Consultation on the European Commission’s Digital Services Act Transparency Database

Google’s Submission
July 17, 2023

Google understands that the European Commission’s goal for the database is to collect data and publicly share it to inform anonymous, statistical insight on content moderation and enforcement in the EU across online platforms, as defined under the DSA. We would like to offer some pragmatic feedback and concrete suggestions for the API, including:

1. Providing latitude for industry to have more lead time (beyond 28 August 2023) to modify technical systems to implement the statement of reasons database requirements and engage with the EC on certain data fields;

2. Introducing flexibility into the API’s validation protocol while very large online platforms continue to engage with the European Commission on certain data fields, including by allowing very large online platforms to submit statements of reasons that have some, but not all, required data fields;

3. Permitting online platforms to transmit data in batches to the European Commission in a pre-agreed frequency;

4. Providing further clarification about potentially overlapping data fields and convening discussion about implementation challenges relating to certain data fields;

5. Clarifying technical details that would be helpful to building workable transmission systems.

We welcome the opportunity to further discuss operational guidance and our proposals to satisfy the European Commission’s goals for creating and managing this new database while safeguarding user privacy, ensuring the database is scalable and
durable, and decreasing potential friction or technical challenges associated with data transmission between systems.

**Implementation Timeline:** Our first concern is around the implementation timeline. The API will likely be finalized within weeks of 28 August 2023, yet the terms of the draft delegated act on the DSA audit requires that all compliance be in place and auditable by 28 August 2023. Google will work to modify technical systems to ensure statement of reasons submissions that are compliant with the requirements of the final API. However, given the time required for modification of technical systems and iterative industry engagement with the European Commission to scope and define data points, we recommend providing industry with an express grace period that takes into account the lead time necessary for implementation once requirements are finalized.

**Validation Protocol:** Likewise, we recommend flexibility around the submission validation protocol to reflect the fact that requirements are still being finalized. Currently, in order for a statement of reasons submission to be “validated” and accepted into the database, all required data fields need to be filled. Some of these data fields present legal, technical, or operational challenges, and the European Commission should host robust stakeholder engagement on these complex topics without the time pressure to conclude all discussions before 28 August 2023. We recommend the European Commission introduce flexibility into the validation process and make certain currently “required” data fields optional.

**Batching:** As we have discussed in response to prior questions, we are also concerned about the database’s ability to intake large volumes of data from online platforms. We believe the API should permit batch uploads, which would (1) make it easier for online platforms to detect and recover from delivery errors or unexpected server interruptions, (2) ensure that large submitters do not overwhelm the European Commission’s transmission system, (3) in the event the European Commission continues to require such expansive data elements, give online platforms more time to remove personal data.

**Data Fields Requiring Further Discussion:** We would also welcome further discussion regarding a number of data fields that may pose legal, technical, or operational challenges.

For example, the draft API requires disclosure of a number of elements that are not specified in Articles 17 or 24 of the DSA, even though the scope of disclosures to the database are bounded, in the language of Article 24(5), by Article 17 of the DSA. Accordingly, the DSA appears to make clear that (very large) online platforms are not obligated to submit to the database data points beyond the information they are
required to provide to users under Article 17 of the DSA. The draft API’s net-new requirements also pose a significant technical and implementation challenge given that online platforms have been building towards the requirements of Article 17 of the DSA for a significant length of time. Additional specifications will require new resourcing allocations and engineering lead time to modify existing technical builds.

Requirements in the API that are not required by the terms of Article 17 of the DSA include:

- **URL:** The draft API requires, for each statement of reasons, “URL to the data that has been moderated.” In addition to the re-identification risks we have discussed above in response to question 23, we have concerns that (1) this requirement will require significant technical and engineering changes, because many services do not currently maintain publicly available URLs for restricted, deleted, or removed content, or URLs may not exist for individual pieces of content (e.g., user reviews), and (2) this requirement may pose significant legal/user safety risks by providing unrestricted access to a trove of publicly accessible material that was removed because it was unlawful and/or harmful.

- **Category:** The European Commission has provided net new subject categories – piracy, discrimination, counterfeit, fraud, terrorism, child safety, non-consent, misinformation, TOS violation, and uncategorized – with which to tag each statement of reasons. These categories are not associated with any requirements in Articles 17 or 24 of the DSA, nor are they based on any industry standard. For this data field to allow for useful comparison among online platforms, we must achieve industry-wide alignment on generally applicable subject categories. Accordingly, more flexibility and discussions with industry are necessary to understand what is feasible by 28 August 2023 and what categories should be used longer-term. Furthermore, the European Commission has indicated that it will likely move to different, revised subject categories in the next version of the database. The significant engineering work to associate statements of reasons with these preliminary subject categories would be disproportionately burdensome if online platforms will, in short order, be required to remap statements of reasons to new or revised subject categories.

- **Whether content is both policy violating and illegal:** The draft API provides a required attribute for SORS to indicate whether “not only was the content incompatible but also illegal.” As a legal matter, the DSA does not mandate that online platforms undertake a review of whether content violates both the law and their policies. Operationally, this requirement is at odds with the current architecture of many online platforms. Typically, if a platform actions content for
a policy violation, it is removed globally—and because no further enforcement action is possible, the platform does not also conduct an analysis of the legality of the content. We recommend the European Commission remove this category, make this category “optional” or provide a third option of “unconfirmed”.

- **Content type**: For “content_type,” the European Commission provides a list of three standard categories (text, video, image) and “other.” There are some standard media categories not accounted for in the European Commission’s list—including audio. Additionally, online platforms may have multiple forms of content that are some combination of the three standard categories (e.g., ads are often image and text, or video and text). The European Commission should provide an option for “audio,” “multiple content types” and/or “combination”.

- **Source type and source**: Article 17(3)(b) of the DSA only requires online platforms to share information on Article 16 notices “where relevant,” but the API requires this information as a mandatory data element. Requiring that online platforms select one of three options (Article 16, Trusted Flagger, Voluntary) goes beyond the requirements of Article 17 of the DSA. Furthermore, the data field “source” requires that online platforms identify the source when they select Article 16. Depending on the level of detail expected, this disclosure could conflict with the requirement under Article 24 of the DSA that no personal data is included in the submissions made to the database. To the extent this field is maintained, it is also important for the European Commission to better define the field’s meaning.

We would also welcome discussion on a number of the European Commission’s proposed data fields that appear to be overlapping. For example, it is not clear how online platforms should understand the delineation between terminating an account and complete termination of the service, and accordingly there could be overlap between:

- “DECISION_PROVISION_TOTAL_SUSPENSION (total suspension of the provision of service)” and “DECISION_ACCOUNT_SUSPENDED (suspension of the account)"

and

- “DECISION_ACCOUNT_TERMINATED (termination of the account)” and “DECISION_PROVISION_TOTAL_TERMINATION (total termination of the provision of the service)”
Lastly, Article 17 of the DSA generally requires that online platforms be “as precise and specific as reasonably possible under the given circumstances” so long as the information provided “reasonably allow the [user] to effectively exercise the possibilities of redress”. Certain text-based fields required in the database may present challenges at scale, such as:

- Facts and Circumstances text field (decision_facts)
- Illegal ground explanation field (illegal_content_explanation)
- Policy ground explanation field (incompatible_content_explanation)

**Additional Technical Details** To allow us to build a workable transmission system, the European Commission should specify additional technical details, including:

- A European Commission point of contact for the database (e.g. to allow notification regarding database maintenance or outages affecting online platforms’ data transmissions and vice versa);
- Protocol for handling outages on both ends – the online platforms’ and the European Commission’s;
- How many statements of reasons may be uploaded simultaneously;
- What protocol online platforms should follow if the server fails; and
- Technical specifications related to retry intervals and parameters for adaptive throttling.

We also note that the API as currently documented is not idempotent (e.g., there’s no protection against sending duplicate records). There may be occasions when duplicate records may be inadvertently transmitted (such as when there is a network error in the middle of a request, before a response is received). The API should be modified to be idempotent to support de-duplication.

Google would welcome an ongoing dialogue to discuss these and other relevant technical details.