Summary of EU-U.S. Workshops on Researcher Access to Online Platform Data and the Role of Online Platform Data for Research on Technology-Facilitated Gender-Based Violence

Under Working Group 5 (Data Governance and Technology Platforms) of the EU-U.S. Trade and Technology Council (TTC), the European Union and the United States cooperate to advance the governance of online platforms. The basis for this cooperation is the shared view that online platforms should exercise greater responsibility in ensuring that their services contribute to an online environment that protects, empowers, and respects its users.

On 25 March 2022, representatives from the European Union and the United States hosted a workshop with participants from academia and civil society to engage in expert-led discussions on data access for researchers. In May 2023, on the occasion of the fourth TTC Ministerial in Lulea, Sweden, the European Union and United States proposed high-level principles on the protection and empowerment of children and youth, and facilitation of data access from online platforms for independent researchers.

On the occasion of the sixth TTC Ministerial in Leuven, Belgium, the European Union and United States delivered a status report on mechanisms for researcher access to online platform data and joint principles on combating gender-based violence in the digital environment. A set of workshops was convened with EU and U.S. experts on data access for researchers and technology-facilitated gender-based violence (TFGBV) on 4 April 2024. Expert participants raised key goals and discussed emerging barriers and possible actions to expand access to public online platform data, and how to use this data to better understand TFGBV.

Opportunities and challenges with public data access mechanisms

In a deep discussion of the state of data access for online platform research, the experts highlighted criteria that data access mechanisms should meet to be useful. This includes ensuring that data is up-to-date, accurate, structured, documented, and available in real time. Researchers explained that public data access mechanisms are helpful for exploratory research and identifying the 'unknown unknowns'.

Participants described ways to make the most of public data access mechanisms. They stressed that if these mechanisms are well designed or used in coordination with open-source tools for analysis and labelling they can be more easily used by non-data scientists and civil society. Additionally, researchers explained that public data can be combined with user data donations (collected via browser plugins and data access and portability provisions under data protection laws such as the General Data Protection Regulation (GDPR)) or qualitative data (surveys, interviews) to answer a broader range of research questions.

Participants noted a few challenges with public data access mechanisms. First, each platform structures their data differently, making it difficult to compare findings or study cross-platform systemic risks – researchers noted a standard schema would be helpful. Additionally, platforms vary in what metadata (e.g. prediction of user age, consent provided

for an image, or generative AI content labels) they make available, much of which is critical for certain research questions. Researchers emphasized that the ability to study recommender systems is 'the ball game' for many research topics. To study recommender systems, researchers often need to combine bot accounts with scraping to view personalized results. Additionally, studying recommender systems often requires metrics describing content exposure which platforms do not always include in their public data access mechanisms. Lastly, researchers mentioned that the data provided through platform Application Programming Interfaces (APIs) is not always complete and that researchers need publicly accessible platform data in order to validate and fill gaps in data provided by platform APIs.

The workshop also included a discussion of causality. Some participants suggested that it is difficult to hold platforms accountable without causal findings and suggested that independent researchers need to be able to collaborate with platforms on A/B testing. A few participants described how publicly available data can be used to illustrate causality when paired with reporting about a platform's policy or feature change (before and after studies). Another participant added, however, that requiring causality is commonly used by platforms as a defense against making changes based on other types of empirical studies. In the context of TFGBV, as with other complex human behaviors that interact with various factors in the online environment, demonstrating causality is extremely difficult. However, access to greater platform data could strengthen the robustness of studies to assess correlation and strength of association between, for instance, the proliferation of violent pornographic content and the growth of TFGBV violence online, such as the non-consensual distribution of intimate images.

Applications for access

Further, researchers highlighted that platforms' standards for evaluating researchers' requests for access are often unclear. Researchers who have applied for access report denials and delays, often without clear reasons for their pending or denied status. This has led to serious concerns about the gatekeeping of access by online platforms, and in response, multiple participants expressed the need for trusted intermediaries to mediate access. Trusted intermediaries can provide an independent review of research proposals assessing ethical risks and societal value of research irrespective of risks to a platform's reputation. Participants described existing examples of these intermediaries such as the Social Media Archive (SOMAR) hosted at the University of Michigan in the United States or iCANDID at KU Leuven in Belgium.

Additionally, the applications often require researchers to describe their data protection and ethical considerations. These considerations can be highly project-specific and new to researchers. Participants suggested that universities could provide support and training to researchers applying for access and that researchers should share learnings within academic communities. The researchers expressed a desire for online platforms to work more collaboratively with universities and research centers to better facilitate, rather than prevent, data access.

Needed guidance and support for the research community

Participants highlighted that many researchers are not aware of the newer data access mechanisms created to comply with the EU's Digital Services Act. They suggested that academic conferences provide a venue to discuss the new mechanisms and share insights regarding how to get the most benefit from the mechanisms. Additionally, funders can include descriptions of public data access mechanisms in their requests for proposals (RFP). Grant solicitations can also promote transatlantic and interdisciplinary collaboration.

Several participants emphasized the ethical risks, especially privacy, of expanded access to personal data held by platforms. They advocated for strong protections for platform users, such as guarantees that research data will not be used for law enforcement, anonymization of personal information and user quotes, and strong data security practices by researchers. These ethical considerations are particularly important in sensitive areas of research, such as TFGVB. Participants emphasized that clear and consistent ethical and data protection frameworks will be helpful.

Participants also highlighted that the research community needs clarity on how datasets of platform data can be preserved and made available to other researchers to increase the transparency and reproducibility of research. Here again, trusted intermediaries were mentioned as a viable solution.

TFGBV context-specific needs

In the second workshop, TFGBV researchers outlined several examples where greater access to platform data would be particularly useful, including to analyze social media posts to better understand help-seeking behaviors among victims and survivors, and to review posts from perpetrators. One participant referred to their forthcoming report detailing how recommender algorithms are amplifying, mainstreaming, and normalizing misogynistic content that would, by most accounts, be considered 'fringe' at scale on popular platforms like TikTok and YouTube disproportionately targeting boys and young men. Another participant echoed these remarks by discussing their own ongoing research documenting how recommender systems disproportionately recommend the most extreme and polarizing content– both on social media platforms and porn sites.

TFGBV researchers explained that public data access mechanisms from the largest online platforms, while critical, are altogether insufficient for studying risk and protective factors for TFGBV. Researchers noted that severe and pervasive forms of TFGBV can also occur on smaller platforms, such as dating apps and messaging applications. Access to data from platforms of all sizes is also critical for the examination of how TFGBV is disseminated across platforms. Some participants raised concerns that the scope of research on TFGBV is not limited to simply studying questions based on the generally greater availability of data from large platforms.

To augment data from platforms, TFGBV researchers have leveraged information from data donations made possible through regulation such as the GDPR's right to data access and portability. Additionally, TFGBV researchers mentioned the importance of combining platform data with qualitative data (e.g. surveys and interviews of victims and survivors, perpetrators, and online bystanders) and even public government datasets. An inherent challenge in studying TFGBV is that the chilling effect on victims and survivors can drive them off platforms, influence them to de-activate their accounts, and therefore reduce their digital footprint. As a result, victims and survivors who go offline are not present in platform data. There are also challenges with tracking repeat offenders who operate across many platforms using unique accounts, and who enjoy the anonymity afforded by online platforms.

Participants discussed ethical challenges of storing, analyzing, interpreting, and publishing data from victims and survivors – especially when gathering sensitive information, such as reports or disclosures of sexual violence. Participants underscored the importance of strong legal and cybersecurity protections for this data and emphasized that TFGBV research – in keeping with best practice in gender-based violence research in general – must be grounded in victim- and survivor-centered, trauma-informed approaches. Participants suggested engaging victims and survivors in research design and dissemination. Experts also emphasized that academic and civil society researchers are at risk of being the targets of online harassment and abuse, due to the nature of their work. Researchers noted adversarial interactions with the perpetrators they are studying online, and reported that they often face threats and harms without institutional support. Participants emphasized the need for researchers in this field to be protected when conducting their work.

Final remarks

White House, U.S. National Telecommunications and Information Administration (NTIA), and European Commission officials thanked the experts for their critical research and advocacy efforts, and further committed to supporting research collaborations and action to facilitate data access, including to further develop the evidence base on TFGBV on both sides of the Atlantic.

Expert Participants:

Theo Araujo (University of Amsterdam)
Lucia Bainotti (University of Amsterdam)
Renée DiResta (Stanford Internet Observatory)
Asia Eaton (Florida International University)
Marc Faddoul (AI Forensics)
Casey Fiesler (University of Colorado Boulder)
Sarah Gilbert (Cornell University)
Debbie Ging (Dublin City University)
Olga Jurasz (Open University)
Margaret Levenstein (University of Michigan)
Philipp Lorenz-Spreen (Max Planck Institute for Human Development)
J. Nathan Matias (Cornell University)

Clare McGlynn (Durham University)
Solomon Messing (New York University)
Dino Pedreschi (University of Pisa)
Elissa Redmiles (Georgetown University)
Silvia Semenzin (University Complutense of Madrid)
Brandon Silverman (George Washington University)
Rebekah Tromble (George Washington University)
Christo Wilson (Northeastern University)
Bieke Zaman (KU Leuven)

Government Representatives
Eleonora Esposito (European Commission)
Silvia Merisio (European Commission)
Julian Ringhof (European Commission
Rita Wezenbeek (European Commission)
Cailin Crockett (White House Gender Policy Council and National Security Council)
Alex Engler (White House Office of Science and Technology Policy)
Anna Lenhart (White House Office of Science and Technology Policy)
Emma Llansó (U.S. National Telecommunications and Information Administration)