

Facebook's data lockdown is a disaster for academic researchers

April 11 2018, by Marco Bastos And Shawn T. Walker



Credit: AI-generated image ([disclaimer](#))

Facebook recently announced [dramatic data access restrictions](#) on its app and website. The company framed the lockdown as an attempt to protect user information, in response to the public outcry following the Cambridge Analytica scandal.

But the decision is in line with [growing restrictions](#) imposed on researchers studying Facebook and its photo-sharing app Instagram, which also began immediately restricting [access](#) to its data on April 4.

In fact, several limitations were put in place in February this year, before the Cambridge Analytica fiasco – in which data was allegedly [harvested from 50m Facebook profiles](#) – erupted publicly. Facebook's API, version 2.5, was scheduled to be retired this month, by – among other things – [preventing access](#) to the ID of users participating in public forums.

Social networks offer two main entry points for the collection of data: they work as interfaces for users and software interfaces designed for consumption by computer programs, known as Application Programming Interfaces ([APIs](#)).

While APIs are intended for programmers building apps that add to the growing ecosystem of services offered by social networks, researchers have also leveraged these interfaces to study social behaviour online.

Given the mammoth size of Facebook's userbase (2.13 billion at the [last count](#)), external scrutiny of the content on the social network is extremely important. In recent years, however, researchers have been fighting an [uphill battle](#) with the company to provide access to data. Now its latest decision has made it virtually impossible to carry out large-scale research on Facebook.

The changes make defunct software and libraries dedicated to academic research on Facebook, including [netvizz](#), [NodeXL](#), [SocialMediaLab](#), [fb_scrape_public](#) and [Rfacebook](#), all of which relied on Facebook's APIs to collect data.

Systematic research on Facebook content is now untenable, turning what

was already a worryingly opaque, siloed social network into a black box that is arguably even less accountable to lawmakers and the public – both of whom benefited from academics who monitored developments on the site.

[Deen Freelon](#), the developer of [fb_scrape_public](#) which analyses large, publicly available datasets on Facebook, told us via email that "the decision to restrict access to the Pages API could severely impair content-based Facebook research going forward, depending on how willing Facebook is to approve access. If it doesn't approve access for most research purposes, that could create incentives for researchers to scrape Facebook directly, which violates its terms of service." [Data scraping](#) or harvesting is a method by which a computer program extracts information from web pages.

[Bernhard Rieder](#), an associate professor at the University of Amsterdam who developed [netvizz](#) – a tool that extracts data from Facebook for research purposes – believes the move was a consequence of the level of unfettered access given to anyone until 2015 and that "there is a real possibility that these services will increasingly be inscrutable and unobservable".

Up until three years ago, Facebook allowed third-party apps to have access to data on the friends of app users. It was this function that was used by Aleksandr Kogan, a researcher at the University of Cambridge.

Kogan – through his Global Science Research startup, which was separate from his academic work – allegedly collected profile information from 270,000 Facebook users and tens of millions of their friends using a personality test app called "thisisyourdigitallife". It's alleged that Cambridge Analytica used that data in an attempt to target political campaigns including the 2016 US presidential election.

[Marc Smith](#), who led the Microsoft team that created [NodeXL](#), which analyses social network data, told us that there was an opportunity to rethink the social networks people choose to use in light of the data scandal.

Why APIs matter

APIs allow researchers to retrieve large-scale data and curate databases associated with meaningful events. Without them, [web interfaces have to be scraped](#) to access the data, which is labour intensive and drastically limits the amount of information that can be collected and processed.

Locking researchers out of the APIs constrains them to human-intensive means of data collection that cannot produce representative samples of real-world events, such as social movements, elections and disinformation campaigns.

Twitter operates three well documented, public [APIs](#) in addition to its premium and enterprise offerings. Twitter's relative accessibility leads it to being vastly overrepresented in [social media](#) research. But public and open APIs are an exception in the social media ecosystem. Facebook's [Public Feed API](#), for example, is restricted to a limited set of media publishers.

Data lockdown

Facebook's API lockdown will widen the gap between industry researchers hired by social networks and researchers working outside corporations. It's a divide [characterised](#) as the gap between "big data rich researchers", who have access to proprietary data and might be working only in the interests of the company they are affiliated with, and the "big data poor" or the broad universe of academic researchers.

Facebook's decision dramatically expands this pool of "big data poor". It limits research to projects sponsored by the network and potentially jeopardises research that is critical of Facebook.

Shortly after the decision to drastically limit API access, Facebook [vowed](#) to help researchers gain access to social media data of public interest, starting with elections. The announcement was met with a mix of [celebration](#) and [subdued support](#) from researchers.

Luca Rossi, who is [associate professor](#) at the IT University of Copenhagen, cautioned that the "data sharing model proposed by Facebook is deeply problematic and it will probably reinforce existing differences in terms of data access". The restriction is likely to continue the trend of researchers doing research they are able to as opposed to research they deem important.

Good news for research on the relationship between social media and society (especially politics) - Facebook and funding foundations to set up a new model for industry- academic partnerships including third party scrutiny and peer review
<https://t.co/87gZm5PyIO>

— Helen Margetts (@HelenMargetts) [April 9, 2018](#)

Fascinating how [#Facebook](#) has such a high barrier of access to its data for legitimate research purposes (a good thing)
<https://t.co/9ZcI4tyrS2> & yet allowed its system to be mined for data, weaponised & monetised by 3rd parties with just good coding skills & API access.

— Sanjana Hattotuwa (@sanjanah) [April 9, 2018](#)

I'm concerned that the right people are not at the table in terms

decision making power and centering the most vulnerable people for research agendas. I'd love to be wrong about this.

<https://t.co/QiKsSf3ZRB>

— Jill Dimond (@jpdimond) [April 9, 2018](#)

The impact on data science education is also considerable. If researchers are unable to access data from social networks, they will be unable to train students in data science, social science, computer sciences and digital humanities on methods of data collection and analysis that are rigorous, critical and ethical.

Facebook's decision to render the API useless for meaningful research is a regrettable departure for collaboration between the social network giant and academics and it's already having an impact.

The [Events API](#), which researchers relied on to retrieve information about public events such as demonstrations, no longer permits access to users or posts on the event wall.

Facebook's [Groups API](#) and [Pages API](#) were the endpoints researchers queried to study public discussions on Facebook, but the recent policy shift seals off those online conversations by restricting access to posts, comments or members participating in a public page or group.

The changes made to Instagram's API are even more radical, with Facebook deciding to deprecate the API – a technical term for killing data access altogether.

Nasty side effect

Facebook's decision to restrict researchers is ironic because academics have long discussed the problems that led to the Cambridge Analytica

scandal. Rieder wrote about the risks of [Facebook API's wide open data door](#) back in 2013.

He cautioned against how much data a third-party app could get from Facebook. Facebook, however, ignored those concerns until 2015, when management and policies regulating the sharing of Facebook data took a sharp turn and became increasingly more restrictive for [researchers](#).

Since then, Facebook has become increasingly more cautious about external scrutiny. In the wake of the Cambridge Analytica scandal, Facebook CEO Mark Zuckerberg told [Wired](#) that the feedback it received was that "having the data locked down is more important to people than having different kinds of experiences".

The public uproar clearly underscores how users' data was poorly handled, but a lockdown is hardly the solution to a problem rooted in the [weaponisation of social networks](#), where people use Facebook, Twitter and so on to spread disinformation.

The Cambridge Analytica scandal has created a worrying side effect: restricting access to [data](#) is likely to facilitate [further weaponisation](#), by turning Facebook into a de facto black box that is largely unaccountable to external oversight.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: Facebook's data lockdown is a disaster for academic researchers (2018, April 11) retrieved 26 June 2024 from <https://phys.org/news/2018-04-facebook-lockdown-disaster-academic.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.