DIGITAL SERVICES ACT (DSA)

Feedback to the Call for Evidence on Data Access Under the DSA

May 2023

Introduction

The Computer & Communications Industry Association (CCIA Europe) is pleased to participate in the call for evidence of the European Commission on the Delegated Regulation on data access[1] provided for in the Digital Services Act (DSA).[2] Article 40 of the DSA lays down the framework that allows authorities and researchers to access data from very large online platforms (VLOPs) and very large online search engines (VLOSEs). The upcoming delegated act will have to outline the technical and detailed conditions and procedures required for accessing such data, with a view to respecting data protection rules as well as the rights and interests of VLOPs and VLOSEs.

Below you will find CCIA Europe's contribution to the main elements raised in the call for evidence:

I. Data access needs
II. Data access application and procedure
III. Data access formats and involvement of researchers
IV. Access to publicly-accessible data

I. Data access needs

1. Types of data (question 1a)

As a preliminary remark, the delegated act should differentiate between the types of data useful for the Digital Services Coordinators (DSCs) and vetted researchers, in line with Article 40 of the DSA. While DSCs can ask for data to verify compliance with the DSA, vetted researchers have a more limited scope, i.e. research on systemic risks and mitigation measures.

The DSCs will already have access to a substantial amount of data thanks to the transparency reporting obligations, risk assessments, and mitigation measures contained in the DSA. DSCs' data-access requests should only cover datasets to which they do not already have access. Data-access requests should clearly state their purpose so that VLOPs and VLOSEs are able to present the most pertinent data. Requests from DSCs for granular

[1] European Commission, Have your say, Call for Evidence, Delegated Regulation on data access provided for in the Digital Services Act, consulted on 09 May 2023, available at: 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] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R2065&qid=1666857517641

data that come without a proper explanation of the situation at hand, or the context in which the data will be used, could end up resulting in a distorted view of the facts.

Vetted researchers will also already have access to a number of publicly-accessible documents (e.g. transparency reports or existing research). Any requests for existing or substantially similar data should therefore be avoided. The vetted researcher's data requests should also demonstrate how the research relates to the systemic risks or mitigation measures pursuant to Articles 34(1) and 35 of the DSA. To ensure that it is the case, the request should explain how the data will contribute to the research outcome and how the research itself contributes to the "detection, identification and understanding of systemic risks in the Union" as mentioned in Article 40.

The request should also assess the balance between the need for data on one hand, and the respect for personal data protection and confidential commercial information on the other. The latter should include any derived data that could expose business or financial information, or expose proprietary information pertinent to the running of products and operations. As part of such an assessment, the requests should explain why a dataset with a narrower scope would not be sufficient. Attention should be given to how data access requests are aligned with VLOPs and VLOSEs' new obligations in the NIS 2 Directive.[3] Among others, researchers should be required to disclose to the service provider any vulnerability found during the course of their research according to companies' procedures laid down in vulnerability handling disclosure programs. Where vulnerability handling programs do not allow third-party vetted researchers to report vulnerabilities, researchers should be required to contact the competent CSIRT in the Member State where the service provider has its main establishment, in line with Article 12(1) of the NIS 2 Directive.

Similarly, requests involving the disclosure of trade secrets should be aligned with existing rules and forthcoming proposals governing third-party access, including the Trade Secrets Directive and the Data Act proposal. In particular, VLOPs and VLOSEs should be able to decline access requests "in exceptional circumstances, when [VLOPs / VLOSEs] can demonstrate that they are highly likely to suffer serious damages" from such access. This should include situations where researchers failed to demonstrate that they have taken appropriate technical and organisational measures to preserve the confidentiality and integrity of the trade secrets.[4] Should researchers choose to contest a refusal before the competent authority, trade secrets should remain protected in the course of dispute resolution proceedings. In doing so, preservation measures under Article 9 of the Trade Secrets Directive should be extended to such proceedings.[5]

In both cases, confidential information should be excluded from the requests of DSCs and vetted researchers. Some elements or documents should be deemed confidential by

[3] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), available at: https://eur-lex.europa.eu/eli/dir/2022/2555/oj

[4] Similar to Council amendments on Articles 4(3a) and 5(8a), available at: https://data.consilium.europa.eu/doc/document/ST-7413-2023-INIT/en/pdf. Similar language in Articles 4(3) and 5(8) of the European Parliament report, available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0069_EN.html

[5] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016L0943

default. These should include but not be limited to source code, privileged data, contractually protected data, data necessary for the security or integrity of the platform, machine learning or other algorithmic model coefficients, and internal documentation (e.g. decisions or memos).

Further guidance on the types of data in the scope of Article 40(4) and (12) is needed. Indeed, there are fundamental differences in the safeguards associated with each of these categories, i.e. vetted researchers and other researchers. As it stands, the scope of both categories seems to be very similar: all researchers can request to access data on risks - while Article 40(12) does not refer to mitigation. Some clarity could also be brought to the concepts of data "publicly accessible" and data requested by vetted researchers. The delegated act should also clarify how VLOPs and VLOSEs are expected to assess compliance with the criteria under Article 40(12), and to what extent DSCs or independent third parties may contribute to that assessment.

Considering the already high amount of data available to researchers, it is key that the requests pursuant to Article 40 of the DSA are limited at least to the following aspects:

- The data exists: it would neither be reasonable nor proportionate to require platforms to change the way they operate their business for the sake of producing data that does not exist or is not collected.
- The data is relevant to systemic risks under the DSA: overly broad data requests from vetted researchers (only to decide at a later stage which subset is relevant) would go beyond the intended scope of this provision.
- The data does not expose vulnerability to malicious actors: it is key that the data shared with vetted researchers does not increase the risk related to a specific risk or create a new one, thus exposing potential vulnerabilities in the systems.
- The sharing of data does not conflict with other regulatory frameworks or legal requirements, or it leads to the collection of sensitive data where companies should not be required to do so.

2. Types of analysis and research (question 1b)

DSCs and vetted researchers should be encouraged to respect best practices in order to ensure that their work reflects state-of-the-art analysis. Analysis of VLOPs/VLOSEs could include correlation statistical techniques, classification of models, machine learning or qualitative coding.

II. Data access application and procedure

1. Vetting process (question 2a)

Digital Services Coordinators (DSCs) will be tasked with assessing the applications of researchers wishing to become vetted researchers under the DSA (Article 40(4) and (8)). That is why the delegated act should provide a comprehensive definition of the decision-making processes and factors that DSCs are required to observe in this respect.

An alternative could be to mandate the Member States to establish unambiguous and transparent national procedures with the force of law. These procedures should be subject to appropriate and effective mechanisms for administrative law challenges to ensure

accountability and fairness. If so, DSCs should be required to publish such procedures in advance of their entry into force and consult stakeholders about them.

The exchange of information on vetted researchers and other researchers between DSCs will be crucial to ensure that requests are valid and that researchers whose access has been revoked cannot request data. This information should also be made available to providers, especially in the case of researchers appointed under Article 40(12). Moreover, a mechanism should be introduced that allows providers to request clarifications or raise concerns about a vetting process. Compliance auditing on the vetting of researchers could also inform if the process is well-designed, or else suggest improvements.

2. Consistency of the vetting process (question 2b)

One of the challenges in putting this into practice will be ensuring there is a consistent process and application of the criteria listed in Article 40(8). Further guidance from the European Commission on the application of the criteria would be welcomed, e.g. through standards. This guidance should help to prevent DSCs from having divergent interpretations of what constitutes a "substantiated application". An oversight mechanism could also ensure that all vetted researchers and researchers verified according to Article 40(9) are trustworthy. Consistency could also be ensured through the coordination of an independent advisory body.

Consistency will also be a challenge in substance. For example, the interpretation of what constitutes a "systemic risk" might vary. To that end, the Commission could include a provision in the delegated act setting out that the vetting process refers directly to the DSA, and in that instance to Article 34(1).

The delegated act should also include provisions that allow for meaningful input from VLOPs and VLOSEs in all decision-making processes of DSCs, prior to the finalisation of relevant decisions. This would contribute to ensuring the consistency of the application of Article 40 across Member States. This should encompass granting the status of vetted researcher, conducting assessments of researchers as specified in Article 40(9), and considering any data access requests submitted.

3. Protection of users' and businesses' rights (question 2c)

To ensure a correct balance between the data access right of Article 40 and the protection of users' and businesses' rights, the delegated act should clearly state that data shared with researchers should be strictly limited to the minimum necessary and pertinent to the research outcome. This would be in line with data minimisation requirements in the General Data Protection Regulation[6]. The delegated act should also clearly state that researchers cannot use or describe data in a way that would undermine data protection, confidentiality, trade secrets, security, or the functioning of an online platform. It should therefore further limit the use of data to the purpose identified in the application.

[6] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj

In order to safeguard the rights of citizens and businesses alike, the delegated act should explicitly state that the publication of confidential company data or personal data of users is strictly prohibited. Failure to do so could result in the exploitation of such data by malicious actors, thereby undermining content moderation efforts in certain cases. Providers should be given the opportunity to provide comments on any research publication that utilises their data and should have the ability to seek legal remedies against such publications if they suspect any misuse. Furthermore, data access requests should be strictly limited to the specific data requested and must not be allowed to evolve into fishing expeditions seeking unrelated information.

The dialogue between providers and researchers, e.g. through an independent advisory mechanism, would help to ensure that access requests are well framed and the correct level of aggregation is provided. Besides, as mentioned previously with regard to the types of data, non-exhaustive examples of confidential information that cannot be disclosed would be useful.

The amendment of the data request foreseen in Article 40(5) is an important part of the process. However, Article 40(5) only addresses the lack of data, vulnerabilities in the security of the service, or the protection of confidential information. Additionally, providers should be able to flag if the request is not within the scope of detection, identification, and/or understanding of systemic risks and the assessment of mitigation measures. This would serve as an additional check to ensure consistency across DSCs and limit abuse.

The delegated act should also give more precisions on the timeframes VLOPs and VLOSEs will have to give access to data to researchers and DSCs. Article 40 already states that the data access should be done within a “reasonable period”. The delegated act could clarify that this period will be decided on a case-by-case basis. Flexibility should be foreseen as long as providers make reasonable efforts to meet the timeframe and justify potential delays. Rushing data access requests to meet arbitrary timeframes could be harmful to the protection of users’ and businesses’ rights.

In case of a data breach (e.g. loss or unauthorised disclosure) or other misuses, the delegated act should clarify that the respective researcher or DSC will be held liable, when necessary, for the data they have requested. The list of damage should cover at least users’ privacy breaches and harm to providers (e.g. financial losses or competitive harm).

4. Safeguards against abuses (question 2d)

Several safeguards should be put in place to assure that data is used for the purposes of Article 40 and in order to minimise the risk of abuse. To that end, the delegated act should clearly articulate these safeguards. General safeguards should include the following:

- All decisions made by DSCs within the framework of the data access regimes should be detailed, reasoned, and public (subject to redaction for the protection of users' and businesses’ rights), so as to enable challenge by providers.
- DSCs should monitor and confirm compliance by vetted researchers with the terms on which any request for access has been granted.
- DSCs should be liable if failures in their vetting procedures or monitoring result in loss of security over data to which access is given and consequent losses or costs to VLOPs/VLOSEs.

- Vetted researchers should be liable for any loss of security over data to which access is given which results in losses or costs to VLOPs/VLOSEs. This can be addressed by insurance, so that this clarification only operates as an appropriate control on overbroad and unnecessary requests.
- In the event of misuse of data, or of failure to make full disclosures in connection with the grant of vetted-researcher status, vetted researchers and the institutions with which they are affiliated should face a range of penalties, including restrictions on future access, exclusion from future vetted-researcher status, exclusion from future EU funding, and in last recourse, fines.
- The delegated act should require meaningful disclosure from researchers applying for vetted-researcher status in relation to their independence from commercial interests, their funding (including the source, nature, scope and conditions of such funding, going beyond the research in question and extending to the funding of the wider research of the researcher or institution), any "affiliation" with research institutions, and the technical and organisational measures they intend to deploy to ensure data protection and data security.
- Researchers should be required to give specific undertakings of non-abuse, refraining from attempting to identify users, and non-combination with other data sets, as part of the grant of access to data.
- VLOPs/VLOSEs should be able to request an investigation into researcher non-compliance, with a view to securing termination of access or vetted-researcher status. Under Article 40(10), third parties able to contribute to the investigation should include providers.
- Data retention limits and data destruction requirements are necessary safeguards, as are limitations making express that the data cannot be shared with any person not vetted. When destruction occurs, notice should be given to DSCs and concerned providers. Specific and appropriate measures should be taken regarding security and privacy.

Some safeguards could be useful on a case-by-case basis and should be mentioned in the delegated act. When vetted researchers collect or host data directly, increased safeguards should be required. Such safeguards should include, among others, limited network access, access controls, encryption, and restrictions on creating copies. It may be appropriate to use access control software that maintains audit logs in case of a breach.

Vetted researchers should also not be permitted to transfer the data outside of the jurisdiction of the EU, as this could result in a loss of control by the researcher and the EU over the data and any accompanying breaches. The pre-publication review of research, especially on custom datasets, could assure that the research outcome relates to the data.

Finally, safeguards can be built within the data access interfaces (such as virtual cleanrooms, physical cleanrooms, CSV files, or APIs). VLOPs and VLOSEs should be able to require registration and to monitor the activity on any data access interface as this could prevent abuse, especially in the case of sensitive data.

5. Independent advisory mechanisms (questions 2e)

An independent advisory body could help with assessing data requests and vetting researchers. The body's independence from providers and researchers would ensure that the balance between data access rights and the rights of users and businesses is protected.

Sufficient expertise should be expected, both in terms of academic practices and EU law (especially data protection), but also in technical and scientific experience, as recommended in the European Digital Media Observatory's report on data access.[7]

This independent advisory body would support the work of DSCs, as they could rely on it to pre-approve researchers and requests, regardless of the DSC's capacity or expertise. As mentioned before, the body could help to ensure that the criteria for vetting researchers are consistently applied in all Member States.

The independent advisory body could serve as a platform to facilitate an informal dialogue between DSCs, researchers, and online platforms – e.g. before a formal reasoned request is submitted to a provider. Likewise, this body could help with managing data requests, by helping researchers check if the data is already available or to better frame their needs prior to filing a formal request. A discussion on a pending request between providers and researchers would help identify relevant and proportionate data sets for the research goals. In this respect, the body could help with the prioritisation of data requests, avoiding duplication (by proposing joint projects instead), and verifying the proportionality of the requests. This would help with the consistency of the assessment of the request regarding the scope of Article 40, i.e. systemic risks and mitigation measures.

The body could also play a role in assessing how the data requests respect the rights of users and businesses, and if the safeguards are appropriate. For online platforms, this would ensure that requests are feasible in time and substance, as well as taking into account the technical aspects. The body would also be able to engage with platforms in advance, so that all relevant details are agreed upon before the 15-day timeframe starts.

Overall, such a process and platform for dialogue would make the decision-making more efficient, benefiting all parties. Indeed, the short timeframes for amending the request under Article 40(5) could lead to hasty decisions and many objections, which would defeat the purpose of this new data access right and potentially endanger the rights of users and companies.

III. Data access formats and involvement of researchers

1. Data access interfaces (question 3a)

The delegated act should state that multiple technical solutions may be used for setting up a data access interface, which should be determined on a case-by-case basis depending on the provider and the requested data. The technical specifications for data access interfaces should be adapted to the risk level associated with the data accessed for research projects.

High-level guidance on how to assess the risk and how to mitigate it through an appropriate data access interface would be useful. Defining best practices, which should be updated on a regular basis if needed, would support the effort of providers and researchers to protect users' and businesses' rights. However, the delegated act should refrain from a prescriptive

[7] European Digital Media Observatory, Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access, 31 May 2022, available here: https://edmoprod.wpengine.com/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf

approach as technical specifications are likely to change as technologies continue to develop. Imposing specific technology choices would prevent the development of more privacy-preserving data access interfaces and technologies in the future. Similarly, the reasoned requests should refrain from specifying which interface is the most appropriate.

The example of virtual cleanrooms previously mentioned could be a pertinent safeguard in some cases where sensitive data is accessed, as they allow data to be disconnected from other sources and the recording of all activity in case of abuse. Likewise, if particularly sensitive data is accessed, a physical cleanroom may also be a pertinent safeguard, as it ensures that the data is not transmitted externally and that copies are not inadvertently created, thus significantly reducing the risk of data breaches.

2. Capacity building measures (question 3b)

The earlier-mentioned independent advisory body could also play an important role in creating a network of researchers and providing regular training. Such training programmes could cover a number of important topics, e.g. how to fill out a vetted-researcher application, how to use existing datasets, or how to respect data protection rules as part of a data request.

As mentioned, the delegated act should also try to provide further guidance on the concepts of “publicly accessible” data and “systemic risks”. This guidance would help researchers to better exercise their data access rights.

3. Standard data glossary (question 3c)

Where feasible given the technical variations and other factors that differ among VLOPs and VLOSEs, a common and precise glossary of language used by all parties involved may be helpful in the application, review, and amendment processes under Article 40. While still allowing for variations across the different VLOPs and VLOSEs, this glossary could ensure that DSCs, vetted researchers, and providers can communicate clearly, avoid misunderstandings, and improve the efficiency of the data request process. The independent advisory body could play a leading role in exploring whether, and in what contexts, such a glossary makes sense and/or is feasible. To that end, the body could consult both researchers and online platforms, taking into account the variations between VLOPs and VLOSEs and the changing nature of their services. To the extent that 'comprehensive' glossaries are suggested, it seems unlikely that such an effort would be practical – or useful – given the significant variations in services and their constantly-evolving nature.

IV. Access to publicly-accessible data (question 4a)

As made clear throughout this contribution to the call for evidence, the delegated act should primarily aim to bring more clarity and guidance on several key concepts in order to ensure access to publicly-accessible data. Indeed, Article 40 does not sufficiently differentiate between the status of “vetted researchers” and “researchers” assessed under paragraph nine. The vetting process foreseen in this paragraph should be further clarified. Similarly, the delegated act should also provide guidance on what exactly constitutes “publicly-accessible” data in this context.

Conclusion

CCIA Europe appreciates the European Commission's efforts to consult stakeholders before drafting the delegated act on data access under the DSA. Our suggestions aim to strike a balance between enabling authorities and researchers to access the data of VLOPs and VLOSEs, while also upholding the rights of users and businesses. We remain available to further discuss our feedback with the Commission.

About CCIA Europe

The Computer & Communications Industry Association (CCIA) is an international, not-for-profit association representing a broad cross section of computer, communications, and Internet industry firms.

As an advocate for a thriving European digital economy, CCIA Europe has been actively contributing to EU policy making since 2009. CCIA's Brussels-based team seeks to improve understanding of our industry and share the tech sector's collective expertise, with a view to fostering balanced and well-informed policy making in Europe.

For more information, visit: twitter.com/CCIAeurope or www.ccianet.org

For more information, please contact:

CCIA Europe's Head of Communications, Kasper Peters: kpeters@ccianet.org