Independent Researcher Access to Social Media Data: Comparing Legislative Proposals

Researchers use data from social media companies and other hosts of user-generated content to study important topics of public concern, such as the efficacy of different content moderation efforts and ideas to improve them, the spread of dis- and mis-information online, ranking and recommendation algorithms, and online advertising. But some researchers have been stymied by the type and amount of data available, the level of control that social media companies exert over researchers’ access, and other barriers.

Lawmakers in both the United States and Europe are increasingly focused on how to meet the needs of independent researchers who want better access to data from social media companies to conduct research in the public interest, while at the same time balancing user privacy and other concerns.

In the last year, members of the US Congress have introduced or published at least four bills or discussion drafts with provisions about researcher access to data held by online services: The Platform Accountability and Transparency Act, Social Media Data Act, Digital Services Oversight and Safety Act, and Kids Online Safety Act.

In Europe, Article 31 of the Digital Services Act is poised to become the first major legislation requiring some online services to make certain data available to researchers. The European Council, Commission, and Parliament have each adopted positions on the DSA and are now engaged in the “trilogues,” through which they will negotiate a joint position, including on researcher access to data. (Because the trilogue process and draft joint positions of the European Council, Commission, and Parliament are not open to the public, the chart below summarizes the European Parliament’s position on Article 31. Many parts of Article 31—such as the specific criteria for the vetting of researchers—are being discussed during the trilogues and may differ in the final version of the DSA.)

CDT has compiled a chart (last updated on April 21, 2022) comparing how these different researcher access to data proposals answer seven key questions.
Table of Contents

I. **Who would have access to data?**

II. **What types of data would be accessible to “researchers,” specifically?**

III. **Are there restrictions on the purpose of the research or research project?**

IV. **Which online services must make data available?**

V. **What privacy and security safeguards would there be for data made available to researchers?**

VI. **What would be the method or mechanism for vetting researchers and providing access?**

VII. **Is there a safe harbor for independent methods of data access?**

---

## I. Who would have access to data?

| Platform Accountability and Transparency Act | “Qualified researchers” =
“a university-affiliated researcher” specifically identified in a research proposal that is approved by the NSF to conduct research as a qualified research project” (Sec. 2) |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Public access to some data:</strong> Gives the FTC rulemaking authority to require covered platforms to report certain other data or information to the public, qualified researchers, or some combination of the two (Sec. 12(a)) and requires the FTC to issue rules requiring platforms to make public reports about content that has been highly disseminated (Sec. 12(b)), advertising (Sec. 12(c)), algorithms (Sec. 12(d)), and content moderation (Sec. 12(e)).</td>
<td></td>
</tr>
<tr>
<td>Social Media Data Act</td>
<td><strong>Academic researchers</strong> and the FTC. (Sec. 2(a)(1))</td>
</tr>
<tr>
<td>Academic researcher = an individual that conducts research in collaboration with an institution of higher education (as defined in section 6 101(a) of the Higher Education Act of 1965) and research is not for commercial purposes. (FTC may update definition as needed) (Sec. 2(d)(1))</td>
<td></td>
</tr>
</tbody>
</table>
| Digital Services Oversight and Safety Act | Researchers affiliated with an institution of higher education or nonprofit whose mission includes developing a deeper understanding of the impacts of platforms on society. Both organizations and researchers must be certified by the Office of Independent Research Facilitation to be established at the FTC. **“Host**
organizations” must meet requirements TBD by the FTC and commit to training researchers, reviewing research projects, and other commitments. **Certified researchers** must meet requirements established by the FTC and make commitments, such as compliance with information or security requirements established by the FTC, agreeing not to attempt to reidentify data, agreeing to publish their research, and more. (Sec. 10(b))

**Public access to some data:** Requires FTC to issue regulations requiring a provider of a hosting service to issue **publicly available transparency reports** relating to content moderation. (Sec. 6(b)) Requires FTC to issue regulations requiring providers of a large covered platform to maintain a public version of an **advertising library** (Sec. 10(f), 10(f)(3)) and a public version of a **high-reach public content stream** (Sec. 10(g), 10(g)(4)).

### Kids Online Safety Act

“Qualified researchers” =  
(1) Affiliated with an **institution of higher education** or a **nonprofit organization**, including any 501(c); and  
(2) **Approved** by Assistant Secretary of Commerce for Communications and Information (NTIA)

To gain approval, a researcher must:

- Conduct the research for **noncommercial purposes**;  
- Demonstrate a **proven record of expertise** on the research topic and related research methodologies; and  
- Commit to fulfill, and demonstrate a capacity to fulfill, specific **data security and confidentiality requirements** corresponding to the application.  
(Sec. 7(a)(2), 7(a)(5), (b))

### DSA Art. 31

**Vetted researchers who are affiliated with academic institutions and vetted not-for-profit bodies, organisations or associations** representing the public interest.

Vetted researchers and not-for-profits must:

- Be **independent from commercial interests**  
- **Disclose the funding** financing the research  
- Be **independent from government/state bodies** (except for public academic institutions)  
- Have **proven records of expertise** in the fields related to the risks investigated or related research methodologies  
- Preserve **data security and confidentiality** requirements.  
(Art. 31 para. 2 & 4)
II. What types of data would be accessible to “researchers,” specifically?

| Platform Accountability and Transparency Act | “Qualified data and information” = “data and information from a platform that the NSF determines is necessary to allow a qualified researcher to carry out the research contemplated under a qualified research project.” (Sec. 2). The criteria for “qualified data and information” is TBD by the NSF, but it must at least be (1) feasible for the platform to provide; (2) proportionate to the needs of the qualified researchers to complete the qualified research project; and (3) not cause the platform undue burden. (Sec. 4).

Qualified data and information could include non-public content data and personally identifiable information. |
| Social Media Data Act | Ad library with certain specified information about any advertiser that purchases $500 or more of advertising in a calendar year: name and unique identification number of the advertiser, digital copy of the ad, targeting method & description of target audience, optimization objective chosen by advertiser, description of the actual audience, number of views, ad conversion, date and time of ad display, amount advertiser budgeted and paid, ad category (such as politics, employment opportunity, housing opportunity, or apparel), ad language, and platform’s advertising policy. (Sec. 2(a)(1))

Would also establish a Working Group for Social Media Research Access at the FTC to study making other data and information available to academic researchers (Sec. 2(c)) |
| Digital Services Oversight and Safety Act | The FTC must issue regs identifying the precise types of information that will be available.

The FTC regs can specify any relevant information, but it must consider particular information: info about internal platform studies; info about content moderation decisions and policies, the people setting the policies and making decisions, & the training of moderators; third party requests to act on a user, account, or content; engagement and exposure data; classification of information sources; archives of removed content and accounts; Advertisements and influencer marketing content; detailed information about a platform’s algorithms. (Sec. 10(c)).

The information required to be disclosed by FTC regulations could |
include non-public content data and personally identifiable information, but the FTC must require platforms to deidentify certain data (non-public data, personal health information, biometric information, and information related to a person under 13 years old), before it may be disclosed and also restricts sharing of precise location information. (Sec. 10(c)(6))

In addition, the FTC must issue regulations requiring covered platforms to submit a data dictionary describing the information that can be provided to certified researchers. (Sec. 10(d))

The FTC must also issue regulations requiring large covered platforms to give researchers and the FTC access to an ad library (Sec. 10(f)) and a “high-reach public content stream.” (Sec. 10(g)).

### Kids Online Safety Act

Data assets that can be used to conduct public interest research regarding harms to the safety and well being of minors, including the following types of matters:

1. promotion of self-harm, suicide, eating disorders, substance abuse, and other matters that pose a risk to physical and mental health of a minor;
2. patterns of use that indicate or encourage addiction-like behaviors;
3. physical harm, online bullying, and harassment of a minor;
4. sexual exploitation, including enticement, grooming, sex trafficking, and sexual abuse of minors and trafficking of online child sexual abuse material;
5. promotion and marketing of products or services that are unlawful for minors, such as illegal drugs, tobacco, gambling, or alcohol; and
6. predatory, unfair, or deceptive marketing practices.

(Sec. 3(b), 7(b)(1))

The term “data assets” is not defined in the statute, and could include non-public content data and personally identifiable information.

### DSA Art. 31

Any data that serves the permissible purposes of research specified in Art. 31 para. 2 [See Section III], unless the ‘very large online platform’ (VLOP) does not have access to the data or giving access would lead to significant security vulnerabilities or reveal confidential information. (Art. 31 para. 6)

Also grants access to “aggregate numbers for the total views and view rate of content prior to a removal on the basis of” orders for removal of illegal content under Art. 8 or content moderation under a provider’s own TOS. (Art. 31, para. 2a)
### III. Are there restrictions on the purpose of the research or research project?

<table>
<thead>
<tr>
<th>Act</th>
<th>Restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platform Accountability and Transparency Act</strong></td>
<td>Only “qualified research projects” approved by NSF. A qualified research project must (1) have IRB approval or be exempt or excluded from IRB approval; (2) <strong>aim to study activity on a platform</strong>; (3) meet other criteria <strong>TBD</strong> by the NSF. (Sec. 4)</td>
</tr>
<tr>
<td><strong>Social Media Data Act</strong></td>
<td><strong>None.</strong></td>
</tr>
<tr>
<td><strong>Digital Services Oversight and Safety Act</strong></td>
<td>Researchers may be certified to gain access to information only for the purposes specified in the Act: “to gain understanding and measure the impacts of the content moderation, product design decisions, and algorithms of covered platforms on society, politics, the spread of hate, harassment, and extremism, security, privacy, and physical and mental health.” (Sec. 10(b)(1) &amp; (2))</td>
</tr>
</tbody>
</table>
| **Kids Online Safety Act**                            | Researchers may access data only to conduct **public interest research** pertaining to **harm to the safety and wellbeing of minors.**  

Public interest research = scientific or historical analysis of information that is performed for the primary purpose of advancing a broadly recognized public interest.  

(Sec. 7(a)(4), (b)(1)) |
| **DSA Art. 31**                                       | Data may be used only for research that contributes to the **identification, mitigation and understanding of specified systemic risks** set out in Art. 26(1) and Art. 27(1).  

In addition, the **Commission must adopt delegated acts** laying down, among other things, the purposes for which the data may be used.  

(Art. 31 para. 5) |
## IV. Which online services must make data available?

<p>| <strong>Platform Accountability and Transparency Act</strong> | “Platforms” = Subject to FTC jurisdiction under section 5(a)(2) of FTC Act; and Is a website, desktop application, or mobile application that allows users to establish accounts to share user-generated content and whose primary purpose is for users to interact with user-generated content and for the platform to deliver ads to users; and has at least <strong>25 million unique monthly users</strong> in the United States for a majority of the months in the most recent 12-month period. (Sec. 2) |
| <strong>Social Media Data Act</strong> | “Covered platform” = any website, desktop application, or mobile application that is consumer-facing; and sells digital advertising space; and has more than <strong>100 million monthly active users</strong> for a majority of months during the preceding 12 months. FTC can update definition as needed. (Sec. 2(d)(3)) |
| <strong>Digital Services Oversight and Safety Act</strong> | “Covered platform” = A hosting service that stores information provided by, and at the request of, users and which, at the request of users, stores and disseminates information to the public; and has at least <strong>10 million monthly active users</strong>. The methodology for determining MAU will be determined through rulemaking. (Sec. 2(11); Sec. 10(c)) In issuing the regulations about the types of information that must be disclosed, the manner of disclosure, and whether disclosure is mandatory or optional, the FTC must “vary the specifications based on the size and scope of a covered platform, including by having different specifications for different services.” |</p>
<table>
<thead>
<tr>
<th><strong>Kids Online Safety Act</strong></th>
<th>“Covered platforms” = a commercial software application or electronic service that connects to the internet and that is used, or is reasonably likely to be used, by a minor. (Sec. 2(2), Sec. 7(b)(3))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DSA Art. 31</strong></td>
<td>Very Large Online Platforms (VLOP) = average monthly active recipients of the service in the EU equal to or higher than 45 million for at least four consecutive months.</td>
</tr>
<tr>
<td></td>
<td>Number of average monthly active recipients can be adjusted based on changes to the EU population. (Art. 25)</td>
</tr>
<tr>
<td></td>
<td>“Online platforms” = a provider of a hosting service which, at the request of a recipient of the service, stores and disseminates to the public information, unless that activity is a minor or a purely ancillary feature of another service or functionality of the principal and cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of the DSA. (Art. 2)</td>
</tr>
</tbody>
</table>
V. What privacy and security safeguards would there be for data made available to researchers?

<table>
<thead>
<tr>
<th>Act</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platform Accountability and Transparency Act</strong></td>
<td>Newly-established FTC Platform Accountability and Transparency Office (Sec. 3) would establish criteria for privacy and cybersecurity safeguards required for qualified data and information related to a qualified research project, and can require reasonable privacy and cybersecurity safeguards for particular data sharing, such as encryption of data; delivery of deidentified data; use and monitoring of a secure environment to facilitate delivery of data. (Sec. 4(j))</td>
</tr>
<tr>
<td><strong>Social Media Data Act</strong></td>
<td>None. The Working Group for Social Media Research Access would study privacy preserving techniques for other data that could be made accessible to academic researchers. (Sec. 2(c))</td>
</tr>
</tbody>
</table>
| **Digital Services Oversight and Safety Act** | Tiered access: More sensitive info has more safeguards and is accessed by fewer researchers than less sensitive info. (Sec. 10(c)(2))

The FTC must issue regulations specifying the manner in which information is to be accessed, including when privacy protecting techniques “such as differential privacy and statistical noise” should be used, what information security standards should be in place, and other privacy and security measures. (Sec. 10(c)(4))

The FTC must issue regulations specifying when the Commission should review research before it is published to protect user privacy or trade secrets. (Sec. 10(c)(5))

FTC regulations must ensure that provision of access to information does not infringe upon reasonable expectations of personal privacy and must require platforms to deidentify certain information before it can be provided: nonpublic information, personal health information, biometric information, information about a child under 13 years old. Also restricts sharing of precise location information. (Sec. 10(c)(6)). |
<table>
<thead>
<tr>
<th><strong>Kids Online Safety Act</strong></th>
<th>Users who do not post public content must be given the ability to <strong>opt-out</strong> of having their information shared with researchers. (Sec. 10(c)(6)(C))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The NTIA must establish standards for privacy, security, and confidentiality</strong> required to participate in the program for a researcher to receive, and a covered platform to provide, data assets. (Sec. 7(b)(4)(C))</td>
<td></td>
</tr>
<tr>
<td>Imposes a <strong>duty of confidentiality</strong> on a qualified researcher with respect to data assets provided by a covered platform. The duty of confidentiality may be defined further by the NTIA. (Sec. 7(b)(5))</td>
<td></td>
</tr>
<tr>
<td><strong>DSA Art. 31</strong></td>
<td><strong>TBD:</strong> The Commission must adopt delegated acts laying down, among other things, “the specific conditions under which such sharing of data with vetted researchers or not-for-profit bodies, organisations or associations can take place in compliance with [the GDPR] taking into account the rights and interests of the very large online platforms and the recipients of the service concerned, including the protection of confidential information, in particular trade secrets, and maintaining the security of their service.” (Art. 31 para. 5.)</td>
</tr>
</tbody>
</table>
## VI. What would be the method or mechanism for vetting researchers and providing access?

| **Platform Accountability and Transparency Act** | The NSF vets the researcher and research project and determines what data a platform must make available; the Platform Accountability and Transparency Office informs the platform and establishes the privacy and cybersecurity safeguards for the particular data at issue.  
Step 1: A researcher submits a research application to the NSF  
Step 2: The NSF determines if it is a “qualified research project” by a “qualified researcher.”  
Step 3: The NSF identifies the “qualified data and information” that platforms will be required to make available to the researcher, and in what form.  
Step 4: The NSF refers the qualified research project to the FTC Platform Accountability and Transparency Office  
Step 5: The Office notifies the platform that it will be required to provide data and establishes reasonable privacy and cybersecurity safeguards for the data.  
Step 6: The platform can comment on the privacy and cybersecurity safeguards; following the platform’s comments, the Office makes a final determination re: the safeguards. (Sec. 4). |
|**Social Media Data Act** | A covered platform must maintain, and grant academic researchers and the Commission access to, an ad library that contains in a searchable, machine readable format. (Sec. 2(a)(1)) |
|**Digital Services Oversight and Safety Act** | The FTC establishes a “research certification process” under which an organization can apply and be qualified as a host organization and an individual associated with a host organization can apply and be certified as a certified researcher. (Sec. 10(b))  
The FTC issues regulations specifying the manner in which researchers will access information from covered platforms. (Sec. 10(c)(1) & 10(c)(4))  
The FTC must consider, among other things, size and sampling techniques used to create data sets, under what circumstances APIs are required, and designate “secure facilities and computers to analyze information through a Federally Funded Research and Development Center” established by the Act. (Sec. 10(c)(4)). |
### Kids Online Safety Act

The NTIA must establish a program under which a researcher can apply for access to data and the NTIA can approve their application. (Sec. 7(b)(1)-(4))

For applications that are approved, a covered platform must provide to a qualified researcher access to data assets through **online databases, application programming interfaces, and data files** as appropriate for the qualified researcher to undertake public interest research. (Sec. 7(b)(3)(A)(ii))

### DSA Art. 31

Researchers would be vetted by the Digital Services Coordinator of establishment or the Commission. (Art. 31, para. 4)

Access to data would be provided through **online databases or application programming interfaces**, as appropriate, and with an easily accessible and user-friendly mechanism to search for multiple criteria. (Art. 31 para. 3)

**More details TBD**: The Commission must adopt delegated acts laying down, among other things, “the technical conditions under which [very large online platforms] are to share data . . . .” (Art. 31 para. 5)
VII. Is there a safe harbor for independent methods of data access?

<table>
<thead>
<tr>
<th>Act</th>
<th>Answer</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Platform Accountability and Transparency Act</strong></td>
<td>Yes</td>
<td>No civil or criminal liability for any person for collecting covered information as part of a newsgathering or research project on a platform. (Sec. 11). Conditions: Only applies to “covered methods of digital investigation”; purpose must be to inform the general public about matters of public concern, and the information in fact is used only that way; the person takes reasonable measures to protect the privacy of the platform’s users; w/r/t research accounts, the person takes reasonable measures to avoid misleading users; and the project does not materially burden the technical operation of the platform. (Sec. 11). “Covered method of digital investigation” = TBD by FTC regulations, but must include collection of data through automated means, through data donation, or through research accounts. (Sec. 11). “Covered information” = publicly available information, information about ads, other information TBD by FTC that does not unduly burden user privacy. (Sec. 11).</td>
</tr>
<tr>
<td><strong>Social Media Data Act</strong></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td><strong>Digital Services Oversight and Safety Act</strong></td>
<td>Yes</td>
<td>Certified researchers granted immunity for liability under state, federal, and local law for violating platform’s TOS for two specified research activities: creating a research account (if researcher takes reasonable means to avoid misleading users and does not burden technical operation of platform) and data donation with informed consent of users. (Sec. 10(c)(10)) Also prohibits a covered platform from discriminating against a certified researcher in the provision of services because of those two research activities. (Sec. 10(c)(10)).</td>
</tr>
<tr>
<td><strong>Kids Online Safety Act</strong></td>
<td>Yes</td>
<td>No cause of action for violating platform’s TOS may be brought based on actions a researcher takes while collecting data assets as part of public interest research regarding harms to minors. (Sec. 7(c))</td>
</tr>
<tr>
<td><strong>DSA Art. 31</strong></td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>