

Dr. Rebekah Tromble – written evidence (DAD0104)

Associate Professor, School of Media & Public Affairs, George Washington University
Associate Director, Institute for Data, Democracy, and Politics (IDDP), George Washington University
Visiting Researcher, The Alan Turing Institute (London)

1. What are the privacy concerns that may prevent data sharing? And what should the Committee be asking Facebook/ Google and Twitter representatives when they appear before the Committee about how to overcome these concerns?

From the tech companies' perspective, the overarching concern with data sharing is that personally identifiable information will be transferred to academic researchers in a way that one or more Member States deems to be in violation of GDPR. These companies have particular questions and concerns about what is permissible under GDPR.

GDPR explicitly seeks to facilitate scientific research and therefore exempts researchers from many of the restrictions on collecting and processing personal data. However, the companies argue that the definition of "research" under GDPR remains unclear.

These companies also suggest that (a) the safeguards that they must take in providing data to researchers and (b) the safeguards that the researchers themselves must take in receiving and processing such commercial data are unclear. For example, they contend that what constitutes user "consent" remains unclear. Or, in the absence of such consent, they question what measures must be taken to remove PII from data transferred to and/or analyzed by external researchers.

2. What is the ideal for data sharing best practice?

Data sharing best practices vary by the level of data sensitivity and privacy protections needed:

- a. For public data, academic researchers should undergo the no more scrutiny, and in turn receive no less data access, than third-party corporate entities. Currently, the platforms give privileged access to other companies. Facebook, for example, has set up review procedures for third-party companies to gain access to the Facebook and Instagram Graph APIs, but the rules for review make it effectively impossible for academics to gain the same access.
- b. The platforms should directly partner with researchers to undertake user surveys and experiments (i.e., research that permits individual-level, causal analyses) *with user informed consent and full ethical review by accredited review bodies*. Researchers should be permitted to publish their findings without restrictions.
- c. For sensitive PII for which direct user consent has not been or cannot be obtained, the platforms should develop so-called "clean rooms" (physical or virtual), wherein researchers can analyze data without needing to transfer said data. Such practices are common in public health and government data research, which should serve

as the model. Researchers should be permitted to publish their findings without restrictions.

Note, however, that in each instance independent review is necessary to ensure that the data provided by companies are full and accurate. Such review is essential for maintaining the integrity of research results.