

23 May 2023

**COMMENTS IN RESPONSE TO
EUROPEAN COMMISSION CALL FOR EVIDENCE ON
“DELEGATED REGULATION ON DATA ACCESS
PROVIDED FOR IN THE DIGITAL SERVICES ACT”**

The Association for Computing Machinery (ACM) is the world’s largest and longest established professional society of individuals involved in all aspects of computing. It annually bestows the ACM A.M. Turing Award, often popularly referred to as the “Nobel Prize of computing.” ACM’s Europe Technology Policy Committee (“Europe TPC”) is charged with and committed to providing objective technical information to policy makers and the general public in the service of sound public policymaking. ACM and Europe TPC are non-profit, non-political, and non-lobbying organizations. Europe TPC is pleased to respond to the European Commission call for evidence on “Delegated Regulation on data access provided for in the Digital Services Act.”¹

As highlighted by the European Commission, online intermediaries (particularly those with millions of users) have become key facilitators of the exchange of information and economic transactions. As such, their actions can have an enormous impact on the safety of citizens or the fairness of businesses’ commercial activities online. Europe TPC offers the following responses to select Commission inquiries regarding researchers’ access to data from very large online platforms (VLOPs) and very large online search engines (VLOSEs), a key feature of the Digital Services Act:

1) Data access needs

a) What types of data, metadata, data governance documentation and other information about data and how it is used can be useful to DSC’s² for the purpose of monitoring and assessing compliance and for vetted researchers for conducting research related to systemic risks and mitigation measures?

Europe TPC notes the DSA’s intent in Article 40.4 to allow vetted researchers who meet the requirements in paragraph 8 of this Article access to data for the purpose of conducting research that contributes to the detection, identification and understanding of systemic risks to assess the adequacy, efficiency, and impacts of risk mitigation measures.

¹ [Consultation re Call for Evidence](https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en) concerning “Delegated Regulation on data access provided for in the Digital Services Act,” opened 25 April 2023. [https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en]

² DSCs are “Digital Services Coordinators,” who can request access for monitoring purposes and to assess compliance with the DSA obligations.

More than data, metadata, and data governance documentation are needed however, to maximize the transparency and accountability of complex and large-scale algorithmic decision-making systems, particularly those constructed with the aid of machine learning. It is equally important to consider such systems' processing architecture and algorithmic components. Increasingly the roots of harmful undesired outcomes may be traced not only to data, but to machine learning models and system level components operative in processing at multiple stages of these complex systems.³ Accordingly, looking only at end-to-end inputs/outputs of these complex large-scale systems may not be enough to obtain meaningful insight into them. Considering intra-system processing steps, including potentially intra-system APIs (with due consideration of privacy and security), also may be necessary.⁴

2) Data access application and procedure

a) Digital Services Coordinators (DSCs) in the Member States will play a key role in assessing researchers' applications and they will act as intermediaries with the platforms. How should the application process be designed in practice? How can the vetting process ensure efficient exchanges between researchers and platform providers?

The vetting process should prioritise the security, privacy and protection of the data made available to researchers. Applicants should be required to certify that they can meet all applicable such requirements once data is made accessible.

b) Article 40(8) exhaustively defines criteria for vetting researchers. How can a consistent assessment across DSCs be ensured, while still taking into consideration the specificities of each request?

Europe TPC recommends that the Commission consider the United Kingdom's GDS Digital Services Framework for guidance.⁵ It enables both new and established companies to provide information about their products or services in a standardised submission format to facilitate their fair and consistent evaluation. A similar approach might also support both DSCs and applicants in this context.

³ The architectures behind VLOP recommender systems as well as the search architectures of VLOSEs go beyond single machine learning models. Instead, these systems consist of large-scale and complex heterogeneous computational components powered by both AI models and software elements that are made up from a broad range of data processing stages. For illustrations of how complex such systems can be see, e.g., [Facebook's Search Architecture](https://arxiv.org/abs/2006.11632) [https://arxiv.org/abs/2006.11632], Alibaba's ["Billion-scale" Recommender Engine Architecture](https://arxiv.org/abs/1803.02349) [https://arxiv.org/abs/1803.02349], or [TikTok's Real-Time Recommendation System](https://arxiv.org/abs/2209.07663) [https://arxiv.org/abs/2209.07663].

⁴ We note that Twitter recently open-sourced the [code for its tweet-recommendation-system](https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm) [https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm], which allowed for public scrutiny and led to meaningful findings. These included the discovery of questionable ["hard-coded logic"](https://github.com/twitter/the-algorithm/blob/7f90d0ca342b928b479b512ec51ac2c3821f5922/home-mixer/server/src/main/scala/com/twitter/home_mixer/functional_component/decorator/HomeTweetTypePredicates.scala#L224-L246) [https://github.com/twitter/the-algorithm/blob/7f90d0ca342b928b479b512ec51ac2c3821f5922/home-mixer/server/src/main/scala/com/twitter/home_mixer/functional_component/decorator/HomeTweetTypePredicates.scala#L224-L246] that captured users' political-leanings]. It would not have been possible to explicitly identify this logic by looking only at the code itself.

⁵ See <https://gds.blog.gov.uk/2013/07/09/digital-services-framework>.

3) Data access formats and involvement of researchers

a) What technical specifications could be considered for data access interfaces, which takes into account security, data protection, ease of use, accessibility, and responsiveness (e.g., APIs, data vaults and other machine-readable data exchange formats)?

It is important that the best practices in terms of Open Data are followed.

b) What capacity building measures could be considered for the research community to take advantage of the opportunities provided by Article 40?

Europe TPC recommends that the Commission create and fund an open consortium with the mission of disseminating best practices for identifying and understanding systemic risks inherent in and associated with the design and operation of VLOPs and VLOSEs.

c) Would it be desirable and feasible to establish a common and precise language for DSCs, vetted researchers, VLOPs and VLOSEs to use when communicating about data access, e.g. by formulating a standard data dictionary and/or business glossary? How might this be implemented?

Establishing a common and precise language for DSCs certainly would be helpful. To that end, Europe TPC observes that a call for proposals regarding how to devise such a language could be effective, particularly if optimized to elicit responses broadly from the European research and industrial communities.