cerre Centre on Regulation in Europe

ACCESS TO DATA AND ALGPORITHMS: FOR AN EFFECTIVE DMA AND DSA IMPLEMENTATION

REPORT

March 2023

Laura Edelson

Inge Graef

Filippo Lancieri

As provided for in CERRE's bylaws and procedural rules from its "Transparency & Independence Policy", all CERRE research projects and reports are completed in accordance with the strictest academic independence.

The project, within the framework of which this report has been prepared, received the support and/or input of the following CERRE member organisations: OFCOM, ARCOM, Google, Booking.com, and TikTok. However, they bear no responsibility for the contents of this report. The views expressed in this CERRE report are attributable only to the authors in a personal capacity and not to any institution with which they are associated. In addition, they do not necessarily correspond either to those of CERRE, or of any sponsor or of members of CERRE.

SUGGESTED CITATION: Edelson, Laura; Graef, Inge & Lancieri, Filippo. "Access to Data and Algorithms: For an Effective DMA and DSA Implementation" (CERRE, March 2023), available at https://cerre.eu/publications/access-to-data-and-algorithms-for-an-effective-dma-and-dsa-implementation

info@cerre.eu – www.cerre.eu

TABLE OF CONTENTS

ABOUT CERRE

Providing top-quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) promotes robust and consistent regulation in Europe's network and digital industries. CERRE's members are regulatory authorities and operators in those industries as well as universities.

CERRE's added value is based on:
Its original, multidisciplinary and cross-sector approach –

- The widely acknowledged academic credentials and policy experience of its team and associated staff members;
- Its scientific independence and impartiality;
- The direct relevance and timeliness of its contributions to the policy and regulatory development process applicable to network industries and the markets for their services.

CERRE's activities include contributions to the development of norms, standards and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological and social environment. CERRE's work also aims at clarifying the respective roles of market operators, governments and regulatory authorities, as well as at strengthening the expertise of the latter, since in many Member States, regulators are part of a relatively recent profession.

ABOUT THE AUTHORS[1]

Laura Edelson is a Postdoctoral Researcher with Cybersecurity for Democracy at NYU's Tandon School of Engineering. Laura studies online political communication and develops methods to identify inauthentic content and activity. Her research has powered reporting on social media ad spending in the New York Times, the Wall Street Journal and the Atlantic. Prior to her current time in academia, Laura was a software engineer for Palantir and Factset. During her time in industry, her work focused on applied machine learning and big data.

Inge Graef is an Associate Professor at Tilburg University. She is affiliated to the Tilburg Institute for Law, Technology, and Society (TILT) and the Tilburg Law and Economics Center (TILEC). Inge holds expertise in the areas of competition law, platform regulation and the governance of data-driven innovation. She is also appointed as a member of the European Commission's expert group to the EU Observatory on the Online Platform Economy.

Filippo Lancieri is a Postdoctoral Researcher at the ETH Zurich Center for Law & Economics and a Research Fellow at the Stigler Center at the University of Chicago Booth School of Business. Filippo's work focuses on the challenges associated with the development of a cohesive regulatory framework for the digital economy. His focus is mostly on understanding the law in action: how different policies and enforcement actions can change behavior on the ground.

[1] Author listed alphabetically per last name.

EXECUTIVE SUMMARY

The Digital Markets Act ('DMA') and the Digital Services Act ('DSA') include a range of obligations to enable data access and transparency for a variety of important objectives: from verifying legal compliance, to increasing market contestability, to enabling a better understanding of how algorithms and advertising systems impact our societies. Yet, these obligations are scattered throughout a long and intricate web of Recitals and Articles. Even more importantly, many of these obligations are significant legal and policy innovations — meaning that regulators, companies and civil society more broadly must develop new processes to ensure their adequate implementation.

This report addresses some of these challenges. We start with a careful mapping and categorisation of all 54 algorithmic transparency and data-sharing obligations that are present in the DMA/DSA package. We then select three of them as case studies: access to online advertisement databases (Article 39 of the DSA), access to data for vetted researchers (Article 40(4) of the DSA), and sharing of click and query data between search engines (Article 6(11) of the DMA). For these selected obligations, the report outlines practical and legal challenges that the involved parties will face when implementing the legal commands, and it provides a combination of legal and technical measures that can help overcome many (though not all) of these challenges. The report concludes with a brief look ahead, focused on understanding whether the categorisation exercise proposed herein can be useful for future European Union ('EU') Regulations such as the draft Artificial Intelligence ('AI') and Data Acts.

More specifically, this report is divided into two main parts, a concluding look ahead, and two technical Annexes:

Part I: Categorising Transparency Obligations in the DMA and the DSA

Part II: Mapping Out Challenges in Three Specific Case Studies

Looking Ahead: Categorising the Proposed Transparency and Data Access Obligations in the AI and Data Acts

Annex I: A Comprehensive Summary of Transparency and Data Access Obligations in the DSA and DMA

Annex II: Fields for an Advertisement Transparency Database

The proper implementation of the novel transparency and data access obligations present in the DMA and the DSA will require close co-operation within and across many fields. We envisioned this report as a cross-disciplinary collaboration that integrates lessons from law, economics and computer sciences.

Categorising transparency and data access obligations

Part I builds on our mapping of all 54 transparency and data access obligations in the DMA and the DSA (**see Annex I to this report**), grouping them in four categories that share important legal and practical characteristics:

1) Regulator Access to Data, which establishes/expands regulatory powers to require access to private data for investigations and other public purposes;

2) Private Party Access to Data, which requires private parties to share data with other previously defined private parties;

3) General Public Access to Data, which requires private parties to provide the general public with certain types of data on their private activities; and

4) Regulatory Transparency, which requires regulators to provide the general public with certain types of data on their public activities.

These are built on five key variables: the target party, the target data, the receiving party, the timeliness of data access (for example, how frequently data access needs to be provided), and the mode of access (for example, the format in which data needs to be provided). Some of these variables, in particular the target and receiving party, are clearly described in the legal provisions, whereas others require further interpretation–for instance regarding the exact scope of the data covered and the timeliness and mode of data access. Beyond this, we also discuss the role of privacy, intellectual property protection, security, and rule of law guarantees in implementing the provisions.

We believe grouping obligations in these four categories can help guide implementation, in particular, because they share common characteristics that can help balance potential conflicts of interest. For instance, while claims regarding trade secret protection can be strong in the context of private party access to data, there is much less room to invoke trade secret protection against disclosure of information to regulators; while rule of law guarantees will be the most relevant for regulator access to data obligations where an enforcer gathers information in order to investigate possible infringements of the rules and to establish potential liability.

A general framework to balance conflicting interests

Indeed, a proportional balance of conflicting interests will be a key challenge in the implementation of the identified data access provisions. To help facilitate this, Part 2.1 develops a balancing framework centred around three key principles:

1) **Legislative purpose as the guiding principle:** the data access request should fit the purpose of the applicable legal obligation and any potential harms invoked by a party to prevent access

to the data should be recognised in the legal framework as relevant harms that need to be weighed against the interest in disclosure;[2]

2) **Data minimisation:** the requested data should fit the stated purpose of the disclosure and should not go beyond what is necessary to achieve that purpose in terms of the scope (the amount and range) and nature (the type) of the data;[3] and

3) **Least intrusive implementation:** the data should be shared in a way that still meets the purpose of the request and the legal obligation while minimising any potential harms, requiring an effort to find the least intrusive manner to implement the request.

This is not to say that all requests should be granted. There will be cases where the disclosure will require important legal and technical safeguards, while others will lead to irreconcilable conflicts between the interest of the receiving party and the interest of the target party. In that case, the relevant enforcer/authority in charge needs to decide whose interest prevails.

Based on these three principles, we develop a seven-step test which can help parties consider specific requests. They are summarised by the flowchart below. A particular challenge is with regards to step seven, the balancing of irreconcilable conflicting interests. In such a case, we believe the decision should be guided by a weighing of the importance of the data access request for achieving the stated purposes of the relevant legal obligation against the strength of the claims for protecting the commercial interests of the target party or of other third parties.[4] This means that there will be situations where it is justified to deny access to specific sets of data.

Below we discuss in more detail what principles should guide the balancing of potential conflicts between data access on one hand, and privacy, intellectual property protection, information security and the rule of law on the other. Before that, we note that while each of the parties has a responsibility to enable data access to the extent necessary and possible,[5] target parties should be particularly proactive and co-operative to facilitate the process of implementation. In particular, target parties must explain and motivate in detail the extent to which the requested data interferes with their interests in commercial confidentiality, trade secret protection, information security or privacy, and under what conditions they believe they could provide access to the data. This will enable the

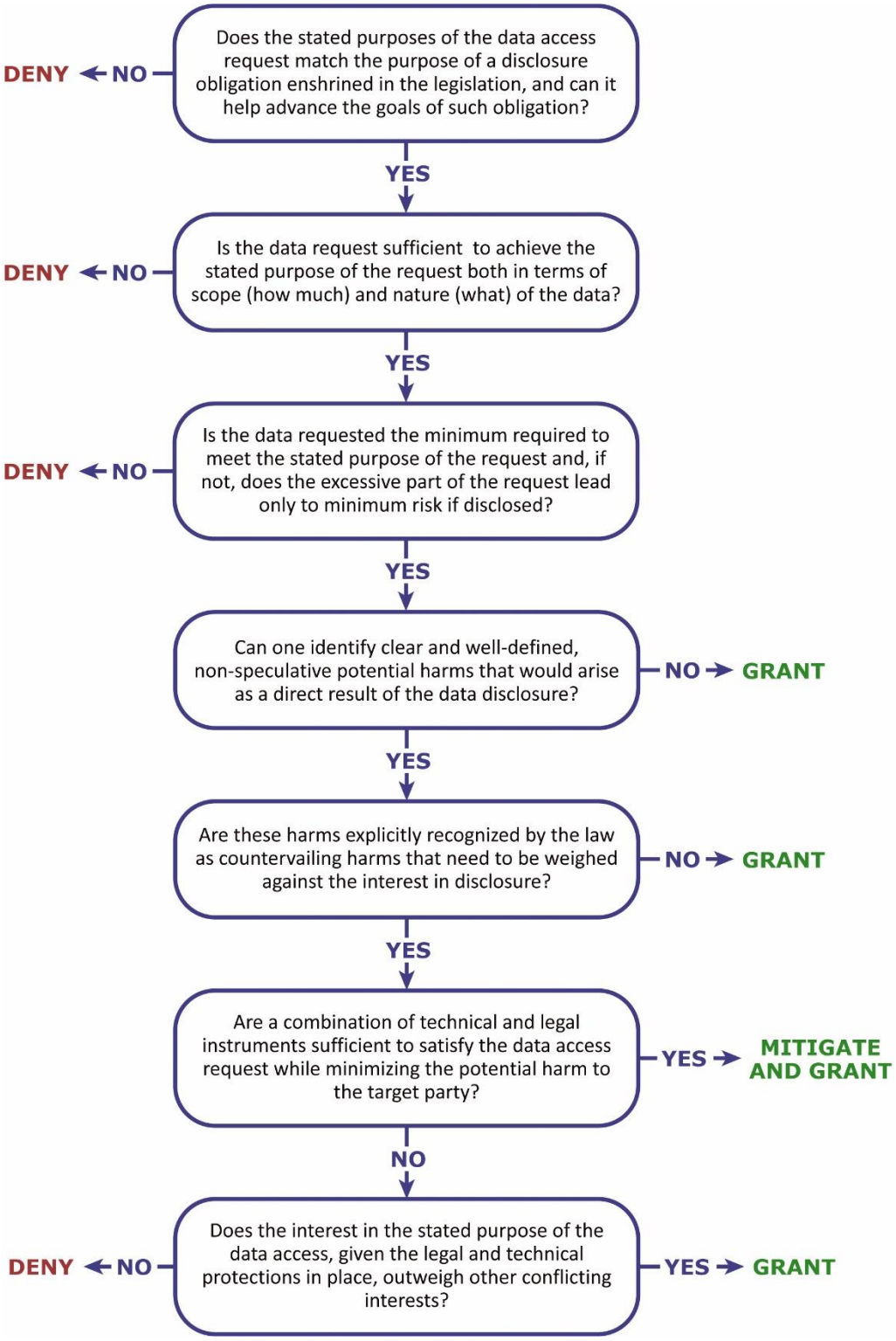
[2] This is a check conducted on the basis of the applicable legal provision. As an example, Article 40(4) of the DSA requires VLOPs and VLOSEs to provide researchers with access to data for the purpose of conducting research into systematic risks and Article 40(13) of the DSA refers to the protection of confidential information, trade secrets, and security as countervailing interests that need to be considered. This means that researchers can only request data relevant to assess systemic risks, while VLOPs and VLOSEs can only invoke the protection of confidential information, trade secrets, and security as interests against disclosure.

[3] This is a more factual check conducted on the basis of the data access request. For instance, requests for accessing data with the aim of developing or improving one's own search engine service (based on Article 6(11) of the DMA) should be restricted to search engine data that is necessary for achieving this purpose.

[4] For example, where the data access request is vital for achieving the underlying purpose of the legal obligation and the claims on the need to safeguard commercial confidentiality, trade secret protections or even user privacy are weak, it will be disproportionate to let the interest in non-disclosure prevail. However, where a dataset is highly sensitive and protected and the importance of data access for achieving the purpose of the relevant legal obligation is low, the interest in non-disclosure is a priority.

[5] And these will likely be a combination of technical and legal safeguards. For instance: offering access to an alternative, but still satisfactory dataset while also requiring the signing of a confidentiality agreement.

receiving parties and/or the relevant authorities to find possible ways to address concerns. We believe that placing this initial burden on the target parties is reasonable because of both the important information asymmetries in this area and the overall intention of the EU legislator to facilitate data access through the relevant provisions in the DSA and DMA.



General considerations on how to balance conflicts between data access and privacy, intellectual property, information security and the rule of law

This framework provides a general roadmap on how to balance the different applicable considerations without predefining which interest should prevail. The outcome of the balancing depends on the particular circumstances in a given case. To determine how to solve tensions between the various interests at stake, the report discusses in the three case studies in more detail how to balance data access and the protection of intellectual property, privacy, security, and rule of law.

With regard to data protection and privacy, the General Data Protection Regulation (GDPR) requires the existence of a lawful ground to process personal data. The sharing of personal data is a processing activity and therefore requires both the target party and the receiving party to have a lawful ground for personal data processing under the GDPR. Whereas target parties can likely rely on the relevant legal obligation in the DMA or DSA as a lawful ground (or rely on the exemption of Article 5(1)(b) of the GDPR when the data is shared for research purposes), the most appropriate lawful ground for receiving parties depends on the circumstances. It could be a legal obligation, public interest, legitimate interests of the data controller, or even consent of the data subject. The four categories of data access outlined herein can provide guidance on what the most suitable lawful basis is for the relevant receiving parties. Beyond the lawfulness of personal data processing, one also needs to pay attention to the necessity of processing, data minimisation, and purpose limitation. In general, data protection and privacy do not block data access as such, but require the implementation of adequate safeguards.

In terms of intellectual property, we focus on the protection of trade secrets and commercial confidentiality. Our analysis of existing case law, decision-making practice and policy documents in other areas shows that neither of these interests are absolute. We find that the nature of the receiving party, the type of data and the specificity of the legal obligation are particularly indicative of the room to claim protection of trade secrecy and commercial confidentiality against data access and transparency.

In the context of security concerns, we distinguish between data security and system security. We believe that legal or technical solutions can provide a proper balance between the need for data access and the interest in security. *Data security* relates to concerns that more data is disclosed than intended or to a wider audience than intended. To address data security concerns, we suggest the implementation of measures such as audits of data availability, robust authentication, the principle of least authorisation, and ensuring the proportionality of protection in relation to the sensitivity of the data. *System security* is at stake when the disclosure of data diminishes the overall security of the system by facilitating the gaming of algorithms and other infrastructure. To address system security concerns, we propose conducting a threat modelling exercise with the receiving party to determine the appropriate and necessary security requirements, and we outline how to conduct such exercise. This requires answering four relevant questions:

1) What would be the change in available information if the data was shared?
2) For which audience would this change be visible?
3) How likely is it that the data might leak to another audience?
4) Can the audience to whom information has been revealed use this information to cause physical, emotional, or financial harm to a well-defined group of individuals?

Finally, rule of law guarantees are particularly relevant for our category of regulator access to data, where the information obtained through provisions of the DMA and DSA can be used to monitor compliance and to establish the liability of a given platform. Relevant protections include an obligation of the relevant enforcer/authority in charge to state reasons for decisions taken during an investigation, a right to be heard, and a right to have access to the file. In addition, the exercise of investigation and enforcement powers needs to be proportionate to the expected harm of the possible infringement. This also entails that the least far-reaching investigation measure should be applied when different measures are equally effective. As a result, a balancing of interests is applicable to the rule of law guarantees as well, where the experience from competition proceedings can guide the relevant enforcers and authorities in how to conduct investigations under the DMA and DSA.

In all sections we reinforce the importance of considering how a combination of *legal and technical measures* can in many cases facilitate the necessary level of data access and minimise harms to the interests of protecting intellectual property, privacy, security, and the rule of law. These measures can consist of limiting the range and detail of the data to which access is provided, signing a non-disclosure agreement (possibly including contractual fines for violating the agreement), relying on anonymisation/pseudonymisation or synthetic data, or using a data clean room. Nevertheless, it is still likely that irreconcilable conflicts between the interests of the target and receiving parties will occur. To decide on these conflicts, we propose to weigh the importance of the data access request for achieving the purpose of the relevant legal obligation against the strength of the claims for protecting the commercial interests of the target party. This implies that there may be cases where it is justified to reject a data access request, for instance where the importance of the particular form of data access at stake is low and the target party has strong claims against disclosure.

Case Study 1: Access to Online Advertisement Databases (Article 39 of the DSA)

Based on our interpretation of the relevant legal and technical conditions, this report then moves to discuss three case studies. Analysing specific obligations is important because it allows us to discuss the challenges of implementing effective data access provisions in concrete contexts.

The first of the case studies is the obligation that requires platforms to provide access to a database that contains advertisement that was displayed on the platform: Article 39 of the DSA. This obligation can be roughly summarised as follows:

1) **Target party:** VLOP/VLOSE;[6]

[6] Very Large Online Platforms and Very Large Online Search Engines, as determined by the DSA.

2) **Target data:** All paid advertisements presented in the platform;
3) **Receiving party:** General public;
4) **Timeliness:** Continuous; and
5) **Mode of access:** Queriable Application Programming Interface (API) and web portal.

Based on these characteristics, one can categorise this obligation as requiring **General Public Access to Data**, a group that requires careful balancing of potentially conflicting priorities in terms of privacy and intellectual property protection in particular.

The report then moves to discuss many specific challenges in the implementation of this obligation: from delimiting what is an advertisement to setting up of concrete thresholds that can protect the privacy of users in an aggregated, public database. In terms of specific balancing challenges, the report argues that the VLOPs/VLOSEs have only limited room to deny access to the data on intellectual property grounds, as the DSA establishes a clear and well-defined data access obligation. In addition, the report proposes a series of criteria to help ensure that the database protects the privacy of individuals. These come mostly from the aggregation of the disclosed information into pools of people that include at least 100 users as a minimum described audience size for any publishable advertisement targeting parameters. With regard to rule of law guarantees and information security, we envision limited concerns as a result of this obligation.

Finally, the **Technical Annex II** to this report provides a list of important fields that should be part of any Transparency Database for Online Advertisements.

Case Study 2: Access to Data for Vetted Researchers (Article 40(4) of the DSA)

Article 40(4) of the DSA requires VLOPs and VLOSEs to provide access to data to previously vetted researchers for the purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the EU and the assessment of the adequacy, efficiency and impacts of risk mitigation measures. This is a much welcome innovation of the DSA, and one that reflects pleas by politicians, civil society representatives, and academics for many years.

Overall, the obligation can be summarised as follows:

1) **Target party:** VLOP/VLOSE;
2) **Target data:** All data;
3) **Receiving party:** Vetted researchers;
4) **Timeliness:** Triggered on action;
5) **Mode of access:** Varies depending on the data requested. There is an express mention of API access.

Based on these characteristics, one can categorise this obligation as requiring **Private Party Access to Data**. This group is particularly suited to a combination of legal and technical measures as a way to facilitate data access in case of conflicts between conflicting interests.

Targeted data: In terms of the targeted data, the report defends an expansive interpretation of what data can be accessed by researchers, including all examples cited in Recital 96.[7] With regards to the territorial scope of the data, we build on OFCOM's experience with the attacks carried out in Buffalo, New York, to argue for a combination of quantitative and qualitative criteria to assess whether a given specific request is under the purview of the DSA (which should be restricted to events linked to the EU — something that is easier said than done).

Another important question is whether the 'targeted data' includes solely what is regularly collected by platforms, or if researchers can request the production of new data. While there is no fixed answer to this question — it will depend on the type of data and the interests at stake — the report builds on the balancing framework described in Part 2.1 to outline which criteria should guide such determination. It also recommends that platforms should publish dataset descriptions and codebooks for their most commonly requested datasets on public archive sites, and that *every request for data access should be made public in a centralised database, as well as the justification given by the company to deny the request and the final decision by the Digital Services Coordinator ('DSC')*. This will not only facilitate the building of a common pool of knowledge on what data is available but also enable the broader community to challenge decisions that are abusive, do not respect user privacy, and so on.

Receiving party: Receiving parties are previously vetted researchers that are affiliated with a research organisation. There are many important outstanding questions on how to implement this vetting in practice. The report partially addresses some of them in different sections. One suggestion is to consider delegating the vetting of researchers to a third party, for instance, national science foundations under Article 40(13) of the DSA. DSCs can focus their attention on assessing the substance of requests.

Timeliness and mode of access: The timeliness and mode of access will depend on what type of data is being requested, as well as what safeguards must be implemented to ensure that the data are accessed in a safe and protective manner. Because this is an open-ended obligation, there is a large room for adaptation that will depend on the need to offset privacy, intellectual property protection, information security and rule of law guarantees — as required by Articles 40(2) and 40(5) of the DSA.

Privacy protection: The report builds on the EDMO working group on 'Platform-to-Researcher Data Access'[8] in making privacy recommendations. Indeed, the design of safeguards will be request-specific, and the EDMO report provides a useful overview. It is worth stressing that in many cases, the best combination will be a solution that relies on a combination of technical safeguards (such as restricted API access, or even safe rooms), with legal safeguards that prevent researchers from abusing their data access (including the conclusion of non-confidentiality agreements to protect the data, mandatory courses to ensure that researchers have the technical skills to protect the data, ethics

[7] Such as data on the accuracy, functioning and testing of algorithmic systems for content moderation, training data and even the code of algorithms.

[8] European Digital Media Observatory, *Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access*, (2022), 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

guidelines, and so on). Again, we outline how the general balancing framework of section 2.1 can help authorities, companies and researchers consider the stringency of these required safeguards.

Intellectual property protection: The implementation of Article 40(4) DSA will trigger a balancing of data access interests against the protection of trade secrets and commercial confidentiality. Based on Recital 97 of the DSA, the report argues that, as a general principle, only *in exceptional circumstances* platforms should be able to preclude any access to a dataset based on grounds of commercial confidentiality or trade secret protection. In many cases, a combination of legal and technical measures should enable the core of the data access request with relevant protections in place. The link between the legal mandate established by Article 40(4), the specific research question of a given data access request and the protections granted to the data by intellectual property laws should guide the balancing of the interests. A practical suggestion is to require platforms to initiate a discussion in response to a researcher data access request of what data they can reasonably provide access to and what in their view would be appropriate contractual and/or technical measures to enable data access without eroding their interests. In turn, researchers would get a chance to react and provide their interpretation of the balancing exercise and of potential other practical mechanisms to facilitate data access.

Information security: In order to assess and overcome information security concerns, the report recommends a threat modelling exercise that can help with the implementation of steps 4-7 of our framework. When conducting such an exercise, platforms should clearly articulate specific scenarios in which the data in question might be a security risk and specify exactly what the outcomes of an adversary using the data might be. Platforms should also clearly describe which adversaries they envision. Clearly articulating the risk situation, risk outcome, and potential threat actors is vital for determining whether the alleged risk is concrete (step 4 of our framework) and determining how the risk might be mitigated (step 6 of our framework). A key threshold is **consistency** between internal and external risk assessments: a claim that the disclosure of a given type of data poses a high external security risk is not credible if a very large number of employees can access the same dataset without significant internal safeguards.

The report envisions only limited concerns with regard to **rule of law guarantees** as a result of this obligation.

Finally, the report addresses a couple of important practical considerations.

Costs: The first is which party bears the cost of compliance and research. These can be roughly divided into two groups: (i) internal platform costs to structure databases, and so on, which should be borne by the platforms themselves; (ii) and the costs to maintain a research team, which are the responsibility of the researchers themselves.

Funding: For researchers to bear these costs, though, they require independent sources of funding. The problem is that the vetting process for the DSA creates a potential mismatch between researchers applying for grants and the decisions to grant access to data (an integral part of a grant application).

To solve this, the report proposes the creation of **'DSA Research Grants',** which are specifically targeted at research based on Article 40(4) of the DSA. To obtain these grants, applicants must explain in their applications which data they must access, how their research contributes to the detection or minimisation of systemic risks in the EU, and how they plan to comply with all the requirements of Article 40(8). Obtaining a DSA Research Grant would provide researchers with strong *prima facie* evidence that they have passed the vetting process, so that their requests should be authorised by the relevant DSC in an expedited time frame. The report discusses some alternatives to conduct this vetting process in an independent manner.

Arbitrating data access disputes: Finally, the report discusses how the DSA grants the DSC of establishment exclusive powers to vet researchers and decide on access requests. This centralisation is worrisome, as the significant problems in the enforcement of the GDPR showcase how some EU regulators — in particular in Ireland and Luxembourg — struggle to effectively enforce laws against Very Large Online Platforms ('VLOPs') and Very Large Online Search Engines ('VLOSE') that are strategically important for national economies. The DSA system may be even worse than the GDPR, as it appears that there is no mechanism to override decisions by the DSC of establishment. Here, an active involvement of the European Commission will be important, in particular in setting up a detailed vetting process that can be applied EU-wide. This would allow, for example, researchers to better rely on the rights established by Article 40(9) of the DSA, which enables researchers to apply for data access with the DSC where they are located, granting them strong prima facie evidence that their request is reasonable and that they are implementing proper safeguards.

Case Study 3: Sharing Click-and-Query Data (Article 6(11) of the DMA)

Our final case study targets Article 6(11) of the DMA, which requires gatekeepers to share with any third-party providing online search engine services ranking, query, click and view data in relation to free and paid searches. It also requires companies to anonymise the data, so that it no longer qualifies as personal data. Overall, it can be summarised as follows:

1) **Target party:** Gatekeepers
2) **Target data:** Ranking, query, click and view data
3) **Receiving party:** Competing search engines
4) **Timeliness:** Upon request, but then likely continuous or in defined intervals
5) **Mode of access:** Queriable or streaming API

This obligation can be categorised as **Private Party Access to Data**. **Target parties** are gatekeepers that are providing online search engine services as a core platform service. **Receiving parties** are all competing search engines that are providing general search services–generally excluding vertical and other specialised search engines. This will require that the Commission establishes some criteria to vet potential entrants that request access to the data, though we do not venture into outlining those. The **target data** is quite precise, facilitating the outlining of a data-sharing obligation and allowing us to provide some more concrete guidance.

Part of our recommendations build on the CMA Report,[9] a previous CERRE report[10] and academic studies in this area. The CMA Report in particular discusses the importance of sharing tail queries, which can be both uncommon queries as well as 'fresh queries', that is, queries that relate to recent events. These pose different challenges in terms of technical solutions to ensure that the data is both useful and anonymised. We then discuss the importance of k-anonymisation as likely the best solution to overcome these challenges and to preserve privacy. In particular, for reasons of utility, we recommend that platforms share query data aggregated by day, the exact search term, users' geographic region to the NUTS 2 level, language, and search platform (desktop or mobile). We believe generalising to this level will allow for a meaningful number of searches to be safely shared, while still conveying the most important context that implicitly defines users' searches. We also recommend that *k* is set between 100 and 500 to ensure that the data is anonymised while maintaining its utility.

In general, the report recommends that the shared data should contain at least the following information.

Proposed Standard Field	Description	Type
search_term	Text of the searched terms/queries	text
search_lang	Language of the search	text
search_region	Inferred location for the aggregate of specific terms/queries	text
query_responses	Set of query responses ordered as they were returned to users, including a boolean field specifying whether the response ranking was affected by paid advertising	json
click_data	Set of user click data in relation to responses (dictionary of response ids to counts)	json

However, because the determination of what exactly is a relevant query is very context-specific, we do not have the information to provide any insights on how to determine what exactly should be the target data.

In terms of timeliness and mode of access, the report differentiates between tail and fresh queries. Tail searches are relatively easier to supply, conceptually speaking, although they pose greater privacy concerns. A reasonable recommendation is that data be made available on a daily basis reflecting data no more than 48 hours old. In terms of mode of access, we recommend that data be made available in a bulk file format. Timely searches**,** however, present other important technical challenges — in particular on determining what exactly qualifies as a timely search that is relevant to increase

[9] Competition and Markets Authority, *Online Platforms and Digital Advertisement - Market Study Final Report*, (2020), https://perma.cc/AJ3F-C44Z

[10] See Kraemer Jan, *Data Access Provisions in the DMA*, (2022), https://cerre.eu/wp-content/uploads/2022/11/DMA_DataAccessProvisions-2.pdf

competition in the search engine market (the overall purpose of this obligation). That is because the setting of parameters is almost a dynamic game: the more popular the search query, the more important it is for competition between search engines, so the faster one would want the data to be shared. These are important challenges that should be carefully considered before search engines are required to share almost real-time data. They also prevent us from providing more detailed guidelines on how to implement such sharing in practice.

Intellectual property protection: Two areas of particular importance to the balancing between data access and intellectual property are the scope of the data to be shared and the remuneration. In terms of scope, the fact that intellectual property protection is not an interest that is explicitly recognised by the DMA as a potential countervailing interest, limits how much gatekeepers can rely on it to block access — a request would fail step 5 of our framework. That is because the framing of the obligation illustrates that the legislator has already conducted a balancing. That is not to say, though, that there should be unlimited sharing of data. Recital 61 clarifies that access to ranking, query, click and view data should allow third-party undertakings to optimise their services and contest the relevant core platform service. This provides an objective function that is reflected in step 3 of our balancing framework: the data to be shared should be the minimum necessary to increase the contestability of general search markets. In terms of remuneration, Article 6(11) of the DMA requires gatekeepers to provide the data on fair, reasonable, and non-discriminatory ('FRAND') terms. Because search query data is collected as a free byproduct of offering a search engine, the marginal cost of obtaining the user information for the gatekeeper is (roughly) zero. It therefore seems undesirable to give gatekeeping search engines the possibility to charge a fee for access to their already collected search query data. Nevertheless, it seems reasonable to let the gatekeeper impose costs for delivering the data in a workable format.

Finally, it is worth stressing that because this provision is addressed to a well-defined set of sophisticated third-parties, we do not anticipate very high information security risks. With regard to the rule of law guarantees, there is no immediate link between access to search query data and potential liability, diminishing any risks in this regard.

Looking ahead

The final Part of the report looks ahead to the proposed AI and Data Acts to confirm the relevance of the four categories of data access distinguished in Part I. Both proposed Acts contain obligations falling within our category of regulator access to data, enabling enforcers to check compliance. The proposed Data Act also includes a novel type of regulator access to data obligations in the form of mandates for private parties to make data available to public authorities in cases of exceptional need. Beyond this, the proposed AI Act includes provisions facilitating transparency towards the general public relating to the use of AI systems. In addition, the proposed Data Act regulates private party access to data in the form of an Internet of Things ('IoT') data access right and sets minimum legal obligations to facilitate switching between data processing services.

While the implementation of these data access and transparency obligations will need to be considered on a case-by-case basis, our general considerations about how to balance conflicting interests and the available measures to address tensions with the protection of intellectual property, privacy, and security are also applicable in the context of the AI and Data Acts. As such, the legal, economic and technical insights brought together in this report also aim at contributing to an effective implementation of current and future data access and transparency obligations for digital markets more generally.

INTRODUCTION[11]

The Digital Markets Act ('DMA') and the Digital Services Act ('DSA') will bring about profound changes in the governance of digital markets. Some of its most noteworthy obligations include a range of rules that mandate digital platforms to share certain types of data and information with regulators, vetted researchers, competitors and society more broadly. Another group of rules also imposes transparency obligations with regard to digital platforms' internal content moderation activities and the algorithms that control much of digital markets.

These are all ambitious obligations that could have a real impact if implemented and enforced effectively.[12] This CERRE Academic Report focuses on the issue of implementation of data access and transparency obligations. Its main goal is to develop insights to help implement the provisions in a way that is both effective in increasing digital transparency, but that also acknowledges and attempts to minimise possible tensions with conflicting interests.

This report is divided into two main Parts, a conclusion and two technical Annexes.

Part I maps out and categorises the multiple data access and transparency obligations that are present in the DMA and the DSA. Technical Annex I complements this section, presenting a detailed mapping and categorisation of all 54 data access and transparency obligations present in both Regulations.

Part II performs a case study of three selected obligations to identify implementation challenges that regulators, companies and civil society must overcome. More specifically, it discusses rules requiring the creation of a database for online advertisement (Article 39 of the DSA); requiring large digital platforms to grant vetted researchers access to internal data (Article 40 of the DSA); and mandating very large online search engines to share click-and-query data with smaller competitors (Article 6(11) of the DMA).

The Conclusion then looks ahead to the future of digital regulation, outlining how our categorisation exercise can be extended to other EU legislative frameworks under discussion, such as the proposals for an AI Act and a Data Act.

[11] The authors are grateful to Thomas Tombal for his comments and suggestions on an earlier draft.

[12] Discussing enforcement challenges, see for example Filippo Lancieri, *Narrowing Data Protection's Enforcement Gap*, 74 MAINE LAW REV. 15 (2022). (Focusing on privacy policies) and OMRI BEN-SHAHAR & CARL E SCHNEIDER, MORE THAN YOU WANTED TO KNOW: THE FAILURE OF MANDATED DISCLOSURE (2014). (Discussing transparency more broadly).

PART I: CATEGORISING TRANSPARENCY OBLIGATIONS IN THE DMA AND THE DSA

The DMA and the DSA introduce a range of data access and transparency obligations addressed at companies and regulators alike. However, these are generally distributed across hundreds of pages of text and their associated recitals–indeed, there are at least 54 different transparency provisions in the DMA/DSA package (32 in the DSA and 22 in the DMA).

This Part sheds light on this complex web by engaging in a comprehensive mapping and categorising exercise. It is divided in two subsections and a technical appendix. Part 1.1 describes the variables that shape our categorisation exercise. Part 1.2 summarises the four categories which we propose to separate these 54 obligations in coherent groups that share similar characteristics:

1) Regulator Access to Data, which establishes/expands regulatory powers to require access to private data for investigations and other public purposes;

2) Private Party Access to Data, which requires private parties to share data with other previously defined private parties;

3) General Public Access to Data, which requires private parties to provide the general public with certain types of data on their private activities; and

4) Regulatory Transparency, which requires regulators to provide the general public with certain types of data on their public activities.

Appendix 1 to this report provides a comprehensive review of all 54 transparency obligations present in both Regulations.

1.1 Understanding the Variables that Can Help Categorise Transparency and Data Access Obligations

Transparency and data access obligations come in multiple shapes and forms. Some are targeted at all private parties, while others only impact large private players or regulatory authorities. Some require general access to data, others some form of restricted access. Some require written reports, while others require Application Programming Interface (API) access or the construction of searchable databases. Indeed, mandated disclosures of information (and transparency obligations in general) are one of the most ubiquitous forms of regulatory intervention, in part because they are usually seen by regulators as reasonably low-cost but still powerful instruments that diminish information asymmetries between companies and third-parties.[13] As digital markets are riddled with large

[13] Even if, at least from a consumer perspective, they are also one of the least successful regulatory tools. See Omri Ben-Shahar & Carl E. Schneider, *The failure of mandated disclosure*, 159 UNIV. PA. LAW REV. 647 (2011).

information asymmetries,[14] it is unsurprising (and generally welcome) that novel regulations such as the DMA and DSA introduce a range of data and algorithmic transparency provisions.

An important next step, then, is a careful mapping and categorisation of the parties impacted by these obligations, what exactly they entail both from a legal and a substantive perspective, and what other private and public interests are impacted by the obligations. In order to do so, one has to first come up with variables that help an eventual categorisation. We propose five variables that can help map and summarise transparency obligations in general:

1) **Target party:** refers to the companies or public bodies that must provide access to the information. It encompasses Very Large Online Platforms (VLOPs) or Very Large Online Search Engines (VLOSE),[15] Gatekeepers,[16] Regulators (usually the European Commission or national Digital Services Coordinators),[17] Companies (a broader category that incorporates providers of online platforms, providers of online intermediary services and other broad categories in the DSA), and others;

2) **Target data:** refers to the type of data that is granted access to as a result of the successful implementation of the obligation. It can include different forms of data: from user-related data, to content moderation data, to commercial data, to algorithmic features, to all ads displayed in the platform, and to all data and algorithms held by platforms (among others). This is a broad category that can help parties better understand what piece of information the obligation targets;

3) **Receiving party:** refers to the party that will access the data. Receiving parties are mostly divided according to regulators, the general public, the recipients of the service, vetted private parties, or competing companies;

4) **Timeliness:** refers to how frequently target parties must provide the target data. Timeliness is mostly divided into continuous access, access triggered on action (or provided upon specific request), annually or at other time intervals;

5) **Mode of access:** refers to the format in which the target party should make the target data available. Mode of access is mostly divided according to a written report (such as, an audit report, or a written explanation of what the company is doing); a data file (such as, a .csv file), an Application Programming Interface (API) or other forms of data access.

[14] Lancieri, *supra* note 11.
[15] As designated according to the procedure institute by Article 33 of the DSA.
[16] As designated according to the procedure instituted by Article 3 of the DMA.
[17] For the purposes of this report, regulators are to be understood as authorities acting independently from private and public bodies. For instance, Article 50(2) of the DSA requires DSCs to act in complete independence. They should 'remain free from any external influence, whether direct or indirect' and should 'neither seek nor take instructions from any other public authority or any private party'.

1.2 A Proposal for Categorisation

Based on these variables, one can create four categories of transparency and data access obligations in the DMA/DSA package. These are:

1) **Regulatory transparency:** this category encompasses a series of obligations that require regulators to make certain types of data publicly available to society at large. Regulatory transparency obligations host multiple commonalities. First, target parties are regulators, which usually have to make certain forms of data available to the general public by means of written reports or the creation of specific data files. Because the obligation targets regulators, the data generally cannot be used to affirm that a given company has violated the law (no potential legal liability). Yet, the usually broad data transparency requirements can raise concerns with regard to privacy and trade secrets/sensitive commercial information, even as the nature of the information and the mode of access (such as reports) diminishes information security concerns. Finally, it is worth noting that regulatory transparency obligations have long been established by other EU regulations and can serve as a benchmark for implementation.[18] However, some transparency provisions that require the creation of databases and other forms of online access can pose new challenges;

2) **Private-party access to data:** is likely the most important and novel group of transparency obligations instituted by the DMA/DSA package. It encompasses Articles that require private parties to provide varied types of data to other previously defined private parties–from recipients of the service, to vetted researchers or competitors. The timeliness of the provision varies significantly, and so does the mode of access. Yet, because the recipients are private parties themselves, technical and legal safeguards need to be instituted to protect privacy, trade secrets/sensitive commercial information and, in many cases, information security. Beyond this, the implementation of the access obligations should be proportional towards the target parties.

3) **Regulator access to data:** encompasses a series of obligations that either require private parties to provide certain types of data to regulators (such as the European Commission or Digital Services Coordinators), or empower these regulators to request data from private parties. These obligations target private parties in their many forms (gatekeepers, VLOP/VLOSE or undertakings more broadly) and the receiving party is, by definition, a regulator. The targeted data varies, but may include all the relevant data held by the companies. The mode and timeliness of access also vary depending on the obligation. Because the data may be used in compliance investigations against the providing companies, rule of law protections/limitations apply. Finally, while some requirements are novel in terms of the extent or type of access, obligations granting regulatory authorities broad access to the data

[18] For example, Article 59 of the GDPR asks each supervisory authority to publish an annual report on its activities.

held by private parties in order to check compliance with regulatory requirements are not new, meaning that past experience may help in the implementation.[19]

4) **General public access to data:** encompasses obligations that require private parties to provide the general public with access to certain types of private data. Different obligations target different forms of data, and the timeliness normally includes either a continuous provision or the provision at fixed points in time (such as annually). The mode of access also varies widely, from data files, to written reports to APIs. The indiscriminate public access to the target data forces parties to ensure the highest level of protection for personal data and trade secrets (relative to the other categories). For some modes of access, such as APIs, data providers also need to account for potential information security risks, both in terms of traditional cybersecurity[20] as well as more specific information security concerns.[21]

Annex I to this Academic Report provides a table that classifies all transparency and data access obligations present in the DMA/DSA Package according to the five variables and then groups them into the four main categories described above. Below, we provide a brief outline of the obligations in each category.

1.2.1 Regulator access to data

There are 8 obligations in the DSA and 5 in the DMA that ensure public regulators' (such as the European Commission or Digital Services Coordinators) access to data held by private parties. These start with a range of traditional obligations that empower regulators to request information, carry out interviews or inspections and perform audits.[22] The DMA and the DSA, however, also introduce more targeted provisions that require undertakings to provide regulators with specific types of data. These include the requirement that certain companies conduct internal risk assessments and prove to regulators that they comply with obligations,[23] requirements that companies/undertakings provide information to the Commission or other regulators whenever they meet specific thresholds,[24] report potential crimes taking place in the platforms,[25] and a novel obligation that requires certain companies to provide regulators with clear explanations on the design, logic of functioning and even allow for the testing of algorithms and recommender systems.[26]

[19] For example, Section V of Regulation 01/2003 has long granted the European Commission broad powers to request information from private parties, take statements and conduct physical inspections.
[20] The risk of unauthorised access to servers and extraction of confidential information. An example here would be a hacker exploiting an API to gain access to the systems of a company.
[21] The risk that by disclosing information on how certain systems work, this information may be exploited by motivated third-parties. An example here would be a company relying on mandated disclosures by an online search engine on how they rank websites (pursuant to a general obligation of algorithmic transparency imposed by the DSA) to increase its rankings and gain more user traffic.
[22] Articles 10, 51, 65-69, 72 and 74 of the DSA, and Articles 13 and 21-23 of the DMA.
[23] Such as requirements that VLOP and VLOSE conduct and preserve information about systemic risk analyses, supplying them to the Commission when requested (Article 34(3) of the DSA) or requirements that Gatekeepers report to the Commission the measures taken to ensure compliance with the DMA (Article 11 of the DMA).
[24] Such as the threshold to qualify as a gatekeeper (Article 3(3) of the DMA), the notification of all intended acquisitions/concentrations by gatekeepers (Article 14 of the DMA) and requirements that out-of-court dispute settlement bodies report to Digital Services Coordinators information on their activities (Article 21(4) of the DSA).
[25] Article 18 of the DSA.
[26] Article 40(1) of the DSA.

These provisions generally allow regulators to request different types of data. Triggers are normally specific requests by a given regulatory authority, and the mode of access depends on the type of data and the type of request. As the information is provided to regulators to carry out compliance assessments, it can certainly be used to assess potential liability, meaning that rule of law protections and limitations must be observed. Finally, there are important concerns with regard to information security, but these are common to most other provisions requiring data access for regulators.[27]

1.2.2 Private party access to data

The most important DMA and DSA innovations in terms of transparency and data access obligations are the provisions that grant private parties access to data held by digital platforms and regulators — there are 7 such provisions in the DMA and 7 in the DSA. These include three obligations that increase the data available to different companies active in the programmatic advertising market (in particular advertisers and publishers);[28] four obligations that establish some form of continuous portability of personal and non-personal data – including sensitive commercial data;[29] and two obligations that provide undertakings with (usual) rights of access to the file the Commission/regulators built in connection with a specific investigation.[30]

The DSA, however, also introduces some significant innovations in terms of data transparency. Article 17 of the DSA requires providers of hosting services[31] to provide recipients of its services[32] with a clear and specific statement of reasons any time they impose restrictions on this recipient. This includes, for example, decisions by platforms to diminish the visibility of certain types of content or the suspension or termination of monetary payments or even entire accounts.[33] Article 26 requires platforms that display ads to provide recipients of the service with information on each ad that is displayed.[34] Articles 30 and 32 require certain platforms to obtain information on traders in their platforms, provide some of the data to recipients of the service and inform consumers when they are aware that they acquired an illegal product or service on the platform.[35] Article 40(4) of the DSA requires providers of VLOP and VLOSE to grant previously vetted researchers access to internal platform data for the purposes of conducting research that contributes to the detection, identification and understanding of systemic risks in the EU.[36] Article 37 requires VLOP and VLOSE to undergo yearly audits, and to provide auditors with all necessary information required for this assessment.

These provisions generally allow a private party to have access to what can be sensitive data of another private party. The timeliness of the access varies, with some obligations being triggered on a

[27] Such as those of Section V of Regulation 01/2003.
[28] Articles 5(9), 5(10) and 6(8) of the DMA.
[29] Articles 6(9), 6(10) and 6(11) of the DMA. Article 6(11), which requires the sharing of click-and-query data between search engines, will be better studied in Part II to this academic report.
[30] Article 79 of the DSA and 34 of the DMA.
[31] That is, generally, an undertaking (or more specific, an information society service) that provides a service consisting of 'the storage of information provided by, and at the request of, a recipient of the service', Article 2(g)(iii) of the DSA.
[32] That is, natural or legal persons who use an intermediary service (which is itself a broadly defined term in the DSA), Article 3(b) of the DSA.
[33] Article 17 of the DSA.
[34] Article 26 of the DSA.
[35] Article 30 and 32 of the DSA, targeting online platforms that allow consumers to celebrate long-distance contracts with traders.
[36] Article 40(4) of the DSA. This obligation will also be better analysed in Part II to this Report.

specific action (such as a data request) while others require either continuous or at least daily access. The mode of access also ranges from different APIs to simpler data files. Because the recipients of the data are usually private parties (even competitors), and the modes of access include continuous access, authorities and undertakings must carefully consider risks to data protection, the leakage of trade secrets or sensitive commercial information and also information security when implementing these mandates.[37]

1.2.3 General public access

The DSA and the DMA also introduce important new obligations that require private parties to provide the general public with access to certain types of data held by the platforms – there are 10 such provisions in the DSA and 5 in the DMA. These obligations range from requirements that platforms publish reports or set up databases that disclose information on some of their content moderation or dispute settlement practices (either voluntary or mandated),[38] explain in some level of detail how their recommender systems and other algorithms (including user profiling algorithms) work,[39] create a public depository of online advertisements displayed in the platforms,[40] publish the average number of monthly active users in the EU,[41] publish general conditions of access to software application stores on FRAND terms,[42] and publish a reference offer with technical details for interoperability for messaging services.[43] The DSA also includes a broad obligation that VLOPs and VLOSEs must provide researchers, nonprofits and independent organisations with access to public data that may help with the detection, identification and understanding of systemic risks in the EU, though it does not explain well what such obligations entails.[44] Finally, the DSA also requires VLOPs and VLOSEs to undergo a detailed audit to assess compliance with legal obligations, publishing results.[45]

These obligations share characteristics. First they are all targeted at private parties — and the majority at very large undertakings that classify as VLOP, VLOSE and gatekeepers under the different Regulations. They also generally require that different types of data are indiscriminately disclosed to the public in general, not to a small subset of previously defined private parties. The obligations usually require either continuous access or well-specified windows for compliance, and the modes of access vary widely: from APIs to web portals to data files or written reports. Finally, because of the type of data shared, the different modes of access and the widespread availability, it is likely that some of these requests will raise concerns with regards to privacy, the protection of trade secret/commercial sensitive information and, in some cases, information security. However, it is worth noting that many obligations are specific in terms of which data must be provided and which parties have access to it.

[37] This does not mean that these rights trump the specific rights of access granted by the legislation, but that obligations should minimise risks to the extent possible.
[38] Articles 15, 22(3), 24 and 42 of the DSA.
[39] Articles 27 of the DSA and Articles 6(5) and 15 of the DMA.
[40] Article 39 of the DSA.
[41] Article 24(2) of the DSA.
[42] Article 6(12) of the DMA.
[43] Article 7(4) and 8 of the DMA.
[44] Article 40(12) of the DSA.
[45] Articles 37(2) and 42(4-5) of the DSA.

This diminishes concerns that the requirements are either disproportional or that they violate other legal frameworks (such as in the area of privacy or intellectual property protection), as these are commands by the EU legislator requiring disclosure of a specific type of data in a specific format.

1.2.4 Regulatory transparency

The final category of relevant transparency obligations in the DMA/DSA require public bodies to publish a range of data that either compile important information provided by private parties or that summarise their own activities in the implementation of both Regulations — there are 7 such provisions in the DSA and 6 in the DMA. These obligations start with requirements that the European Commission and Digital Services Coordinators maintain updated lists of trusted flaggers, VLOPs/VLOSEs and gatekeepers;[46] a requirement that Digital Services Coordinators ('DSCs') publish an annual report of their activities;[47] a requirement that the Commission publishes preliminary and non-confidential versions of specifications for DMA obligations, DSA decisions and its assessment of codes of conduct;[48] and a requirement that DSCs publish a report on the functioning of the out-of-court dispute settlement body.[49] A final noteworthy obligation is the one that requires the European Board for Digital Services to publish comprehensive annual reports that identify the most important systemic risks reported by VLOPs and best practices to mitigate such risks.[50] The Commission is also empowered to issue general guidelines on the best practices in the detection and mitigation of such systemic risks, a moment when it is also required to organise public consultations on the draft guidelines.[51]

These obligations share similar characteristics: they require regulators to either disclose information to or engage with the general public with regard to the implementation of the DSA/DMA package; most of the disclosure comes in the form of written reports, and these data are not usually disclosed in the context of an ongoing investigation (with some exceptions). However, they also share some important distinctions. One group of obligations generally targets the work of regulators (publish annual reports) and, as such, raises more limited concerns in terms of privacy, trade secret protection and information security. Another group, however, requires regulators to publish public versions of otherwise confidential documents in their possession (such as preliminary or final decisions, or summaries of best practices). In these cases, regulators must be careful not to publish confidential and strategic corporate information.

[46] Articles 22(5) and 33(6) of the DSA and 4(3) of the DMA.
[47] Article 55 of the DSA.
[48] Article 8(6) of the DMA and Articles 45(4) and 80 of the DSA.
[49] Article 21(4) of the DSA.
[50] Article 35(2) of the DSA.
[51] Article 35(3) of the DSA.

PART II: MAPPING OUT CHALLENGES IN THREE SPECIFIC CASE STUDIES

The categorisation exercise in Part I of this Report can help with the visualisation of the breadth and scope of the new transparency and data access obligations in the DMA/DSA package. This part complements our high-level analysis with three case studies, allowing for a better discussion of how technical and legal solutions can help minimise implementation concerns. We focus on three novel obligations:

1) The obligation that VLOPs/VLOSEs provide additional online advertising transparency, as established by Article 39 of the DSA (*general public access to data obligation*);

2) The obligation that VLOPs/VLOSEs provide access to data to vetted researchers for the detection, identification and understanding of systemic risks, as established by Article 40(4) of the DSA (*private party access to data, where the private party is a previously defined civil society representative*); and

3) The obligation that gatekeepers provide competing online search engines with access to ranking, query, click and view data on fair, reasonable and non-discriminatory grounds, as established by Article 6(11) of the DMA (*private party access to data, where the private party is a competitor*).

First, though, we present some general considerations on how to balance competing interests that can help guide this exercise.

2.1 General Considerations on Balancing Competing Interests between Disclosure, Privacy, Intellectual Property Protection and Information Security

The DMA and the DSA emphasise the importance of transparency and data access obligations to promote an effective implementation of both Regulations and help increase the competitiveness of some digital markets. However, societal interests in promoting transparency and data access are not absolute. Rather, they occasionally clash with citizens' fundamental rights to privacy, businesses' rights to the protection of intellectual property and to freely conduct a business, society's general interests in ensuring adequate data security, and even general requirements that regulators respect the rule of law when wielding their normative and adjudicative powers.

Proportionality can be a key guiding principle in balancing these conflicting interests. Below, we present a structured proportionality test for data access obligations, centred around seven relevant questions that can help balance out conflicting rights:

Q1: Do the stated purposes of the data access request match the purpose of a disclosure obligation enshrined in the legislation, and can it help advance the goals of such an obligation?

Q2: Is the data requested sufficient to achieve the purpose of the request both in terms of scope (how much) and nature (what) of the data?

Q3: Is the data requested the minimum required to meet the stated purpose of the request and, if not, does the excessive part of the request lead only to minimum risk if disclosed?[52]

Q4: Can one identify clear and well-defined, non-speculative potential harms that would arise as a direct result of the data disclosure?

Q5: Are these harms explicitly recognised by the law as countervailing harms that need to be weighed against the interest in disclosure?[53]

Q6: Are a combination of technical and legal instruments sufficient to satisfy the data access request while minimising the potential harm to the target party (what is the least intrusive manner to implement the right)?

> Technical: Sometimes the adoption of certain technical protocols such as anonymisation/pseudonymisation, k-anonymity or differential privacy protocols, the use of synthetic data and so on, can help mitigate risks associated with access to the data without preventing the receiving party from achieving the purpose of the legal obligation. In that case, these technical measures should be employed before the data sharing.[54]
>
> Legal: In addition, in some cases, technical measures alone are not enough, but they can provide enough guarantees when coupled with legal protection against the disclosure of the information by violating parties. These may include express legal commands,[55] non-disclosure agreements and other forms of binding contracts, requirements that receiving parties provide adequate notice to the target party before the data is disclosed to the public,[56] and others.

[52] This notion of minimal risk slightly expands the concept of data minimisation, but can greatly facilitate access from a technical perspective. The idea is to prevent the creation of new databases for new requests, when the increased access does not pose risks. For example, imagine that a database of social media posts includes the content of posts as well as their reach and the number of likes each individual content received (in the aggregate), but that a given data access request requires only the content of the posts and their reach to fulfill the purpose of the disclosure obligation. However, because the number of likes per post is aggregate, disclosing the information only poses a minimal risk to privacy and other interests. In that case, one could grant the data access request including also the number of likes, as the alternative would be to create a new database that excludes the number of likes per post just for this specific request — something that would be burdensome without any significant gain to protected interests.

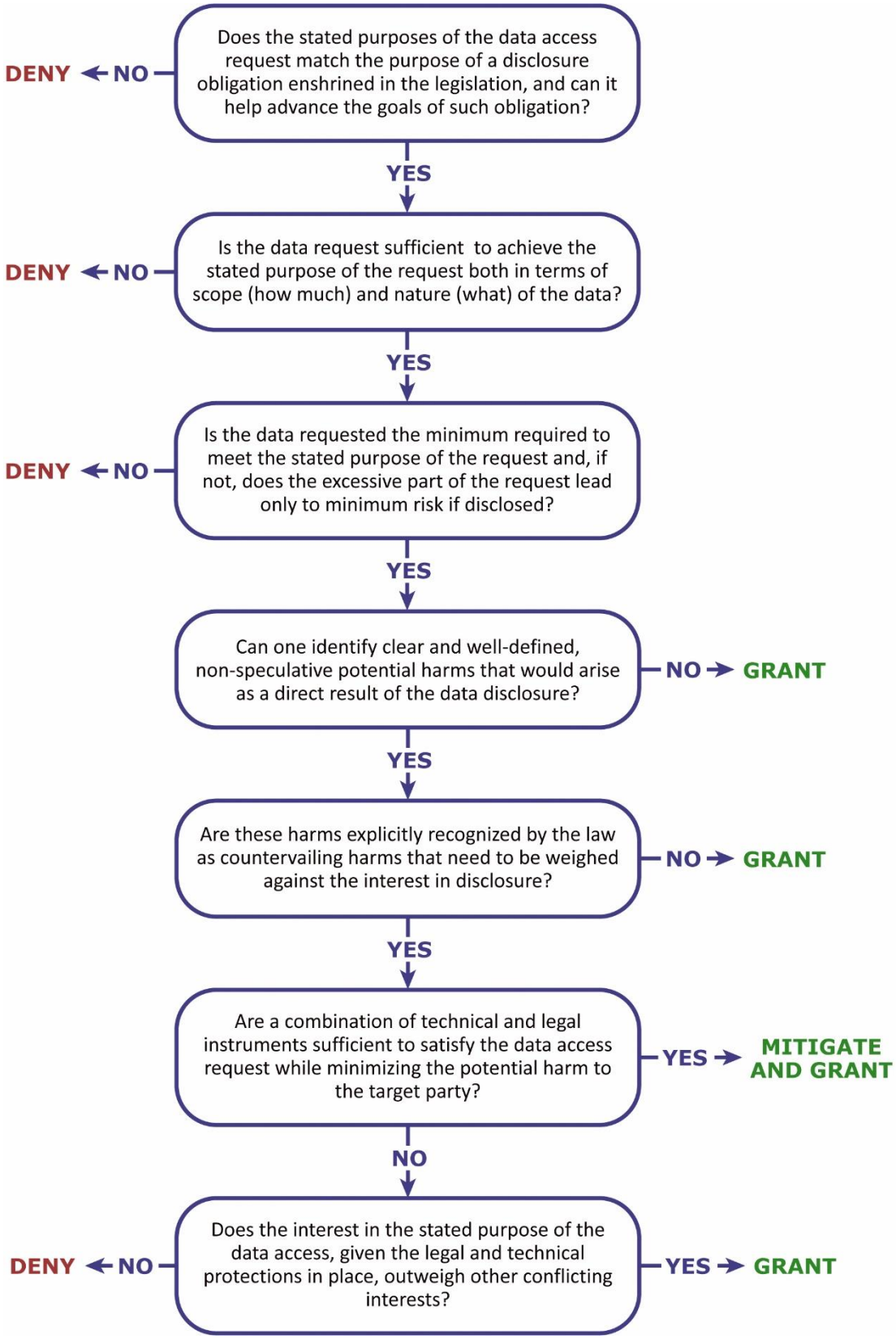
[53] For example, in the case of the DSA Researcher Access to Data mandate, Article 40(5) expressively mentions trade secrets and data security as a basis that can potentially offset data access requests. But similar protections are not applicable to requests by regulators (as discussed below).

[54] For example, if a given research question can be answered with anonymised data, then the data should be anonymised before sharing.

[55] For example, many laws criminally punish public employees that share corporate confidential information obtained as part of their work obligations.

[56] As is done, for example, with the disclosure of security vulnerabilities, enabling companies to patch the vulnerabilities before they are released to the broader public.

Q7: Does the interest in the stated purpose of the data access, given the legal and technical protections in place, outweigh other conflicting interests?



These seven steps, contained in the above flowchart reflect three main principles that should instruct any adequate balancing test for data access provisions.

1) Legislative purpose as the guiding principle:

The data access request should fit the purpose behind the applicable legal obligation and any potential harms invoked by the party having to grant access (the target party) should be recognised in the legal framework as relevant harms that need to be weighed against the interest in disclosure. This is a check conducted with reference to the scope of the applicable legal obligation. As an example, Article 40(4) of the DSA requires VLOPs and VLOSEs to provide researchers with access to data for the purpose of conducting research into systematic risks, and Article 40(13) of the DSA refers to the protection of confidential information, trade secrets and security as countervailing interests that need to be considered. This means that researchers can only request data relevant to assess systemic risks, and VLOPs and VLOSEs can only invoke the protection of confidential information, trade secrets and security as interests against disclosure.

2) Data minimisation:

The requested data must fit the stated purpose behind disclosure and should not go beyond what is necessary to achieve that purpose in terms of the scope (the amount and range) and nature (the type) of the data. This is a more factual check of whether the framing of the data access request is appropriate. For instance, requests for access to data in order to provide a search engine (based on Article 6(11) of the DMA) should be limited to data that is necessary for that purpose. This also means that whenever there are technical measures that can tailor or limit the scope and nature of the disclosed data without significantly compromising the purpose of the access, they should be implemented. In other words, the data access should be restricted to what is strictly necessary to achieve the purpose underlying the data access request.

3) Least intrusive implementation:

The data needs to be disclosed in a way that still meets the stated purpose of the request and the legal obligation, while minimising any potential harms. At this stage, parties must find the least intrusive manner to implement the core of the data access request. In case there are irreconcilable conflicts between the interests of the receiving party and the interests of the target party, the relevant enforcer/authority in charge needs to decide whose interests prevail, while minimising the harm to the best extent possible.

In that case, we believe that weighing the importance of the data access request for achieving the purpose of the relevant legal obligation against the strength of the claims for protecting the commercial interests of the target party should be the key considerations in these decisions. Where the data access request is vital for achieving the underlying purpose of the legal obligation and the claims for instance commercial confidentiality or trade secret protection are weak, it will be disproportionate to let the interest of the platform in non-disclosure prevail. However, where a

dataset is highly sensitive and protected, and the importance of data access for achieving the purpose of the relevant legal obligation is low, the interest of the platform deserves to get priority. This implies that there may be situations where it is justified to deny access to specific sets of data.

Beyond this, we believe the parties involved have a responsibility on each of their sides to take measures to still enable data access to the extent necessary and possible. Such measures may include the preparation of an alternative dataset by the target party that is less commercially sensitive but can still satisfy the purpose underlying the data access request, the signing of non-disclosure agreements by the data access seeker (possibly with contractual fines in case of violations of the agreement) or the use of a data clean room, where possible and appropriate. At the start of the implementation of the DSA and DMA, the exact boundaries of data access will need to develop and will gradually become clearer through experience.

As an important context in conducting the balancing, we believe it is reasonable to expect target parties to be proactive and co-operative in implementing data access, placing the burden on them to justify why they cannot grant access to specific requests (in particular at the beginning of the implementation process). There are important information asymmetries in this domain, and these clash with the intention of the EU legislator to facilitate data access through the relevant provisions in the DSA and DMA. This requires, for example, an expectation on the part of target parties to explain how and to what extent the requested data interferes with commercial confidentiality, trade secret protection, security or privacy. These insights will help the receiving party and/or the relevant enforcer/authority in charge to adjust the data access request to address the concerns of the target party, who will then either need to explain why the adjusted data access request still cannot be facilitated or agree to provide the data access and explain the conditions under which it deems the data access appropriate. This cycle of interaction between the target and receiving parties can continue until all relevant matters are resolved, with the relevant enforcer/authority or court in charge as the ultimate resolver of disputes.

While this general balancing framework is relevant for all categories of data access, the determination of what is proportionate in a particular scenario will depend on the circumstances of the case. We will discuss this balancing exercise in the context of our three case studies in sections 2.2, 2.3 and 2.4. Before that, the remainder of this section 2.1 analyses more general legal considerations of how to balance data access with, respectively, privacy (section 2.1.1), intellectual property protection (section 2.1.2.), security (section 2.1.3), and the rule of law (section 2.1.4). These can help instruct considerations in steps 6 and 7 of our framework above.

2.1.1 Balancing data access with privacy and data protection

When data access involves personal data, privacy and data protection rules apply. The GDPR defines personal data as ‘any information relating to an identified or identifiable natural person (‘data

subject')'.[57] The concept of personal data is broad.[58] As soon as information 'by reason of its content, purpose or effect' is linked to a particular person, it qualifies as personal data.[59]

Many instances of data access will therefore involve personal data, triggering privacy concerns.[60] Still, the goals of promoting transparency, data sharing and privacy can be aligned – or, at least, tensions can be minimised.

2.1.1.1 Lawfulness of personal data processing

At the outset, it is worth emphasising that the data protection and privacy rules do not ban the sharing of data. The GDPR does, however, regulate how personal data can be processed. The sharing of data or the provision of access to data to a third-party is a form of processing.[61] A key requirement under the GDPR is that all processing of personal data must be based on a lawful basis. The available lawful grounds depend, among others, on the parties involved (both the target party and the receiving party must have their own, separate lawful grounds) and the purpose of the processing.

For example, in the context of Article 40(4) of the DSA that mandates data access to researchers, the 2022 report of the European Digital Media Observatory's (EDMO) Working Group on Platform-to-Researcher Data Access states that the target party can rely on Article 5(1)(b) of the GDPR – which authorises further processing of personal data for research purposes.[62] This means that VLOPs and VLOSEs can share data with vetted researchers (per Article 40(4) of the DSA) without the need to identify a separate lawful ground for processing. This, however, is not the case in relation to other provisions regarding data access beyond researchers in the DMA and DSA. In these situations, the target parties must be able to point to a self-standing lawful basis for sharing personal data.

For targeted parties, the most relevant lawful basis in the GDPR is 'a legal obligation to which the controller is subject'.[63] The GDPR and case law of the European Courts, however, require that the legal obligation is formulated in a clear and precise manner that is also predictable and accessible.[64] It is beyond the goals of this Report to study whether all DMA and DSA transparency obligations meet

[57] By its turn, an identifiable natural person is 'one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'. See Article 4(1) of the GDPR.

[58] Nadezhda Purtova, *The law of everything. Broad concept of personal data and future of EU data protection law*, 10 LAW INNOV. TECHNOL. 40 (2018).

[59] Case C-434/16 *Nowak*, ECLI:EU:C:2017:994, par. 35.

[60] European Data Protection Supervisor, *Opinion 3/2020 on the European strategy for data*, (2020), https://edps.europa.eu/sites/edp/files/publication/20-06-16_opinion_data_strategy_en.pdf At 8-12.

[61] Article 4(2) of the GDPR defines 'processing' as 'any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'.

[62] More specifically, the provision stipulates that further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes is to be considered as compatible with the initial purposes, so that the further processing in the form of the data access to researchers can be based on the same lawful basis as the initial purpose for which platforms collected the data. See European Digital Media Observatory, *supra* note 8. At par. 57-61.

[63] Article 6(1)(c) of the GDPR. 48 THOMAS TOMBAL, IMPOSING DATA SHARING AMONG PRIVATE ACTORS: A TALE OF EVOLVING BALANCES (2022). At 354-356.

[64] See EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS & COUNCIL OF EUROPE, HANDBOOK ON EUROPEAN DATA PROTECTION LAW (2014). At 82, mentioning as examples of cases in which a legal obligation has been accepted as a lawful ground for processing the legal duty of employers to process data about their employees for reasons of social security and the legal duty of businesses to process data about their customers for reasons of taxation. See also 48 TOMBAL, *supra* note 93. At 356.

these thresholds. For our purposes, it suffices to say that 'compliance with a legal obligation' provides a possible lawful ground for the sharing of personal data in **all four of our categories of data access obligations** (irrespective of the identity of the addressee) considering that the DMA and DSA mandate these forms of data access.

The receiving parties need to have their own lawful basis for processing personal data under the DMA and DSA.[65] This, again, will depend on the nature of the receiving party and the purpose of processing. For the category of **regulator access to data**, regulators as receiving parties can rely on public interest as a lawful basis.[66] The categories of **regulatory transparency** and **general public access to data** aim to create transparency for public interest purposes. Because **general public access to data** obligations are not targeted at a particular receiving party, this category can cover instances where there is no further processing of personal data after the addressee of the access obligation discloses the data. In such situations, the requirement of a lawful basis only applies towards the addressee of the data access obligation. Beyond these situations in the category of **general public access to data** and for the category of **regulatory transparency**, the grounds of legal obligation and public interest, can provide a lawful basis for further processing.[67]

Finally, for the category of **private party access to data**, the lawful grounds of legitimate interests of the data controller or consent are likely the most relevant ones.[68] For researcher data access under the DSA, public interest is also a suitable lawful ground considering the link with the identification of systemic risks.[69] For private parties beyond researchers, the link with the public interest is less strong– even though the fact that the entitlement to the data stems from EU legislation may plead for the relevance of the lawful ground of public interest.[70] Nevertheless, legitimate interests and consent may be more suitable as lawful grounds. Legitimate interests as a lawful basis require a balancing with the interests or fundamental rights and freedoms of data subjects.[71] As long as the receiving parties use the data in line with the purposes that justify the access under the DMA and the DSA, it will be hard to claim that their legitimate interests in using the personal data are outweighed by the rights and freedoms of data subjects.[72] It is worth noting, though, that the processing of sensitive personal data (such as personal data revealing racial or ethnic origin, political opinions, religious or philosophical

[65] For a discussion of the requirement of lawful ground in the context of the European data economy more generally, see Christiane Wendehorst, *Of Elephants in the Room and Paper Tigers: How to Reconcile Data Protection and the Data Economy*, *in* TRADING DATA IN THE DIGITAL ECONOMY: LEGAL CONCEPTS AND TOOLS (2017). At 334-337.

[66] As the processing can be considered as 'necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in' them. See Article 6(1)(e) of the GDPR.

[67] Respectively, Article 6(1)(c) and (e) of the GDPR.

[68] Respectively, Article 6(1)(f) and (a) of the GDPR. See 48 TOMBAL, *supra* note 93. At 357-360.

[69] Article 6(1)(e) of the GDPR. See also the European Digital Media Observatory, *supra* note 8. par. 69-70.

[70] Note, however, that Recital 45 of the GDPR clarifies that the processing should have a basis in Union or Member State law when it is based on public interest as a lawful ground and that this law has to meet certain conditions: it should 'determine the purpose of processing [...], specify the general conditions [...] governing the lawfulness of personal data processing, establish specifications for determining the controller, the type of personal data which are subject to the processing, the data subjects concerned, the entities to which the personal data may be disclosed, the purpose limitations, the storage period and other measures to ensure lawful and fair processing'.

[71] Article 6(1)(f) of the GDPR.

[72] For a discussion of this balancing exercise, see also 48 TOMBAL, *supra* note 93. At 360 noting the need for the data recipient to be very specific about the intended use of the shared data and the possibility for the data subject to object to the processing based on legitimate interests later on by relying on Article 21 of the GDPR.

beliefs)[73] and personal data concerning children[74] require special protection, which may affect the outcome of the balancing under the lawful ground of legitimate interests.[75]

As stated in the 2022 EDMO report, consent can be a suitable lawful basis in cases where the purpose of the data processing by the receiving party is clear and can be communicated to data subjects as such.[76] The standards for valid consent are strict and require a 'freely given, specific, informed and unambiguous indication of the data subject's wishes'.[77] Because reliance on consent requires an affirmative action of all individuals involved, it is not appropriate for bulk access.[78]

The three case studies below will provide more concrete discussions on the identification of lawful bases for the processing of personal data.

2.1.1.2 Other data protection and privacy requirements beyond the lawfulness of data processing

The existence of a lawful ground for data processing is but one of the relevant data protection rules that impact the implementation of data access obligations. Another important consideration is that the processing can only take place as long as it is necessary to achieve the purpose behind the lawful basis at stake. Necessity requires an assessment of whether the same purpose can be achieved through less intrusive means.[79] Furthermore, the GDPR also establishes general principles of purpose limitation and data minimisation. Purpose limitation requires that personal data is collected for specific purposes and not further processed in a manner that is incompatible with those purposes.[80] Data minimisation limits personal data collection to what is necessary in relation to the purposes for which it is processed.[81] As a result, where the data access obligations imposed by the DMA and the DSA require the processing of personal data, parties will need to ensure that this processing is limited to what is necessary to achieve the purpose of the relevant provisions. Other important requirements include the adequate protection and security of personal data and respect of data subject rights.[82] We will discuss some of these issues in more detail below in the context of our three case studies.

Beyond the GDPR, the more general right to privacy protected by Article 7 of the European Charter on Fundamental Rights and Article 8 of the European Convention on Human Rights (ECHR) is also relevant — in particular in the context of the exercise of powers by public authorities. In *Deutsche Bahn*, the General Court held that the exercise of the powers of inspection in competition cases under

[73] Article 9(1) of the GDPR.
[74] Article 6(1)(f) of the GDPR.
[75] See also European Digital Media Observatory, *supra* note 8. par. 82-87.
[76] *Id.* par. 71-75.
[77] Article 4(11) of the GDPR.
[78] Vikas Kathuria & Jure Globocnik, *Exclusionary conduct in data-driven markets: limitations of data sharing remedy*, 8 J. ANTITRUST ENFORC. 511 (2020). At 529-530.
[79] See also European Digital Media Observatory, *supra* note 8. par. 76-81.
[80] Article 5(1)(b) of the GDPR. Article 6(4) of the GDPR lists criteria data controllers have to take into account to consider whether processing for another purpose is compatible with the purpose for which the personal data are initially collected.
[81] Article 5(1)(c) of the GDPR.
[82] For a detailed discussion of how to apply all of these GDPR requirements in the context of platform-to-researcher data access, see European Digital Media Observatory, *supra* note 8.

Regulation 1/2003[83] is 'a clear interference with the latter's right to respect for its privacy, private premises and correspondence'.[84] More recently, the General Court applied a similar reasoning to requests for information sent out by the Commission in competition cases under Regulation 1/2003 in the context of a 2020 application for interim measures brought by Facebook. As argued by the General Court, interferences with or limitations of the exercise of the right to privacy have to comply with Article 52(1) of the Charter and Article 8(2) of the ECHR.[85] According to the General Court, Facebook's argument that the documents it is required to provide go beyond what is necessary for the Commission to establish the presumed infringements does not appear to be unfounded.[86] This was particularly the case for documents containing sensitive personal data such as 'documents containing private correspondence of employees concerning medical and autopsy reports and correspondence of employees at times of great personal distress'.[87] Because of the 'extremely personal and sensitive nature of medical data', their treatment requires an especially rigorous examination in the General Court's view.[88] While the nature of the personal data at stake may have raised additional scrutiny, it is clear from the interim order that the proportionality of the processing of employees' personal data should be taken into account in competition cases more generally. Similar considerations will apply to the enforcement powers of the Commission and the DSCs under the DMA and DSA in our category of **regulator access to data**. This also connects with the rule of law requirements, considered in section 2.1.4 below.

2.1.2 Balancing data access with intellectual property protection

2.1.2.1 The different legal mechanisms to protect commercially sensitive information in the EU

Although there is no specific property right for data as such, data can fall within the scope of existing protection regimes like copyright, the *sui generis* database right, and trade secrets.[89] Copyright protects the original expression of an idea by providing authors with temporary exclusive rights.[90] Although facts or data will hardly qualify as copyrightable material in themselves, databases set up by platforms may benefit from protection under copyright or *sui generis* database protection. Copyright protection is available for databases which by reason of the selection or arrangement of their contents constitute the author's own intellectual creation.[91] A relevant issue in this regard is whether the

[83] Council Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in Articles 81 and 82 of the Treaty (Regulation 1/2003) [2003] OJ L 1/1.
[84] Joined Cases T-289/11, T-290/11 and T-521/11 *Deutsche Bahn*, ECLI:EU:T:2013:404, par. 65. This part of the judgment was upheld by the Court of Justice in Case C-583/13 P *Deutsche Bahn*, ECLI:EU:C:2015:404.
[85] Case T-451/20 R *Facebook*, ECLI:EU:T:2020:515, par. 57.
[86] Case T-451/20 R *Facebook*, ECLI:EU:T:2020:515, par. 61.
[87] Case T-451/20 R *Facebook*, ECLI:EU:T:2020:515, par. 62.
[88] Case T-451/20 R *Facebook*, ECLI:EU:T:2020:515, par. 63.
[89] See Josef Drexl, *Designing competitive markets for industrial data*, 8 J INTELL PROP INFO TECH ELEC COM L 257 (2017). At 267-270 and Josef Drexl, *Data access and control in the era of connected devices*, REP. BEUC (2018), https://www.beuc.eu/sites/default/files/publications/beuc-x-2018-121_data_access_and_control_in_the_area_of_connected_devices.pdf at 59-106.
[90] Articles 2 and 3 of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (Information Society Directive) [2001] OJ L 167/10 as amended by Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market [2019] OJ L 130/92.
[91] Article 3(1) of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

setting up of the database is dictated by technical considerations, rules or constraints, in which case there is no room for an author's creative freedom and thus no copyright protection available.[92] *Sui generis* database protection, in its turn, applies to the contents of databases in so far as a substantial investment has been made in either the obtaining, verification or presentation of the contents.[93] The proposed Data Act clarifies that databases containing data obtained from or generated by the use of products or related services cannot benefit from *sui generis* database protection.[94]

Beyond copyright and the *sui generis* database protection, data may qualify for trade secret protection if: (1) it is secret; (2) has commercial value because it is secret; and (3) has been subject to reasonable steps to keep it secret.[95] Unlike copyright and *sui generis* database protection, trade secret protection does not provide rights holders with an exclusive right to prevent third parties from using the subject matter of protection. Trade secrets only protect against unlawful acquisition, use and disclosure,[96] but not against the use of information obtained through legitimate means such as own observation or independent creation.[97] Trade secret protection is particularly relevant for our purposes of balancing the interests of platforms with the public interest in data access, because a trade secret loses its value once the information is openly available. In this regard, the stakes for platforms to prevent disclosure of the information through data access are likely higher in the case of trade secrets than for other forms of intellectual property that grant exclusive rights. For these reasons, our analysis here focuses on the protection of trade secrets while acknowledging that copyright and *sui generis* database protection could be relevant in parallel. It is worth mentioning that trade secrets and intellectual property have already been invoked by platforms as a justification not to facilitate data access requests.[98] In terms of the nature of trade secret protection, it is worth illustrating its interaction with other forms of protection for confidential information available in the EU regulatory framework.

The most important for our purposes likely is Article 339 of the Treaty on the Functioning of the European Union (TFEU), which requires members of EU institutions and committees as well as officials and other EU servants not to disclose information covered by the obligation of professional secrecy, 'in particular information about undertakings, their business relations or their cost components'. This principle of professional secrecy is extended to the Commission, national authorities as well as auditors and experts in the DMA and the DSA.[99] The General Court has referred to three criteria to determine whether information falls within the scope of the obligation of professional secrecy: (1) the information is known only to a limited number of persons; (2) it is information whose disclosure is

[92] Case C-604/10 *Football Dataco*, ECLI:EU:C:2012:115, par. 39.
[93] Article 7(1) of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.
[94] Article 35 of the proposed Data Act. This is particularly relevant in the context of data collected by sensors in Internet of Things devices. By clarifying that the requirements for protection under the sui generis database right are not met in such cases, the EU legislator wishes to ensure that sui generis database protection does not interfere with the rights for businesses and consumers to access and share data under the proposed Data Act's data access right.
[95] Article 2(1) of 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 (Trade Secrets Directive) [2016] OJ L 157/1.
[96] Article 4(2)-(5) of the Trade Secrets Directive.
[97] Article 3 of the Trade Secrets Directive.
[98] See for instance: Facebook, *Correspondence to Max Schrems*, (2011), http://www.europe-v-facebook.org/FB_E-Mails_28_9_11.pdf
[99] Respectively: Article 36(4) DMA and Article 84 DSA.

liable to cause serious harm to the person who has provided it or to third parties; and (3) the interests liable to be harmed by disclosure must objectively be worthy of protection.[100]

In the context of the publication of a competition decision where the firm did not agree with the extent of information included in the non-confidential version of the Commission's decision, the General Court argued that the assessment of the confidentiality of a piece of information requires weighing the legitimate interests opposing disclosure and the public interest in transparency.[101] In particular, the General Court clarified that information cannot be considered to be covered by the obligation of professional secrecy where the public has a right of access to documents containing certain information.[102] As a result, the protection of professional secrecy is not absolute.

In another competition case, the General Court clarified that the Commission does not infringe on its duty to observe professional secrecy by failing to prohibit the disclosure to national courts of documents containing confidential information and business secrets. This would only be the case if the Commission allows such documents to be transmitted to national courts without taking the necessary precautions, for instance by informing the latter of the documents or passages of documents containing confidential information or business secrets.[103] A refusal to disclose documents to national courts would in the view of the General Court only be justified 'where it is the only way of ensuring 'protection of the rights of third parties', which in principle is a matter for the national courts, or 'where the disclosure of that information would be capable of interfering with the functioning and independence of the Community', which, in contrast, is a matter exclusively for the Community institutions concerned'.[104] This again points to the restrictive interpretation of professional secrecy when it interferes with public interests, in this case, the effective co-operation between the Commission and national courts in competition cases.

Beyond references to the obligation of professional secrecy and the protection of confidential information,[105] the DMA and DSA also refer to the legitimate interests of gatekeepers in the protection of their business secrets.[106] The DSA mentions trade secret protection as well, but the DMA does not refer to it. This may, however, not be the result of a conscious choice and merely be a matter of semantics. Nevertheless, trade secret protection can be seen as a different or specific form of confidential information or business secrets. In this sense, information covered by professional secrecy can include both confidential information and business secrets,[107] while trade secrets benefit from a concrete protection framework set out by the Trade Secret Directive.

According to the General Court, business secrets constitute 'information of which not only disclosure to the public but also mere transmission to a person other than the one that provided the information

[100] Case T-198/03 *Bank Austria Creditanstalt*, ECLI:EU:T:2006:136, par. 71.
[101] Case T-198/03 *Bank Austria Creditanstalt*, ECLI:EU:T:2006:136, par. 71.
[102] Case T-198/03 *Bank Austria Creditanstalt*, ECLI:EU:T:2006:136, par. 74.
[103] Case T-353/94 *Postbank*, ECLI:EU:T:1996:119, par. 90 and 92.
[104] Case T-353/94 *Postbank*, ECLI:EU:T:1996:119, par. 90 and 93.
[105] Articles 34(4), 36, 44(2), and 50(4) DMA, and Articles 37(2), 40(2), 40(5), 40(13), 42(5), 79(4), 80(2) and 84 DSA.
[106] Articles 14(4), 15(3), and 34(4) DMA and Article 79(4) DSA.
[107] See Case T-353/94 *Postbank*, ECLI:EU:T:1996:119, par. 86.

may seriously harm the latter's interests'.[108] Because business secrets are afforded 'very special protection',[109] the transmission of information containing business secrets requires to be 'subject to an appropriate procedure intended to safeguard the legitimate interests of the undertakings concerned in not having their business secrets disclosed'.[110] This already indicates that there are ways to transmit information and protect business secrets at the same time.

For our category of regulatory transparency more specifically, Regulation 1049/2001 regarding public access to European Parliament, Council and Commission documents[111] is relevant. Article 4(2) of this Regulation states that the institutions shall refuse access to a document where disclosure would undermine the commercial interests of a natural or legal person including intellectual property protection, unless there is an overriding public interest in disclosure. According to the Court of Justice, when applying Article 4(2) of Regulation 1049/2001, the institutions first have to consider whether the exception of the right of public access covers the document, then examine the existence of a reasonably foreseeable and not purely hypothetical risk of the commercial interests being undermined by disclosure, and lastly assess whether there is any overriding public interest justifying disclosure.[112] This indicates the need to conduct a balancing exercise between the commercial interests in intellectual property protection and the public interest in the disclosure of documents held by regulators.

2.1.2.2 Balancing trade secret protection and data access

The protection afforded by the DMA and the DSA to confidential information or business secrets does not by definition stand in the way of facilitating data access requests. For example, Article 40 of the DSA itself creates a broad data transparency obligation, and Article 34(4) of the DMA and Article 79(4) of the DSA, in regulating the right of access to files, explicitly state: 'nothing in this paragraph shall prevent the Commission from disclosing and using information necessary to prove an infringement'. While the independent audits to be conducted by VLOPs and VLOSEs under the DSA should ensure 'an adequate level of confidentiality and professional secrecy', Article 37(2) of the DSA also requires this protection to 'not adversely affect the performance of the audits and other provisions of this Regulation, in particular, those on transparency, supervision and enforcement'.[113] Article 40(5)(b) of the DSA does, however, provide VLOPs and VLOSEs with the option to request the Digital Services Coordinator to amend the data access on the ground that 'giving access to the data will lead to significant vulnerabilities in the security of their service or the protection of confidential information,

[108] Case T-353/94 *Postbank*, ECLI:EU:T:1996:119, par. 87.
[109] Case 53/85 *Akzo*, ECLI:EU:C:1986:256, par. 28.
[110] Case T-353/94 *Postbank*, ECLI:EU:T:1996:119, par. 87.
[111] Regulation (EC) No 1049/2001 regarding public access to European Parliament, Council and Commission documents [2001] OJ L 145/43.
[112] Joined Cases C-39 & 52/05 P, *Sweden and Turco v Council*, ECLI:EU:C:2008:374, par. 38-44. For a discussion in the context of access to documents concerning chemical substances, food and medicinal products, see Päivi Leino & Emilia Korkea-aho, *Who owns the information held by EU agencies? Weed killers, commercially sensitive information and transparent and participatory governance*, 54 COMMON MARK. LAW REV. (2017). At 1067-1068, 1075-1081.
[113] Recital 92 DSA further clarifies that the guarantee of confidentiality, security and integrity of the information, including trade secrets, provided by auditors 'should not be a means to circumvent the applicability of audit obligations'.

in particular trade secrets'. These provisions thus point to the balancing of interests that needs to be conducted.

The targeted data, the receiving party and the specificity of the legal obligation are important factors impacting the balancing exercise.[114] For example, the Trade Secrets Directive itself contains provisions to balance the need for protection with the need for openness in the case of regulatory access and transparency. Article 1(2)(b) of the Trade Secret Directive states that its provisions do not affect the application of EU or national rules 'requiring trade secret holders to disclose, for reasons of public interest, information, including trade secrets, to the public or to administrative or judicial authorities for the performance of the duties of those authorities'. This can be interpreted to imply that limited to no trade secret protection is available to information required by regulatory authorities to fulfil their enforcement mandates under the DMA and the DSA and to create transparency to the general public — our categories of **regulator access to data** and **general public access to data**.[115]

Similarly, Article 1(2)(c) of the Trade Secret Directive clarifies that its provisions also do not affect the application of EU and national rules requiring or allowing EU institutions and national authorities to disclose information submitted by businesses that they hold following EU and national law. This points at our category of **regulatory transparency,** whereby regulators are required by the DMA and the DSA to disclose certain information to the general public – possibly including information obtained from platforms. It thus seems there is limited room for companies to rely on trade secret protection vis-à-vis regulatory authorities for information they need to fulfil their tasks, including duties of disclosure imposed on them by EU and national rules.

For the remaining category of **private party access to data,** there is more leeway for platforms to invoke trade secret protection. Inspiration for how to balance the different considerations may be drawn from the right to data access and data portability of the General Data Protection Regulation (GDPR). While recognising the need of ensuring that the right to data access does not adversely affect the rights or freedoms of others, including trade secrets or intellectual property, Recital 63 of the GDPR explicitly states that 'the result of those considerations should not be a refusal to provide all information to the data subject'. With regard to the GDPR's right to data portability, the 2017 Guidelines of the Article 29 Working Party claim that a potential business risk cannot 'in and of itself serve as the basis for a refusal to answer the portability request' and that data controllers can find ways to transmit personal data in a form that does not disclose information protected by trade secrets or intellectual property rights.[116]

As a result, trade secrets and professional secrecy need to be protected when they interfere with obligations of data access but they cannot lead to such requests being denied completely. There is a responsibility of target parties as well as receiving parties to find ways in which data access can be facilitated without foregoing protection. The proposed Data Act is more specific than the DMA and

[114] See also Inge Graef, Martin Husovec & Nadezhda Purtova, *Data portability and data control: lessons for an emerging concept in EU law*, 19 GER. LAW J. 1359 (2018). At 1379.
[115] In particular if the mandated disclosure is well defined.
[116] Article 29 Working Party, 'Guidelines on the right to data portability', WP 242 rev.01, 5 April 2017, p. 12.

the DSA in this regard. For instance, in the context of a user's right to share data with third parties, the proposed Data Act lays down that trade secrets shall only be disclosed to third parties 'to the extent that they are strictly necessary to fulfil the purpose agreed between the user and the third party'. In addition, the third party has to take the measures that are agreed between the data holder and the third party to preserve the confidentiality of the trade secret.[117]

This general approach is also relevant for the balancing of trade secret protection and access to data rights under the DMA and the DSA. In the case of **private party access to data** (where the receiving party is previously known)**,** a balancing exercise should reflect the steps we outlined above, and should rely on a combination of legal and technical requirements as a way to facilitate data access. For example, if some data is both sensitive for a given company and essential for the understanding of a systemic risk, the best solution will likely be the establishment of technical safeguards that ensure that the data does not spread easily (such as, a restricted API, or, for certain types of sensitive data, even a clean room for data access) combined with a confidentiality agreement in which the private parties receiving access to the data commit to a restricted use and are held liable for any proven misuse — with the potential sanctions being proportional to the damage done by information disclosure. This allows for a better fine tuning of the different interests in a way that maximises access and protects rights. It is important to stress that the room for invoking trade secret or another form of intellectual property protection also depends on the legal mandate in question. In some areas, the DMA, the DSA or other laws establish a clear and targeted mandate,[118] diminishing the scope for reliance on trade secret or intellectual property protection. Because of the specific focus of the legal obligation, one can assume that the legislator already conducted a balancing exercise in advance and decided that disclosure prevails over the need for protection and secrecy in the area at stake. However, the broader the mandate and the more general the legal obligation is, the more room there will be for platforms to require stronger safeguards to ensure that the data sharing also protects their legitimate interests.[119]

This combination of technical and legal protections as a way to balance conflicting interests, though, is harder to implement in the case of **general public access to data**. That is because the exact purpose is the disclosure and availability of information to all. Still, in analogy to the reasoning of the Article 29 Working Party discussed above and to ensure the purpose of transparency is achieved, the need to protect trade secrets and other confidential information should, however, not stand in the way of enabling any form of general public access to data. As discussed above, Article 1(2)(b) of the Trade Secret Directive indicates that there is limited room to rely on trade secret protection in case of well-defined EU or national transparency and data access rules targeted at the general public for reasons of public interest.

What can be concluded from this discussion is that the target data, the nature of the receiving party and the specificity of the legal obligation matter for determining how trade secret protection should

[117] Article 5(8) of the proposed Data Act.
[118] E.g. the disclosure of advertisement data, our first case study below.
[119] Something we will also outline when discussing researcher data access in our second case study.

be balanced with the need for transparency and data access. While there may be cases where a particular form of data access cannot be facilitated because of intellectual property concerns, solutions should be found as much as possible to limit the range or detail of the data to be shared to protect the rights of businesses and still facilitate the necessary transparency. The scope for reliance on trade secrets is limited whenever legal obligations mandate the sharing of a specific type of data with a specific party. The strength of trade secret protection will be weaker towards regulators and the general public, but stronger for private party access to data–in this case, a combination of technical solutions and confidentiality agreements may provide a workable compromise.

2.1.3 Balancing data access with information security

A final (for our purposes) important balancing consideration is with regard to the trade-offs involved between granting access to data and ensuring the security of the data and of the systems involved. This gives rise to (at least) two concerns: the first is with regards to data access leading to accidental or even intentional leakage of data not intended for release — a concern we refer to as data security.[120] A second is with regards to the disclosure of information that, once known, diminishes the overall security of the system by facilitating the gaming of algorithms and other infrastructure — a risk we will call system security. Both the DMA and DSA refer to issues of security and integrity that need to be balanced against the obligations imposed on market players.[121] Similarly to the protection of trade secrets, privacy and rule of law, data security and information security are not absolute values. Security concerns need to be weighed against the interest in data access and, in this area above all, practical ways/technical solutions can be found to reach a proper balance. As there are fewer legal checks and balances to be followed, the two discussions below focus on practical guidance that can help mitigate concerns.

2.1.3.1 Data security

A full description of best practices for data security in potentially relevant scenarios is beyond the scope of this document, and we refer the reader to guidance from established computer security authorities.[122,123] Because data security is a common concern for nearly all APIs and other data-sharing mechanisms, best practices for controlling access to data securely are well established and adhering to these best practices should diminish concerns associated with the implementation of data-sharing mandates.

Data access concerns generally fall into two categories: data can be accessed by a wider or different audience than was intended, and data can be made available that was not intended to be shared. The first concern — ensuring that only authorised parties have access to data and that data is stored and

[120] That is, the more APIs, doors and so on, the higher the risk that the system may be breached by hackers.
[121] See for instance Article 42(5) of the DSA in the context of transparency reporting obligations and Article 7(9) DMA as regards interoperability of number-independent interpersonal communications services.
[122] NIST, *Access Control Policy and Implementation Guides*, (2023), https://csrc.nist.gov/Projects/Access-Control-Policy-and-Implementation-Guides
[123] DG Comp, *Best Practices on the disclosure of information in data rooms in proceedings under Articles 101 and 102 TFEU and under the EU Merger Regulation*, https://ec.europa.eu/competition/mergers/legislation/disclosure_information_data_rooms_en.pdf

accessed securely — can be managed with the use of a robust authentication and authorisation scheme, a practice of the principle of least authorisation, and regular audits of systems and user access. Protections to secure data should be proportional to the sensitivity of the data in question and the nature of the receiving party. We discussed privacy concerns specifically in section 2.1.1, but data may be sensitive or of particular interest to adversaries for reasons other than the privacy of particular users. The second concern, that data not intended to be shared is inadvertently made accessible, is a risk that must be managed with regular data audits verifying that data being shared does not exceed the current authorisation.

Some small number of security errors and data breaches are inevitable, and the goal of these measures is to minimise the frequency and severity of their occurrence. We recommend regular audits of user activity, systems, and data, and it is to be expected that some of these audits will uncover security incidents. For this reason, we additionally recommend that parties develop reporting plans for notifying relevant parties of breaches that occur and incorporate these reporting plans into their larger data access plans. These can help further minimise risks.

2.1.3.2 System security

Some obligations require platforms to increase the transparency of their algorithms or other systems to regulators, vetted researchers or even, in some cases, to the public more broadly. However, in some cases, there are key pieces of information that are integral to the security of those systems that would cease to function if they were widely known. This is a cybersecurity strategy generally known as security through obscurity, because the security guarantee offered is proportional to the obscurity of the system in question. An example of this are keyword lists. If a platform seeks to detect hate speech with a simple list of keywords, then if that list was widely known, it would be trivial for an adversary to engage in the targeted hate speech while simply avoiding those exact words.[124]

The DSA text manages this concern by limiting what data must be provided if it would pose a risk to 'Information Security'.[125] In some cases, it is obvious that the data being requested poses security concerns.[126] However, in other cases, the security issues may be more subtle.[127]

In cases where platforms argue that disclosure may diminish security through diminished obscurity,[128] our recommendation is that the involved parties engage in a threat modelling exercise that can facilitate the balancing of conflicting interests. Such an exercise should, at minimum, involve answering the following questions:

1) What would be the change in available information if the data was shared?

[124] For example, by using a not covered synonym, switching a for @, e for & or spacing the letters (depending on how the list is built).
[125] For example, Article 40(5b) allows platforms to request an amendment to a data access request by vetted researchers if such access: 'would lead to significant vulnerabilities in the security of their service'.
[126] Access keys are a good example.
[127] For example, information on how a given platform selects and presents search results may both facilitate the understanding of that platform's impact on society and facilitate the gaming of its services by motivated adversaries, which could exploit for both commercial (e.g. better rankings) or strategic gains (e.g. better spreading of disinformation).
[128] It is beyond the goals of this study to map all possible cases in which such allegations would be credible.

2) For which audience would this change be visible?
3) How likely is it that the data might leak to another audience?
4) Can the audience to whom the information has been revealed use this information to cause physical, emotional, or financial harm to a well-defined group of individuals?

Although platforms may seek to avoid disclosures that would pose more than a minimal risk of harm to individuals, we believe platforms may be able to comply with requests for access to even extremely sensitive data in certain situations by limiting exactly *what* they disclose or *how* they disclose it. As mentioned above, in particular in the case of **private party access to data,** the best solution is likely a combination of technical and legal means. The above threat modelling exercise can also guide decisions about which of these requirements are necessary. Additionally, platforms must consider the relative security of the parties with whom they are sharing data. For example, a relatively high degree of security can be offered by regulators in **regulator access to data** obligations, while virtually no security can be guaranteed when data is released to the public in **general public access** obligations.

As always in matters of security, there is a trade-off between the risk of harm and functionality. Platforms are in a difficult position in understanding how to make this trade-off in response to requests for data, since they lack the context to judge the relative importance or utility of the data being requested. Instead, we recommend that they pursue **consistency**. Platforms already have to make judgments about the security considerations of various data under their control when determining which employees have access to information. Internal measures taken to protect the security of information can form an indication of how to interpret security concerns of disclosing data to third parties. If data is widely internally available to many employees of a particular company with limited protective measures in place (technical and legal), it will be harder to claim that information security concerns are serious if the relevant data need to be disclosed to third parties. However, if platforms can demonstrate that access to a given database/data point is internally restricted to only a selected group of collaborators and/or is subject to major limitations and controls to safeguard the data, they will have a stronger claim to require stricter protections when revealing these data to those outside of their organisation.

Even in that case, however, regulators must still assess whether the arguments against the data sharing pass muster in a balancing test such as the one we articular in this Part II. Otherwise, platforms would have strong incentives to game the system by simply restricting internal access to data and then use this as an excuse to restrict external access to data as well.

2.1.4 Balancing data access with rule of law guarantees

Rule of law guarantees are particularly important in the context of our category of **regulator access to data**, where the information obtained can be used to establish the liability of a platform. To understand the applicable rule of law requirements under the DMA and DSA, an analogy can be made with competition cases where the European Commission and national competition authorities hold enforcement competences under Regulation 1/2003. Rule of law guarantees do not stand in the way

of facilitating any form of data access but they require a balancing, especially in terms of the extent and type of data access.

In order to enable platforms to assess the scope of their duty to co-operate and to safeguard their rights of defence, the Commission is under a duty to state reasons for a decision ordering an investigation to prevent so-called fishing expeditions.[129] Recital 116 of the DSA requires Member States to guarantee that DCSs take final decisions after a prior, fair and impartial procedure, including 'the right to be heard of the persons concerned, and the right to have access to the file, while respecting confidentiality and professional and business secrecy, as well as the obligation to give meaningful reasons for the decisions'. More specifically, Articles 67 and 69 of the DSA as well as Articles 23 and 23 of the DMA require the Commission to state the legal basis for requests for information and orders for inspections. In the context of data access, the duty to state reasons enables companies to assess whether the extent of access requested is proportionate – for instance, because of considerations relating to professional secrecy, trade secrets and privacy.

In addition, as stated in Recital 116 of the DSA, the exercise of powers by DCSs should be 'proportionate to, inter alia the nature and the overall actual or potential harm caused by the infringement or suspected infringement'. Recital 29 of the DMA states that the implementing measures imposed by the Commission on gatekeepers 'should be designed in an effective manner and in compliance with the principle of proportionality and the fundamental rights of the undertakings concerned, as well as those of third parties'. This requirement of proportionality also entails that the least far-reaching investigation measure should be applied when different measures are equally effective. In terms of technical arrangements for data access, it would thus imply that the easiest and least costly way of providing data should be used.

A final guarantee that is worth mentioning here is that information obtained during an investigation should not be used for reasons other than those indicated in the decision. This usually requires the information to be used in a particular proceeding, so that if regulators need the same information in an associated but different investigation, they must request it again.[130] In this regard, Article 38(5) of the DMA requires information exchanged between the Commission and national competition authorities to be used only for the purposes of co-ordination of the enforcement of the DMA and the rules referred to in Article 1(6) of the DMA (including national competition rules, the EU Merger Regulation, and national merger rules). According to the Court of Justice, the undertakings' rights of defence 'would be seriously endangered if the Commission were able to rely on evidence against undertakings which was obtained during an investigation but was not related to the subject-matter or purpose thereof'.[131] However, the Commission is not barred from initiating an inquiry to verify or supplement information it obtained in a previous investigation if that information indicates an

[129] Case C-583/13 P *Deutsche Bahn*, ECLI:EU:C:2015:404, par. 56; Case C-37/13 P *Nexans*, ECLI:EU:C:2014:2030, par. 34; and Case C-94/00 *Roquette Frères*, ECLI:EU:C:2002:603, par. 47.

[130] See Alexandre de Streel et al., *Enforcing the Digital Markets Act: Institutional Choices, Compliance, and Antitrust*, YALE TOBIN CENT. ECON. POLICY DISCUSS. PAP. NO 7 (2022). At 21-22; and OFCOM Policy Statement: Information gathering under section 145 of the Communications Act 2003 and section 13B of the Wireless Telegraphy Act 1949, 2005, at 4.4 https://www.ofcom.org.uk/__data/assets/pdf_file/0019/46045/policy.pdf

[131] Case C-583/13 P *Deutsche Bahn*, ECLI:EU:C:2015:404, par. 58; Case 85/87 *Dow Benelux*, ECLI:EU:C:1989:379, par. 18.

infringement of the competition rules. According to the Court of Justice, such a bar 'would go beyond what is required to safeguard professional secrecy and the rights of the defence and would thus constitute an unjustified hindrance to the Commission in the accomplishment of its task of ensuring compliance with the competition rules'.[132] This shows a balancing of interests is applicable to the rule of law guarantees as well, where the experience from competition proceedings can guide the Commission and DSCs in how to conduct investigations under the DMA and DSA.

Having at least sketched how to perform balancing exercises in relation to data access and privacy, intellectual property protection, information security, and rule of law, we can now move to our three selected case studies: (i) access to online advertisement databases (section 2.2); (ii) vetted researchers access to data (section 2.3) and (iii) the sharing of click and query data (section 2.4).

2.2 Access to Online Advertisement Databases – Article 39 of the DSA

Article 39 of the DSA requires providers of VLOP and VLOSE that display advertising in their online interfaces 'to compile and make publicly in a specific sector of the online interface, through a searchable and reliable tool that allows multicriteria queries, and through application programming interfaces, a repository' containing advertisement that was displayed in the platform. The Article requires that the 'repository' does not contain any personal data of the recipients of the services to whom the advertisement was or could have been displayed, and requires VLOP to 'make reasonable efforts to ensure that the information is accurate and complete'.

The displayed information, which must be kept for at least one year after the advertisement was presented for the last time, includes seven specific types of data that must be made available (Article 39(2) of the DSA):

(a) the content of the advertisement, including the name of the product, service or brand and the subject matter of the advertisement;

(b) the natural/legal person on whose behalf it was presented;

(c) the natural/legal person who paid for the advertisement;

(d) the period during which the advertisement was presented;

(e) whether it was intended to be presented to a particular group of recipient, and the main parameters used for targeting;

(f) an identification of whether the advertisement contains 'commercial communications',[133]

[132] Case C-583/13 P *Deutsche Bahn*, ECLI:EU:C:2015:404, par. 59; Case 85/87 *Dow Benelux*, ECLI:EU:C:1989:379, par. 19.

[133] Commercial Communications are basically direct forms of email and other marketing. More specifically, the eCommerce Directive defines commercial communications in Article 2(f) as 'any form of communication designed to promote, directly or indirectly, the goods, services or image of a company, organisation or person pursuing a commercial, industrial or craft activity or exercising a regulated profession'. The Article excludes 'information allowing direct access to the activity of the company, organisation or person, in particular a domain name or an electronic-mail address' and 'communications relating to the goods, services or image of the company, organisation or person compiled in an independent manner, particularly when this is without financial consideration'. Article 6 of the same Directive also requires that this

(g) the total number of recipients of the service reached and, if possible, the aggregate numbers in each member state.

If a VLOP/VLOSE removed an advertisement because it was illegal or incompatible with its terms of service, the Article then only requires the undertaking to maintain a subset of this information.[134] It also allows the Commission to issue specific guidelines on how to implement this obligation.[135]

Overall, this obligation can be summarised as follows:

1) **Target party:** VLOP/VLOSE
2) **Target data:** All paid ads presented in the platform
3) **Receiving party:** General public
4) **Timeliness:** Continuous
5) **Mode of access:** Queriable API and web portal

Based on these characteristics, one can categorise this obligation as requiring **General Public Access to Data**, a group that requires careful balancing of potentially conflicting priorities in terms of privacy, intellectual property protection, information security and rule of law guarantees.

Also important, this obligation has a clearly delineated **purpose:** to increase transparency with regards to which types of advertisement are displayed in VLOPs and VLOSEs, and how such companies target these ads. Indeed, Recital 68 clarifies that the provision aims to facilitate supervision and research into emerging risks brought about by the distribution of targeted digital advertising. Understanding the specific purpose/goal of each obligation is key for step 1 of the balancing exercise we propose. As discussed in our general notes on the balancing exercise (Part 2.1), it provides an objective function, so that regulators and other stakeholders can check whether the stated purpose of the data access request matches the purpose of the disclosure obligation enshrined in the legislation, and whether it helps advance the goals of such obligation. Beyond this, the legislative purpose can help shape alternatives that achieve the stated goals of the obligation in a manner that is least restrictive of other protected rights.

The implementation of these commands will require a range of specifications. While the list of information in Article 39(2) is taxative, other considerations are relatively open-ended and involve trade-offs.[136] In this section, we fill some of the gaps by providing implementation recommendations

communication be clearly identifiable, and Article 7 requires companies providing these services to check and respect opt-out databases, though it gives Member States some leeway to regulate these provisions. Article 26(2) of the DSA then requires online platforms to give recipients of the service a functionality to declare whether their content contains commercial communications. Article 39 requires companies to compile those in a searchable database.

[134] As per Article 39(3), in that case, the platform should maintain information on why it removed the advertisement (Article 17(3) of the DSA) and, if it was a result of an official order, information about that order specific in Article 9(2) of the DSA.

[135] After consultation with the Board and relevant vetted researchers.

[136] For example, what exactly qualifies as an advertisement? What is the territorial scope of the database? At what level of anonymisation can one be certain that the database does not contain any personal data of the recipients of the services?

that follow the five key variables identified in Part I above, to which we also add a short section 'other important practical considerations'.[137]

As mentioned before, we do not purport this to be an exhaustive exercise. Rather, the goal is to provide a roadmap of challenges and alternatives that can help authorities and companies consider trade-offs in this and other similar cases.

2.2.1 Targeted party

Article 39(2) of the DSA is clear that the obligation applies to all 'Providers of very large online platforms or of very large online search engines that display advertising on their online interfaces'.

2.2.2 Targeted data

A basic but important question is: what is an advertisement?

Article 3(r) of the DSA defines 'advertisement' as 'information designed to promote the message of a legal or natural person, irrespective of whether to achieve commercial or non-commercial purposes, and presented by an online platform on its online interface against remuneration specifically for promoting that information'. This definition is focused on: (i) the promotion of the message of a natural or legal person; that is (ii) presented by an online platform on its online interface, against specific remuneration. As such, we interpret it as focusing only on ads for which the platform controls the display and is directly paid for running them. This would exclude, for example, the growing share of influencer advertising (for which/or when platforms do not receive direct remuneration) as well as any ads run without payment.[138] It also seems to exclude ads run by the platforms themselves, which are not subject to direct remuneration.

While these omissions may be considered unfortunate from a transparency perspective, they substantially simplify the technical implementation by defining a clear target: Platforms must only identify ads they received direct payment for, a relatively straightforward task.

Another important question is with regard to territorial scope. Article 2 affirms that the DSA applies to 'intermediary services offered to recipients of the service that have their place of establishment or are located in the Union, irrespective of the place of establishment of the providers of those services'. This seems to imply that the relevant advertisement should be geolocated to the Union. For example, if a non-EU citizen accesses the social networking services of a VLOP/VLOSE inside the EU, any ads displayed to this person should be made available in the database. However, it also appears that if an

[137] This final section covers questions such as which party will be responsible for deciding conflicts or fund implementation costs.
[138] Such as, potentially mandated regulatory campaigns, charity campaigns, and so on.

EU citizen travels abroad and accesses the same services abroad, the VLOP/VLOSE is exempt from adding the advertisement to the database

In practical terms, this means that platforms can rely on a combination of IP addresses and the country in which the user is registered (if available) to identify the recipients of the service that are physically located in the EU.

2.2.3 Receiving party

Article 39(1) affirms that the information should be made 'publicly available in a specific section of [the VLOP/VLOSE] online interface'. We interpret this publicly available as information that is available to the general public at large, with no discrimination in terms of recipients. This is aligned with the requirements that the information is stripped of the personal data of the recipients.

2.2.4 Timeliness and mode of access

The DSA requires continuous access through APIs and a reliable tool that allows for multicriteria searches. In practical terms, this means the development of both an API with specific characteristics for bulk access and a searchable web portal for ad-hoc access by less technically sophisticated parties. This is because while APIs are flexible formats that are excellent mechanisms for requesting and serving large volumes of data, they are only accessible to users who can write code. The DSA's overall purpose of increasing transparency over which ads are displayed, how they are targeted and whether there are any hidden risks in this targeting process is better served by a mechanism such as a web portal that is accessible to a wider range of European citizens.

Article 39(1) of the DSA specifies that data must be made available via an API. In addition, there is also a requirement that parties develop 'a searchable and reliable tool that allows multicriteria queries'. This also strengthens the case for the development of queriable APIs, which are a combination of both. One of us authored a technical specification describing an implementation of such an API[139], and here we recommend an identical structure. Such an API should be queryable by date range, the geographic area down to the NUTS 3 level,[140] the identity of the advertiser, and a keyword match of the ad message.

APIs typically have 'rate limits' that constrain how often they can be queried. While we do not feel it would be appropriate to specify an exact rate limit in this document, we recommend they should be set such that a year of all data (the minimum period of time data is required to be retained and accessible) be collectable in 1 week or less. This threshold should be seen as a floor, rather than an exact target, leaving the parties with some flexibility on how exactly to implement the standard in a

[139] Laura Edelson et al., *A Standard for Universal Digital Ad Transparency*, KN. FIRST AMEND. INST. OCCAS. PAP. (2021).
[140] Eurostat, *National structures - NUTS - Nomenclature of territorial units for statistics*, https://ec.europa.eu/eurostat/web/nuts/national-structures

way that minimises potential downsides. Ultimately, though, this is a clear and direct legal command, and one that should be implemented effectively.

In terms of message format: We recommend that data returned by these queriable APIs be returned in JSON format, our only significant recommended departure from the standard proposed by Edelson et al. 'A Standard for Universal Ad Digital Ad Transparency', which recommends the HDF5 format (developed to store very large datasets). We believe this format is less appropriate for the queriable API specified by Article 39 than the time series data recording structures that the standard proposes. There are two reasons for this. First, for practical reasons APIs must return data in frames, or small chunks of the response that the receiver will collect serially, in case of network connection interruption. These serially served frames do not lend themselves to the advantages HDF5 format. Second, in practice, most users will query the API for the specific records they are looking for, rather than the bulk data that the HDF5 format was designed to accommodate.

In addition, **Technical Annex II** builds on Edelson *et al.* to provide a more detailed overview of important fields that should be present in this database, and how to populate them.[141]

2.2.5 Offsetting privacy, intellectual property protection, information security, and rule of law guarantees

2.2.5.1 Balancing privacy concerns

Ensuring that advertising databases do not compromise user privacy is a key challenge, and Article 39 of the DSA is clear in requiring that platforms 'shall ensure that the repository does not contain any personal data of the recipients of the service to whom the advertisement was or could have been displayed'.

While natural persons that buy the advertisement may want to invoke data protection to hide the fact that they acquired the ads, Article 39(2)(b) and (c) of the DSA are clear in stating that the published information shall include data on the natural persons on whose behalf the advertisement is presented and the natural persons who paid for the advertisements. In addition, Article 39(1) of the DSA only mentions the removal of information about the recipients of the service. In this case, the express command for disclosure will likely qualify as a legal obligation under the GDPR and thereby in itself form a lawful ground to share personal data. In line with step 7 of our proposed balancing exercise, the legislator considered the two conflicting interests and opted for transparency in this case.

The same applies to Ad targeting information, although in a more complex consideration. Ad targeting information — one of the categories required to be made transparent — is the key mechanism that may lead to the exposure of personal data through advertising-related databases. Whether a set of targeting parameters for any particular ad is privacy-compromising comes down to the number of

[141] Also available at this link.

users described by those parameters, and whether data published about ad targeting can be connected to other datasets to narrow the number of users to whom a set of targeting criteria might apply. This poses a clash between the legal obligation to disclose the data and privacy protections. A question, then, becomes whether legal and technical safeguards can ensure the disclosure of the data while minimising harms — step 6 of our framework.

In this case, the answer is largely yes. First, platforms must ensure that the targeting information they publish about each ad describes a large enough pool of people that any individual who may have seen the ad is not identifiable. In previous work, Edelson et al. proposed the threshold of 100 users as a minimum described audience size for any publishable ad targeting parameters,[142] and we reinforce this recommendation. However, this concern is only relevant for certain categories of ad targeting. For example, many ads are targeted on the basis of customer lists: advertisers provide VLOPs with a list of email addresses they have previously sold products to, and ask the platform to match the email addresses provided to known users of the platform. Similarly, some VLOPs offer 'lookalike audiences' to advertisers, where advertisers can upload customer lists and platforms serve their ads to audiences that are similar to but not the existing customers on the list. In list-based cases such as these, platforms can meet the requirement to make transparent 'the main parameters used for that (targeting) purpose' by simply disclosing the targeting mechanism (advertiser's list, lookalike based on advertiser's list, 3rd party list with list owner name, and so on). In cases of lookalike lists, platforms should additionally disclose at least the audience features with the highest degree of commonality (such as geography, gender, or income).

In other cases, advertisers provide characteristics that describe the audiences they wish their ads to be served, and ask platforms to match the characteristics to the correct audiences. We are aware of platforms that allow advertisers to define audiences by geographic area, age, gender, race & ethnic affiliation, income, hobbies, life events, health considerations, education, profession, and many other categories.[143] Much attention has been paid to the practice of advertisers targeting users based on behavioural characteristics, but for purposes of understanding how privacy revealing any particular targeting parameter is, the central concern remains how many people are described by that parameter. We are not aware of any VLOPs that allow advertisers to target audiences smaller than 100 users, but there is (apparently) no technical or legal barrier to such a practice. This means that in the future, VLOPs may need to genericise targeting criteria before making them public. For example, a VLOP may allow an advertiser to specify a single street address as a targeting parameter, which would describe potentially only a single user. In cases such as these, Edelson *et. al* recommend making the type of parameter (in this case, geographic) transparent at a lower level of specificity. This would mean that instead of publishing the exact street address targeted the platform can reveal the associated (but broader) postal code.

[142] Edelson et al., *supra* note 138.

[143] It is important to notice that this should change in the EU, as the DSA blocks online platforms from targeting advertisement based on profiles that rely on special categories of personal data defined by Article 9(1) of the GDPR. See DSA Article 26(3) and Recital 69.

This is always a dynamic game, one that should reflect considerations on how privacy revealing a *set* of targeting parameters are given the overall size of the audience that is described by ***the combination of all the criteria***. In cases where the combination of targeting parameters describe an audience smaller than our recommended threshold even if individual criteria do not, we recommend that the most specific criteria be listed as their category (rather than their exact parameter) iteratively until the set describes an audience larger than 100 users. For example, if the targeting criteria of a particular zip code, the gender female, and an educational level of PhD described fewer than 100 people, the criteria of 'educational level of PhD' can be listed as 'educational attainment'. This provides no additional information about the audience described, but still retains some information about advertiser targeting intent.

The other factor to be considered when evaluating how privacy-compromising a dataset might be is whether one set of data can be connected to another, and if so, if that combined dataset might be identifying. We have so far only discussed the combination of different targeting parameters, but Article 39 of the DSA calls for information about 'aggregate numbers for the group or groups of recipients to whom the advertisement was targeted specifically'. In cases where an ad was served to fewer than 100 members of a listed targeted group, we recommend that platforms instead list '<100 impressions' for the group in question.[144]

Given the rapid change in ad technology that has occurred over the last decade, we are aware of the impossibility of foreseeing every category of advertising or ad targeting that may be created. Therefore we further recommend two general principles, which we have used in developing these specific recommendations. First, care should be taken to ensure that the size of the audience who can be inferred to have seen an ad does not fall below the threshold we recommend of 100. We have focused on targeting parameters as the most likely way this might happen, but future forms of advertising may be linkable to users in other ways we don't yet know how to define. Second, we recommend that care be taken to ensure that no two ad records are linkable as having been seen by the same user. Currently, this means avoiding the use of anonymised user ids or cohort ids, but again we recognise that other ways of doing this type of 'linking' may be possible in the future, so we offer this as a general, rather than a specific, recommendation.

2.2.5.2 Balancing concerns of intellectual property protection

Article 39 of the DSA does not itself refer to any balancing with interests in commercial confidentiality, trade secrets or intellectual property protection. This means that any claims regarding these interests would fail in step 5 of our proposed balancing framework, according to which only harms explicitly recognised by the law can be weighed against the interest in disclosure. In addition, the Trade Secret Directive specifies that its protection does not affect the application of EU rules requiring the disclosure of information to the general public for reasons of public interest.[145] In other words, it is

[144] This is to avoid small groups being identified, which may be privacy compromising when combined with other data.
[145] Article 1(2)(b) of the Trade Secret Directive, as discussed in section 2.1.2.2 above.

hard to claim trade secret protection against disclosing information that the law requires to be openly available to the public.

Beyond this, it is also worth noting that the extent of data access is well-defined by law because the scope of the information to be provided in the repository is laid down by Article 39 of the DSA itself. Public access implies that confidentiality of the data is hard to achieve. While public access is the most far-reaching form of data access from the perspective of ensuring commercial confidentiality, the availability of information is precisely the aim the legislator had in mind. Ultimately, the legislator did the balancing, and opted for transparency.

2.2.5.3 Information security concerns and rule of law guarantees

With regard to information security and rule of law, we envision limited concerns as a result of Article 39 of the DSA.

In terms of establishing potential liability, one cannot exclude that the repository published by VLOPs and VLOSEs gives rise to insights that can inspire the start of an investigation. Once an investigation is opened, rule of law guarantees will apply. In the competition context, the Court of Justice clarified that the Commission is not overstepping its powers when initiating an inquiry to verify or complement information it received in a previous investigation and that alerted it to a possible competition violation.[146]

A similar reasoning seems applicable here, where the information made publicly available under Article 39 of the DSA could form the basis for a decision of the relevant authorities to start an investigation into certain behaviours of VLOPs or VLOSEs. Rule of law guarantees would not stand in the way of doing so, as the platforms can defend themselves during the course of the investigation and the relevant protections apply, including an obligation of the Commission/DCSs to state reasons for decisions, a right to be heard and a right to have access to the file. Indeed, in some jurisdictions it is customary that regulators issue a new request for information when formally opening an investigation, even in relation to data that they already hold. This provides a formal opportunity for the investigated parties to challenge the scope and the applicability of the data and ensure their rights of defence and right to a fair trial.

2.2.6 Other important practical considerations

According to Article 39(3) of the DSA, the Commission can 'issue guidelines on the structure, organisation and functionalities of the repositories'. As a result, the practical implementation of the obligation is under the control of the Commission, which can also decide on conflicts between different interests. For the data protection aspects, the competent data protection authority will need to be involved.

[146] Case C-583/13 P *Deutsche Bahn*, ECLI:EU:C:2015:404, par. 59; Case 85/87 *Dow Benelux*, ECLI:EU:C:1989:379, par. 19. As discussed in section 2.1.4 above.

Finally, because the provision is framed as an obligation incumbent on VLOPs and VLOSEs, it is reasonable to let them bear the costs of compliance as they usually bear with any other legal obligation. This means that the ad databases should be freely accessible to the public both through the web portal and through the API. It is worth noting that this obligation only applies to VLOPs and VLOSEs, so very large companies with enough resources to bear compliance costs.

2.3 Access to Data for Vetted Researchers – Article 40(4) of the DSA[147]

Article 40(4) of the DSA requires VLOPs and VLOSEs to provide access to data to previously vetted researchers 'for the sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the Union, (...) and to the assessment of the adequacy, efficiency and impacts of risk mitigation measures'. This is a significant and welcome innovation by the DSA,[148] one that reflects years of concerns that platforms' opacity prevented a proper understanding of their broader impacts on societies.[149] It also partially comes as a result of a general conclusion that voluntary data-sharing measures have been imperfect at best.[150]

More specifically, the obligation enables the Digital Services Coordinator of establishment to require VLOPs/VLOSEs to provide vetted researchers with access to internal platform data. Platforms have 15 days following receipt to request an amendment if they: (i) do not have access to the data; or (ii) believe that giving access to the data would lead to significant vulnerabilities in terms of information security or the protection of confidential information (such as trade secrets).[151] In such cases, the VLOP/VLOSE must propose one or more alternatives that would provide effective access in a way that is still appropriate and sufficient for the purpose of the request but that is less intrusive.[152] The DSC of establishment issues a final decision on what exactly the platform is required to do, and it must also inform the Commission and the Board about any requests.

It is possible for researchers to file requests with the DSC of the Member State of the research organisation they are affiliated with. In that case, this 'auxiliary' DSC conducts an initial assessment and, upon approval, sends the application and supporting documents to the DSC of establishment. The latter, however, has the power to issue final decisions.[153] The DSC that awarded the status of

[147] Note that we focus on Article 40(4) of the DSA and do not consider Article 40(12) of the DSA, which requires VLOPs and VLOSEs to provide researchers in certain circumstances with real-time data that is publicly accessible in their online interface.
[148] Mathias Vermeulen, *Researcher Access to Platform Data: European Developments*, 1 J. ONLINE TRUST SAF. (2022). At 1-2.; Alex Engler, *Platform data access is a lynchpin of the EU's Digital Services Act*, BROOKINGS (2021), https://www.brookings.edu/blog/techtank/2021/01/15/platform-data-access-is-a-lynchpin-of-the-eus-digital-services-act/
[149] Stigler Committee on Digital Platforms, *Stigler Committee on Digital Platforms: Final Report*, (2019), https://perma.cc/RWV9-KRL5 Robert Gorwa & Timothy Garton Ash, *Democratic transparency in the platform society*, SOC. MEDIA DEMOCR. STATE FIELD PROSPECTS REFORM 286 (2020). At 2; Jef Ausloos, Paddy Leerssen & Pim ten Thije, *Operationalizing Research Access in Platform Governance What to learn from other industries?*, (2020), https://algorithmwatch.org/de/wp-content/uploads/2020/06/GoverningPlatforms_IViR_study_June2020-AlgorithmWatch-2020-06-24.pdf At 8-10; 13-16.
[150] Ausloos, Leerssen, and ten Thije, *supra* note 148. At 17-21. See also, Davey Alba, *Facebook sent flawed data to misinformation researchers.*, THE NEW YORK TIMES, Sep. 10, 2021, https://www.nytimes.com/live/2020/2020-election-misinformation-distortions/facebook-sent-flawed-data-to-misinformation-researchers?smid=url-share Nathaniel Persily, *A proposal for researcher access to platform data: The platform transparency and accountability act*, 1 J. ONLINE TRUST SAF. (2021). At 1-2.; Engler, *supra* note 147.
[151] Article 40(5(a) and (b) of the DSA.
[152] Article 40(6) of the DSA.
[153] Article 40(9) of the DSA.

vetted researcher is also responsible for ensuring that the vetted researchers maintain their status, and should terminate access if this is no longer the case — though only after a formal investigation and after giving the vetted researchers the opportunity to react to the findings of violation.[154] The DSCs must communicate the name of the vetted researchers as well as the purpose of their research to the Board. The DSA also grants the Commission the power to adopt delegated acts laying down technical conditions for the sharing of data.

Finally, it is noteworthy that the DSA requires VLOPs to give researchers, organisations and associations[155] access to publicly accessible data, including real-time data where possible.[156] However, this obligation differs in terms of the target data and the receiving parties–which are no longer fully vetted. As such, it is in between a **General Public Access to data** and a **Private Party Access to Data** provision and many of our propositions below may or may not apply. We therefore *treat Article 40(12) as a separate obligation that is **not** the object of this case study,* meaning that we do not directly address it.

Overall, Article 40(4) of the DSA can be summarised as follows:

1) **Target party:** VLOP/VLOSE
2) **Target data:** All data
3) **Receiving party:** Vetted researchers
4) **Timeliness:** Triggered on action
5) **Mode of access:** Varies depending on data requested. There is an express mention to API access.

Based on these characteristics, one can categorise this obligation as requiring **Private Party Access to Data**, a group that also requires careful balancing of many potentially conflicting priorities in terms of privacy, intellectual property protection and information security. The implementation of these commands will require a range of specifications and the balancing of conflicting legal protections, which we discuss below.

2.3.1 Targeted party

Article 40(4) targets all VLOPs and VLOSEs.

[154] Article 40(10) of the DSA.
[155] These must meet certain criteria, laid out in Article 40 such as: (b) independence from commercial interests; (ba) disclosure in the sources of funding; (c) the capacity to preserve data security and confidentiality requirements of the data, including the ability to describe measure in place for this end; and (d) a justification of the necessity and proportionality of the data access requested (given the purposes of the research), the access timeframes and the expected results. This provision has been called the 'CrowdTangle Provision', and aims to facilitate access to already public data by protecting data scraping and, potentially, requiring platforms to facilitate APIs and other forms of access to this data. See Paddy Leerssen, *Platform research access in Article 31 of the Digital Services Act: Sword without a shield?* (2021), https://verfassungsblog.de/power-dsa-dma-14/
[156] Article 40(12) of the DSA.

2.3.2 Targeted data

Article 40(4) of the DSA calls for access to data 'for the sole purpose of conducting research that contributes to the identification and understanding of systemic risks' without any limitation on what kinds of data undertakings may be required to provide. Recitals 79-83 list four categories of systemic risks that are certainly the subject of the Regulation as defined whenever: (i) the service leads to the dissemination of illegal content;[157] (ii) the service negatively impact fundamental rights;[158] (iii) the service has actual or foreseeable negative effects on democratic processes, civic discourse, electoral processes or public security; and (iv) the service has general negative impacts on public health.[159]

In addition, Article 40(4) also calls for access to data to help with the 'assessment, efficiency and impacts of risk mitigation measures' that VLOPs and VLOSEs adopted to diminish systemic risks pursuant to Article 35 of the DSA. In other words, access to data has a dual purpose of facilitating the identification of systemic and an understanding of whether measures adopted by digital platforms have achieved their risk mitigation goals.[160]

As can be seen, these are very broad provisions that can (in theory) enable access to data on a wide variety of topics.

Recital 96 lists certain types of data that can be accessed through the provisions of Article 40, namely: data on the accuracy, functioning and testing of algorithmic systems for content moderation, training data and, apparently, even the code of algorithms. It is not clear from the text whether the listed types of data are only relevant for the monitoring of compliance by DSCs and the European Commission (established by Articles 40(1) and 40(2)), or whether they also fall within the range of the data to which researchers can ask access (Article 40(4)). Still, the reference in the recital to the importance of investigations of researchers on systemic risks 'for bridging information asymmetries and establishing a resilient system of risk mitigation, informing providers of online platforms, providers of online search engines, Digital Services Coordinators, other competent authorities, the Commission and the public' may indicate that the legislator intended to enable researchers access to a broad range of data as well. An expansive read of the scope of the available data is also aligned with previous calls for researcher data access that justified the creation of the obligation in the first place.[161] In our opinion, this *expansive read of the scope of the available data is the best interpretation of the provision* — not only because Article 40(4) calls for researchers to help assess systemic risks and the impact of risk

[157] Including child sexual abuse material, illegal hate speech, the sale of articles that are either prohibited, dangerous or counterfeited, and even the systematic dissemination of misleading or deceptive content (including disinformation).
[158] Such as those protected by the EU Charter, including human dignity, freedom of expression and information, the right to a private life and to data protection, the right to non-discrimination and the right of the child and consumer protection.
[159] Including physical and mental wellbeing or gender based-violence.
[160] Mathias Vermeulen, Researcher Access to Platform Data: European Developments. Journal of Online Trust and Safety, September 2022, at 2-4.
[161] See Persily, *supra* note 149. At 2, stressing that any obligation that grants researchers access to internal platform data should be based on three key principles: '(1) access by researchers not chosen by the to (2) the same data the firms' own data analysts can analyse but (3) in a secure environment that minimises any risks of disclosure of user private data'. See also Caitlin Vogus, *Improving Researcher Access to Digital Data: A Workshop Report. Center for Democracy and Technology Report*, (2022), https://cdt.org/insights/improving-researcher-access-to-digital-data-a-workshop-report/ (stating that data necessary for proper research on digital platforms includes not only advertising data and public/semi-public data, but also information on content moderation, engagement, historical content data and deleted data, ranking and recommendation algorithms, and real time data, among others).

mitigation measures, but also because a too narrow read would risk hollowing out this important obligation.

This is not to say that access is unlimited: the DSA establishes important safeguards that protect undertakings and users from abuse, and these should be respected (which we discuss in more detail below). Indeed, many questions will likely arise during the implementation of Article 40(4) of the DSA about the range of data included in the scope and the extent of co-operation required of platforms. While we believe that no group of data should be excluded from the scope of the obligation *ex ante*, considering what is proportionate in the circumstances of a particular request will be key to achieve an effective and workable implementation of Article 40(4) of the DSA for all parties involved.

In addition, we stress the importance that *every request should be made public in a centralised database, as well as the justification given by the company to deny the request and the final decision by the DSC*. This will not only facilitate the building of a common pool of knowledge on what data is available and what is not, but also enable the broader community to challenge decisions — either granting access or denying — that are abusive, do not respect user privacy, and so on.

2.3.2.1 Territorial scope

Another first order question is what is the territorial scope of the target data. As discussed in Part 2.2.2 above, the DSA is only applicable to 'intermediary services offered to recipients of the service that have their place of establishment or are located in the Union'.[162] This limitation, together with general principles of non-extraterritoriality of laws and regulations, indicates that the targeted data must be connected to European recipients of the service. Therefore, one cannot use Article 40(4) to obtain information about events not linked to the EU.

Because of the cross-border and fluid nature of the internet, this is much easier said than done. Content moves around freely, and the link to the EU may be more or less clear. A good (albeit now non-EU), concrete example of some of these challenges was OFCOM's determination on whether the attacks carried out in Buffalo, New York, in May 2022[163] were within the purview of the agency's powers to regulate the content displayed by Video-Sharing Platforms, or to 'protect all users from material that is likely to incite violence and hatred and material relating to terrorism, racism and xenophobia'.[164]

In that case, OFCOM determined that it was, despite the fact that all events took place in the US. This was due to the fact that the attack: (i) was live-streamed on Twitch, which is a UK-established

[162] Article 2(1) DSA.

[163] As OFCOM summarises: 'On 14 May 2022, an 18-year-old far-right extremist allegedly undertook a violent terrorist attack on a supermarket in a predominantly Black neighbourhood in Buffalo, New York. During the attack, he killed ten individuals and injured three others, the majority of whom were Black. The attack was livestreamed1, recorded and disseminated on several online services along with a manifesto and 'diary'. Based on subsequent hate crime charges brought against the alleged attacker by the US Justice Department, the attack appears to have been racially motivated and drew significant inspiration from previous far-right attackers, including one carried out in Christchurch, New Zealand, in 2019. It was perpetrated by an individual seeking to maximise the spread of footage of their livestreamed attack'. See OFCOM, *The Buffalo Attack: Implications for Online Safety*, (2022), https://www.ofcom.org.uk/__data/assets/pdf_file/0019/245305/The-Buffalo-Attack-Implications-for-Online-Safety.pdf at 4.

[164] *Id.*, at 4.

platform; (ii) copies of the content where then shared on other UK-based platforms; (iii) the attack triggered discussions on the role of online platforms in radicalisation, which are discussions also taking place in the UK; (iv) the attack was an important event that led to significant dissemination and discussion.[165]

Ultimately, DSCs will need to determine specific links and the limits of borderline content on a case-by-case manner. This means that stakeholders must develop a combination of *objective* and *subjective* criteria to help assess whether a given specific request is under the purview of the DSA. We believe that the objective criteria should include at least whether the platform in question is a designated VLOP/VLOSE and the number of European users who have been exposed to or engaged with the content and the language (among others). Subjective criteria should include the relevance of the topic under investigation and its connection to the specific systemic risk that justified the data access request, whether the data is absolutely necessary for the research question, and the strength of the link or if alternatives exist (also among others).

A final question is on whether there are territorial limitations in terms of where the data is stored — meaning that an undertaking can refuse to provide access to data on European recipients of the service[166] because the data is stored in servers outside of the EU. There is a large body of scholarship discussing the US CLOUD Act, the limits of cross-border data requests and the merits and demerits of data localisation.[167] We will not enter this discussion other than to point out that allowing undertakings to decline to provide data because it is not physically located in the EU would be a major loophole and produce perverse incentives that can significantly weaken the DSA. This interpretation is also in line with Article 2(1) of the DSA, according to which the DSA applies to EU recipients of intermediary services 'irrespective of where the providers of those intermediary services have their place of establishment'.

2.3.2.2 Current data versus new data

Another important question is whether the 'targeted data' includes solely data that is regularly produced and stored by the platforms, or whether researchers can request the VLOPs and VLOSEs to produce new data that can help with the identification and understanding of systemic risks.[168] While Article 40(5a) states that undertakings may ask DSCs to amend data access requests because of a lack of data access, Article 40(6) also requires companies to provide alternative means which are appropriate and sufficient for the purpose of the request.

Overall, a careful balance needs to be struck keeping in mind what is proportional to the objective of enabling research that contributes to the detection, identification and understanding of systemic risks. On the one hand, it does seem unreasonable to impose on VLOPs and VLOSE a general obligation to

[165] *Id.* and conversations with OFCOM.
[166] Meaning undertakings and natural persons that are physically located in the European Union.
[167] See, for example, Paul Schwartz & Karl-Nikolaus Peifer, *Data Localization Under the CLOUD Act and the GDPR*, 20 COMPUT. LAW REV. INT. 1 (2019).; Anupam Chander & Uyen P. Le, *Breaking the Web: data localization vs. the global internet*, EMORY LAW J. (2014).
[168] For example, in order to answer a given research question researchers may need the results of a series of A/B tests that the platform could implement, or may need different types of data that the platform does not collect.

acquiesce to open-ended requests to produce new types of data—not the least because these requests can be financially costly, burdensome for the platforms to implement, and depending on the nature, may expose recipients of the service to experimentation that decrease the quality of the product, are unethical, and so on.

On the other, Recital 97 of the DSA is explicit in stating that 'consideration of the commercial interests of providers should not lead to a refusal to provide access to the data'. In addition, enabling providers to simply claim as a defence that the data are not regularly collected and would be costly to produce creates perverse incentives in which VLOPs and VLOSE diminish their data collection efforts in order to reduce the extent of disclosure on possible systemic risks. Whilst the DSA does not have an explicit anti-circumvention provision like the DMA,[169] it does require platforms to provide access to internal data in a way that effectively enables researchers to identify and understand systemic risks.[170] This means that platforms are expected to be proactive as well in gathering relevant data themselves, and it will be incumbent on DSCs and the Commission to ensure that platforms collect and provide access to data necessary that would be expected in the regular course of business. The DSCs and the Commission should also ensure, likely after engagement with the research community and with platforms, that VLOPs/VLOSEs start collecting new data that is vital to answer certain questions at the heart of the systemic risk assessment that the DSA mandates.

The overall balancing framework we proposed in Section 2.1 can provide guidelines on how to consider requests to access new data. Ultimately, such a request would depend on whether the data is necessary to answer a very important research question associated with the detection or minimisation of a specific systemic risk recognised by the DSA (steps 1 and 2); whether it is the minimum necessary to assess that risk, and if the questions cannot be answered by using another database (step 3); whether the collection of new data may harm users or lead to decreases in privacy, security vulnerabilities or violations of intellectual property (steps 4 and 5); and whether a combination of technical or legal measures can mitigate these risks (step 6). If not, then regulators would need to weigh the importance of the research in question against the concrete harms that it may cause after legal and technical safeguards are implemented (step 7).

A final order question is with regard to timing. Here also there are important considerations on the proportionality of requests. For example, imagine that a major event has taken place and that a given VLOP or VLOSE has important data pertaining to the understanding of why and how this event took place, but that these data are not structured as a proper database.[171]

To ensure the effectiveness of researcher data access, it seems reasonable to require platforms to structure the data they already have under their control into a database if this is necessary for

[169] Article 13 of the DMA.

[170] Article 40 of the DSA and Recitals 96 and 97. In the past, others have asked that regulators be given the powers to 'require the production of datasets deemed reasonably necessary for providing answers to questions researchers ask' as a way to disencourage platforms from simply stopping the collection of some forms of data in response to data access mandates. Persily, *supra* note 149., at 4.

[171] Building on the Buffalo Shooter example above, an example would be researchers trying to understand the role of social media platforms in the context of increasing extremism linked to a particular event or protest – something can be within the purview of the systemic risks to democracy established by Recital 82 of the DSA. The question is what can be expected of social media platforms if they do not have a structured database with some form of specific information on recipients of the service that could be useful to understand this link.

understanding a certain systemic risk unless the platform can provide the DSC with reasons specifying that:

> (i) the request is technically impossible;
>
> (ii) complying with the request would lead to significant violations of privacy or intellectual property protections that cannot be mitigated by a combination of technical and legal protections; or
>
> (iii) the request would be totally disproportional and there are alternative data that could help researchers achieve similar results (as better discussed below).

In that case, regulators will also need to balance conflicting interests, and our framework also provides relevant guidelines.

2.3.2.3 Understanding what types of data are available to researchers

One of the main challenges in ensuring effective access to data for researchers will be understanding what types of data are available and what are not. This will be particularly true in the beginning, when researchers, companies and regulators will have limited experience with such requests.

To ease this burden, our recommendation is that platforms should publish dataset descriptions and codebooks for their most commonly requested datasets on public archive sites such as Zenodo (managed by CERN) or one of the many Dataverse instances that are hosted in Europe. Indeed, the need for the development of a system that helps researchers understand what types of data are available for research in the first place was one of the key recommendations of the CDT Workshop report on researcher data access as well.[172] Public dataset descriptions and codebooks can ease the burden for regulatory bodies as well. A major component in evaluating vetted researcher access to data will be evaluating proposals to ensure that requested data can answer specific research questions, and understanding the privacy concerns associated with requested data — something that allows for appropriate protections can be put into place. Such descriptions will help regulators avoid replicating work when making determinations for common datasets.

In addition, and as we stressed above, *every request for data access should be made public in a centralized database, as well as the justification given by the company to deny the request and the final decision by the DSC*. This should also facilitate the building of a common pool of knowledge on the available data and the safeguards needed to access such data.

[172] See Vogus, *supra* note 160. At 21-23. Their recommendation was the creation of codebooks for covered and not-covered data.

2.3.3 Receiving party

Receiving parties are previously vetted researchers that are affiliated with a research organisation within the meaning of Article 2 of the Directive on Copyright in the Digital Single Market.[173] That list is composed of universities, research institutes or any other entity which has as a primary goal to conduct scientific research or to carry out educational activities that also involve scientific research. In addition, these parties must also: (i) be not-for-profit entities, or reinvest all profits in scientific research; (ii) pursue a public interest mission recognised by a Member State; and (iii) not directly promote the interests of an undertaking that has influence over such organisation.[174] Recital 97 makes it clear that civil society organisations that conduct research to support a public interest mission are also eligible for vetting.

Article 40(8) establishes further requirements for vetted researchers:

> (i) they must be independent from commercial interests, and specific application requests must clearly disclose sources of funding;
>
> (ii) they must show that they can maintain data security and confidentiality requirements, describing in detail appropriate technical and organisational measures to further this;
>
> (iii) they must justify the necessity and proportionality of a given request, as well as how the results will contribute to the detection, identification, understanding and mitigation of systemic risks; and
>
> (iv) they must commit to making the results of the research available free of charge within a reasonable period.

While these guidelines provide an important benchmark, there are numerous important outstanding questions on how to implement this vetting in practice. It is beyond the scope of this report to flesh out these requirements in full. Still, we partially address items (i), (ii) and (iii)) in other parts of this section when we discuss the target data, the offsetting of privacy and security guarantees and the importance of DSA Researcher Grants.

Another important outstanding item is with regards to whether a researcher meets the requirements of being affiliated to an institution that 'has as a primary goal to conduct scientific research or carry out educational activities that also involve scientific research'. This seems to be a Member State-specific determination, requiring some level of deference to the determination of the DSC where the researcher is based. DSCs would likely do well to maintain an updated list of the institutions that meet this criterion.

Another possibility is to delegate the vetting of researchers to a third party, for instance national science foundations. Article 40(13) of the DSA foresees the adoption of delegated acts to lay down

[173] Directive 2019/790.
[174] Vermeulen, *supra* note 147. At 3.

'independent advisory mechanisms in support of sharing of data'. This could include outsourcing the vetting to a third party.[175] Involvement of national science foundations seems appropriate, because they have experience in checking the eligibility and quality of researchers. Beyond this, DSCs can focus attention on assessing the substance of requests when the vetting process is handled by a third party.

2.3.4 Timeliness and mode of access

The timeliness and mode of access will depend on what type of data is being requested, as well as what safeguards must be implemented to ensure that the data are accessed in a safe and protective manner. Initially, the answer to the timeliness question is 'triggered upon request', as undertakings are only required to provide data after receiving a specific order from the DSC. However, this order may require a continuous form of data access that is enabled by an API (for example).

The same applies to mode of access, which is defined above in multiple forms: Article 40(7) requires VLOPs and VLOSES to set up any necessary interface–including online databases and APIs. As this is an open-ended obligation as the mode of access will vary depending on the target data, the type of protections required and the specific research needs.

In both cases, the final configuration will depend on the safeguards that will need to be implemented to protect the interests of the recipients of the service and of the VLOPs/VLOSEs supplying the data, which we turn to next.

2.3.5 Offsetting of privacy, intellectual property protection, information security and rule of law guarantees

Articles 40(2) and 40(5) DSA state that DSCs and the Commission must take into account both the interests of platforms and of the recipients of the service — in particular in terms of privacy, the protection of intellectual property/trade secrets and information security — when determining the types of data that can be accessed as well as the required protections.

Recital 97 provides some further guidelines. First, it stresses how commercial interests alone should not lead to refusals. Rather they should help modulate how to obtain access to the data: for example by requiring the signing of non-disclosure agreements and/or the creation of data vaults. Second, it also stresses how platforms should do their best to anonymise or pseudonymise data, unless doing so would make it impossible for researchers to accomplish their goals. As Article 40(4) imposes an open-ended data access obligation — meaning that it grants access to data which may be sensitive or personally identifying of users, as well as potential trade secrets and other forms of protected data — authorities will need to balance conflicting interests.

[175] As also mentioned by Caitlin Vogus, *Defending Data: Privacy Protection, Independent Researchers, and Access to Social Media Data in the US and EU*, (2023), https://cdt.org/wp-content/uploads/2023/01/2023-01-23-CDT-Defending-Data-Independent-Researcher-Access-to-Data-report-final.pdf at 17.

The considerations below build on the more general balancing exercise outlined in Part 2.1 above — meaning that we recommend following those general steps, and below we discuss specific considerations with regard to conflicts involving privacy, intellectual property, and information security.

2.3.5.1 Privacy

Managing vetted researcher access to platform data poses numerous challenges for the protection of user privacy and compliance with the GDPR. The EDMO working group on 'Platform-to-Researcher Data Access'[176] already tackles many of these issues. In particular, the report proposes a draft Code of Conduct for practical compliance with GDPR (as was originally specified in Article 40 of the GDPR) and envisions the creation of a body to oversee compliance on the part of Code signatories (as is specified in Article 41 of the GDPR).

In addition, as discussed above in section 2.1.1, compliance with data protection rules requires the existence of a lawful ground for processing as well as regard for purpose limitation and data minimisation. While platforms do not need a separate legal basis but can rely on Article 5(1)(b) of the GDPR to share data under Article 40(4) of the DSA for research purposes in the public interest, researchers will need to have a lawful ground for processing and using the personal data to which they get access. Depending on the type of research, different lawful grounds are available including public interest, legitimate interests of the data controller, and, in some cases, consent.

Beyond this, researchers have to make sure the extent of data they process is limited to what is necessary for the purposes of their research in order to comply with the principles of data minimisation and purpose limitation. Because it may sometimes be difficult for researchers to know in advance exactly what data is necessary for their research, it may be reasonable to apply the principle of data minimisation less restrictively at the start of a project but require researchers to keep monitoring their research needs and to take action once it becomes clear that certain data is not necessary for their research. Relevant measures will then include immediate erasure of any irrelevant data already obtained[177] as well as the discontinuation of any ongoing access requests for data not necessary for the specific purposes of the research. To this end, the EDMO report includes practical guidance for researchers on how they can implement the necessary safeguards.[178] The example of the Finnish system for access to medical data (Findata) is also interesting.[179] In that case, researchers rely on an exemption to purpose limitation justified on the promotion of scientific research, and more specific data is solely provided if more aggregate data cannot answer the relevant research question.

The design of safeguards will be request-specific, and the EDMO report provides a useful overview. It is worth stressing that in many cases, the best combination will be a solution that relies on a

[176] European Digital Media Observatory, *supra* note 8.
[177] In the context of concerns about law enforcement access to social media data via researchers, the January 2023 Center for Democracy & Technology report on 'Defending Data: Privacy Protection, Independent Researchers, and Access to Social Media Data in the US and EU' also suggests requiring researchers to destroy data when it is no longer needed. Vogus, *supra* note 174. At 63.
[178] See in particular the Annexes attached to the European Digital Media Observatory, *supra* note 8.
[179] See Ausloos, Leerssen, and ten Thije, *supra* note 148. At 69-78.

combination of technical safeguards (such as restricted API access, or even safe rooms), with legal safeguards that prevent researchers from abusing their data access (including the conclusion of non-confidentiality agreements to protect the data, mandatory courses to ensure that researchers have the technical skills to protect the data, ethics guidelines, and so on). The universities and other research organisations with which the researchers are associated should also be part of the agreements, adding another layer of protection to prevent abuses, and others have even proposed going as far as imposing criminal liability for intentional abuses.[180] Indeed, the advantage of vetting researchers is that regulators and platforms know in advance who has access to the data, can impose limits on the types of data that are available and how the data can be accessed, and can hold individuals accountable in case of mistakes. At the same time, this should facilitate access to data that is confidential or more protected in nature.[181]

The general balancing framework of section 2.1 can help authorities, companies and researchers consider the stringency of required safeguards, which will most likely be a function of: (i) the stated purpose of the data access request, and how well it matches to a relevant systemic risk recognised by the DSA (step 1); (ii) whether the request is targeted at the minimum data necessary to answer the research question (step 3); (iii) whether the alleged harms potentially caused by the data access are recognised by the legislation as potential offsetting criteria (step 5); (iv) whether a combination of legal and technical solutions can minimise these harms (step 6); and, if not, (v) which interest ultimately prevails given a consideration of the importance of the research question and the harms caused after these mitigation measures are in place (step 7).

Finally, one interesting proposal to be considered is the development of a safe harbour that would protect the VLOP/VLOSE from potential lawsuits for the violation of data protection laws if: (i) the VLOP/VLOSE followed all safeguards required by the DSC to ensure that the data sharing protected the personal data of the users involved; and (ii) the data that was improperly accessed or shared by researchers was made available explicitly as part of the data sharing mandate.[182] This would encourage the companies to ensure that proper privacy safeguards are in place and would also encourage data sharing by diminishing risks.[183]

2.3.5.2 Intellectual property protection

Article 40(5) and Recital 97 of the DSA explain that requests for data access by researchers should also consider the legitimate interests of VLOPs and VLOSEs, including in the area of trade secrets and confidential information. In addition, while Article 1(2)(b) of the Trade Secret Directive states that its provisions do not affect the application of EU rules requiring disclosure of trade secrets to the public for reasons of public interest, the range of data to which researchers can request access is potentially

[180] As the University/Research Organisation can punish abuses independently of the platforms. Persily, *supra* note 149. At 5.
[181] A good example is what the French CASD and the U.S. Census do to facilitate unrestricted access to researchers while requiring that researchers abide to many safeguards to protect privacy and confidentiality. See, for example, https://www.casd.eu/wp/wp-content/uploads/casd_user_guide-5.pdf and https://www.census.gov/about/adrm/linkage/guidance.html
[182] This was originally proposed by Persily, *supra* note 149. At 5.
[183] This safe harbor would not of course cover other violations of privacy laws by the undertakings, nor exempt the companies from liability in case the researchers help expose systemic risks or inappropriate safeguards as a result of the data access.

very broad. As such, the implementation of Article 40(4) of the DSA will trigger a balancing of data access interests against the protection of trade secrets and commercial confidentiality — one, however, that is not made clear by the text of the DSA itself.

Recital 97 does state that 'commercial interests of providers should not lead to a refusal to provide access to data necessary for the specific research'. Therefore, as a general principle, we believe that only *in exceptional circumstances* platforms should be able to preclude access to a dataset based on grounds of commercial confidentiality or trade secret protection. Using our balancing framework as a basis, this would be recognised by step 5: the protection of trade secrets is a countervailing interest recognised by the DSA. However, the legislation itself also imposes limits on the strength of this countervailing interest in contrast to the more general data access mandate imposed by the Regulation.

Here, however, one should stress the link between the legal mandate established by Article 40(4), the specific research question of a given data access request and the protections granted to the data by trade secret and intellectual property protection (Steps 1-3 of our balancing framework). Even if a certain dataset is highly commercially sensitive and trade secret protected, it should still be possible to mandate its disclosure to a researcher under Article 40(4) of the DSA — however, only if the specific request targets an equally important systemic risk, and such risk cannot be effectively studied by accessing other information.[184] In such a case, it would be disproportionate to let the interests of the relevant VLOP or VLOSE prevail, something recognised by step 7 of our balancing framework. This is, of course, contingent on researchers meeting DSA requirements that they demonstrate their ability to protect the confidentiality of the data they are being granted access to (either through the use of encryption, clean rooms, and so on) and agree to contractual/legal guarantees[185] that are proportional to the risk of disclosure.[186]

On the other hand, if the purpose of the data access request is of limited importance to assessing systemic risks and the claim of the VLOP or VLOSE in protecting commercial confidentiality, trade secrets and security is strong,[187] the interest of the VLOP or VLOSE should be given priority (also in step 7 of our framework).

The specific circumstances under which a platform may reject to facilitate data access for reasons of trade secret or intellectual property protection will need to be determined on a case-by-case basis during the implementation of researcher data access. As experience grows and precedent develops, the boundaries of the requirements will gradually become clearer. A practical suggestion to facilitate the implementation process is to require platforms to initiate a discussion in response to a researcher data access request of what data they can reasonably provide access to and what in their view would

[184] One can for instance think of a highly protected and sensitive dataset to which a researcher requests access in order to understand whether and how a particular recommendation algorithm contributes to suicidal feelings among teenage girls, as an example.
[185] Outside of the DSA, liability may apply on the basis of national law for practices violating legal rules in the relevant areas of civil and criminal enforcement.
[186] On this, see also Mathias Vermeulen, *The Keys to the Kingdom*, KN. FIRST AMEND. INST. (2021), https://knightcolumbia.org/content/the-keys-to-the-kingdom Section IV.
[187] See the general consideration on balancing data access and intellectual property, outlined in section 2.1.2 above.

be appropriate contractual (such as confidentiality agreements) or technical measures (such as data vaults) to enable data access without eroding their interests. In turn, researchers would get a chance to react and provide their interpretation of the balancing exercise and of potential other practical mechanisms to facilitate data access. By requiring platforms to take the initiative to present an offer and letting platforms and researchers enter into a procedural negotiation-like framework monitored by DSCs, parties may achieve workable results.[188] As long as the experience with platform-to-researcher data access is limited, the use of a procedural mechanism can be a way to achieve fair outcomes on the merits. Once the experience grows and lessons are available from the initial implementation of platform-to-researcher data access, more concrete substantive guidance regarding the balancing exercise can be developed by the Commission or DSCs.

2.3.5.3 Information security

Some of the data that is of greatest importance to independent researchers also poses the greatest potential security risks to VLOPs and VLOSEs. This triggers another balancing exercise: how should companies meet their disclosure obligations while protecting information that, if public, would compromise the security of their systems? Again, there is no single answer and the modelling of each specific circumstance is beyond the goal of this report.

Instead, in section 2.1.3 above we recommended the adoption of a threat modelling approach, as it can help shed light on how to implement steps 4-7 of our general balancing framework with regard to information security risks. In this approach, a series of questions are posed about the nature of the security threat, the threat actors, and the potential disclosure as a way of clarifying the nature and severity of the risk, as well as the ways in which researchers' objectives might be achieved while minimising the likelihood and consequences of the security risk. The questions are:

1) What would be the change in available information if the data was shared?
2) For which audience would this change be visible?
3) How likely is it that the data might leak to another audience?
4) Can the audience to whom the information has been revealed use this information to cause physical, emotional, or financial harm to a well-defined group of individuals?

When conducting a threat modelling exercise, platforms should clearly articulate specific scenarios in which the data in question might be a security risk and exactly what the outcomes of an adversary using the data might be. They should also clearly describe which adversaries they envision. Clearly describing the risk situation, risk outcome, and potential threat actors is vital for determining whether the alleged risk is concrete (step 4 of our framework) and determining how the risk might be mitigated (step 6 of our framework). For example, if a company foresees a risk only if data is disclosed to the public, then non-disclosure agreements and controlled modes of access might be an appropriate

[188] Inspiration is drawn here from the framework set up by the Court of Justice in *Huawei/ZTE* for determining when the seeking of injunctive relief for an alleged infringement of a standard essential patent amounts to abuse of dominance under Article 102 TFEU. See Case C-170/13 *Huawei/ZTE*, ECLI:EU:C:2015:477, par. 60-69. For an application of the negotiation framework in the context of implementing data access under the proposed Data Act, see Erik Habich, *FRAND Access to Data: Perspectives from the FRAND Licensing of Standard-Essential Patents for the Data Act Proposal and the Digital Markets Act*, 53 IIC-INT. REV. INTELLECT. PROP. COMPET. LAW 1343 (2022). At 1365-1370.

mitigation to allow otherwise sensitive data to be shared — as in other cases, it is very likely that solutions will involve a combination of technical and legal restrictions. Similarly, if the perceived threat actor already has access to the data in question, as might be the case with a state actor, then even if the data was inadvertently disclosed to that actor, it would not meaningfully change the threat calculation. Finally, platforms must also be able to demonstrate **consistency**: that is, that access to that specific dataset is equally limited internally. A claim that the disclosure of a given type of data poses a security risk is not credible if a very large number of employees can access the same dataset without significant safeguards.

If a threat modelling exercise determines that sharing data poses an unacceptable level of risk that cannot be mitigated through access controls alone, then regulators will need to consider the importance of the research question as well as the viability of alternative datasets (step 7 of our framework). Articles 40(5) and (6) of the DSA state that it will be up to the companies to propose alternative data disclosures that may meet researchers' goals, placing on them the burden to propose the least intrusive manner which still enables researchers to answer a given, important research question.

Overall, while platforms certainly have real security reasons to limit access to certain datasets, these should be the exception and very well corroborated. General allegations that the disclosure of a given dataset poses security risks without a clear outlining of what this risk may be and a demonstration that access to such data is subject to important internal safeguards (and why such safeguards cannot be duplicated in the case of vetted researchers) are not enough to prevent access to data under a strong legal mandate to do so.

2.3.5.4 *Rule of law guarantees*

There is no immediate link between researcher access to data and potential liability. Nevertheless, the data obtained by researchers may later on feed into investigations by the Commission or DSCs — both in terms of the detection of potential systemic risks and the assessment of risk mitigation measures. At that point, rule of law guarantees become applicable and VLOPs and VLOSEs will have the disposal of the relevant protections to ensure their rights of defence are safeguarded. As we mentioned before, in many jurisdictions it is usual that regulators issue new requests for information (even re-requesting data they already have) when they open an investigation as a way to enable undertakings to fully exercise their rights of defence — including challenging the use of researchers' conclusions for liability purposes.

2.3.6 Other important practical considerations

Finally, there are at least three other important practical considerations that will shape the effective implementation of the obligations contained in Article 40 of the DSA: 1) who bears the costs of compliance and of research?; 2) how to tie the data access mandate with independent sources of

research funding; and 3) how to arbitrate data access and conflicts between the researchers and DSCs and between DSCs. We outline some of the challenges and a roadmap for potential solutions below.

2.3.6.1 Which party bears the research costs?

Starting with funding. The implementation of many data access provisions will require significant resources. These can be roughly divided into two groups:

> 1) the costs that VLOPs and VLOSEs will have to incur to *internally structure* data sources, maintain databases and APIs, process datasets to remove confidential or other protected information, set up safe rooms (if necessary), train employees and researchers on how to access the data,[189] and so on; and
>
> 2) the costs of building and maintaining multidisciplinary research teams that combine lawyers, computer and data scientists, political scientists, economists, sociologists and other researchers that will be responsible for conducting the research itself.[190]

It is our general understanding that VLOPs and VLOSEs are solely responsible for covering the costs associated with (i), as these are no different than costs associated with compliance with other legal obligations — which are normally under the responsibility of the undertaking.[191] This interpretation is reinforced by the fact that Article 40(5) does not list 'overly-burdensome' or some similar reason as a basis for an undertaking to request an amendment of a data access request issued by a DSC. Recital 97 also clearly states that 'commercial interests of providers should not lead to a refusal to provide access to data necessary for the specific research'.

What this means in practice is that platforms should not be able to charge researchers for the costs of providing access to the data. Even now, at least some data that researchers wish to access for the purpose of studying systemic risks are already public, albeit at a financial price that researchers are unable to pay. For example, public tweets on Twitter were accessible via the Twitter Firehose. However, even then, researchers who have attempted to extract meaningful quantities of data have been faced with price quotes that would be out of reach of even the wealthiest institutions,[192] something that will likely become worse as platforms increasingly charge for access (as Twitter has recently decided to). Once the DSA comes into force, undertakings should provide the receiving party with free access to the data targeted by a valid data access order from a DSC.[193] Companies would

[189] These are not the costs of training researchers how to code, something they should bear on their own. Rather, these are costs associated with access to a specific type of data held by that specific undertaking (for example, a training that may be required for researchers to be able to understand the specific internal interfaces of a given company).

[190] These costs can include, for example, researchers' salaries, research assistants to help handle the data, access to computing infrastructure, trips to access safe rooms located in other countries and more.

[191] See, for example, the discussion on costs for compliance with the E-PRTR regulation in Ausloos, Leerssen, and ten Thije, *supra* note 148. At 45-46.

[192] Something that also helps propagate inequalities in research and can prevent the study of topics that impact under-represented minorities that may have smaller access to resources. See Vogus, *supra* note 160. At 24.

[193] Importantly, this does not mean that VLOPs/VLOSEs have no control over what data they should provide. Recital 97 also requires that access requests are proportionate and appropriate. In addition, VLOPs/VLOSEs may generally challenge a data access request as disproportionately burdensome in the same way that they may challenge other legal obligations. What this obligation requires is that DSCs ensure that requests are tailored and targeted solely at the data necessary to accomplish the goals of that specific project. This will likely require some exchange between the DSCs, the researchers and the VLOPs/VLOSEs, so DSCs will likely do well in setting up channels to

remain free to charge researchers (vetted or not) the price for data they deem reasonable outside of these official requests (as they do now).

A different rationale exists in relation to (ii). Article 40(8) of the DSA states that in order to pass the vetting process, researchers must be fully independent from commercial interests and must disclose the sources of funding for their research. These are important requirements. However, a question remains on which party will be responsible for covering the costs associated with these large multidisciplinary research projects, which can quickly become burdensome.[194] *The requirement of independence reinforces the conclusion that these costs should be borne by the researchers themselves.* A key challenge, then, will be in establishing proper and independent sources of funding, something we turn to next.

2.3.6.2 Connecting authorisation and funding requests: DSA Research Grants

The current system creates a gap between two crucial research sources: data and funding. The DSA correctly requires researchers to obtain and report independent sources of funding to pass the vetting process. For those, a leading source will likely be research agency grants. However, many grant review processes take months or even years, and awards can last for multiple years. At the same time, to obtain grants, researchers must normally indicate what data they require and how they plan to obtain it. One can easily foresee a situation where researchers obtain grants that are dependent on accessing internal platform data, only to have their requests later denied by DSCs because the data does not exist, is protected for privacy or intellectual property reasons, and so on — wasting valuable time and resources across the board.

A proper implementation of this obligation, therefore, will likely require a matching of the grant authorisation process with the researcher vetting process. One possibility is for public research funding agencies (like the European Research Council or national science foundations) to develop grants that are specifically targeted at Article 40(4) of the DSA — **call them 'DSA Research Grants'.** The 'DSA Research Grants' would require applicants to explain in their applications which data they must access, how their research contributes to the detection or minimisation of systemic risks in the EU, and how they plan to comply with all the requirements of Article 40(8), with a particular emphasis on the requirements around the protection of personal data and intellectual property. DSA Research Grants could then be assessed by specific review committees that incorporate not only other researchers and representatives of the funding agencies, but also representatives of the DSCs and, in a consulting role, representatives of the VLOP/VLOSEs.[195]

facilitate this tripartite exchange. Ultimately, though, once DSCs issue an official data access order, it will be up to platforms to ensure that researchers can freely access the required data.

[194] Many multi-year research grants quickly go in the hundreds of thousands or even millions of Euros.

[195] Further thinking is necessary to determine what the ideal composition and rules of such a Review Committee would be. For example it could be composed of a majority of ERC/researchers representatives, combined with representatives of different DSCs (potentially on a rotating basis) and a consulting group of VLOP/VLOSE representatives. The ERC/researchers would be given the power to decide on the awards by simple majority, with the group of DSCs being granted a justified 'veto' power in case they believe that researchers would not be able to comply with the requirements of Article 40(8), in particular in terms of their ability to protect the privacy and the security of the data. Platforms could then be required to issue a formal opinion on grant proposals that target data under their control, which should be taken into account by the review committee when deciding whether the proposal is feasible and what types of privacy and security

Obtaining a DSA Research Grant would provide researchers with strong prima facie evidence that they have passed the vetting process, so that their requests should be authorised by the relevant DSC in an expedited time frame. It is possible that this joint-funding and review committee would be part of the new independent advisory mechanism that is foreseen by Article 40(13) of the DSA and which was also recommended by the EDMO report,[196] accepted by platforms (in some form) in the EU's 2022 Strengthened Code of Practice on Disinformation[197] and also recommended by other reports on the topic.[198]

Importantly, to ensure that research also continues to be conducted in a fully independent manner, the DSA Research Grants should be an addition to the roll of grants currently awarded by research councils and funding agencies without any form of DSC or VLOP/VLOSE interference. Researchers awarded Starting/Consolidator/Advanced Grants, Veni/Vidi/Vici, Marie Curies, and so on, would be equally entitled to apply for access to internal platform data under Article 40(4) of the DSA whenever they wish to do so, undergoing the traditional vetting process. Both systems would run in parallel.

A final important question is on the origin of the resources to fund the DSA Research Grants. One of us has written elsewhere about how regulators should rely on the money collected through fines for violations of the GDPR to help fund grants that advance research on privacy violations — helping with the development of an enforcement ecosystem that is not restricted to regulators but also incorporates civil society more broadly.[199] A similar system could be implemented here: part of the money collected by the European Commission and DSCs through fines for violations of the DSA would be used to fund the DSA Researcher Grants, enabling independent, vetted researchers to help detect and mitigate systemic risks in the EU — something that would help increase deterrence while at the same time helping decentralise the DSA's enforcement system.

2.3.6.3 Determining limits and arbitrating data access disputes

A final aspect of Article 40(4) is worth calling attention to: the potential for significant disputes between DSCs in data access requests.

The DSA grants the DSC of establishment exclusive powers to: (i) vet researchers; and (ii) determine which types of data these researchers can access, as well as the safeguards to be employed (such as data anonymisation, safe rooms, and so on). While this system may be rational from a purely administrative point of view — otherwise companies may have to comply with requests from dozens of different authorities with varied levels of sophistication — this form of centralisation is also worrisome.

safeguards must be implemented. In this way, the independence of the review process is guaranteed while also including checks to ensure the feasibility of the proposed research.

[196] See European Digital Media Observatory, *supra* note 8. At page 12.

[197] see Commitment 27 of the EU's 2022 Strengthened Code of Practice on Disinformation, available at https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

[198] For example, Ausloos, Leerssen, and ten Thije, *supra* note 148. at 83-84; Vermeulen, *supra* note 185. Section IV.

[199] Lancieri, *supra* note 11. A somewhat similar system is currently foreseen by the recently enacted California Privacy Rights Act of 2020. The CPRA establishes that 3% of the fines collected by the California Privacy Protection Agency should be distributed as grants to nonprofit agencies to help promote and protect consumer privacy.

The significant problems in the enforcement of the GDPR showcase how some EU regulators — in particular in Ireland and Luxembourg — struggle to effectively enforce laws against VLOPs and VLOSE that are strategically important for national economies.[200] Still, these regulators retain sole responsibility for deciding almost everything connected to researcher data access under Article 40(4). It is worth noting that the system put in place by the DSA is worse than the one put in place by the GDPR. That is because Chapter VII, Section II of the GDPR at least establishes a dispute resolution mechanism that allows for a majority of data protection authorities to overrule the decision by the authority of establishment.[201] This is a burdensome and inefficient system, but it provides a partial escape valve that has already been used to force the Irish Data Protection Commissioner to adopt stricter decisions in some important cases.[202]

No such mechanism, however, appears to exist to overrule decisions taken by the DSC of establishment in the case of vetted researcher access to data,[203] with a potential narrow exception associated with the EC's powers to adopt delegated acts that lay out the technical conditions for effective researcher access.[204] That is because, while not totally clear, it seems that the referral and joint investigation system that is set up by Chapter IV, Section II of the DSA focuses solely on potential infringement procedures that impact another Member State, but not against administrative decisions that are within the powers of the DSC of establishment — as is the case with the vetting of researchers.[205] The large information asymmetries between companies, regulators and civil society representatives, when combined with the particular nature of many data access requests (which, in some cases, may be thwarted by small changes in access or anonymisation protocols, for example), increase the risks that this obligation to enable researcher access to data ends up being hollowed-out in practice by the very authorities that are in charge of implementing it.[206]

For these reasons, an effective implementation of researcher data access depends on both the professionalism of DSCs of establishment as well as the active engagement of other DSCs, the Commission and civil society more broadly — which we strongly encourage. In particular, it would be important that the Commission's guidelines set up a detailed vetting process that can be applied EU-wide. This would allow, for example, researchers to better rely on the rights established by Article 40(9) of the DSA, which enables researchers to apply for data access with the DSC where they are located. While the final decision always lies with the DSC of establishment (as per Article 40(9)), the approval of a request by another DSC should provide researchers with strong prima facie evidence that their request is reasonable and that they are implementing proper safeguards — similar to the prima facie evidence granted by the DSA Research Grants. A subsequent denial of a forwarded request

[200] See *Id.* and the Communication from the Commission to the European Parliament and the Council: Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the General Data Protection Regulation, COM/2020/264 final, 24 June 2020, at 6. The matter, and the lack of sufficient oversight by the European Commission, is now under investigation by the EU Ombudsman (among many other parties engaged in this discussion).
[201] See GDPR, Chapter VII, Section II, and in particular Article 65.
[202] See European Data Protection Board, *Record fine for Instagram following EDPB intervention*, (2020), https://edpb.europa.eu/news/news/2022/record-fine-instagram-following-edpb-intervention_en
[203] Some of the drafts of the DSA empowered the Commission to issue such decisions in relation to VLOPs and VLOSEs. However, these provisions did not make it to the final text.
[204] Article 40(13) DSA.
[205] This, however, is a matter that requires more careful evaluation.
[206] See Lancieri, *supra* note 11. For a similar argument with regards to GDPR enforcement.

by the DSC of establishment should require a careful, well-justified and public decision that explains the exceptional circumstances that led to the specific decision.

2.4 Sharing of Click and Query Data with Competitor Search Engines - Article 6(11) of the DMA

Article 6(11) of the DMA requires gatekeepers to share with any third-party providing online search engine services, upon request, 'access on fair, reasonable and non-discriminatory terms to ranking, query, click and view data in relation to free and paid search generated by end users on its online search engines'.[207] Recital 61 stresses how search engines are characterised by network externalities that end up constituting an important barrier to entry and expansion of competitors in online search markets. By forcing the sharing of such data, the idea is to help increase the contestability of those markets.[208] This obligation is rooted in several academic and policy studies on the competitiveness of online search markets, which largely reached the same conclusion and encouraged the sharing of such data.[209]

Article 6(11) also requires companies to anonymise the data, so that it no longer qualifies as personal data.[210] Recital 61 stresses how the overall goal is to protect user privacy (including from the risks of re-identification) while at the same preventing data quality degradation. A previous market investigation conducted by the UK Competition and Markets Authority (CMA) had also concluded that click and query data possess particular characteristics that enable effective anonymisation (removing any form of personal identifier) while maintaining data quality for competition purposes, largely by means of aggregating the queries and resulting clicks.[211] The UK CMA investigation provided some recommendations with regard to the sharing of this type of data. A former CERRE report has also issued a series of recommendations on how to effectively share this type of data.[212] We incorporate most of them here.

The sharing of click and query data is another novel data-sharing obligation enacted by the DMA. Overall, it can be summarised as follows:

1) **Target party:** gatekeepers
2) **Target Data:** ranking, query, click and view data
3) **Receiving party:** competing search engines
4) **Timeliness:** upon request, but then likely continuous or in defined intervals
5) **Mode of access:** queriable or streaming API

[207] Article 6(11) DMA.
[208] Recital 61 DMA.
[209] See Filippo Lancieri & Patricia Sakowski, *Competition in digital markets: A review of expert reports*, 26 STANF. J. LAW BUS. FINANCE (2021). Academic studies on the topic include: Jens Prüfer & Christoph Schottmüller, *Competing with big data*, 69 J. IND. ECON. 967 (2021). And TOBIAS J. KLEIN ET AL., *How important are user-generated data for search result quality? Experimental evidence*, (2022).
[210] As Recital 26 of the GDPR explains: 'The principles of data protection should [...] not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable'.
[211] Competition and Markets Authority, *supra* note 9. par. 8.38 and its Appendix V: assessment of pro-competition interventions in general search, par. 101-107 and 117-120.
[212] Jan, *supra* note 10. At 15-26.

Based on these characteristics, one can categorise this obligation as requiring **Private Party Access to Data**. The implementation of these commands will require a range of specifications and the consideration of relevant trade-offs. We discuss this below — and as this is a targeted obligation, we attempted to provide more concrete recommendations.

2.4.1 Targeted party

Article 6(11) of the DMA targets all gatekeepers providing core platform online search engine services.[213]

2.4.2 Targeted data

Article 6(11) of the DMA requires companies to share 'ranking, query, click and view data in relation to free and paid search generated by end users on its online search engines'. Ranking is further defined as 'the relative prominence given to goods or services (...) or the relevance given to search results by online search engines',[214] while Recital 61 affirms that the obligation targets 'information about what users searched for, and how they interacted with, the results with which they were provided' for both free and paid searches.

Somewhat similar to considerations on advertisement databases and researcher access to data, a first question is with regards to geographic scope. Article 1(2) of the DMA establishes that it applies to core platform services offered by gatekeepers to business users established in the European Union or end users established or located in the Union, irrespective of the place of establishment or residence of the gatekeeper. As in the DSA examples discussed above, this should restrict the data to that collected in the EU or that involves EU citizens — something that can be done by relying on IP and location information or other forms of registration data (such as, for searches where consumers are logged in).

A second question is with regards to the specific data to be shared. The CMA report is the most detailed investigation into this matter. The agency concluded that there are important economies of scale in obtaining click-and-query data, but that the marginal benefit of additional types of data depends on the type of query.[215] In particular, 'where a search engine sees a search query very frequently (sometimes referred to as 'head queries'), then the marginal benefit from seeing that query more often is relatively lower. Conversely, the marginal benefit of seeing a query more often is higher for uncommon queries (sometimes referred to as 'tail queries')'[216]. Tail queries could include both uncommon queries as well as what Microsoft called 'fresh queries', that is, queries that relate to recent events.[217]

[213] As defined by Article 2(b) and 2(6) of the DMA.
[214] Article 2(22) DMA.
[215] Competition and Markets Authority, *supra* note 9. Annex I, page 18.
[216] *Id.*
[217] *Id.* I7 to I9.

It is beyond the scope of this report to provide insights on what constitutes a tail query or a fresh query, and how to determine which of those are relevant for competition purposes. This will require detailed data from different market participants — exactly the type of data that allowed the CMA to conclude the above. We can, however, provide some other insights on how to share the data and how to consider different trade-offs.

Modern search engines return different results for the same search term based on a user's browser language, inferred location, search platform (desktop vs. mobile) and session history. However, Article 6(11) requires companies to anonymise data so that it no longer qualifies as personal data, which poses a difficulty given that much of this search context is identifying of users. There are additional practical considerations relating to the sheer volume of data that is the target of this obligation. To address both these privacy-related and practical considerations, a useful technique is aggregation.[218] Indeed, we recommend that this category of data be provided aggregated by day, search term, and a limited set of search context parameters.

For aggregated results to be practical, useful, and most importantly, privacy-preserving, an important determination is the level of aggregation. To select this level of aggregation in a way that minimises the privacy risks to users, we recommend a *k*-anonymisation[219] approach. This approach ensures that users cannot be distinguished from *k*-1 other users by ensuring that no record in a dataset represents fewer than *k* users. That is to say, for any record in a *k*-anonymised dataset, all the information in that record relates to at least *k* users. This can be achieved in multiple ways, either by *generalising* data or *suppressing* data, and in this case we recommend a combination of these two approaches.

For reasons of utility, we recommend that platforms share query data aggregated by day, the exact search term, users' geographic region to the NUTS 2 level, language, and search platform (desktop or mobile). We believe generalising to this level will allow for a meaningful number of searches to be safely shared, while still conveying the most important context that implicitly defines users' searches.[220]

We do not recommend that data be shared on a per-search basis for two reasons: first, the volume of data that would result would be so large as to be nearly unusable, and the privacy concerns would become very difficult to manage because each record would directly relate to a single user's action. However, aggregating search data poses practical difficulties as well, because search query responses are non-deterministic — even the same user is not guaranteed to receive the same response to the same query if they search twice in a row. Ranking of responses to queries will similarly need to be aggregated by search term and context. This could be done as a cumulative ranking, by returning a list of responses weighted by their total ranking overall of all the searches within each aggregation grouping.

[218] See discussion on access to advertisement data, Part 2.2. above.
[219] Pierangela Samarati & Latanya Sweeney, *Protecting privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression* (1998).
[220] For example, location can be provided based on zip codes instead of specific addresses. See our discussion of ad databases in section 2.2 for references for more general considerations on the importance of aggregation for privacy preservation.

Search engines commonly return web links, but do not do this exclusively. In addition to the web links they believe are most relevant to a user's search, a search engine might respond with a dictionary definition of a word, an excerpt from an encyclopedia entry, related searches, or some other piece of relevant context. Article 6(11) of the DMA states that ranking data should also be provided, so search engines can provide an ordered list of all query responses recommended to the user, regardless of the category of response. The provision does not specify how many ranked responses must be provided, and we recommend that search engines provide a list that corresponds to the first webpage of responses to the user's query. This recommendation is based on two factors. First, the total list of all ranked responses to searches can be very long and may become unwieldy. Second, since very few users ever go past the first page of search results, rankings past this threshold are likely of low information value.[221]

Another question with both practical and privacy implications is the minimum number of search records required before an aggregate group must be reported. In *k*-anonymity terms, this is the *k*-value. We discuss this and other privacy considerations in 2.4.5.

In general, the shared data should contain at least the following information.

Proposed Standard Field	Description	Type
search_term	Text of the searched terms/queries	text
search_lang	Language of the search	text
search_region	Inferred location for the aggregate of specific terms/queries	text
query_responses	Set of query responses ordered as they were returned to users, including a boolean field specifying whether the response ranking was affected by paid advertising	json
click_data	Set of user click data in relation to responses (dictionary of response ids to counts)	json

Finally, there is a question on timing and the amount of data to be provided, which are not clearly specified in the DMA. The CMA study on this topic concluded that click and query data was particularly relevant to improve the quality of uncommon (and potentially fresh) search queries.[222] The problem is that these searches are particularly salient to users, who rely on them to compare otherwise unknown quality standards between search engines. This poses a challenge in terms of timing because

[221] See, for example, Johannes Beus, *Why (almost) everything you knew about Google CTR is no longer valid*, SISTRIX (2020), https://www.sistrix.com/blog/why-almost-everything-you-knew-about-google-ctr-is-no-longer-valid/
[222] Lancieri and Sakowski, *supra* note 208. at 86, Competition and Markets Authority, *supra* note 9. at t 93-94, app. I, at 16, 18.

it encourages almost real-time data sharing between companies — something that could become both unfeasible and unreasonable. We discuss this in more detail in section 2.4.4. below.

2.4.3 Receiving party

Article 6(11) of the DMA defines receiving parties as competing search engines.[223] Search engines are then defined in Article 2(5) of the Platform-to-Business regulation as a 'digital service that allows users to input queries in order to perform searches of, in principle, all websites, or all websites in a particular language, on the basis of a query on any subject in the form of a keyword, voice request, phrase or other input, and returns results in any format in which information related to the requested content can be found'. Recital 61 expands the obligation to also include 'third parties contracted by a provider of an online search engine, who are acting as processors of this data for that online search engine'. This definition is focused only on general search engines, meaning that specialised or vertical search engines[224] are not entitled to request the sharing of data.

An important discussion is what to do with regard to potential entrants in this market. Here, we follow a past CERRE report affirming that the European Commission will have to establish some criteria to vet potential entrants that request access to the data,[225] though we do not venture into outlining those.

2.4.4 Timeliness, mode of access

Article 6(11) DMA establishes that the gatekeeper has the obligation to provide the data at the request of the interested competing online search engine. As such, the initial timeliness of the obligation is 'upon request'. A question remains, though, on how often the gatekeeper should provide the data after this initial request, and what is the mode of access that can guarantee the effective implementation of this obligation.

As mentioned above, implementation decisions should also be made with regard to what is the final objective of the obligation. Purpose provides a guiding north star against which to assess potential trade-offs. In this case, the obligation has a clear goal of increasing the contestability of online search engine services. Building on the CMA study, one can consider the potential provision of both 'tail searches' (meaning non-common queries) and 'timely or fresh searches', meaning queries relating to recent events. These lead to different implementation challenges, so they are discussed separately below. These mostly reflect considerations to be made in steps 3 and 6 of our framework: what is the minimum amount of data that must be shared to achieve the stated purpose of the obligation, and how different legal and technical solutions can help mitigate concerns but still enable data access.

[223] Or 'any third-party providing online search engines'.
[224] Such as Online Travel Agencies, online marketplaces, app stores and other applications or websites that also provided (limited) search functionalities.
[225] See Jan, *supra* note 10. At 22-23.

Tail searches are relatively easier to supply, conceptually speaking, although they pose greater privacy concerns. Four of the main questions are on determining what qualifies as a tail search, how aggregate should search results be, on what frequency should parties provide the data and what is the mode of access. In Part 2.4.2 we explained why we cannot provide insights on determining what qualifies as a tail search and discussed levels of aggregation. Here we focus on timeliness and mode of access.

Given that tail searches are, by definition, not 'fresh', there is no urgency to immediately share data. Still, unjustified delays may lead to losses in terms of increased market competition without significant gains in other dimensions. As such, a reasonable recommendation is that data be made available on a daily basis reflecting data no more than 48 hours old. We base this timeliness recommendation on the fact that we have previously recommended that results are aggregated by day, making the minimum possible delay for making results available 24 hours. We further believe that it is reasonable to allow target parties up to 24 hours to coalesce data into a format that can be provided in response to requesting parties, leading to a maximum delay of 48 hours. We further expect that parties will request this data for all searchers in all countries subject to this requirement.

In terms of mode of access, we recommend that data be made available in a bulk file format. This mode of access is economical for both the creating and receiving parties and is likely the most useful format for competing search engines attempting to leverage the provided data. This mode has several advantages over 'in-situ data access', which has been generally proposed as a potential mechanism to enable access for data sharing obligations.[226] In the context of access to search query data, this would entail that competing search engines bring their algorithms to the gatekeeper's data without the data leaving the gatekeeper's platform.

In-situ data access is claimed to have several advantages, namely the data being directly actionable because it is not separated from its context as well as stronger protection of privacy and security due to the fact that the data do not leave the gatekeeper's platform.[227] However, we believe that in-situ data access is not a suitable way to implement Article 6(11) of the DMA, in particular because it brings additional complexities in terms of effective monitoring and enforcement.[228] As the training of competing search algorithms happens inside of the gatekeeper's dataset, there will be a degree of uncertainty and a need to trust the gatekeeper that all the relevant data is made available for training the competitors' search algorithms. And, as concluded by the UK CMA in its advertising sector inquiry, there are ways to share search query data without eroding privacy and security..[229] Beyond this, the least privacy-compromising search query data can still allow for maximum value and insights for competing search engines. We therefore believe that the advantages of in-situ data access do not outweigh the extra monitoring complexities it brings.

Timely searches, however, present other important technical challenges — in particular on determining what exactly qualifies as a timely search that is relevant to increase competition in the

[226] Bertin Martens et al., *Towards efficient information sharing in network markets*, PUBL. TILEC DISCUSS. PAP. NO DP2021-014 (2021). At 4.
[227] *Id.*
[228] See also the discussion in Jan, *supra* note 10. At 17-18.
[229] Competition and Markets Authority, *supra* note 9. At par. 8.32-8.43

search engine market (the overall purpose of this obligation). One could imagine that the determination of relevant timely searches is tied to a determination of search results that scale fast within a given set of parameters — for example, a timely search query is a query which has a frequency above X thousand in a given location and in a given language within a period of X hours. The setting of these parameters, however, becomes almost a dynamic game — as the more popular the search query, the more important it is for competition between search engines, so the faster one would want the data to be shared. This, however, requires almost constant monitoring of all search terms by the gatekeeper, a cost that rises the more 'real-time' the monitoring has to be. This also impacts the mode of access, as this may require almost real-time data sharing between competing companies. These are important challenges, and they should be carefully considered before search engines are required to share almost real-time data. They also prevent us from providing more detailed guidelines on how to implement such sharing in practice in a reasonable manner.

2.4.5 Offsetting of privacy, intellectual property protection, information security, and rule of law guarantees

2.4.5.1 Offsetting privacy concerns

A foremost concern in designing our recommendations lies in protecting the privacy of users' searches, which can be highly sensitive.[230] The problem of protecting the privacy of users who contribute to public datasets is one that has been grappled with by both regulators and companies more frequently in recent years. Differential privacy[231] is an approach that is increasingly employed to balance the competing interests of user privacy on one hand and data utility on the other.[232] Differential privacy, generally speaking, works by applying 'noise' to data, such that information about any particular individual cannot be inferred from the more aggregate dataset. The problem is that the amount of 'noise' that needs to be added is a function of how unique is the individual being protected. In the case of click and query data, the data that poses the greatest risk of identifiability are infrequently searched query terms and extremely specific data about the context in which users search and click on terms. However, search terms are not a data type that can be permuted with 'noise' as other categories of data can be. As such, it is not clear how to apply differential privacy in these circumstances such that the utility of the data might be preserved. This is because the search query *itself*, rather than any query data returned in response to the search, may be identifying if made by a small number of persons.

For this reason, as well as the practical difficulties of providing data in anything other than aggregated form, we have pursued *k*-anonymisation as an alternative means of minimising the privacy risks to users. Ensuring that users have as robust a privacy guarantee as possible drives us to select a high *k*,

[230] See EDPS, Opinion 2/2021 on the Proposal for a Digital Markets Act, 10 February 2021, p. 12, available at https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_markets_act_en.pdf

[231] Cynthia Dwork, *Differential privacy: A survey of results*, *in* THEORY AND APPLICATIONS OF MODELS OF COMPUTATION: 5TH INTERNATIONAL CONFERENCE, TAMC 2008, XI'AN, CHINA, APRIL 25-29, 2008. PROCEEDINGS 5 1 (2008).

[232] US Census Bureau, *Differential Privacy and the 2020 Census*, CENSUS.GOV, https://www.census.gov/library/fact-sheets/2021/differential-privacy-and-the-2020-census.html

where *k* is the smallest aggregate grouping of reported data. However, concerns of practicality also constrain *k*. A lower threshold for minimum grouping size, in combination with more generalised parameters for geographic range may yield an unwieldy volume of data from the largest search engines, while too high a threshold may obscure a meaningful portion of the distribution of search terms. Therefore, we generally recommend that *k* is between 100 and 500, but be determined by the target party given their knowledge of the volume and distribution of their own data. This range is high enough to offer users robust privacy guarantees given the data that is being made available, while also being likely to yield useful aggregate grouping sizes.

Again, the purpose of the obligation can provide a direction that helps evaluate trade-offs in an eventual step 6 of our balancing framework: in this case, balancing between user privacy and the utility extracted from the data with a general stated purpose of increasing competition in search engine markets. To keep data volumes reasonable, we have recommended sharing click data aggregated by specific common user search contexts. These contexts will need to be kept fairly large, to the point where these groups would be expected to contain many searches with at least thousands of requests per day. Large numbers of requests within a context are inherently a high barrier to identifiability. That is why we recommended a minimum level of aggregation — at least 100 searches per aggregate group be reported — as a better protocol that provides adequate levels of privacy protection through 'anonymity of the crowd' while at the same time protecting the utility of the data.

As concluded by the UK CMA, avoiding the disclosure of any personal data appears possible, so that compliance with the GDPR can be preserved by providing access to a more limited range of search data that does not involve personal data.[233] This is the preferred approach and the one ordered by Article 6(11) of the DMA, which requires any personal data to be anonymised.[234]

2.4.5.2 Offsetting intellectual property, information security and rule of law concerns

A second question is with regard to balancing potential conflicts with intellectual property rights. Indeed, Google affirmed to the CMA that the sharing of such click and query data could lead to a decrease in incentives to innovate in indexing technologies and ranking algorithms.[235] Indeed, considering that the receiving parties of the search query data are competitors, the scope for gatekeeping search engines to rely on intellectual property protection is larger than in scenarios where data needs to be shared with regulators for purposes of checking compliance. As acknowledged by the CMA, the sharing of search query data may enable free riding by rivals and could reduce a gatekeeper's incentives to innovate.[236] The agency also suggested that one way to ensure that the data access enhances instead of reduces incentives to innovate is to focus on the sharing of observed

[233] Competition and Markets Authority, *supra* note 9. par. 8.38 and its Appendix V: assessment of pro-competition interventions in general search, par. 101-107 and 117-120.
[234] Irrespective of these anonymisation efforts, competing search engines might be able to rely on their legitimate interests as data controllers as a lawful ground for data processing if any personal data turns out to be inadvertently included in the dataset shared by the gatekeeper.
[235] Competition and Markets Authority, *supra* note 9. par. 8.39 and its Appendix V: assessment of pro-competition interventions in general search, par. 123.
[236] *Id.* par. 8.40 and its Appendix V: assessment of pro-competition interventions in general search, par. 124.

and inputted data, rather than the results of a gatekeeper's analysis.[237] By requiring the sharing of raw data only, rivals can maintain their incentives to innovate and develop their own models to analyse search queries, increasing consumer choice.[238]

Two areas are particularly important to this balancing between data access and intellectual property: the scope of the data to be shared and the remuneration.

Scope of the data: In this case, the balancing framework proposed in section 2.1 above can provide guidance on how to trade the relative interests: a duty to share search query data and the incentives to innovate afforded by intellectual property protections. This would trigger concerns as of step 4 of our general balancing framework in section 2.1. That is, one can identify risks to intellectual property as a well-defined, non-speculative potential harm that arises as a result of the obligation. The problem, however, is that this is not a type of harm that is explicitly recognised by the law as a potential countervailing interest that can block a well-defined obligation to disclose the data. Therefore, the protection of intellectual property would not be a basis to fully prevent the sharing of the data considering the clearly defined legal obligation, failing step 5 of our framework. That is because the framing of the obligation illustrates that the legislator has already conducted a balancing.

That is not to say, though, that there should be unlimited sharing of data. Recital 61 clarifies that access to ranking, query, click and view data should allow third-party undertakings to optimise their services and contest the relevant core platform service. This provides an objective function that is reflected in step 3 of our balancing framework: the data to be shared should be the minimum necessary to achieve the purpose of the obligation – in this case, to increase the contestability of general search markets. As mentioned, the focus should be on tail and, potentially, fresh queries that are above the minimum thresholds for the protection of user privacy.

Remuneration: With regard to the conditions of access, Article 6(11) of the DMA requires gatekeepers to provide the data on FRAND terms. Because search query data is collected as a free byproduct of offering a search engine, the marginal cost of obtaining the user information for the gatekeeper is (roughly) zero. It therefore seems undesirable to give gatekeeping search engines the possibility to charge a fee for access to their search query data.[239] Nevertheless, it seems reasonable to let the gatekeeper impose costs for delivering the data in a workable format. A similar approach is taken in the Directive on Open Data and Re-Use of Public Sector Information, where Article 6(1) requires the re-use of documents to be free of charge but allows for the 'recovery of the marginal costs incurred for the reproduction, provision and dissemination of documents as well as for anonymisation of personal data and measures taken to protect commercially confidential information'.[240] There are reasonable benchmarks for the price of data delivery. Cloud services, for example, have storage products where they charge rates for data writes, storage, and reads. Different products are optimised

[237] *Id.* par. 8.42.
[238] See also Inge Graef & Jens Prüfer, *Governance of data sharing: A law & economics proposal*, 50 RES. POLICY 104330 (2021). At 4.
[239] *Id.* At 5. See also the discussion in Jan, *supra* note 10. At 24-26.
[240] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast) [2019] OJ L 172/56.

for different use cases, either read-heavy usage or write-heavy usage. These prices can serve as a 'market price' for data access.

Finally, it is worth stressing that because this provision is to a well-defined set of sophisticated third parties, we do not anticipate very high information security risks. With regard to the rule of law guarantees, there is no immediate link between access to search query data and potential liability. Nevertheless, the data obtained by competing search engines may later on feed into investigations by the Commission or national competition authorities — for instance when the search query data provides insights regarding other practices like self-preferencing or the combination of data across services. The opening of an investigation triggers rule of law guarantees, allowing gatekeeping search engines to rely on the relevant protections to ensure their rights of defence are safeguarded.

LOOKING AHEAD: CATEGORISING THE PROPOSED TRANSPARENCY AND DATA ACCESS OBLIGATIONS IN THE AI AND DATA ACTS

This report aimed to facilitate an effective implementation of the transparency and data access mandates that are present in the DMA and the DSA package. We did so by first proposing a categorisation exercise, and then by relying on three case studies to help address some practical implementation challenges that regulators, companies and civil society more broadly will face when trying to transpose the legal commands to the real world.

These challenges, however, do not stop here. Beyond the DSA and the DMA, the EU legislator is also in the process of adopting an Artificial Intelligence (AI) Act and a Data Act. These Acts also lay down obligations relating to transparency and access to data. To conclude this report, and in order to provide a complete overview of the developments in this area, this Part looks ahead to the future of EU Regulations by introducing a short discussion of the relevant provisions in these Acts. Note that the AI Act and the Data Act have not yet been adopted and are still undergoing discussions at the European Parliament and the Council. For this reason, our review of the relevant provisions is based on the legislative proposals introduced by the European Commission.

The goal here is not to provide an equally detailed analysis of each specific obligation in these draft Regulations — not least because they are still subject to the changes that will take place as part of the legislative process. Rather, the goal is to showcase how our categories and variables can also help rationalise obligations under discussion in other complex EU Regulations and help facilitate potential discussions around conflicts and harmonisations between different rules that apply to digital markets in an overlapping way. The same exercise is present in our more detailed table of obligations in Annex I.

In comparison with the DSA/DMA Package, the AI and Data Acts have a broader scope of application. While the DSA and DMA are targeting platform services, the AI Act and the Data Act apply more broadly to the use of artificial intelligence and data across various industries. The AI Act and the Data Act can therefore be seen as horizontal instruments with a more general scope. In terms of provisions regarding transparency and access to data, the AI Act and the Data Act have requirements in place to enable competent authorities and regulators to check compliance. The approach behind these provisions is similar to the one of the DMA and DSA and falls within the category of **regulator access to data**. Beyond this, the Data Act also contains a more far-reaching form of regulator access to data whereby private parties are required to share data with regulatory authorities who can use the data to pursue public interest objectives on grounds of exceptional need. Whereas the access to data in the AI Act is limited to the category of regulator access to data for purposes of compliance, the Data Act also includes provisions that regulate **private party access to data** through requirements of data holders to share data with users and third parties. Furthermore, the AI Act contains provisions enabling **transparency towards the general public** about when AI systems are used.

3.1 Proposed Artificial Intelligence (AI) Act

In April 2021, the Commission published its proposal for an AI Act[241] laying down rules regarding the implementation and use of AI systems. The rules include prohibitions of certain artificial intelligence practices (Article 5), requirements and obligations relating to high-risk AI systems (Articles 6-29), and transparency rules for certain AI systems (Articles 52-55).

Article 52 of the proposed AI Act creates a form of **transparency towards the general public** by requiring providers of AI systems to inform natural persons when they are interacting with AI systems. The extent of transparency is limited to a duty to inform natural persons about the use of AI systems and does not cover transparency about the functioning of the AI system.

In terms of **regulator access to data**, the AI Act contains a number of relevant provisions. First, Article 23 of the proposed AI Act requires providers of high-risk AI systems to provide national competent authorities 'with all the information and documentation necessary to demonstrate the conformity of the high-risk AI system' with the relevant requirements upon request, including 'access to the logs automatically generated by the high-risk AI system, to the extent such logs are under their control by virtue of a contractual arrangement with the user or otherwise by law'. The same obligations apply to representatives of importers of AI systems in line with Articles 25(2)(b) and 26(5) of the proposed AI Act. The access to the logs generated by the high-risk AI system is the more novel element of these obligations, because it requires providers and importers to give access to part of the output of their AI systems. Under Article 27(5) of the proposed AI Act, distributors of high-risk AI systems must provide national competent authorities upon request 'with all the information and documentation necessary to demonstrate the conformity of a high-risk system with the requirements'. No obligation to give access to the logs generated by the AI system applies to distributors.

Second, a more extensive and novel form of data access can be found in Article 64 of the proposed AI Act. Article 64(1) requires providers to grant market surveillance authorities 'full access to the training, validation and testing datasets used by the provider, including through application programming interfaces ('API') or other appropriate technical means and tools enabling remote access'. Where necessary to assess the conformity of the high-risk AI system with the relevant requirements, this includes access to the source code of the AI system as noted in Article 64(2). Beyond this, Article 64(5) entitles the market surveillance authority 'to organise testing of the high-risk AI system through technical means' at the request of a national public authority supervising or enforcing fundamental rights under EU law when such a national public authority otherwise does not have sufficient documentation under the AI Act to perform its tasks. Considering the extensive range of data to which access can be mandated, risks regarding the protection of trade secrets and information security will likely need to be taken into account in the implementation.

[241] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (AI Act) and amending certain Union legislative acts, COM(2021) 206 final, 21 April 2021.

These provisions relating to our category of **regulator access to data** involve different types of target data and share a common purpose in helping authorities verify compliance with the various obligations contained in the AI Act.

3.2 Proposed Data Act

The Commission published its proposal for a Data Act in February 2022[242] as the 'last horizontal building block of the Commission's data strategy [that] will play a key role in the digital transformation'.[243] The proposed Data Act lays down harmonised rules on the making available of data in a number of areas.[244] The new measures can be divided into a number of pillars, of which the following are the most relevant for our purposes: a right of access for the user to data generated by the use of a product or related service (Articles 3-7); a set of minimum legal obligations for providers of data processing services to allow customers to switch effectively to another provider (Articles 23-26); and an obligation for private data holders to make data available to public sector bodies on grounds of exceptional need (Articles 14-22). These are analysed in more detail below.

3.2.1 Right of access to data generated by the use of a product or related service (Articles 3-7)

The right to access data under the Data Act is a form of **private party access to data** and mainly focuses on the 'Internet of Things' (IoT), where manufacturers of smart devices have so far been able to limit the access to and the transfer of data through technical restrictions in the design of products and services. Such restrictions can prevent users from obtaining the data they need to use, repair and access complementary services from other providers. The target data is data generated by the use of a product or related service.[245] The data access right is addressed to consumers as well as business users.[246] According to the explanatory memorandum of the proposed Data Act, the purpose of the right to data access is twofold: (1) to empower consumers and business users 'to meaningfully control how the data generated by their use of the product or related service is used'; and (2) to enable innovation by more market players by allowing 'for a competitive offer of aftermarket services, as well as broader data-based innovation'.[247]

The data access right of the proposed Data Act is novel because it introduces for the first time a more general right to access data for both consumers and business users and including personal as well as non-personal data. The right to data portability of the General Data Protection Regulation only targets

[242] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (proposed Data Act), COM/2022/68 final, 23 February 2022.
[243] Press release European Commission, 'Data Act: Commission proposes measures for a fair and innovative data economy', 23 February 2022, available at https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113
[244] Article 1(1) of the proposed Data Act.
[245] See the definitions in Article 2(1), (2), and (3) of the proposed Data Act.
[246] Article 2(5) of the proposed Data Act defines 'user' as 'a natural or legal person that owns, rents or leases a product or receives services'.
[247] Explanatory memorandum of the proposed Data Act, p. 13.

personal data of individuals,[248] while other mandates are focused on certain industries (such as, the payment and the electricity sectors).[249]

Article 3(1) of the proposed Data Act requires products and related services to be designed and provided 'in such a manner that data generated by their use are, by default, easily, securely and, where relevant and appropriate, directly accessible to the user'. This implies that manufacturers of IoT devices will have to enable users to access data by default. When users cannot directly access data, Article 4(1) of the proposed Data Act obliges the data holder to make available to the user the data generated by its use of a product or related service 'without undue delay, free of charge and, where applicable, continuously and in real-time'. Upon the request of a user, the data holder needs to make available the data generated by the use of a product or related service to a third party.[250] In terms of the mode of access, the data should be made available 'without undue delay, free of charge to the user, of the same quality as is available to the data holder and, where applicable, continuously and in real-time'.

With regard to the timing of access, the user's request triggers the data access. Once his/her use of a product or related service generates data, a user can invoke the data access right. As a result, the receiving party can either be the user, namely a consumer or business user, or a third party to whom the user wishes to transfer his/her data. However, the proposed Data Act puts a limitation on who can qualify as a third party. Article 5(2) of the proposed Data Act namely makes undertakings designated as a gatekeeper under the Digital Markets Act ineligible as third parties. This means that gatekeepers cannot receive data from a data holder or a user under the proposed Data Act, nor can a third party make available the data it receives to a gatekeeper.[251] Considering the broad range of data access, the relevant provisions refer to the need to consider data protection interests as well as trade secret protection.[252]

3.2.2 Switching between data processing services (Articles 23-26)

The set of minimum legal obligations that the proposed Data Act imposes on providers of data processing services includes in Article 23(1)(c) an obligation to remove commercial, technical, contractual and organisational obstacles inhibiting customers from porting 'data, applications and other digital assets to another provider of data processing services' as the target data. These requirements can be categorised as **private party access to data** obligations. Because the purpose of the provisions is to let customers switch services in order 'to establish fair and competitive market

[248] Article 20 of the 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) [2016] OJ L 119/1.
[249] Respectively, Articles 66 and 67 of Directive (EU) 2015/2366 of 25 November 2015 on payment services in the internal market [2015] OJ L 337/35 and Article 23 of Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU [2019] OJ L 158/125.
[250] Article 5(1) of the proposed Data Act.
[251] Article 6(2)(d) of the proposed Data Act. The recitals explain that 'given the unrivaled ability of these companies to acquire data, it would not be necessary to achieve the objective of this Regulation, and would thus be disproportionate in relation to data holders made subject to such obligations, to include such gatekeeper undertakings as beneficiaries of the data access right'. Recital 36 of the proposed Data Act.
[252] Article 4(3) and (5) and Article 5(6), (8) and (9) of the proposed Data Act.

conditions for the internal market in cloud, edge and related services',[253] the receiving parties are providers of data processing services.

The obligations regarding cloud switching build on the self-regulatory approach that the Free Flow of Non-Personal Data Regulation[254] introduced to the problem of vendor lock-in and that resulted in the industry-developed 'Switching Cloud Providers and Porting Data (SWIPO)' Codes of Conduct.[255] The novelty is the switch to a regulatory approach.[256] Compared to the data access right in Articles 3-5 of the proposed Data Act, the provisions on cloud switching are narrower and more specific. As such, they can be expected to raise fewer concerns with regard to the protection of personal data and trade secrets as compared to the more broadly formulated data access right. However, considering the objective of Articles 23-26 to ease switching between cloud providers, trade secret protection may still be a relevant factor in their implementation. Beyond this, information security considerations remain important.

3.2.3 Access for public sector bodies to private sector data on grounds of exceptional need (Articles 14-22)

Finally, the Proposed Data Act also introduces a range of novel and important obligations that expand **regulator access to data**. Article 14 of the proposed Data Act requires data holders, with the exception of small and micro enterprises, to make data available upon request to a public sector body or to a Union institution, agency or body (the receiving parties) demonstrating an exceptional need to use the data requested.[257] With regard to the target data, Article 17(2)(d) states that requests for data need to concern non-personal data as far as possible. This mitigates data protection risks to some extent. Where compliance with data access requests requires the disclosure of personal data, Article 18(5) requires data holders to take 'reasonable efforts to pseudonymise the data, insofar as the request can be fulfilled with pseudonymised data'. Trade secret protection and the protection of commercially sensitive information also play a role. According to Article 19(2), disclosure of trade secrets to a public sector body is only required to the extent it is strictly necessary to achieve the purpose of the request. In such cases, the public sector body has to 'take appropriate measures to preserve the confidentiality of those trade secrets'.

According to the explanatory memorandum of the proposed Data Act, the purpose of the obligation is to 'enhance the capacity of public authorities to take action for the common good, such as to

[253] Explanatory memorandum of the proposed Data Act, p. 7.
[254] Article 6 of Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union [2018] OJ L 303/59.
[255] See https://swipo.eu/
[256] According to the explanatory memorandum of the proposed Data Act, 'the self-regulatory approach seems not to have affected market dynamics significantly'. Explanatory memorandum of the proposed Data Act, p. 14.
[257] An exceptional need is deemed to exist in line with Article 15 of the proposed Data Act: (1) where the data requested is necessary to respond to a public emergency; (2) where the data request is limited in time and scope and necessary to prevent a public emergency or to assist the recovery from a public emergency; and (3) where the lack of available data prevents the relevant body from fulfilling a specific task in the public interest that has been explicitly provided by law and the body has either been unable to obtain such data by alternative means and the adoption of new legislative measures cannot ensure the timely availability of the data, or obtaining the data from the data holder would substantively reduce the administrative burden for data holders or other enterprises.

respond, prevent or assist in the recovery from a public emergency'.[258] This form of data sharing is novel in EU law, because other forms of EU data access so far concerned data sharing between businesses or between businesses and consumers, and not between businesses and public sector bodies as foreseen by the proposed Data Act. Rule of law and proportionality requirements will need to be considered in the implementation and the relevant provisions in the Data Act already contain a balancing with the interests of data holders by specifying the situations and conditions for data access.

Overall, this outlook on the AI and Data Acts illustrates the relevance of the categories of data access and the variables we distinguished in Part I for other legislative initiatives beyond the DMA and DSA. Because transparency and data access provisions are likely to be at the core of policy and regulatory approaches for digital markets, this report hopes to have laid a foundation for how to balance conflicting interests and effectively implement such obligations now and in the future.

[258] Explanatory memorandum of the proposed Data Act, p. 14.

ANNEX I: A SUMMARY OF TRANSPARENCY AND DATA ACCESS OBLIGATIONS IN THE DMA/DSA

See the spreadsheet available at this link.

ANNEX II: FIELDS FOR AN ADVERTISEMENT TRANSPARENCY DATABASE

Proposed Standard Field	Description	Type
platform_archive_id	Id associated with the ad while being presented in a platform archive	bigint
texts	Text(s) of ads displayed	text[]
images	Image(s) of ads displayed	[]
videos	Video(s) of ads displayed	[]
links	Link(s) of ads displayed	[]
captions	Caption(s) associated with links displayed	text[]
disclosure_string	Text "aid for by" disclosure displayed to the user	text
advertiser	Advertiser name displayed to the user	text
ultimate_payer	Legal name of the advertiser who paid for the ad	text
platform_advertiser_identifier	Platform specific id associated with the advertiser	bigint
national_advertiser_identifier	Tax or election commission id associated with the advertiser, if applicable	text
ad_creation_date_time	Date, time when ad was created	datetime
ad_active_dates	Dates when ad was active	date[]
displayed_creative_combination	The combinations of ad creative elements as they were displayed to the user	json
budget	Daily budget amount for ad	float
spend_by_day	Amount paid by the advertiser for the ad	float[]
currency	Currency of the spend	text
removed	Whether the ad was removed	boolean
removed_reason	Reason the was was removed, if applicable	text
removed_date	Date, time when the ad was removed	datetime
total_impressions_by_day	All paid and organic ad impressions broken down by day	json

paid_impressions_by_day	Paid ad impressions broken down by day	json
engagement	Total ad engagement by type, if applicable	json
total_impressions_by_geography	All paid and organic ad impressions broken down by geographic region such as zipcode	json
paid_impressions_by_geography	Paid ad impressions broken down by geographic region such as zipcode	json
total_impressions_by_demo_group	All paid and organic ad impressions broken down by age and gender group	json
paid_impressions_by_demo_group	Paid ad impressions broken down by age and gender group	json
total_impressions_by_platform	All paid and organic ad impressions broken down by platform, if applicable	json
paid_impressions_by_platform	Paid ad impressions broken down by platform, if applicable	json
placement_details	Other details about where an ad appeared	text
targeting_type	Targeting categories used	text
targeting_details	Textual description of targeting	json
targeting_inclusive_parameters	Parameters of inclusion in ad targeting	json
targeting_exclusive_parameters	Parameters of exclusion in targeting	json
delivery_platform	Platform on which the ad was shown to the user	text
delivery_platform_opmisations	Targeting optimisations performed by the platform	text
deliverty_advertiser_input	Input from advertiser which is used by the platform to optimise delivery	text
delivery_proxy_or_external_data	Proxy optimisation via influencers or additional external data	text

cerre Centre on Regulation in Europe

Avenue Louise 475 (box 10)
1050 Brussels, Belgium
+32 2 230 83 60
info@cerre.eu
www.cerre.eu
@CERRE_ThinkTank
Centre on Regulation in Europe (CERRE)
CERRE Think Tank