Platform research access in Article 31 of the Digital Services Act

Sword without a shield?

Leerssen, P.

DOI
10.17176/20210907-214355-0
Publication date
2021
Document Version
Other version
License
CC BY-SA

Platform research access in Article 31 of the Digital Services Act

Paddy Leerssen 2021-09-07T10:21:19

The issue of research access is becoming ever more urgent in platform governance. Over the past year, dominant platforms such as Facebook have repeatedly interfered with independent research projects, prompting calls for reform. The matter went mainstream in October 2020, when, only weeks before the US elections, Facebook tried to shut down an independent audit of their political advertising by NYU. Last month, they tightened the screws even further by suspending the researchers' Facebook accounts, stripping them of access to the Ad Library API and Crowdtangle research tools. And closer to home, Facebook also retaliated against data collection by the Berlin-based NGO AlgorithmWatch, sending them "thinly veiled threats" of legal action on the grounds that independent data collection violated the platform's Terms of Service. Platforms are shaping up as gatekeepers not only of online content and commerce, but of research into these phenomena.

As self-regulation flounders, researchers are increasingly looking to government to secure platform research access. In particular, their sights are set on Article 31 of the proposed Digital Services Act (DSA), on "Data Access and Scrutiny". A highly ambitious plan, it is to my knowledge the first legislative framework for researcher access to platform data.

What does Article 31 DSA do, and how does it constrain gatekeeper power over public interest research, and how will it help the likes of AlgorithmWatch and NYU? There are some important limitations in the current draft, and it won't actually resolve the scraping disputes we've seen over the past year. Researchers will welcome Article DSA 31 as a tool to compel access to certain data, but they also need a shield to protect them against interference with their independent projects.

Article 31 DSA in short

In short, Article 31 DSA creates a procedure for the European Commission and national authorities ('Digital Service Coordinators') to compel confidential access to platform data.

Under this framework, regulators can order access for their own monitoring and enforcement purposes (Paragraph 1) or for use by third-party researchers (Paragraph 2). Access is limited to so-called "vetted researchers", subject to various conditions such as a university affiliation, independence from commercial interests, and compliance with confidentiality and security requirements (Paragraph 4). Another important limitation is that researchers may only use this data for purposes of research into "systemic risks" as defined in Article 26 DSA. Platforms may object to data access requests in cases where they do not have the data, or access would

pose “significant vulnerabilities” to security or “protection of confidential information, in particular trade secrets”. This regime applies only to Very Large Online Platforms (VLOPs) with more than 45 million average monthly active recipients in the EU (Article 25(1) DSA).

All this is covered in one rather brief provision. Many technical and procedural details are left for the Commission to sort out in delegated acts (Paragraph 4), including compliance with the GDPR, and the protection of platform security and trade secrets.

There is much to like here for researchers, who have been pushing for this kind of confidential access frameworks for a while now. Still, the current draft leaves many loopholes and uncertainties that could undermine its impact in practice. And it does little to address the contractual powers that platforms wield over researchers through their Terms of Service.

Research topics: ‘systemic risks’ only

Article 31 DSA only applies to research related to “systemic risks” per Article 26 DSA. Admittedly this category is broad and open-ended, including catch-all concepts such as the as fundamental rights to privacy, freedom of expression and information alongside more specific issues such as “dissemination of illegal content” and “intentional manipulation of the service”.

One wonders why the legislator did not opt for a more neutral, open-ended purpose such as scientific or public interest research. The present approach seems to treat research access solely as a means to enable better enforcement of the DSA. But scientific interest in platform data is by no means limited to these types of regulatory concerns. Thankfully, the concept of “systemic risks” is so broad that many researchers will still manage to fit the bill, but ideally such box-ticking exercises would not be necessary.

Research actors: academics only

Article 31 DSA only benefits “vetted researchers”, defined as follows in paragraph 4:

“In order to be vetted, researchers shall be affiliated with academic institutions, be independent from commercial interests, have proven records of expertise in the fields related to the risks investigated or related research methodologies, and shall commit and be in a capacity to preserve the specific data security and confidentiality requirements corresponding to each request.”

There is much to unpack here, but the most important point is that access is limited to university-affiliated academics. This approach has the downside of ruling out usage by other valuable watchdogs in platform governance, such as journalists and NGOs (unless they partner with academics, of course). Critics including AlgorithmWatch have already called for the university affiliation rule to be dropped. Mathias Vermeulen proposes an amendment from academic to *scientific* researchers. Comparative research by Jef Ausloos, Pim ten Thije and I has also

shown that data access frameworks in other industries such as public health have made do with actor-neutral approaches, focused on scientific research *purposes* rather than actors.

But an academics-only approach also has an important upside: academics are relatively straightforward to accredit via the university system, whereas journalists and NGOs are more amorphous categories more open to abuse. The Commission will likely have less trouble deciding who qualifies as an academic, than as a journalist or NGO. Consider also how attractive Article 31 DSA might be for commercial parties, such as IP lobbyists collecting ammunition in their war against platforms, or professional advertising or financial analysts. Without proper safeguards, there is a real risk of such commercial usage crowding out public interest applications, as has already happened in other areas of transparency regulation such as US public records laws. The DSA already requires researchers to be independent from commercial interests, but this test is relatively difficult to enforce in practice, especially as regards NGOs. Limiting access to universities throws up an additional barrier against co-optation by private interests.

In my view, the correct answer here depends on other aspects of Article 31 DSA that are still unclear. As I'll discuss below, important procedural aspects still need to be decided on, such as whether Article 31 DSA will produce automated and scalable solutions or instead will take a slower, smaller-scale approach focused on bespoke data grants. If barriers to access are low, and application times are short, the Article 31 DSA framework will be more attractive to non-academic watchdogs, while also being less sensitive to overcrowding from their additional usage. But if Article 31 DSA remains smaller in scale, it makes more sense to prioritize university researchers.

At this stage, my main criticism of the "vetted researchers" category is that it is too detailed and inflexible. Precisely because so much else about Article 31 DSA still needs to be worked out in delegated decision-making, this definition is uncharacteristically, unhelpfully specific. Once the dust settles, requirements such as university affiliation and a 'proven track record' may well prove overly restrictive, or too administratively cumbersome. Why legislate on these choices now?

Will it scale? Bespoke grants versus programmatic access

Important procedural aspects of Article 31 DSA remain unclear. At present, there is no way for researchers to apply for access, and the initiative instead relies entirely on regulators to request data on their behalf. How responsive will regulators be to researcher demands? Ideally, researchers and academic institutions will be closely involved in setting the data access agenda. But in the current draft, government calls all the shots. Combined with the topical restriction to 'systemic risks', one gets the impression that the Commission sees Article 31 DSA primarily as a means to outsource regulatory monitoring burdens to universities, rather than supporting independent research for its own sake.

A related question is how repeated usage of the same resources will be handled. Once a given dataset or tool has been accessed by one researcher, does it remain available for access by others? Or must each instance of data access be decided on *de novo* in a separate procedure? Paragraph 3 stipulates that platforms "shall provide access to data […] through online databases or application programming interfaces", suggesting that the DSA envisages the creation of automated, scalable access solutions. However, APIs and databases must only be used "as appropriate", leaving room for alternative interpretations. Overall, it remains to be seen whether Article 31 DSA will mainly produce bespoke data grants for specific recipients, or instead automated, scalable tools available to a larger pool of researchers.

Procedural delays and logjams are a central problem in other areas of transparency regulation, such as Freedom of Information laws. And they could be especially pronounced with dominant platforms, as they are technically complex, highly adverse to transparency and notoriously litigious. Judging by their recalcitrance to earlier attempts at transparency legislation, platforms will contest compelled disclosures vigorously in and outside of court. All the more important for regulators to prioritize access to general-purpose resources that serve many comers, as each victory will be hard-fought.

The Commission's delegated acts could make or break these issues, since the current draft barely specifies any procedural aspects. And that's of course presuming the Commission ever gets around to these tasks all. From earlier episodes like the GDPR we already know that the Berlaymont's eyes are often bigger than its stomach, and that delegated rulemaking announced in legislation often fails to materialize in practice.

Carveouts: security, trade secrets, and "confidential information"?

One of the hardest problems created by data access regulation is managing the risk of abuse. Article 31 DSA does this by restricting access to vetted researchers, but also by creating carveouts. Platforms can refuse an access request in cases where "giving access to the data will lead to significant vulnerabilities for the security of its service or the protection of confidential information, in particular trade secrets" (Paragraph 6). We can expect platforms to litigate these carveouts to their limits. Facebook has already abused privacy law as a pretext to refuse data access, and security and trade secrets considerations can be put to similar ends.

Worryingly, the exemption for commercial interests doesn't just cover trade secrets as such but all "confidential information, in particular trade secrets". This protection of the confidential is almost paradoxically broad; is it not the very purpose of a research access framework to provide access to information that has not yet been disclosed – that is, confidential? Even an exemption for "trade secrets" alone is problematic; under recent CJEU case law, transparency exceptions for trade secrets have been read so broadly that they already function, in the words of Emilia

Korkea-aho and Päivi Leino, as a “general presumption of non-disclosure” against transparency requests toward EU agencies.

Arguably, confidentiality conditions obviate the need for such exemptions. Vermeulen points out that independent auditors, regulated in Article 28 DSA, have more far-reaching access rights, covering trade secrets so long as they guarantee their confidentiality. “Pre-vetted researchers must live up to the same standards and their vetting process should be conditional upon their ability to live up to those standards,” Vermeulen argues, “but security reasons and trade secrets should not be a ground for a platform to refuse access to data a priori”.

What about scrapers? Protecting independent data collection

It is important to note that Article 31 DSA doesn’t provide any clear answers for disputes like those between Facebook and NYU or AlgorithmWatch. These disputes revolve around the independent ‘scraping’ of data collected with the help of volunteer-installed browser extensions. Legally, the main problem is that platforms prohibit such practices in their Terms of Service. These provisions grant platforms the power to arbitrarily restrict access and shut down unwelcome research.

What scrapers need is a guarantee that Terms of Service won’t be used to shut down privacy-compliant public interest research. In the United States, the Knight First Amendment Institute is advocating for a self-regulatory solution where platforms add a so-called “safe harbor” clause for public interest research in their Terms. In Europe, AlgorithmWatch is now looking to the DSA to “ensure that Terms of Service cannot be weaponized against individuals or organizations that attempt to hold large platforms to account”.

Of course, scraping can also be abused. Amelie Heldt, Matthias Ketteman and I have argued in an earlier blog post that platforms should still be able to take action against unlawful and unethical scraping. Matthias Vermeulen has argued for a GPDR Code of Conduct that clarifies the application of data protection law to independent scraping, which should help to distinguish the good from the bad and minimize any chilling effects on legitimate research.

Will we still need scraping once – if! – Article 31 gets up and running? Yes, I argue. Scraping is an important supplement to regulated access, certainly for the time being. As should be amply clear by now, disclosure regulation is highly complex and may take years or decades to succeed, while scraping is something that already happens every day. Moreover, scraping is entirely independent of platforms and can help to fact-check the official data they provide under a regulated framework. For instance, scraped data has been used to detect political advertisements that platforms failed to include in their official disclosures. Regulated access may be more powerful on the longer term, since it applies systemically to all platform data, whereas scraping only observes what platforms reveal to their users. But for the time being we depend on scraping. To that end, the DSA should not only strike

at platforms to compel disclosure, but shield researchers to protect independent collection.

A more theoretical account would observe that the DSA's current approach fits neatly into existing patterns of platform regulation as described in Julie Cohen's landmark account of informational capitalism, *Between Truth and Power*. Policymakers are eager to construct complex new regulatory duties on and with platform services, but remain largely blind to the role of existing legal institutions in determining the baseline allocation of entitlements around platform data, such as trade secrets and Terms of Service contracts. This is how our legislators arrive at baroque new transparency rules, while leaving unquestioned the legal strictures that brought us to this problem in the first place.
