like Telegram or Discord continue to grow, then they may also face additional regulatory obligations that require greater investments in content moderation, transparency and data access. In the case of Telegram, its considerable growth in recent years could soon make it subject to the same requirements as other more established large platforms, compliance with which would require significant changes to how it currently operates.

[xxv] See for example: Hammer, Dominik, Gerster, Lea and Schwieter, Christian, 'Inside the Digital Labyrinth: Right-Wing Extremist Strategies of Decentralisation on the Internet & Possible Countermeasures' (February 2023)

[xxvi] See: Annex - Ethical, Legal & Security Risks (July 2022)

[xxvii] European Parliament, 'Digital Services Act'

[xxviii] Tech Policy Press, 'Twitter API Changes Set to Disrupt Public Interest Research' (February 2023);
The Verge, 'Meta reportedly plans to shut down CrowdTangle, its tool that tracks popular social media posts' (June 2023)

Smaller alternative platforms, like Odysee, that have positioned themselves as ‘free speech’ alternatives to larger, more mainstream tech companies, typically have existing userbases that have chosen to adopt these platforms precisely because of their lax approaches or opposition to content moderation. If they seek to placate these types of users or communities by refusing to comply with regulatory or legal requirements, they may no longer be able to operate in certain jurisdictions and/or they may face legal action or financial penalties from regulators looking to reign in the levels of harmful content or behaviours they facilitate. Alternatively; if they want to grow and move towards long-term profitability, they are likely to face higher regulatory requirements and scrutiny, as well as need additional technical and human resources to ensure their platform can handle increased activity. Separately LBRY, the company that provides the underlying technology on which Odysee is built as well as the cryptocurrency (LBRY Credit or LBC) that allowed users to monetise their content, announced in July 2023 that it would be winding down following an US Security and Exchange Commission (SEC) ruling and fine. While the ruling focuses on LBRY Credits as an unregistered security, it nonetheless demonstrates that so-called ‘alt-tech’ platforms are facing increasing levels of regulatory scrutiny.[xxix]

[xxix] The Guardian, ‘Extremist-friendly tech company closes after legal fine’ (16 July 2023)

Recommendations

Based on the findings and implications for researchers identified during this project, our two roundtable discussions, and the policy context outlined above, this section outlines a series of overarching recommendations for online platforms, policy-makers and regulators, and civil society and academic researchers. For policy-makers, we have provided different recommendations according to the varying policy contexts across jurisdictions outlined above. We hope these recommendations will help to address the range of technological, ethical and legal, and fragmentation barriers we have encountered during our research.

We acknowledge that some of our recommendations will create additional work for platforms, especially smaller platforms with fewer resources or technical capabilities. However, at present, platforms provide effective functionalities for enabling the communications of harmful actors while only offering restricted or poor functionalities for public interest research into these spaces. We argue this imbalance must be addressed.

Platforms

Outside of the context of incoming or potential future legislation and regulation to mandate greater data access from platforms, below we provide a series of recommendations for steps platforms could take to become more open to independent researchers. We would argue that increasing data access can be a positive step that allows platforms to benefit from any research produced. For example, research identifying illegal activity, or research in contexts or languages where the platform lacks internal expertise, would help platforms to mitigate risks to user safety and comply with any incoming regulatory obligations.

Technological Barriers

Research into online platforms faces substantial challenges due to the speed and scale at which content can be created. A technological inability to access relevant data – whether the result of deliberate barriers, complicated platform structures or poorly documented technologies – substantially impacts public interest research, particularly for organisations with limited access to technical capabilities. The proliferation of novel technologies like blockchain will only increase this complexity and further limit the availability of technical tools and expertise.

Platforms should therefore provide permissioned data access (e.g. via APIs) to third-party researchers conducting public interest research at a reasonable cost. These should be accompanied by clear, consistent and accessible documentation that includes guidance on the types of data that can be collected and sufficient limits on the collection of sensitive personal data to protect user privacy. Where possible, platforms should also streamline the range of tools they directly provide to access data in order to reduce the burden on researchers needing to use multiple tools with overlapping capabilities.

Fragmentation Barriers

Platforms should provide data access in a systematic way that enables researchers to reliably access accurate data from across public areas of the platform, rather than having to manually explore the platform for spaces or communities of interest. This would allow for greater scale but also more targeted research as researchers would be able to identify relevant communities (and therefore data) more easily. Tools for systematic search, whether platform-native or third-party, should demonstrate reliability, coverage and accuracy, which could be supported and assessed, for example, via third-party reviews conducted by researchers, regulators or an intermediary body responsible for managing data access. The recommendation for addressing technological barriers set out above would also help to address the fragmentation barriers we have identified.

Legal Barriers

Platforms should provide clear Terms of Service for both users and researchers that outline what data is collected, how it is used by the platform, and what is made available to external researchers. These should then be enforced consistently. Greater clarity in TOS would also set clear expectations regarding the use of data collected by third-parties; this would provide more certainty for researchers as well as help users to understand how their data may be used. Increasingly this may be a requirement for platforms under incoming regulation (if not already under existing data protection regulation), even for smaller platforms that enable user-to-user communications and the sharing of content online.

If platforms are unable to provide researchers with access to data in a systematic and structured way, then at a minimum their TOS should also allow for privacy-respecting public interest research. TOS should not prohibit third-party methods (e.g. scraping of publicly accessible data) for non-commercial, public interest research purposes where researchers take appropriate steps to protect users' privacy. TOS should not be used to effectively insulate a platform from public scrutiny.

Ethical Barriers

One of the key challenges raised by our research is how to protect users' rights to privacy while conducting public interest research on harmful online communities and behaviours. Unfortunately, we lack a common definition of private spaces online, making it difficult for researchers to know precisely which spaces to treat as private versus public. At the same time, the lack of a collective understanding of what constitutes a private space online allows platforms to limit transparency and access to spaces that are easily accessible and arguably public.

As we have previously argued, factors such as size, purpose, accessibility and the nature of relationships between users of an channel or a community should be taken into account when making assessments about public or private spaces online.[xxx] There is also a danger that illegal or harmful online activity openly conducted in nominally private online spaces creates uncertainty and could undermine the case for stronger privacy safeguards (e.g. via encryption) for genuinely private online spaces and means of communication.

Platforms should determine a reasonable limit for the number of members that can participate in private groups and channels, and declare online spaces with audiences over a certain threshold as public. This should also help to provide greater clarity to both platform users and third-party researchers around what is acceptable in these spaces, as well as which types of data may be accessible to third parties with appropriate privacy and data protection safeguards. Users should benefit from a clearer understanding of the nature and privacy expectations of the online spaces in which they communicate. Content with no reasonable privacy expectations (e.g. content posted on public pages) should be made available via vetted API access, including relevant metrics on reach, impressions and engagement.

Providing Support & Creating Incentives for Platforms to Provide Greater Access to Data

Despite the recommendations above, we recognise that currently there may be insufficient incentives, expertise and/or resources for many smaller or even medium-sized platforms to take these steps. To address this, there are several possible approaches that could be employed to better support or incentivise smaller or medium-sized platforms

[xxx] ISD, Researching the Evolving Online Ecosystem: Telegram, Discord & Odysee (April 2023);
ISD, Extracts From ISD's Submitted Response to the UK Government Online Harms White Paper (July 2019)

that may not be subject to regulatory data access requirements to voluntarily collaborate more closely with external researchers to address technological, fragmentation and legal and ethical barriers.

Existing multistakeholder initiatives could play an important role in providing support to platforms with limited internal experience or expertise. In the countering terrorism and violent extremism space, initiatives such as Tech Against Terrorism, who have experience working with smaller platforms, the Christchurch Call, or the cross-industry Global Internet Forum to Counter-Terrorism (GIFCT) would be well placed to share best practices and offer expertise or resources.[xxxi] There are also comparable initiatives working to address other forms of online harms, such as the WeProtect Global Alliance, which brings together governments, the private sector, civil society and intergovernmental organisations to develop policies and solutions to protect children from sexual exploitation and abuse (CSEA) online.[xxxii] Alternatively, there are an increasing number of coalitions or associations made up of current or former tech industry staff, such as the Digital Trust & Safety Partnership, the Trust & Safety Professional Association, or the Integrity Institute, that could provide more informal expertise and advice as a starting point without platforms having to formally join multistakeholder initiatives or make firmer commitments.[xxxiii] Finally, emerging intermediary bodies set up to establish data access mechanisms for larger platforms would also be well placed to contribute (discussed further below).

For platforms that are not receptive to outreach, engagement and offers of support, other approaches could be employed to generate pressure to be more transparent and open up options for external data access for researchers. **For more established medium-sized platforms or fast-growing smaller platforms, this could include engaging with investors and advertisers to create financial incentives for platforms to provide greater access to data**, as well as further invest in content moderation to address harmful activity on their services. Access to data can also be of particular importance to advertisers wishing to understand where and alongside which content their brands are appearing on a platform. As a last resort for platforms that openly position themselves as venues for harmful online communities and activity displaced from elsewhere across the online ecosystem, pressuring app stores, payment services or hosting and security providers could also be an option. However, there are considerable risks with this type of approach, which may not be proportionate beyond exceptional cases, given the potential to set dangerous precedents for freedom of expression and the open internet.

Policy-makers

Where possible, policy-makers should provide regulators with flexible powers to require access to public data that is accurate, reliable and sufficient for public interest research from any platform that presents high levels of risk in future regulation. This would include the ability to require access to data from the wide range of smaller- and medium-sized online platforms that can still pose significant risks, rather than an exclusive focus on the largest platforms. This is vital to better understand how illegal or harmful online activity is increasingly conducted or coordinated on smaller- and medium-sized platforms that may present considerable risks despite their size. As far as possible, policy-makers should also work across jurisdictions to ensure consistency in data access requirements for platforms (which would help avoid over-burdening smaller businesses) and help develop best practices for both platforms and researchers over time.

Legal Barriers

At a minimum, policy-makers should introduce legal protections for researchers conducting public interest, privacy-respecting online research to protect them from unreasonable legal threats from platforms. Where researchers use proportionate methods and approaches and make sufficient efforts to protect user privacy, they

[xxxi] Tech Against Terrorism; Christchurch Call; Global Internet Forum to Counter-Terrorism
[xxxii] We Protect Global Alliance
[xxxiii] Digital Trust & Safety Partnership; Trust & Safety Professional Association; Integrity Institute

should not face the potential risk of legal action from platforms when researching important societal questions around the extent and impact of illegal or harmful activity in public spaces online. This is particularly relevant where platform TOS are incomplete, unclear or ambiguous. Platforms should not be able to completely evade scrutiny by placing a blanket ban on data access in their TOS, including those that actively position themselves as venues for online hate and extremism, or have insufficient systems in place to address illegal online activity. In-lieu of governments introducing legal protections for researchers, responsible platforms should establish voluntary exemptions in their TOS to permit research via methods like crowdsourced data collection, where researchers have secured the informed consent of users.

Ethical Barriers

Policy-makers could also consider introducing requirements for platforms to clarify which areas are truly public or private, and set reasonable thresholds for the number of users that can participate in private online spaces if platforms are unwilling or disincentivised to do so voluntarily. Rather than introducing blanket regulatory thresholds in this area, regulation could require companies to set clear and reasonable limits based on the nature of their platforms and risk assessments that consider a platforms' specific features or functionalities (e.g. encryption), risks and potential vulnerabilities. Regulation could also encourage companies to make clear to users which aspects of their platform are more public or more private, as well as the consequences of this for user privacy and researcher or third-party data access. An independent regulator could then assess, based on the risk assessment, whether the limit set by the platform is appropriate and sufficiently mitigates any risks or harms identified.

Jurisdictions with Existing Legislation and Regulation

In contexts where regulation is already in place that contains a fixed user number threshold for in-scope platforms or services to be subject to data access requirements, such as the EU's DSA, smaller platforms will not be required to formally provide researchers with access to data. However, given that mechanisms for data access will need to be established for larger platforms, there may be opportunities to leverage the same mechanisms and infrastructure to encourage smaller platforms to provide greater or more formalised means of access. An intermediary body, if properly resourced, could assist by providing access to resources and expertise, identifying and sharing best practices, or safeguarding or validating the data provided.

There may also be added incentives for medium-sized platforms approaching the VLOP user number threshold to put such mechanisms in place to ensure compliance once they reach it. Voluntary co-regulatory initiatives such as the EU's Code of Practice on Disinformation that also include data access commitments should be used as a starting point for platforms that may not be subject to stricter regulatory requirements. As such, the European Commission should continue to engage with non-signatory platforms to encourage them to participate and make commitments to greater transparency and data access.

In contexts such as Australia, where regulation is already in place but does not include specific data access requirements, existing legislation would need to be amended to add such provisions which may not be likely in the immediate future. However, given the eSafety Commissioner's responsibility "to support, encourage, conduct and evaluate research about online safety for Australians", greater data access for researchers could be recommended as a form of risk mitigation. Improved data access would assist both platforms and the regulator to identify existing and emerging online risks across a wider range of platforms, explore other potential mitigations, and assess mitigations that have already been put in place to establish cross-industry best practices.

Jurisdictions with Incoming Legislation and Regulation

In other contexts where legislation is planned but is still being finalised or has yet to be formally introduced, there is potentially scope to include smaller platforms within any proposed data access provisions. In jurisdictions such as the UK (where the Online Safety Bill is nearing its final passage through Parliament), New Zealand (where initial proposals have been released for consultation), or Canada (where a draft bill is expected later in 2023), there may be opportunities to ensure that smaller but high-risk platforms are also required to provide greater access to data for public interest research.

As outlined in the section above, the UK OSB considers both a platforms' size as well as its functionalities or levels of risk (as assessed by the regulator Ofcom). The New Zealand proposal also follows a similar risk-based approach without a strict user-number threshold, while we are yet to see the approach that the Canadian government will take. This type of more flexible approach opens up the possibility to include stricter obligations for smaller or medium-sized platforms that present higher levels of risk, in terms of both mitigating those risks, but also providing greater transparency and access to data to monitor them. However, as outlined in the previous section, the UK OSB would need to be further amended to strengthen the existing provisions around data access to ensure the regulator would be able to require rather than just recommend greater data access for small but high-risk platforms. With proposals at an earlier stage in Canada and New Zealand, there may be greater scope to include more flexible data access provisions at an earlier stage, or at least legal protections for privacy-protecting public interest research into online safety that would support the work of regulators.

Jurisdictions Still Considering Potential Legislation or Regulation

In other jurisdictions where legislation or regulation is not imminent, or there is a lack of consensus over whether it is even required or appropriate, there are fewer options to address data access challenges regardless of the size of the platform. However, while regulation may not be imminent in contexts like the US, there are various existing proposals such as DSOSA, KOSA or PATA that do offer some promise in terms of their provisions on data access for independent researchers in both academia and civil society. If agreement cannot be secured for legislation that combines online safety provisions with greater data access, the proposals like PATA should be supported to at least introduce baseline transparency standards and start to address data access related challenges. Even if proposals such as PATA do primarily focus on larger platforms, the additional protections for researchers via the creation of a 'safe harbour' to prevent platforms from taking legal action against researchers are a welcome development.

Beyond legislative and regulatory options, there may also be opportunities to make more limited progress through non-regulatory approaches, including those outlined in the platform recommendations section above. For example, while many have expressed scepticism over the effectiveness of voluntary initiatives, recent efforts by the White House to secure voluntary commitments from the private sector on mitigating the risks posed by AI technologies do include promises from industry to share information and collaborate with academia and civil society.[xxxiv] There may be opportunities to encourage online platforms to extend the data access they will be required to provide in jurisdictions where regulation is in place (such as the EU) to those where it has yet to be introduced, given that the necessary infrastructure and processes will have to be developed in order to comply.

[xxxiv] White House, FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI (July 2023); Politico, Digital Bridge (July 2023)

Researchers, Research Organisations and Funders of Online Research

Drawing from our roundtable discussions to further develop our previous recommendations, below we identify a series of avenues for researchers, research institutions and funders of online research to create greater economies of scale across the field. Regardless of progress made by either platforms or policy-makers in reducing barriers to data access, the following recommendations can be adopted to help progress towards a more equitable, diverse and global online research community.

Technological Barriers

Academic and civil society researchers should share effective data collection approaches or tools, as well as any lessons learnt, for accessing the growing range of platforms across the evolving online ecosystem. On platforms that require more technical expertise to access data, considerable time can be spent exploring and testing potential options. Wherever possible, they should pursue economies of scale and reduce disincentives for sharing methods, tools and approaches. The research community (and its funders) should seek to avoid duplicative or proprietary approaches, for example by adopting Open Science approaches to accessing, storing and sharing data, or creating virtual sandbox environments for access to sensitive data. This could be coordinated by cross-sectoral initiatives, such as the Coalition for Independent Tech Research, existing funding and collaboration mechanisms such as Horizon Europe, or via nascent independent data access bodies, such as those proposed by EDMO or the Carnegie Endowment for International Peace (CEIP).[xxxv]

Ethical Barriers

The research community should work to formalise ethical approaches to researching public, semi-private and private online spaces, in line with the potential severity of the risks such spaces could pose. Such approaches must balance rights to privacy with the rights of those that may be negatively impacted if these spaces, especially the largest and least moderated, remain out of the reach of researchers. These efforts should build on the existing and well-established field of internet research ethics. Coordination could occur through existing initiatives, such as the Association of Internet Researchers (AoIR), or the potential future independent data access bodies mentioned above (e.g. EDMO).[xxxvi] In an academic context, institutions should ensure that their ethical research boards (or equivalents) have sufficient expertise in and understanding of online research to be able to effectively assess and evaluate ethical risks and mitigations.

Legal Barriers

Academic and civil society researchers should also consider opportunities to share or pool expertise on the legal implications of platform data access. The legal review of a platform's TOS and other related conditions or policies can be time consuming, resource intensive and often jurisdictionally specific. Given that many researchers may not have access to outside legal support (spanning data protection and contract law), greater coordination and sharing of resources and expertise could help to reduce legal risks and barriers to entry for online research, thereby increasing equity among researchers. Again, as with the technological and ethical recommendations above, this type of cross-sectoral cooperation should be facilitated by existing initiatives where possible.

[xxxv] Coalition for Independent Technology Research; European Commission, Open Science; European Commission, Horizon; European Digital Media Observatory, May 2022, op cit.; Wanless, Alicia and Shapiro, Jacob N., 'A CERN Model for Studying the Information Environment', Carnegie Endowment for International Peace, November 2022.

[xxxvi] AoIR, 'Ethics'

ISD | Powering solutions to extremism and polarisation

Amman | Berlin | London | Paris | Washington DC

www.isdglobal.org