Google's answer to ARCOM's public consultation on data access for research purposes

---

On 25 May 2022, the French Regulatory Authority for Audiovisual and Digital Communication (ARCOM) decided to conduct a public consultation on access to online platforms' data for research purposes.

The aim of this public consultation is to "*learn initial lessons regarding the implementation of an operational framework for accessing data from online platforms and thereby contribute to the various stakeholders' general reflection concerning these issues, particularly researchers and the public sphere*."

This document constitutes Google's answer to this public consultation.

For all intents and purposes, submitting this document is not an acknowledgement from Google of the qualification of any of its services as "online platform operator" under Article L. 111-7 of the French Consumer Code or of "(very large) online platform" under the provisions of the Digital Services Act (DSA) recently adopted.

* * *

Google not only supports but has a track record of strong engagement with the research community in exploring the dissemination and impact of illegal and harmful online content, as well as mitigating risks associated with such content. We receive frequent inquiries from academic researchers who investigate these issues as they seek to better understand our products and policies or have questions with regards to public data they've collected as part of their independent research efforts. We respond to these requests as diligently as possible, and aim to inform the work of these researchers to the best of our ability.

In this context, we're always looking for new ways to deepen transparency and collaborate with academic researchers from around the world. A recent example is the launch of the YouTube Researcher Program, which contributes to our continuing efforts to assess risks and determine efficient mitigation strategies for our services.

We welcome the introduction of Article 31 of the DSA as a means to harmonize requesting and granting access to data for conducting research into systemic risks and their mitigation. We understand that the DSA aims to bring about maximum harmonization at EU level, given the inherently cross-border nature of such issues.

We also supported the inclusion of safeguards within Article 31, and in that regard, the DSA provides:

- The ability for online platforms to object to a request where they do not have access to data, or 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.
- Requirements that researchers be affiliated with a research organization, are independent from commercial interests, and disclose funding of the research;
- Requirements that researchers are in a capacity to preserve the specific data security and confidentiality;
- Requirements for requests to be duly substantiated, including justification of the necessity and proportionality for the purpose of their research of the data requested; and
- A commitment to make research results publicly available free of charge, within a reasonable period after the completion of the research and taking into account the rights and interests of the recipients of the service concerned.

We examine some of these elements in more detail below.

We also understand that Article 31 of the DSA will be further specified through implementing measures, including mandatory delegated acts, that the Commission will adopt, and into which authorities such as ARCOM will feed via the Board of Digital Services Coordinators. We expect those implementing measures to address a number of important concerns which require the implementation of different safeguards to avoid risks of misuse or abuse of data requested from online platforms. We elaborate on those below.

1. **Substantiating requests for access to data**

It is important that any request online platforms receive to share data with researchers be specific and appropriately reasoned. The request should, in particular, explain what the research is and how the requested dataset will contribute to it.

Article 31 of the DSA recognises this and, among other things, requires researchers to prove to the Digital Services Coordinator vetting them that the data requested is necessary for and proportionate to the purpose of the proposed research. It also requires researchers to prove that the expected results of the research will actually contribute to the purpose of assessing and mitigating systemic risks in the EU.

The data access request that the Digital Services Coordinator will eventually address to the online platform (after the vetting of the researcher) should transmit this information. This will allow online platforms to confirm the validity of the request. It will also allow them to better understand the context in which the request is made and what kind of alternative data they may be able to offer, if the requested dataset is not available.

2. Ensuring requests for access to data do not undermine safety on online platforms

Access to online platforms' data should be scoped appropriately, bearing in mind the risk of potential circumvention of safety measures put in place by online platforms to protect Internet users and their products. The regulatory framework should ensure sensitive and confidential data remain so and are not compromised in a way that could enable bad actors to circumvent safety measures put in place.

It is in recognition of this risk that Article 31 of the DSA[1] provides that a very large online platform may request the Digital Service Coordinator of establishment to amend a data access request within 15 days following its receipt, if the platform is unable to give access to the data because this will lead to "*significant vulnerabilities*" for the security of the service, or in order to protect confidential information (in particular trade secrets).

Not giving overly broad access to data that would infringe on data confidentiality is of paramount importance. We would welcome more detailed guidance on how regulators intend to carve out and protect platforms' sensitive and confidential data. We urge regulators to interpret the notion of "*significant vulnerabilities*" in a manner that would allow this concern to be adequately addressed.

3. Setting the right framework to protect shared data

In addition, all stakeholders would benefit from more clarity on safeguards to be taken by researchers prior to their accessing platforms' data. Robust guarantees should be given to platforms when it comes to the data they give access to following a researcher's request.

Legal frameworks and technical protocols related to data collection and retention by researchers should be created to avoid any risks of misappropriation of the said data and to contribute to ensuring confidentiality protection. In particular, it should be clarified which party has liability for breaches of personal data protection rules that could result from sharing data with researchers.

4. Focusing data requests on accessible data

Data requests should focus on data that is easily identifiable and already available. They should be restricted and proportionate to the research project's objectives, starting with data that is already made public.

In particular, Google regularly publishes a wealth of information directly on its platforms or as part of transparency reports. In line with the proportionality principle, this is the place where

[1] References to the Digital Services Act (DSA) in this document relate to the text of the provisional agreement DSA published on 16 June 2022

researchers should start looking for information before making additional data requests to the extent needed.

In any case, it should be acknowledged that platforms can't give access to data that is not reasonably accessible to the online platform (e.g. because it does not exist in the form requested). Hence, specific datasets or data points should not have to be created for the purpose of meeting a researcher's isolated data request.

This being said, as mentioned above and illustrated by the below YouTube example, Google remains fully open to consider partnerships with research organizations which may require to co-develop data to meet the research project objectives. Such partnerships represent significant efforts and are based on a mutual understanding between Google and the research organization. This is therefore a distinct path from the researchers' possibility to request access to specific data under legal statutes enabling them to do so.

Recent example on how Google shares data with research community: the YouTube Researcher Program

On July 12, 2022, YouTube launched the *first phase* of a new program called YouTube Researcher Program aiming at equipping researchers from around the world with data, tools, and supporting them to advance the public's understanding of our platform and its impact.

In particular, YouTube offers participants the following:

- scaled access to YouTube's public data corpus via our Data API with as much quota as required for their research,
- opportunity to derive insights from global YouTube data, and
- support and technical guidance from YouTube.

Eligible researchers from diverse disciplines can apply to use YouTube data to study a variety of topics. You can click here for more information on eligibility requirements and conditions of the program and here for more information on how it works.

Google is pleased to be working more closely with the research community and to receive feedback on additional data or features that may be integrated in future stages of the Program.

***