

Analysis

Researchers fear tech firms will sever their access to data

If social media information is locked up, scholars can't study big questions, writes David Matthews

Step on to an underground train full of silent commuters glued to their smartphones, and one thing is obvious: much of human life now occurs online.

Americans spend an average of 22.5 hours a week on the web, according to a study last year by the Center for the Digital Future, based at the University of Southern California.

In theory, this means that social scientists should be skipping through a data paradise, delving deeper than ever before into the workings of our parallel, online world using billions upon billions of likes, shares, comments and emojis.

But researchers are sounding the alarm that the opposite is happening. They fear that their freedom to access and study this global data explosion is being steadily narrowed by the social media companies and platforms that hold the information.

The restrictions means that academics – and by extension regulators, the public and politicians – have little idea what is really going on online, be it fake news, extremist propaganda or Russian disinformation.

“Research access is bad – and getting worse,” warned Philip Howard, director of the Oxford Internet Institute at the University of Oxford.

Particularly in the wake of last year's Cambridge Analytica scandal, during which it emerged that millions of Facebook users had unwittingly had their data harvested and passed on to a political consultancy that once worked for Donald Trump's presidential election campaign, some social media companies have been turning off the data taps.

“Most people don't appreciate [that] when they read in newspapers that Facebook is cracking down on bad apps, what it is actually doing is shutting down a lot of researchers' apps,” said Richard Rogers, professor of new media and digital culture at the University of Amsterdam, whose group builds tools that extract data from the web and social media.

The most common way for researchers to get their hands on social media data is through application programming interfaces (APIs), which are often designed to allow third parties to build apps on

top of another system's data. For example, a city subway system might use an API to make live travel information available to a developer who wants to incorporate it into a map app.

But API access to social media data has become more restrictive or has been cut off altogether, researchers warn. “Most of the major social media companies are making it increasingly difficult for academics and journalists to obtain comprehensive access to their data,” according to a report released at the end of August by the Social Observatory for Disinformation and Social Media Analysis (Soma), launched by the European Commission to fight fake news.

Researchers have suffered an “APIcalypse”, in the view of Axel Bruns, a professor of media and communication at the Queensland University of Technology and president of the Association of Internet Researchers.

Facebook is seen as the most restrictive platform, according to researchers who spoke with *Times Higher Education*. For example, last month it shut down an API that allowed researchers to collect information about Facebook pages. Several of the most crucial apps through which researchers harvested data are now defunct. “Now we can no longer see which posts have received a lot of reach or attention,” for example, Russian disinformation pages, said Amsterdam's Professor Rogers.

Asked what questions he cannot explore because of the unavailability of Facebook data, he reeled off some of the most vexing questions of modern politics: What is the reach of Russian disinformation efforts? What kinds of posts animate alt-right groups the most? What kind of extremist content

Offline Facebook, led by Mark Zuckerberg

(below left), is at the centre of vexing question such as Russian disinformation, but it's closing some doors to researchers

tends to go viral? To what extent are human moderators involved in taking down Facebook content?

All these topics are now very difficult to properly investigate, he said.

Facebook has introduced some new tools as it has shut down others – for example, in 2018 it launched one that tracks political adverts, for example – but, so far, researchers remain unconvinced. Social media companies' recently released tools are “for the most part insufficient and limited in scope and information richness”, according to the Soma report.

What's more, researchers said, Facebook-owned Instagram was all but inaccessible, its previous API having been closed down last year. “We think a lot of the interesting action is on Instagram, but there's no access at all,” said Oxford's Professor Howard.

Researchers said that data from WhatsApp (also owned by Facebook) were impossible to obtain because messages are end-to-end encrypted.

Some platforms, however, are far

more open. Reddit allows researchers to mine data from the website without restrictions; and since 2015 it has given academics a complete copy of the site. YouTube is “very generous” with access, said Professor Rogers, allowing researchers to investigate, for example, whether the site directs users to ever more extremist content through video recommendations.

Although Twitter, too, has tightened access in recent years, it has historically been more open, researchers say. But this may have skewed what research takes place. Nearly two-thirds of the papers analysed by Soma examined Twitter data, despite the platform being only the 12th most used social network globally. Just one in 10 studies looked at Facebook, the world's most popular, with more than 2 billion users.

And none of the major platforms discloses what content it has removed, making it hard to study “the politics of deletion”, explained Professor Rogers. “If you want to study the extent to which social

media companies are arbiters of speech...that's something that you cannot do with all of them,” he said.

As APIs have become more restrictive, other access models have emerged. Perhaps the best known is Social Science One, a partnership launched last year by academics, funders and Facebook that aims to facilitate researchers' access to data without breaching user privacy.

Gary King, director of the Institute for Quantitative Social Science at Harvard University and a co-founder of the project, said that academics had previously had to sign contracts with social media companies to dive deeply into their data, thereby compromising their academic freedom.

Social Science One aims to act as a third party that brokers data access without giving the platform a say over the resulting research findings. A series of research projects “on the effects of social media on democracy and elections” have been commissioned.

But the initiative has run into trouble. In August, funders threat-

ened to pull out unless Facebook handed over the data it had promised. Professor King said there would be “less than they thought” in some areas, as initially Facebook had unintentionally over-promised on what it would be possible to provide and had hit privacy and regulatory hurdles.

Check the terms and conditions

Researchers who spoke to *THE* expressed hope that the project would succeed, but they nevertheless considered it to be far from ideal.

“The issue for us was that the terms and conditions for playing with the data meant that it wasn't worth our time and effort to participate,” said Professor Howard. If researchers have already jumped through the peer review and ethics scrutiny of a traditional funding body such as the European Research Council, it was unclear why they needed subsequent approval from Social Science One to gain access to Facebook data, he argued.

Researchers involved in the project do not actually get to play with

Facebook's data on their own systems. Instead, they are allowed to work on Facebook's servers to extract information at an aggregate level. Professor Bruns calls the Social Science One model “corporate data philanthropy”.

The project gives “way too much power and discretion to Facebook, whose policies – eg, regarding privacy protection – and operations – eg, use of algorithms – are often and rightly the object of academic research that is often critical of these”, said Charles Melvin Ess, professor of media studies at the University of Oslo. “It's the equivalent of letting the fox guard the henhouse.”

Andrew Chadwick, professor of political communication at Loughborough University, said he deliberately did not apply to work with Social Science One because he feared it could hurt academic freedom.

“My view was that Social Science One was part of a strategy of crisis management public relations at Facebook,” he said. “My concern, which many other academics also had, was that Facebook would use Social Science One as a means of legitimising its arguments to policymakers in the US and the EU that it was dealing with its problems.”

If API access is disappearing, and projects such as Social Science One leave too much power in the hands of companies such as Facebook, what other solutions are there?

As *THE* reported, the European Union is considering forcing big technology platforms to allow researchers access to their data.

The ultimate, long-term goal is to store social media data outside companies' control, allowing researchers “stable access to social media data in controlled and safe spaces”, concludes the Soma report.

“It shouldn't be up to Facebook to decide how this is to be done,” said Anja Bechmann, one of the report's authors, who is director of the Datalab Center for Digital Social Research at Aarhus University in Denmark.

But until a better solution is found, many social scientists are worried. Professor Howard said he “cannot think of a parallel moment” in history when information about the workings of society has been so hidden from the public; for comparison, he pointed to the Second World War, when governments concealed national economic and health data.

“The future of the social sciences is dependent on what happens here,” he said.

david.matthews@timeshighereducation.com

