

## COMMENTS IN RESPONSE TO THE DRAFT DELEGATED REGULATION ON DATA ACCESS PROVIDED FOR IN THE DIGITAL SERVICES ACT<sup>1</sup>

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.<sup>2</sup> Europe TPC has responded to the European Union (EU) stakeholder’s consultations in the past in the context of the AI Act<sup>3</sup>, the Data Act,<sup>4</sup> the Digital Services Act<sup>5</sup>, the Digital Citizen Principles<sup>6</sup>, the Cyber Resilience Act<sup>7</sup>, amongst others<sup>8</sup>. ACM and Europe TPC are non-profit, non-political, and non-lobbying organisations.

Europe TPC is pleased to provide feedback on the draft Delegated Regulation on data access provided for in the Digital Services Act (DSA) supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council by laying down the technical conditions and procedures under which providers of Very Large Online Platforms (VLOPs) and of Very Large Online Search Engines (VLOSEs) are to share data pursuant to Article 40 of Regulation (EU) 2022/2065.<sup>9</sup>

### Overview

Europe TPC broadly agrees with the provisions of the regulation set out in the draft act. The procedures and technical conditions appear well-considered and identify the roles and responsibilities of the key stakeholders involved, including the researchers, data providers, and Digital Services Coordinators (DSCs). The time constraints identified for each step of the procedure are likely to play an important role in ensuring that the regulation enables researchers to study systemic risks in the Union and assess the effectiveness of measures to reduce those risks.

---

<sup>1</sup> The authors of this document are: Advait Deshpande, Luigi Di Biasi, Maciej Zuziak, Jože Rožanec, Ahmed Nagy, and Alejandro Saucedo. Additional inputs and assistance were provided by Francisco Medeiros and Jolanta Mizera-Pietraszko.

<sup>2</sup> <https://www.acm.org/public-policy/europe-tpc>

<sup>3</sup> <https://www.acm.org/binaries/content/assets/public-policy/europe-tpc-comments-ai-consultation.pdf>

<sup>4</sup> <https://www.acm.org/binaries/content/assets/public-policy/acm-eur-tpc-data-act-comments-13may22a.pdf>

<sup>5</sup> <https://www.acm.org/binaries/content/assets/public-policy/europetpc-digital-services-act-comments.pdf>

<sup>6</sup> <https://www.acm.org/binaries/content/assets/public-policy/europetpc-comments-digital-principles.pdf>

<sup>7</sup> <https://www.acm.org/binaries/content/assets/public-policy/acm-europe-tpc-cyber-resilience-comments-pdf>

<sup>8</sup> <https://www.acm.org/public-policy/public-policy-statements>

<sup>9</sup> [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](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)

Europe TPC offers the following recommendations specific to the provisions in the draft act.

## Recommendations

- As part of any publicly accessible interfaces, and the public dashboard provided through the DSA data access portal, the European Commission should aim to ensure wider reach and proliferation of the datasets made available for research, any completed research based on the available datasets, and outcomes relevant to the stated objective of identifying systemic risks to the EU.
- Objective and consistent criteria should be specified for assessing access modalities, reasoned requests, amendment requests, and granting extensions for duly justified cases to ensure uniform implementation of the draft act across member states, and to enable the DSCs to exercise due discretion in a consistent manner.
- As part of the reasoned request, the Commission should seek to establish the intent and planned utility of the data access application in clear, research-specific terms, and how it can lead to the identification of systemic risks.
- Subject to considerations of security, end-user anonymity, and assessments of unforeseen systemic risks to the EU, the Commission should encourage the data providers to grant access to the data in the spirit of open data principles.

## Detailed comments

These comments elaborate the above recommendations, and are aimed at further strengthening the delegated regulation and enabling robust, rigorous research of high quality and relevance.

## Section II Information and contact obligations

### Article 3 DSA Data access portal

As part of the DSA data access portal and the process to vet the researchers, Europe TPC encourages the European Commission to consider the use of existing repositories / databases of researcher and institutional identification such as Open Researcher and Contributor ID (ORCID)<sup>10</sup>, Research Organization Registry (ROR)<sup>11</sup> or member-state specific systems such as Digital Author Identifier (DAI)<sup>12</sup> to enable researchers to establish their credentials in an expedient yet effective manner.

Article 3.6 states that the DSA data access portal will consist of a publicly accessible interface. As part of such an interface, a description of the type of content should be made available on the public dashboards in plain language to ensure wider reach and accessibility of the content.

### Article 5 Processing of personal data in the DSA data access portal

As part of Article 5.2(c), further consideration should be paid to ensure that the logs about the flow and movements of the exchanged data are immutable and trustworthy. Distributed ledger technologies (including blockchain) may prove to be of consequence to ensure immutability and trustworthiness of the logs.

Article 5.3 states that the storage of data will use information technology infrastructure located within the European Economic Area. Europe TPC recommends that any such infrastructure be required to

---

<sup>10</sup> <https://orcid.org/>

<sup>11</sup> <https://ror.org/>

<sup>12</sup> <https://wiki.surfnet.nl/display/standards/DAI>

comply with appropriate safeguards and security standards (including but not limited to SOC 2<sup>13</sup>, ISO/IEC 27001:2022<sup>14</sup>, and the NIST Cybersecurity Framework 2.0<sup>15</sup>).

#### Article 6 Points of contact and public information on the data access process

Pursuant to Articles 6.2, 6.3, and 6.4, Europe TPC recommends the Commission to consider publishing an 'overview of data inventory, services, and examples of available datasets with brief descriptions' for each data provider. For the benefit of future researchers, this dashboard could also incorporate information on completed research linked to data providers, along with brief descriptions of the data accessed. An inventory of any research based on DSA data, and a public Application Programming Interface (API) which would enable third parties to process and analyse the information provided as part of the public dashboard would further strengthen the public dashboard's functionality.

As part of establishing points of contact and public information on the data access process, Europe TPC recommends the Commission to devise a process of escalation including the line-of-responsibility for identifying any breaches of data or security breaches to the data access portal. Such a process should include provisions for enabling members of the public or researchers not directly engaged in the use of the data access portal to identify such breaches in a discreet and confidential manner with the express intent of minimising the risk of public disclosure of either the data covered in the reasoned request or any security vulnerabilities in the data access portal.

### **Section III Requirements for formulating and assessing reasoned requests**

The time constraints imposed on assessing reasoned requests should be periodically reviewed to ensure that these are operationally feasible so that the reasoned requests are efficiently and expediently processed. Additionally, the Commission should also consider a mechanism to differentiate the requests, and to ensure that requests which are closely aligned with the core objective of identifying systemic risks to the Union are prioritised.

The current descriptions of processes for assessing access modalities, reasoned requests, and amendment requests appear to be reliant on the discretion of DSCs concerned. Europe TPC recommends that specific objective criteria are identified to make the implementation of these processes consistent across all member states.

#### Article 7 Formulation of reasoned request

The current description of the process of granting extensions for "duly justified cases" lacks any specification of time-constraints. Europe TPC suggests that in due time, the time constraints are identified to ensure that the implementation of this process is consistent with other parts of the draft act.

#### Article 8 Prerequisites for formulating a reasoned request

As part of the summary of the data access application, Europe TPC suggests that the researchers are required to provide details of how the datasets will be managed. For example, depending on the nature of reasoned request, such details may include (but not be limited to) strategies for imputation, detecting/managing outliers, managing bias-variance trade-offs, approach to 'fit' the data, any planned hypothesis testing, and the approach to model training and validation. The goal in requesting such details would not be to judge the validity of the planned approach itself or hold the researchers to account on the basis of the description provided. The aim in seeking such details would be to establish the intent and planned utility of the data access application in clear, research-specific terms, and how it can lead to the identification of systemic risks. In the long-term, such a provision could

---

<sup>13</sup> <https://www.aicpa-cima.com/topic/audit-assurance/audit-and-assurance-greater-than-soc-2>

<sup>14</sup> <https://www.iso.org/standard/27001>

<sup>15</sup> <https://www.nist.gov/cyberframework>

enable the Commission to establish clear boundaries vis-à-vis the use cases for the datasets. This will also provide researchers the necessary freedom to analyse the datasets within such boundaries.

Article 8 (6) can serve as a ground for (hypothetically) arbitrary dismissal of the request and therefore should be clarified to minimise any perceived ambiguity. The information described by the principal researcher, pursuant to Article 8 Points (4), (5), (7) and (9) should serve as a viable justification of the request - specifying (inter alia), research project, systematic risks or mitigations measures and proposed safeguards. The requirement to provide evidence of the inability to obtain data from alternative sources has two drawbacks. Firstly, it may be challenging to provide reasonable evidence that different measures could not collect the requested data. Secondly, it could prohibit the use of the data sharing mechanism based on Article 40 of Regulation (EU) 2022/2065 - which seems contrary to the general assumptions of that mechanism. Europe TPC suggests that the Commission focus on the requirements concerned with the general research proposal and assessment of the outlined security safeguards rather than introduce requirements for proving that the data cannot be obtained from different sources.

#### Article 9 Appropriate access modalities

Determination of appropriate access modalities as per Article 9 appears reliant on the discretion of the DSCs. Europe TPC recommends that clearer criteria are established to ensure that the DSCs' approach is consistent across member states.

#### Article 10 Content of the reasoned request

Aligned with the stated aims of Articles 9 and 10, Europe TPC recommends that there should be provision in the 'reasoned request' to consider the 'ethics' of the proposed research (including for example, how it would not lead to any online or societal harm).

#### Article 13 Dispute settlement procedure

The dispute settlement procedure, as presented in Article 13 of the draft act, could be further enhanced. Below are two significant concerns:

1) The request for mediation shall include the identity and contact details of the proposed mediator. Subsequently, the DSC of the establishment shall verify that the proposed mediator meets the requirements of impartiality and independence and holds appropriate qualifications. It could be beneficial for both parties if there were an existing list of mediators that fulfil the criteria outlined in Article 13(2) and Article 13(3). The existence of such a list could help strengthen the success rate of the mediations conducted under Article 13 of the draft act. In the absence of the personal identification of the mediator in the request for mediation, parties could select a mediator from the list of qualified mediators. Such practices have been routinely and successfully used in commercial mediation. The draft act should consider such (or a similar) set of practices regarding the selection of the mediator.

2) According to Article 13(4) of the draft act, the data provider shall be solely responsible for covering the costs of the mediation. This provision, combined with Article 13(6), which states that the principal researcher should act - only where relevant - in good faith - could amount to a situation where the mediation process is requested without a just reason - leading to a time-consuming and expensive process. Europe TPC suggests that the Commission incorporate an exception in Article 13(4), according to which the exemption of costs will apply only to the principal researcher who requested the mediation in good faith. Alternatively, the Commission could adopt a different phrasing in Article 13(6) - requiring the principal researcher to act in good faith throughout the mediation (without limiting the scope of applicability by 'where relevant' wording).

## **Section IV Conditions for providing the data requested to vetted researchers**

### *Article 15 Data format and data documentation*

Subject to the stated objectives of the reasoned request, Europe TPC recommends the Commission to encourage and facilitate the data providers to grant access to the data in the spirit of open data principles.<sup>16</sup> Since the data may be commercially sensitive, such provision of open data should be at the discretion of the data provider. Any provision of open data would also need to consider security implications for the VLOP and/or VLOSE in question, and end-user anonymity implications depending on the relevant business-to-consumer (B2C) or business-to-business (B2B) use cases it covers. The Commission should consider whether any open data provision would pose unforeseen systemic risks to the Union and whether the open data provision should take place after a time-delay in consultation with the data provider and the researchers.<sup>17</sup>

Given that data providers “shall be allowed to limit vetted researchers’ use of standard analytical tools, including relevant software libraries,” Europe TPC recommends that provisions are made in the DSA data access portal to enable researchers to understand the processing architecture and algorithmic components of the VLOPs and VLOSEs acting as data providers subject to privacy and security considerations.<sup>18</sup>

Generation of synthetic data is a key, well-established practice as part of most of the current-day data science, computational statistics, and machine learning research projects. Used correctly, such a practice is a proven mechanism for training and testing models in a time-bound and expedient manner. As part of the documentation provided by data providers to vetted researchers, a clear indication should be provided by data providers where synthetic data is included (including, for example, for imputation purposes, or for reasons of security, confidentiality, or end-user anonymity).

Given the large quantities of data generated by most VLOPs and VLOSEs on a daily basis, the data providers should be required to identify the age of the data in accordance with the timelines for the submission of the data access request and details requested by the researchers.

Data pre-processing is often an essential and yet time-consuming part of most data-driven research projects requiring 60% to 80% of the data scientists and researchers’ time.<sup>19</sup> Given the technical and time-debt this can create in any research project and the challenges it can pose to timely identification of systemic risks to the Union, Europe TPC recommends that the data providers are required to include provisional estimates for how ‘clean’ the data is and provide possible indications of the extent of pre-processing required to make the data effectively useful for the planned research purposes. Europe TPC emphasises that this recommendation does not mandate the data providers to provide ‘clean’ or pre-processed data. Subject to the established intent of the reasoned request, raw data can prove to be an asset to a research project, and this recommendation does not preclude the provision of raw data.

To facilitate optimal use of the provisions of Article 40 for data access and enable timely research of systemic risks, the Commission should consider the inclusion of a list of indicators for Data Readiness Levels (similar to the concept of Technology Readiness Levels<sup>20</sup>) on the DSA data access portal. When

---

<sup>16</sup> <https://data.europa.eu/en/dataeuropa-academy/what-open-data>; and <https://opendatacharter.org/principles/>

<sup>17</sup> This reiterates ACM Europe TPC’s recommendation on the use of best practices for open data as part of the Call for evidence on Delegated regulation on data access provided for in the Digital Services act. See <https://www.acm.org/binaries/content/assets/public-policy/acm-europe-tpc-dsa-comments.pdf>.

<sup>18</sup> This reiterates ACM Europe TPC’s prior recommendation submitted as part of the Call for evidence on Delegated regulation on data access provided for in the Digital Services act. See <https://www.acm.org/binaries/content/assets/public-policy/acm-europe-tpc-dsa-comments.pdf>.

<sup>19</sup> <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>

<sup>20</sup> See: [http://www.artemisinnovation.com/images/TRL\\_White\\_Paper\\_2004-Edited.pdf](http://www.artemisinnovation.com/images/TRL_White_Paper_2004-Edited.pdf)

considered in conjunction with the time commitments required by data pre-processing in a research project, such a list of indicators could potentially enable the Commission to evaluate not only the extent to which the data providers fulfilled the data access request effectively but also the extent to which the researchers could deliver their planned research objectives.