



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/165426/>

Version: Accepted Version

---

**Article:**

Dommett, K. (2020) Researching for democracy? Data access and the study of online platforms. *Political Insight*, 11 (3). pp. 34-36. ISSN: 2041-9058

<https://doi.org/10.1177/2041905820958822>

---

Dommett K. Researching for Democracy? Data Access and the Study of Online Platforms. *Political Insight*. 2020;11(3):34-36. Copyright © 2020 The Author(s). DOI: <https://doi.org/10.1177/2041905820958822>

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

## Researching for Democracy? Data Access and the Study of Online Platforms

What impact does misinformation have on user behaviour? How is data used and sold in the online world? Do political adverts change our minds? Are foreign actors deliberately spreading disinformation to undermine society? These questions are critical to our understanding of the impact of the internet on democratic society. And yet, whilst seen to be important, academics' ability to provide answers is fundamentally hindered by the issue of data access.

As researchers it can be easy to take our ability to study our interests for granted. Whatever method we use, be it interviews, surveys, ethnography, time diaries or experimental research, its often relatively easy to generate data on our interests. When it comes to the Internet, however, its often far from simple for researchers to gain access, gather data or conduct analyses. Whilst by no means facing a unique challenge, those studying the online sphere have confronted ongoing difficulties relating to data availability. With online platforms – such as Facebook, Instagram, YouTube, Google, Snapchat, TikTok and Reddit – dominating the online world, access to information is often highly curtailed. Far from being open and transparent, information is often given at the grace of a private corporations, with access defined on terms of their choosing. As a result it's exceedingly difficult to determine what is happening online, meaning scholars can often provide only partial or limited answers to the questions above. As pressure grows around the need to protect democracy from problematic trends on the internet, there is an urgent need for further research about the online world. In this context, it's interesting to note growing interest from policy makers in the issue of researcher access, with specific moves from within the UK Parliament and European Commission to improve data access and online transparency.

Engaging with these debates, I draw on my recent experience of working as a Special Advisor for the House of Lords Committee on Democracy and Digital Technology to reflect on current debates around data access for researchers. Published in June 2020, the report *Digital Technology and the Resurrection of Trust* argues that 'transparency of online platforms is essential if democracy is to flourish' and makes explicit recommendations for greater researcher access. Reflecting on the experience of producing this report, I consider what has been recommended, and whether any change is likely to come about soon.

### Studying the online world

Since the early days of the internet, researchers in academia and beyond have played an important role in building our understanding of what is happening online. Originating from different disciplines, an array of methods and approaches have been used, many of which have been affected by issues of data access.

Perhaps the biggest challenges have been faced by those using computational methods. Scholars in this tradition often gather large data sets (often of social media data) to analyse practices online. Identifying patterns in online interactions, profiling types of user and producing 'sentiment' analyses that look at the expression of emotion online, computational approaches have provided a range of insights into the online world. However, in recent years access to computational data - secured largely through Application Programming Interfaces (APIs) – has been heavily limited or withdrawn. Sparked by the Cambridge Analytica scandal, these changes were justified in relation to concerns over user privacy, consent and inappropriate data use, leading many platforms to curtail the information available to scholars (or other outsider actors). Notably, Facebook dramatically restricted its data access, rolling back API access and substantially limiting scholars' ability to gather data about activity on its platform. Lynge Asbjørn Møller and Anja Bechmann (2019) have also traced trends across Twitter, YouTube and Instagram to show that 'the methods for data exchange provided by the social media platforms are subject to increasingly strict restrictions of data access, making it difficult - if not impossible - to extract substantial social media data for thorough investigations'.

For computational analysts, these changes mean that systematic research is now often untenable.

In spotlighting issues of data access, it is, however, not only computational analysts that have been affected. With limited transparency around how platforms' internal processes work, a culture of non-disclosure agreements and a tendency to refuse interviews and access for ethnographic studies, it's also challenging for researchers using interviews, observation, textual analysis or surveys to generate different kinds of insight. Whilst a few exceptions exist, many researchers have therefore been forced to adapt their methods, or to tailor their research questions to available data, limiting the insights on offer. Whilst these developments are of course frustrating for academics, they also raises more substantial questions about our understanding of what is happening online. With researchers no longer able to provide computational or qualitative scrutiny of the online world, it becomes harder to identify problematic practices and to hold those responsible to account.

This is not to say that no methods for analysis exist. In terms of computational analysis in particular, there have been some attempts to broker access. In addition to small research grants offered by WhatsApp and Instagram, Facebook have been working to facilitate academic research. Under the auspices of the Social Science One initiative launched by Gary King and Nate Persilly, Facebook have agreed to provide access for some scholars to some social media data. This programme has, however, been defined by a series of delays and legal complexities. It therefore took nearly two years for the first URL dataset to be made available (King and Persilly, 2020), and whilst additional datasets are due to be added (for example, on political advertising) they have a limited scope, curtailing the extent of academic scrutiny. Other avenues for inquiry have similar limitations. Some companies have chosen to make data about certain parts of their activity available. For example, in 2018 Facebook and Google launched online advertising archives to facilitate public scrutiny – an initiative that Snapchat has also adopted. These schemes provide some insight, but they have been widely criticised as lacking detail and functionality. Moreover, they place power in the hands of companies to decide what and how they disclose information, allowing them to direct the kind of research that can be done. Some scholars have taken a different approach to trying to generate insights. Research has been done by creating mock online environments to simulate engagement online, and surveys have been used to gather data on practices online. Whilst research is not impossible, and many scholars have taken inventive approaches to overcome data limitations, it is clear that the breadth and detail of available information is far from ideal.

### The politics of data access

Against this backdrop, it's notable that policy makers and civil society organisations have been voicing growing support for the idea of compelling online platforms to provide data access for researchers. Organisations such as Algorithm Watch, for example, have argued that 'Ensuring adequate research access should...be a paramount priority' (2020, p.3). Whilst the EU Commission is currently consulting on 'non-personal data access obligations' as part of its Digital Services Act. Elsewhere, the Stigler Centre in the US have called the Federal Trade Commission to moderate independent researchers' access to databases. These calls have drawn attention to the need for independent research in order to provide oversight and enable accountability. Yet they have also stressed the role policy makers can play in brokering access for researchers to online data. This step is seen to be particularly important because, as the Kofi Annan Foundation has argued, 'Even when the platforms have promised to make available data for independent academic research, those promises have often gone unfulfilled' (2020, p.38). As such, there is seen to be a need to compel companies to comply with data access requests.

The latest call for greater access to data was made in the newly published report from the House of Lords Democracy and Digital Technology Committee. Set up in September 2019,

and chaired by Lord David Putnam, the Committee conducted an expansive inquiry looking to restore trust in digital technology. A key part of its argument contended that 'For the public to trust individuals with power there must be transparency', leading the report to consider 'the information which platforms should share with researchers, what platforms should share about the algorithms that govern them, and how open companies should be about the decision-making process regarding what can stay on their platforms' (2020, p.13). In regards to data sharing in particular, the importance of action was stressed, with oral evidence from Ben Scott from Luminata, Paddy McGuinness, Alex Krasodomski-Jones, Professors Helen Margets, Cristian Vaccari and many others emphasising the need for further insight into the online world.

Within the final report, the Committee argued for 'independent researchers to have greater access to data from technology platforms' in order to allow them to 'verify their activities and effects' and promote public trust (House of Lords, 2020, p.60-1). Reflecting on previous calls from Bruns (2018) and Asbjørn Møller and Bechmann (2019), the Committee recommends that:

Ofcom should be given the power to compel companies to facilitate research on topics that are in the public interest. The ICO should, in consultation with Ofcom, prepare statutory guidance under Section 128 of the Data Protection Act 2018 on data sharing between researchers and the technology platforms. Once this guidance is completed, Ofcom should require platforms to:

- (a) Provide at least equivalent access for researchers to APIs as that provided to commercial partners;
- (b) Establish direct partnerships with researchers to undertake user surveys and experiments with user informed consent on matters of substantial public interest;
- (c) Develop, for sensitive personal information, physical or virtual 'clean rooms' where researchers can analyse data.

This recommendation seeks to secure greater access for researchers to platform data and to expand industry partnerships between academics and companies, yet it also aims to develop recent proposals for data analysis infrastructure in a safe space across and outside platforms (Asbjørn Møller and Bechmann, 2019, p.31). The idea of 'clean rooms' seeks to recognise the privacy and data protection concerns raised by companies by providing infrastructure for researcher access. This is an idea that has garnered significant interest in the EU, and which offers a fruitful path for data access (for more see Lineate, 2019).

Reflecting on my experience as special advisor to the Committee and the likelihood that these recommendations will bring about change, it is clear that there is growing pressure for increased greater data access. There is now a concerted effort to use policy change to secure academic access, recognising that digital companies are not inclined to share insights that are democratically vital. And yet, whilst the pressure is growing, it became clear through the course of the inquiry that it will not be easy to secure change. When speaking to social media companies there was little appetite to open up, and especially to offer more expansive data access. Whilst Parliamentary Committees and Governmental bodies are therefore voicing increased support for these policies, it is likely that any attempt to legislate for change will face considerable opposition from online companies. It may, therefore, be some time before academics secure greater data access, but it is reassuring to know that policy makers are committed to fighting for this idea.

## **Bibliography**

Asbjørn Møller, L. and Bechmann. A. (2019) *Research Data Exchange Solution*, available here: <https://www.disinfobservatory.org/download/26541>, accessed 13 May 2020.

Bruns, A. (2018) 'Facebook shuts the gate after the horse has bolted, and hurts real research in the process', *Internet Policy Review*, available here: <https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786>, accessed 29.6.2020.

House of Lords Democracy and Digital Technology Committee. (2020) *Digital Technology and the Resurrection of Trust*, London: House of Lords.

King, G. and Persily, N. (2020) 'Unprecedented Facebook URLs Dataset now Available for Academic Research through Social Science One', *Social Science One*, available here: <https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one>, accessed 29.6.2020.

Lineate. (2019) 'What data clean rooms are, and why they matter', available here: <https://www.lineate.com/technology-insights/what-data-clean-rooms-are-why-they-matter>, accessed 30.6.2020.

#### Notes:

Peter, there is a good diagram of how the Social Science One scheme works in this article; <https://gking.harvard.edu/files/gking/files/partnerships.pdf>

page 3

I'm not sure if you'd need permission to reproduce it, but could be useful for the article?