16
Transparency

16.1 Introduction

Light has disinfectant properties. Although it might not be the best disinfectant for everything, its power is universal once cast. The same can be said about transparency. While the Digital Services Act's (DSA) transparency rules are framed as obligations observing the business of platforms, they are not actually about platforms. They are about *us*—the individual users of those platforms. When people voice their views in pubs or public squares, anyone can observe. If they do so in pseudo-anonymous chatrooms, it is less likely. Such data is suddenly *proprietary*, as it is amassed by the digital gatekeepers that operate chatrooms, newsfeeds, social media, and other services. Without transparency, we are in the dark about what is happening within our societies.

Consider the recent United States (US) lawsuit initiated by X (formerly Twitter) against the Center for Countering Digital Hate (CCDH), a non-profit organisation that tracks hate speech on social platforms.[1] X complains about researchers scraping its proprietary data to study societal risks. Its real purpose is to shut down any independent monitoring of its services to be able to woo back its concerned advertisers. Amazon's legal case against its designation as a very large online platform (VLOP) was potentially motivated by the possibility of temporarily suspending the obligation to create advertising archives.[2] There is no doubt that meaningful transparency undermines some of the practices that the companies would engage in otherwise. If companies are allowed

[1] *X Corp v Center for Countering Digital Hate Inc*, No 3:23-cv-03836 (ND Cal)
[2] Case T-367/23 R *Amazon Services v Commission* ECLI:EU:T:2023:589.

Principles of the Digital Services Act. Martin Husovec, Oxford University Press. © Martin Husovec 2024.
DOI: 10.1093/law-ocl/9780192882455.003.0016

to protect their commercial interests via data ownership aggressively, a lot of important research can be put at risk, and everyone will be in the dark about what is happening in the digital environment.[3]

To protect research, the DSA must think about the data access questions extraterritorially. If global non-governmental organisations (NGOs) can be sued for their DSA-endorsed analytical work outside of Europe, researchers might be discouraged from undertaking legally risky evaluations of how digital services pose risks to society at large. While the DSA transparency reports are a step forward, they do not provide the granular data needed to assess product-specific problems. Transparency reports only provide a rear-view mirror of what has transpired in society. To see forward, researchers' access to data is key.

Prior to the DSA, civil society and providers tended to negotiate the scope of transparency. The Santa Clara Principles were the most influential.[4] The principles largely envisaged what the DSA now delivers as binding rules. Its success is reflected in the fact that all major US platforms today issue some form of transparency reports.[5] Apart from the Santa Clara Principles, some other transparency tools, such as Harvard's Lumen Database,[6] directly inspired the European Commission in its DSA proposal.[7] Thus, undoubtedly, the DSA builds upon many years of voluntary disclosure and even attempted regulation in this area. For instance, the German hate speech law (NetzDG) that preceded the DSA also mandated certain transparency reporting, although with mixed results as to its actual usefulness, partly due to reliance on idiosyncratic methodologies.[8]

The early European legislation tended to omit rules on transparency. Only recently have the newly introduced Terrorist Content Regulation (TCR) (2021) and Child Sexual Abuse Regulation (2021) envisaged some transparency reporting by relevant actors.[9] The lack of transparency was sometimes remedied by voluntary reporting by companies, such as when Google published reports about the newly created right to be delisted from the search results.[10]

[3] Vittoria Elliott, 'How X is Suing Its Way Out of Accountability' *Wired* (15 August 2023) <https://www.wired.com/story/twitter-x-ccdh-lawsuit-data-crackdown/> accessed Joshua A Tucker and Nathaniel Persily (eds), *Social Media and Democracy: The State of the Field, Prospects for Reform* (CUP 2020).5 September 2023.

[4] 'Santa Clara Principles on Transparency and Accountability in Content Moderation' (*Santa Clara Principles*)<https://santaclaraprinciples.org> accessed 3 September 2023.

[5] For the pre-DSA situation, Daphne Keller and Paddy Leerssen, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation' in

[6] 'The Lumen Database' (*Lumen Database*) <www.lumendatabase.org/> accessed 3 September 2023.

[7] Commission, 'Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC' COM(2020) 825 final, art 17(5): ('Online platforms shall ensure that the decisions, referred to in paragraph 4, are not solely taken on the basis of automated means.')

[8] Heidi Tworek and Paddy Leerssen, 'An Analysis of Germany's NetzDG Law' [2019] Working paper of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression <https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf> accessed 6 September 2023.

[9] Arts 7 TCR, and 8–9 European Parliament and Council Regulation (EU) 2021/1232 on a temporary derogation from certain provisions of Directive 2002/58/EC as regards the use of technologies by providers of number-independent interpersonal communications services for the processing of personal and other data for the purpose of combating online child sexual abuse [2002] OJ L274/41 (Child Sexual Abuse Regulation).

[10] Theo Bertram and others, 'Five Years of the Right to be Forgotten', *Proceedings of the Conference on Computer and Communications Security* (2019).

Going forward, the DSA will likely transfer reporting behaviour into one of the general safeguards of any delegated enforcement in the digital environment. The reason is that as a horizontal measure, it applies in all areas of the law where illegality leads to content moderation. Thus, even if specific areas do not think of transparency, the DSA serves as a back-up set of default safeguards.[11]

The DSA transparency obligations create important data points for scrutiny regarding content moderation and risk mitigation practices by digital service providers. In their totality, they unveil many parts of previously impenetrable services for outsiders. The DSA increases the obligations incrementally according to a services' categorisation. The lowest level of universal transparency obligations is expected of all hosting, caching, and mere conduit services. The next level of advanced obligations is expected from hosting services that qualify as online platforms. Finally, VLOPs and very large online search engines (VLOSEs) are subject to special transparency obligations.

16.2 Access to Data of VLOPs/VLOSEs (Article 40)

Researchers are given an important role in the DSA. They help identify what counts as a risk. In effect, they formulate the agenda for the regulators and providers of digital services. They also monitor and assess those risks, their causes, and contributing factors, and suggest ways to mitigate them. Their suggestions are directly relevant to providers' compliance with the DSA.

To be able to do so, they get special access to data under the DSA, which cannot be easily refused. The DSA has changed the norm: now researchers pick their projects and platforms, not the other way around. Once their research demonstrates evidence of particular risks and their causes or suggests a way forward, it cannot be ignored (by providers or regulators) either. However, researchers need time and money to do this properly. Moreover, researchers act as a check on the abuse of power by the regulators. They hold regulators' legal demands and actions to account in line with what science says, not what politics finds convenient.

Unlike in other sectors, under the DSA, the relevant risks are to individuals, communities, and society at large—basically, everything we cherish. Starting from scratch can overwhelm and disorient regulators as to what the enforcement priorities should

[11] At the time of writing, transparency is hotly debated in other jurisdictions, such as the US. For instance, the 'Platform Transparency and Accountability Act' is a bipartisan bill that is currently being discussed. See Laura Edelson, 'Platform Transparency Legislation: The Whos, Whats and Hows' (*Lawfare*, 29 April 2022) <www.lawfareblog.com/platform-transparency-legislation-whos-whats-and-hows> accessed 6 September 2023. Many state and federal bills are trying to introduce some form of transparency to content moderation or other practices of online platforms. The US Supreme Court might soon hear a case about the permissible scope of general and individual transparency rules that can be imposed on platforms in the context of content moderation in several pending cases concerning two state bills (*NetChoice & CCIA v Paxton* No 1:21-cv-00840 (WD Tex) (1 October 2021) and *NetChoice & CCIA v Moody* No 4:21-cv-00220 (ND Fla) (30 June 2021)). It is likely that at least some of the DSA-style rules will be seen as unconstitutional if the Court is taken up, see Daphne Keller, 'Platform Transparency and the First Amendment' [2023] Stanford Cyber Policy Center <https://www.ssrn.com/abstract=4377578> accessed 6 September 2023.

be and how to deal with them. And this is where researchers are key. Researchers' toolkits and literature reviews[12] can become incredibly important.

Researchers have two broad data access tools: (a) project-based access to data held by providers (privileged access to data), and (b) general access to public data held by providers. In addition, regulators have their own routes for data access to assess the DSA compliance. Table 16.1 depicts the options.

The DSA does not explicitly limit access to either type of data to only European researchers. However, because of the requirement to assure data protection, security, and confidentiality, the data flows from outside of the European Union (EU) can be complicated by the General Data Protection Regulation (GDPR).[13]

16.2.1 Privileged access to data for vetted researchers (Article 40(4)–(11))

For privileged access to data, the DSA creates a system of vetting researchers who are granted project-based access to data (Article 40(8)). The access requests are issued by the Digital Services Coordinators (DSCs) of the establishment upon a duly substantiated application from the researchers.[14] Given that many VLOPs and VLOSEs are established in Ireland, this will be often the Irish regulator.[15]

To qualify as a vetted researcher, the persons must be:

a) affiliated with a research organisation,[16] including an NGO;[17]
b) independent from commercial interests;
c) disclose any relevant funding;
d) have capabilities to assure data security, confidentiality, and personal data protection;

[12] For an example of a toolkit for disinformation, see Anastasia Kozyreva and others, 'Toolbox of Interventions Against Online Misinformation and Manipulation' (16 December 2022) <https://interventionstoolbox.mpib-berlin.mpg.de/index.html> accessed 6 September 2023.

[13] On the latest situation, see Robert Maddox and others, 'Take Three: New European Adequacy Decision Gives Green Light to EU-U.S. Data Protection Framework' (*Debevoise Data Blog*, 31 July 2023) <https://www.debevoisedatablog.com/2023/07/31/take-three-new-european-adequacy-decision-gives-green-light-to-eu-u-s-data-protection-framework/> accessed 6 September 2023.

[14] The application can be also made in the country of the researcher's institution but it will be transmitted to the DSC of the establishment anyway, see the DSA, art 40(9).

[15] Mathias Vermeulen, 'Researcher Access to Platform Data: European Developments' (2022) 1(4) Journal of Online Trust and Safety 4.

[16] Art 40(8)(a) together with art 2, point 1 Copyright in a Digital Single Market (CDSM) Directive: '"research organisation" means a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation;'

[17] Rec 97 seems to interpret the definition for the purposes of the DSA as allowing the NGOs to qualify as research organisations: 'a research organisation within the meaning of Article 2 of Directive (EU) 2019/790, which may include, for the purpose of this Regulation, civil society organisations that are conducting scientific research with the primary goal of supporting their public interest mission'.

Table 16.1 Types of data access

Article 40 DSA	Privileged access to data (Art 40(4)–(11))	Public data (Article 40(12))	Regulator's access (Art 40(1)–(3))
Beneficiaries	Vetted researchers	Any 'researchers'	DSA authorities
Sources of data	VLOPs and VLOSEs	VLOPs and VLOSEs	VLOPs and VLOSEs
Flexibility of access	Project-based	General-purpose access	Necessity for regulators
Speed of access	Slower (in months)	Faster	Faster
Eligible purposes	(1) detection, identification and understanding of systemic risks in the EU; (2) assessment of adequacy, efficiency and impacts of risk mitigation measures	detection, identification and understanding of systemic risks in the EU	Monitoring and assessment of compliance with any DSA obligation
Required affiliation	Affiliation with an entity anywhere in the world devoted to scientific research that operates: (a) on a not-for-profit basis or reinvests all the profits in its research; or (b) pursuant to a public interest mission recognised by an MS	Researchers from anywhere in the world, including those working for non-profits, charities, think tanks, consumer groups, journalists, etc.	Digital Services Coordinator of provider's establishment, or the European Commission
Procedure to obtain access	DSC approval process for projects	None	Reasoned request
Conditions for use of data in teams	Use within the scope of the approved project and among its members	Use for any own research purpose, and sharing with any other eligible researchers	Use to monitor and assess compliance within the regulator's powers
Explicit funding, access and publication requirements	GDPR, data security and confidentiality safeguards; independence from commercial interests; funding disclosure; free of charge publication;	GDPR, data security and confidentiality safeguards; independence from commercial interests; funding disclosure;	Balancing the regulatory needs with the rights of others

e) explain how the data access is necessary and proportionate given their methodology and research questions;
f) fulfil at least one of the DSA's accepted purposes (detection, identification, and understanding of systemic risks or assessment of the risk mitigation measures); and
g) committed to making the research results publicly available free of charge.

The access to researchers can be granted only 'for the sole purpose of conducting research that contributes to the detection, identification and understanding of systemic

risks in the Union', including 'the assessment of the adequacy, efficiency and impacts of the risk mitigation measures' (Article 40(4)). Vermeulen referred to these two functions of researchers as 'pathfinders' and 'quasi-auditors'.[18]

Given the scope of the DSA's covered threats, the provision has a much broader impact than offering only an opportunity to conduct some parallel external audits. It invites scientists to come and scrutinise the most significant platforms to better understand the systemic risks and how to effectively approach them. The provision is a giant subsidy to social sciences which otherwise have to collect their own data, and find research subjects. While the scope of access to data is limited to the DSA's regulatory remit, it is clear that through understanding risks to individuals, communities, and society at large, we also advance our understanding of people in their social interactions. The close linkage to regulatory purposes also means that the research must be limited to issues within the EU. It is thus not possible to only gain access to study risks specific to the Global South. However, the issues studied can be universally applicable, and the insights gained are also useful for other jurisdictions. It should also be possible, for instance, to use non-European settings as control groups for the EU-focused studies. Finally, the Codes of Conduct can extend the reach of the data access provisions and 'non-EU' issues.

With the privileged access, the hope is that researchers eventually formulate an agenda for the risk-assessments, methods of capturing various risks, and reflect upon various ways to mitigate them. Article 40 is thus designed as a crucial accountability tool because it makes providers accountable not only to the state authorities but also to science. While the authorities might be temporarily deaf to scientific evidence, providers should listen to it very carefully for the sake of their own compliance; science and evidence, after all, is also a check on authorities and auditors.

The access to data can take various forms. Recital 97 highlights that data access requests could cover questions like 'the number of views or, where relevant, other types of access to content by recipients of the service prior to its removal by the providers'. The DSA does not limit the compliance only to the type of data that the platform already has in a ready-made dataset. Thus, a certain amount of data collection and preparation is implied in the DSA's data access regime.

Each data access request is submitted by researchers to the DSCs. They can either directly submit to the DSC of the provider's establishment or the DSC of their affiliation. In either case, the DSC of the provider's establishment decides about the request. The DSC then forwards each data access request to providers who can propose to modify it within 15 days. However, they can propose amendments only on two grounds: (a) if they 'do not have access to the data', or (b) if 'giving access would lead to significant vulnerabilities in their security or the protection of confidential information, in particular trade secrets' (Article 40(5)). In both situations, they must be constructive and propose alternative means of access which the researchers could use in

[18] Vermeulen (n 15) 3.

their projects. Ultimately, the DSC of the establishment decides whether to modify the access request or not.

For the ground of modification under Article 40(5)(b) regarding confidential information, the compromise is that while access is allowed, the researchers are required to observe applicable laws, including trade secrecy and data protection laws, when reporting their results. However, these laws often include considerations of legitimate research interests in their statutory designs. As such, the trade secrecy defence cannot be relied upon to trump legitimate research projects. Indeed, Recital 97 also makes it clear that purely 'commercial interests of providers should not lead to a refusal to provide access to data necessary for the specific research objective'. The research of the relevant risks is thus given priority over absolute commercial secrecy.[19]

Moreover, access to data according to Article 40(4) does not immediately mean that trade secrecy rights are endangered. Trade secrecy does not mean no one outside the company can see the information. If this were true, licensing agreements with business partners would be impossible without destroying secrecy. Trade secret rights are preserved as long as 'it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret'.[20] In practice, this usually means asking such parties to sign non-disclosure agreements.

Thus, to even have a conversation about rejection based on Article 40(5)(b), VLOPs/VLOSEs must explain why the non-disclosure agreements are not sufficient to solve the risk of public disclosure. Only if this is convincing can the provision be relied upon. As a result, it should be expected that VLOPs/VLOSEs might require researchers to sign non-disclosure agreements regarding certain aspects to preserve their protection of trade secrets. Such agreements are, however, very sensitive because they can endanger academic integrity. The regulators, including the European Commission, should try to standardise such agreements to prevent VLOPs/VLOSEs from using their bargaining position to impose undue conditions. One can borrow from existing practices in the area of court-appointed experts (eg in competition cases) who are often asked to receive sensitive information from firms to produce their assessments. However, the DSCs should be aware that protection claims can sometimes be overstated. The companies' reflex is to try to protect factual secrecy over algorithms, data or components; however, often those components have little chance to be also protected legally by trade secrecy laws.[21]

If the vetted researchers fail to live up to the above requirements of trust, the DSC of the provider's establishment can terminate their status (Article 40(10)). The European

[19] For a comprehensive report on data access under the DSA, see Laura Edelson, Inge Graef, and Filippo Lancieri, 'Access to Data and Algorithms: For an Effective DMA and DSA Implementation' (CERRE 2023) <https://cerre.eu/publications/access-to-data-and-algorithms-for-an-effective-dma-and-dsa-implementation/> accessed 6 September 2023.

[20] European Parliament and Council Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use, and disclosure [2016] OJ L 157/1, art 2(1)(c).

[21] See Sharon K Sandeen and Tanya Aplin, 'Trade Secrecy, Factual Secrecy and the Hype Surrounding AI' in Ryan Abott (ed), *Research Handbook on Intellectual Property and Artificial Intelligence* (Edward Elgar (forthcoming)) <https://papers.ssrn.com/abstract=3929928> accessed 6 September 2023.

Commission and the Board will have been collecting the names of vetted researchers and descriptions of their projects, including those whose status was previously terminated.

The European Commission is also empowered to adopt a delegated act within five years of the DSA's entry into the force (Article 87). Such delegated acts can further specify technical conditions addressing access to data and addressing important compliance questions (Article 40(13)). The data access rules also form parts of the Codes of Conduct, such as the Code of Conduct on Disinformation, which have regulatory relevance too.[22]

The clear disadvantage of privileged access to vetted researchers is its limitation to project-based investigations. Without prior access to data, it might be hard to properly formulate research projects. The risk is that the vetting process is conceived too formalistically and pushes researchers to construe very narrow research questions, thus not allowing any changes in the research trajectory. If the researchers are forced to resubmit numerous requests to reflect the inherently iterative character of the research, it can not only slow down the researchers but also discourage them from using the tool. Thus, the delegated acts must carefully calibrate the requirements to avoid being too prescriptive. The DSCs might borrow some procedures from the elaborate European grants systems that allow more risk-taking by the researchers, such as European Research Council grants.

Any disputes about the scope of Article 40(4) requests are going to be losing battles for researchers. If VLOPs/VLOSEs start seeking judicial review of the data access grants, the system can be slowed down to the point of complete uselessness. To avoid this, amicable and fast resolution of disputes should be preferred. Article 40(13) allows the Commission to create 'independent advisory mechanisms in support of sharing of data'. This mechanism could also help with the informal resolution of disputes. It should be composed of specialised lawyers, ethicists, and leading social scientists, including data scientists, who can provide amicable resolution of any disputes by providing advisory opinions. The Commission might encourage its use in the delegated act by allowing parties or regulators to initiate such opinions in controversial or difficult cases before proceeding to the final decision, according to Article 40(8).

Article 40(13) empowers the European Commission to specify 'technical conditions' regarding the privileged data access by vetted researchers, including on issues such as 'relevant objective indicators, procedures and, where necessary, independent advisory mechanisms in support of sharing of data'. Thus, the DSA foresees the emergence of additional structures of advisory mechanisms that could help the DSCs when assessing research projects, including compliance with data protection, security, and confidentiality.

The Achilles heel of researchers as agents of public interest lies in funding their work. The researchers engaging in such data-intensive projects need money to conduct them properly and independently. If the only funding available to do research comes from

[22] For more detail, see Vermeulen (n 15).

the industry, even remotely, we have a problem. European academia needs specific grants for researchers who want to use the data opened up by the DSA. However, the financial support must be equally independent of the authorities that act as regulators. Researchers cannot act as a check on the abuse of state power exercised by regulators if they also need funding from the same authorities.

Researchers also play another role in the DSA's design. The more robust their evidence, the less need for regulators to demand data access. Because digital platforms store a lot of sensitive information about individuals and their mutual interactions, keeping the state from accessing such data protects individuals' privacy. This is another reminder of why digital services regulation cannot employ the same approach to data handling that regulators take when dealing with chemicals or food. There, the key consideration is trade secrecy. With online platforms, it is user privacy, data protection, and freedom of expression. The DSA's design empowers researchers to reduce the need for state powers to exercise as close oversight directly as is customary in other areas. This is not only a result of the resource constraints of authorities, but also of the sensitive nature of the data and risks being studied.

16.2.2 Access to public data (Article 40(12))

To complement the privileged access tool, the DSA also considers the role of publicly available data that is already often mined and studied by researchers. This data already disclosed to the public by platforms and their users is usually accessible through various interfaces. The provision was inspired by Crowdtangle,[23] a tool that independently tracks how user-generated content spreads around the web. Such data aggregation tools often serve various stakeholders enforcing their rights or by providing insights into how content travels from one service to another. For researchers, the tool helps to aggregate what is going on in the digital public sphere, such as on social media, thus providing a kind of real-time pulse of the web.[24]

Article 40(12) foresees access to such publicly accessible data, which often includes real-time data from social media or video-sharing platforms. The provision grants access to a broader set of 'researchers' than those eligible to become 'vetted'. Recital 98 explains the provision as follows:

> In addition, where data is publicly accessible, such providers should not prevent researchers meeting an appropriate subset of criteria from using this data for research purposes that contribute to the detection, identification and understanding of systemic risks. They should provide access to such researchers including, where technically possible, in real-time, to the publicly accessible data, for example on aggregated

[23] ibid 5.

[24] Casey Newton, 'Facebook Buys CrowdTangle, the Tool Publishers Use to Win the Internet' *The Verge* (11 November 2016) <https://www.theverge.com/2016/11/11/13594338/facebook-acquires-crowdtangle> accessed 6 September 2023.

> interactions with content from public pages, public groups, or public figures, including impression and engagement data such as the number of reactions, shares, comments from recipients of the service.

The goal is thus clearly to extend the scope of access to data provisions to researchers who cannot be become vetted for specific research projects but who build or use real-time tools from the already publicly accessible data, such as feeds of social networks. While Article 40(12) is worded as a 'shall' provision, the recital seems to emphasise mostly the defensive dimension, ie that providers may not shut down or obstruct such analytical tools. Article 40(12) does not seem strong enough to force the companies to build such a tool for the research community. However, if advertisers already benefit from some such tools, Article 40(12) would seem strong enough to ask providers to extend such tools to researchers. In practice, where data is publicly accessible, researchers can build their own tools to scrape and analyse the content. However, doing so might be very expensive; thus, ready-made tools others provide are crucial. Providers often try to impose technical constraints on scrapping tools operated by researchers or other companies or use legal threats based on intellectual property rights to stop them.

Article 40(12) acts as a *sword* and a *shield*. As a shield, it provides a legal defence to unjustified attempts to shut down access because it acts as a lex specialis to the rights of providers.[25] As a sword, it provides researchers with a claim against obstructing access to publicly available data.

The use as a shield is important for any litigation that providers might initiate against researchers, such as on the basis of violations of sui generis database protection rights, copyright law, technical protection measures, unfair competition law or contract law.[26] Article 40(12) of the DSA here goes beyond Articles 3–4 of the Copyright in the Single Digital Market (CDSM) Directive that equally permits some forms of data mining, including for research purposes.[27] Any litigation trying to shut down these activities or tools is pre-empted by the provision, as long as its conditions are met. Article 40(12) thus acts as a safe harbour against other areas of law.

The use as a sword is crucial when providers use technical restrictions, such as blocking Internet Protocol (IP) addresses or reliance on captchas, that are clearly meant to frustrate access by researchers. It allows them to rely on the DSA to lift technical restrictions that are seen as illegitimate obstacles to data access, or at least to obtain special 'scraping exceptions' from these technical restrictions.

[25] A similar analogy is presented by Paddy Leerssen, 'Seeing What Others Are Seeing: Studies in the Regulation of Transparency for Social Media Recommender Systems' (PhD thesis, University of Amsterdam 2023) <https://pure.uva.nl/ws/files/125112610/Thesis.pdf> accessed 6 September 2023.

[26] See Martin Husovec, 'The End of (Meta) Search Engines in Europe?' (2014) 14(1) Chicago-Kent Journal of Intellectual Property 145. For after the change in case law, see Estelle Derclaye and Martin Husovec, 'Sui Generis Database Protection 2.0: Judicial and Legislative Reforms' [2021] European Intellectual Property Review (Forthcoming) <https://papers.ssrn.com/abstract=3964943> accessed 6 September 2023.

[27] See generally, Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71(8) GRUR International 685.

The eligible persons building or using the tools must be 'researchers' who use the data solely for studies that contribute to the detection, identification and understanding of systemic risks in the Union. Article 40(12) does not address how these researchers can further provide access to others. Presumably, any researcher could invoke the provision for access to data that is provided to them by other researchers. However, potential access by citizens is probably not possible. Article 40(12) incentivises researchers to build and maintain the tool only for other researchers. Thus, the public can learn from using such tools only through research publications.

The privileged access is limited to those eligible to become vetted researchers. The access to public data allows access by a broader set of researchers. Who exactly are such researchers under Article 40(12) is not further specified anywhere in the DSA. Thus, they can include persons who are *not* affiliated with a 'research organisation' within the meaning of the CDSM Directive.[28] This creates legal risks for any unaffiliated researchers, such as NGOs, who might decide to maintain or build a tool with access to public data of platforms. The European Commission should thus further clarify the concept.

The reliance on application programming interfaces (APIs) offered by online platforms has not been without problems.[29] The tools have their internal limitations and have been increasingly either shut down or made subject to high fees. The online platforms have been found to restrict access to APIs to selected researchers.[30] Some academics have also accused online platforms of potentially manipulating the data in APIs, and argued extensively that scraping techniques are necessary to identify potential instances of manipulation.[31]

Thus, while API access based on Article 40(12) should be further developed through Codes of Conduct, the researcher-initiated scraping should remain central. Scraping provides alternative access to public data regardless of what APIs or other tools are offered by providers. It is a technique important for the validation of any API-provided data sets. Moreover, scraping puts pressure on firms to create properly functioning APIs, and facilitates researchers explore the relevant risks before they formulate their requests pursuant to Article 40(4).

This fact is also underscored by the General Product Safety Regulation, which states that online marketplaces must 'allow the scraping of [their platform] data only for product safety purposes based on the identification parameters provided by the requesting market surveillance authorities', including by removing the technical obstacles.[32]

[28] Art 40(8)(a) DSA, read together with art 2(1) CDSM Directive.

[29] Axel Bruns, 'After the "APIcalypse": Social Media Platforms and their Fight against Critical Scholarly Research' (2019) 22(11) Information, Communication & Society 1544.

[30] See Nicolas Kayser-Bril, 'AlgorithmWatch Forced to Shut down Instagram Monitoring Project after Threats from Facebook' (*AlgorithmWatch*) <https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/> accessed 16 September 2023.

[31] Jade Garcia Bourrée and others, 'On the Relevance of APIs Facing Fairwashed Audits' [2023] arXiv2305.13883 [cs.LG] 1.

[32] European Parliament and Council Regulation (EU) 2023/988 on general product safety, amending Regulation (EU) No 1025/2012 and Directive (EU) 2020/1828, and repealing Directive 2001/95/EC and Directive 87/357/EEC [2023] OJ L135/1, art 22(12)(i) (General Product Safety Regulation).

The European Commission should encourage the research in the following ways:

- The delegated act should clarify that Article 40(12) is the backstop for any qualified researcher and that it covers scraping and highlights its exploratory value for Article 40(4).
- The European regulators should develop a detailed safe harbour for scraping according to Article 40(12), which specifies under what conditions scraping is manifestly legal and cannot be legitimately opposed by VLOPs and VLOSEs. This would not serve as an exhaustive statement of the law but would make IRB approvals much easier. The safe harbour could be adopted as an EC Guidance or a Code of Conduct.

The expanded scope of beneficiaries is particularly relevant to non-profit organisations, consumer groups, charities, investigative journalists, security researchers, etc. However, even researchers are subject to the requirement for their 'independence from commercial interests'. This term, which can be further specified by the delegated act, limits both types of data access. The hint that this term should be construed narrowly is also clear from the fact that the broader set of researchers, according to Article 40(12), do not have to publish their results in freely accessible publications.[33] Thus,a journalist who use the tools to write an investigative story for their media outlet that is subject to a paywall would seem to still qualify. However, given that media owners are increasingly in legal rows with platforms about the licensing of their content, such journalists could also be seen as not independent from commercial interests.

In contrast to privileged access, access to public data remains a faster, more horizontal but less in-depth scrutiny tool. The main challenge here is to create tools which decrease the costs for researchers to access the data. Clearly, not every researcher has the skills and resources to set up their own tools or even comply with data protection, security, and confidentiality requirements. There is hope that universities will build specific infrastructure that researchers can use similarly to libraries that facilitate access to books and databanks. The same approach can be taken by consortia of investigative journalists, non-profit organisations, and think tanks. There is thus room for collaboration on building some intermediaries who set up tools to be used by different eligible individual researchers.

Finally, the obvious unresolved issue concerns data protection laws and how they will shape access to data by researchers. Recital 97 reminds us of the GDPR's applicability and that 'providers should ensure appropriate access for researchers, including, where necessary, by taking technical protections such as through data vaults'. Similarly, Recital 98 says that 'providers and researchers should pay particular attention to the protection of personal data, and ensure that any processing of personal data complies with Regulation (EU) 2016/679. Providers should anonymise or pseudonymise personal data except in those cases that would render impossible the research purpose

[33] Art 40(12) drops the requirement art 40(8)(g).

pursued'. In other words, the DSA does not resolve exactly how data protection interests shall be resolved in practice. The data protection authorities will thus have to develop approaches that do not compromise the goals behind Article 40 which sufficiently safeguard the data protection rights of affected individuals.

16.2.3 Access to other data enabled by the DSA

It should be emphasised that data access is not limited to Article 40 DSA. Various transparency obligations discussed in this chapter further increase the scope of public data about regulated services. Transparency reports (Articles 15, 24, and 42) include rich and valuable information about content moderation decisions and the internal workings of companies running them. Archives for advertising (Article 39) track information about advertisers, their commercial communications, the duration of individual campaigns, and their audience, including ways in which they are targeted by advertisers. The fair design requirements, such as codification of content moderation rules, including notification of significant changes (Article 14), central collection of statement of reasons for each content moderation decision (Article 24(5)), or disclosure of the main parameters in the recommender systems, further improve the data collection capabilities of researchers. The auditing requirements also publish reports concerning how VLOPs and VLOSEs identify and mitigate risks. The DSA often anticipates researchers and their bots when upgrading obligations of various kinds to 'machine-readable'.[34]

16.3 Transparency Reports

The transparency obligations in the DSA deal with three basic areas: content moderation, advertising, and user base. Content moderation is at the centre of transparency obligations. Advertising and user-base information are relevant for players who potentially qualify as VLOPs or VLOSEs.

16.3.1 Annual reports on content moderation (Article 15)

All intermediary service providers must issue annual reports that 'engage in any content moderation'. The DSA, however, uniquely exempts micro and small enterprises *if* their services do not qualify as VLOPs or VLOSEs.[35] It equally exempts those services which did not engage in any content moderation, both on its own initiative, and

[34] Arts 14(1), 14(5), 15(1), 22(5), 24(5), 31(3), and 55(1).

[35] Commission Recommendation of 6 May 2003 concerning the definition of micro, small, and medium-sized enterprises [2003] OJ L124/36.

Table 16.2 Universal transparency reports (Article 15 DSA)

	Content moderation activity	Mere conduits	Caching	Hosting
Universal obligations (Article 15)	Any orders from the EU Member States (Art 15(1)(a))	(1) absolute number of orders, categorised by: (2) type of illegal content, (3) issuing Member State, (4) median time for handling, and (5) taking the action		
	Notices sent by third parties (Art 15(1)(b))	Voluntary use and reporting		(1) absolute number of notices, (2) divided per types of illegal content, (3) absolute numbers for trusted flaggers, (4) actions taken based on ToS or law, (5) absolute number of automated decisions, and (6) median time for taking the action
	Any content moderation based on own investigations (Art 15(1)(c))	(1) absolute number of relevant restrictions, categorised by: (2) each restriction type, (3) detection method, (4) type of illegal content, (5) basis in ToS or law, and providing (5) other information concerning automated tools, training, and assistance to content moderation staff		
	Any internal appeals of content moderation decisions (Art 15(1)(d))	*At least:* (1) absolute number of internal appeals, (2) median time for taking the action, (3) absolute number of reversals; *for online platforms only*: categorisation by (4) the restriction type, and (5) basis in ToS or law		
	Any use of automation for content moderation (Art 15(1)(e))	*At least:* (1) qualitative description, (2) specification of the precise purposes, (3) indicators of the accuracy, (4) error rates and (5) any safeguards applied.		
	Any orders from outside the EU	No, voluntary reporting (however, if they affect accessibility in the EU, they should be disclosed under their own investigations).		

because it had not receive any requests to do so. Thus, an internet access provider, or messaging service that does not store any messages, could avoid issuing reports only if it received no requests to block websites or notices concerning such websites.[36] Engaging in content moderation, therefore, means both refraining from moderating content and also not being exposed to any requests to conduct content moderation by others. All the reports must be made publicly available in a machine-readable and easily accessible format on an annual basis. Given that content moderation not only relates to individual content but also user accounts and privileges, it is thus safe to say that most of the relevant regulated services will have to issue annual transparency reports (see Table 16.2).

[36] Given that providers of browsers could also qualify as mere conduits (ch 6), laws that allow blocking orders against browser operators would be subject to art 15 transparency disclosures. For a French example, Udbhav Tiwari, 'France's Browser-Based Website Blocking Proposal Will Set a Disastrous Precedent for the Open Internet' (*Open Policy & Advocacy*, 26 June 2023) <https://blog.mozilla.org/netpolicy/2023/06/26/france-browser-website-blocking> accessed 16 September 2023.

The *orders* caught by Article 15(1)(a) are only orders issued by the Member State authorities. Orders issued by authorities outside the EU are not covered because they are generally presumed not to impact citizens in the EU. However, a special case might be orders with extra-territorial effects from outside of the EU. These orders produce effects in the EU although they were not issued by local authorities.

The *notices* received must be published only by hosting providers (Article 15(1)(b)). However, if mere conduits or caching providers act upon notices, from the DSA's perspective, they act without being obliged to do so. Thus, they must report such actions under their investigations. Notices are only an input for these investigations. The notices covered includes both notices sent on the basis of illegality or terms of service violations. The share of notices processed by automated means must also be noted.

What counts as a notice is not specified by the DSA. However, inferring from other uses of the term across the regulation, it is clear that the term is not about individual claims which are notified but rather the *message* in which they are notified. Thus, a single notice, can may put providers on notice concerning many items of content. This fact alone, which is hard to avoid, makes any simple benchmarking of numbers akin to the comparison of apples and oranges. For online platforms, any data reported in transparency reports can be analysed together with the EU database of statements of reasons that should give the public further insights into specifics of industry notification practices. The same insights, however, are not available for services other than online platforms, as they are either not obliged to issue a statement of reasons at all (caching and mere conduits), or do not have to report it (non-platform hosting services).

Similar obligations regarding notices likewise apply to the *internal appeals* concerning content moderation decisions submitted to providers (Article 15(1)(d)). However, here, the obligation applies equally to other intermediary service providers—those being caching and mere conduits. The reason is that these providers can equally engage in content moderation, although such behaviour is voluntary. If they do so, they must report it as their own investigations and the complaints to such decisions under the section dedicated to complaints. Naturally, online platforms subject to the most elaborate rules concerning content moderation have the most extensive obligations to report.

Any content moderation based on provider's own investigations—ie investigations where the provider exercises some discretion—must be further elaborated by qualitative description that offers meaningful and comprehensible information about the content moderation practices (Article 15(1)(c)). This includes information about content moderation professionals employed and their training. The qualitative information is complemented by numerous statistics, as summarised above. In this context, it is worth noting that the scope of relevant activities that qualify as content moderation is vast. It captures any activities 'aimed, in particular, at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service' (Article 3(t)). Thus, demotion, de-ranking,

Table 16.3 Examples of transparency reports

Examples for each service	Mere conduits	Caching	Hosting
Any orders from the EU Member States (Art 15(1)(a))	*Identity disclosure orders or website blocking orders*	*Preservation orders, or service blocking or mandates*	*Identity disclosure and content removal orders, staydown-orders, or user-blocking-orders*
Notices sent by third parties (Art 15(1)(b))	N/A	N/A	*Removal requests sent to webhosts, social media, cloud-based messaging apps, or app stores*
Any content moderation based on own investigations (Art 15(1)(c))	*Detection activities concerning child abuse or piracy*	*Termination of services to some clients by CDNs*	*De-ranking of users or content in the recommendation systems or demonetisation of accounts*
Any internal appeals to content moderation decisions (Art 15(1)(d))	*Appeals to account suspensions*	*Appeals to account terminations*	*Appeals of notifiers or content creators on video-sharing platforms*
Any use of automation for content moderation (Art 15(1)(e))	*Detection tools for abuse on networks*	*Detection tools for abuse of services*	*Detection tools for abuse, or decision-assistance for human moderators, prioritisation tools*

demonetisation, removal, suspension, or termination all qualify as relevant activities, regardless of whether their basis is terms and conditions or illegality.

Finally, to round up the process of content moderation, all automated tools regardless of their content (detection, initial assessment or assistance, decisions, or complaints), must be qualitatively and, to the extent possible, also quantitatively described. The DSA's significant innovation is that it includes horizontal safeguards for many of these tools by forcing disclosure of error and accuracy rates, safeguards, and specification of purposes for which such technologies are used. This information will make important contributions in many areas, including in copyright law, where filtering technology is mandated in some cases (Chapter 18.3).

Table 16.3 provides an overview of typical orders falling under each of the sections.

Given that transparency obligations form part of due diligence obligations, for VLOPs, they form part of the audits. To enhance the usefulness of data points, the Commission is empowered to adopt implementing acts[37] laying down templates in which the above information is to be reported.[38] The first such implementation can be adopted only following the advisory procedure, and thus is unlikely to enter into effect before providers issue the first round of reports.

[37] TFEU, art 291.
[38] DSA, art 88(3).

16.3.2 Advanced reporting for online platforms and search engines (Article 24)

Online platforms are subject to elaborate procedural norms regarding content moderation. Transparency is thus understandably extended for these services to cover other aspects of such procedures, such as disputes using out-of-court dispute settlement (ODS) and mandatory suspensions. These obligations are complemented by separate reporting obligations of the ODS bodies (Article 21(4)), trusted flaggers (Article 22(3)), and potentially also state authorities.

Online platforms also have an enhanced reporting obligation concerning their monthly active users. This information should establish whether they are designated as very large. Although search engines have no other reporting obligations unless they qualify as one of the intermediary services, they do have an obligation to report their monthly users regardless of their size. This disclosure of active users must follow the methodology adopted by the European Commission through a delegated act. The disclosure must take place via the provider's dedicated online interface and updated every six months.[39] If requested by the DSC of establishment or the European Commission, the provider must provide the latest data at a point in between these periodic disclosures. They can also request further explanations regarding the exact underlying calculations.

Table 16.4 provides an overview of specific transparency reporting based on tiers.

The most far-reaching transparency requirement concerns the statement of reasons to the EU's central databases.[40] These statements must be first stripped of any personal data. As explained by Recital 66, 'in order to keep the database continuously updated, the providers of online platforms should submit, in a standard format, the decisions and statement of reasons without undue delay after taking a decision, to allow for real-time updates where technically possible and proportionate to the means of the online platform in question'. The volume of these individuals' statements is likely to be massive.

The challenge will be how to apply this reporting obligation in a way which does not discourage more nuanced forms of content moderations due to the compliance costs considerations of providers.[41] Keller argues that '[h]ighly prescriptive transparency obligations might also drive de facto standardisation and homogeneity in platform rules, moderation practices, and features'.[42] The concern about the cost of standardisation to

[39] At time of writing, we already see the first stages of such reporting playing out in February 2023 and August 2023, but at present, without guidance from the delegated act as it has yet to come into force. For the full discussion and examples of such disclosures, ch 9.

[40] DSA, art 17(5).

[41] This point was raised by Daphne Keller, 'Some Humility About Transparency' (*The Center for Internet and Society – Stanford Law School*, 19 March 2021) <https://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency> accessed 6 September 2023.

[42] See ibid.

Table 16.4 Platform transparency reports

	Content Moderation Activity	Mere Conduits	Caching	Mere Hosting	Online Platform	Search Engines
Platform transparency obligations (Article 24)	ODS (Art 24(1)(a))	Voluntary use and reporting			(1) absolute number of disputes, (2) their outcomes, (3) median time for completion, and (4) the share of implemented decisions	Voluntary use and reporting
	Mandatory suspensions (Art 24(1)(b))				(1) absolute numbers of mandated suspensions, categorised by: (2) the grounds (illegal content, manifestly unfounded notices, or manifestly unfounded complaints)	
	Monthly active users (Article 24(2))	Voluntary use and reporting			at least every six months: (1) average monthly active recipients of the service, categorised by: (2) each relevant service operated by the provider	
	Statement of reasons (Article 24(5))	Voluntary use and reporting			continuous submission of statements of reasons to the EU database	Voluntary use and reporting

innovative content moderation practices indeed should not be lost on regulators, as the potential trade-off between two pursuits—innovation and standardisation—is well established.[43] However, standardisation can equally act as a catalyst of innovation.[44] Standardisation does not only lock in suboptimal solutions, as was argued in the debate around QWERTY keyboards,[45] but can actually also encourage innovation through newly gained scale, such as standardised shipping containers.[46] There is no reason to believe that content moderation standardisation will be any different. That said, the innovation concerns should still be on the radar of the regulators to avoid lock-ins into content moderation practices of the past.

[43] Knut Blind, 'Standards and Innovation—What Does the Research Say?' (ISO 2022) ISO research and innovation papers <https://www.iso.org/publication/PUB100466.html> accessed 6 September 2023.

[44] For overview of the academic debate, Knut Blind, 'Standardisation as a Catalyst for Innovation' [2009] ERIM Report Series Reference No EIA-2009-LIS <https://papers.ssrn.com/abstract=1527333> accessed 6 September 2023.

[45] The QWERTY keyboard is a famous example of suboptimal solutions that lock in everyone. However, the situation seems a bit more complex, even from the innovation standpoint: Neil M Kay, 'Rerun the Tape of History and QWERTY Always Wins' (2013) 42(6) Research Policy 1175.

[46] Steve Saxon and Matt Stone, 'Container Shipping: The Next 50 Years' (*McKinsey & Company*, October 2017) <www.mckinsey.com/~/media/mckinsey/industries/travel%20logistics%20and%20infrastructure/our%20insights/how%20container%20shipping%20could%20reinvent%20itself%20for%20the%20digital%20age/container-shipping-the-next-50-years-103017.pdf> accessed 6 September 2023.

16.3.3 VLOPs' and VLOSEs' content moderation reports (Article 42)

VLOPs and VLOSEs have enhanced obligations to issue content moderation annual reports in one of the official languages of the EU. They must be issued more frequently (at least every six months) and be more elaborate when setting out:

- content moderation staff dedicated to each relevant official EU language;
- qualification and linguistic expertise of content moderation staff;
- accuracy of automated tools for each relevant official EU language; and
- average monthly recipients of the service for each EU Member State.

The content moderation staff must only include numbers about those members whose work is relevant for the EU single market. Thus, employees or contract staff in Asia or the Americas are irrelevant. Information about the staff must also be reported 'in respect of the service offered' (Article 42(2)(a)). This means that each regulated service must be subject to a specific report. Thus, information about the content moderation staff of an app store can be copied and reported for a video-sharing service if the same staff worked on both services.

Article 42(2)(a) also implies that the breakdown should specifically reveal what resources are dedicated to internal complaint mechanisms, and enhanced support for trusted flaggers.

The breakdown per the relevant official EU language is meant to unveil structural differences between how platforms allocate their resources. The providers must report their average monthly users per the Member State individually. At the time of writing, it is impossible to find out how many content moderators even bigger services employ for each language in which they operate. This is reinforced by the requirement that such numbers are accompanied by a qualitative description of their training and linguistic expertise.

The reported information is an indication that helps to understand the attention given by providers to individual societies within the EU and its corresponding effect. After the DSA, they can be contrasted with the reported amount of advertising and content moderation decisions related to individual countries to unveil interesting connections and correlations. Indeed one of Frances Haugen's important revelations about Facebook concerned how providers under-invest in some countries, even though they have a substantial presence in them.[47] These revelations informed policymaking.

Furthermore, automated tools must be specifically tested for their accuracy in each language. This obligation is particularly important as the training of technologies often

[47] Rebecca Staudenmaier, 'Whistleblower: Facebook stirs global conflicts' *DW News* (Berlin, 6 November 2021) <https://www.dw.com/en/facebook-whistleblower-warns-company-is-neglecting-languages-other-than-english/a-59739260> accessed 6 September 2023.

takes places in the cultural and linguistic context of some larger countries, and thus potentially neglects the effects of the tools on smaller countries when scaled up to serve the same needs.

Finally, VLOPs and VLOSEs must publish their audit documentation (Article 42(4)). This includes the following documents:

- self-assessment regarding the results of the risk assessment;
- report outlining the specific risk mitigation measures;
- audit reports by auditors;
- follow-up audit implementation report;
- any consultations held to support risk assessment or risk mitigation.

The providers are allowed to redact confidential information. The latter can only be redacted if it is likely to 'cause significant vulnerabilities for the security of its service, undermine public security or harm recipients' (Article 42(4)). Any such redaction must be accompanied by a public statement explaining why such information is omitted. Unredacted reports must be sent to the European Commission and the DSC of the establishment.

16.3.4 Transparency of other actors

Content moderation and advertising concerns are not limited to behaviour of providers of digital services. Content moderation activities are shaped by content creators, notifiers of content, such as victims, NGOs or enforcement agencies, and authorities which issue relevant orders. Ills of the advertising ecosystem are also shaped by advertisers and their agencies. The DSA places some transparency obligations on these other actors.

Trusted flaggers (Article 22) are given a privileged position among notifiers. Their notices carry more weight because they ought to represent notifications that are well-considered. Although trusted flaggers are subject to certification by regulators, in order to shine some light on their performance, the DSA obliges them to publish annual transparency reports. The reports cover only illegal content.

The mandated reports must be 'comprehensive and detailed' and include information about the absolute number of notices with a breakdown according to providers, type of illegal content, and the follow-up action taken. They must also detail how the trusted flagger retains its independence from any provider. While the text does not mandate reporting for each service, to facilitate general comprehension of the reports, they should not make only general statements about the number of notices sent to large companies that operate hundreds of different services but distinguish between individual services. In addition to Article 22(3), whenever a trusted flagger is operated by a public entity, such as Europol, the applicable freedom of access legislation can offer additional data points to scrutinise its practices.

Publication of reports about notices concerning content that 'only' violates terms and conditions is not mandated. They can be, however, introduced by the relevant Codes of Conduct.

Despite the importance of the *ODS bodies* (Article 21) in the DSA, their transparency is arguably underdeveloped. The ODS bodies do not have to publish public annual reports or their individual decisions. They only have to annually report to the DSC who certified them, specifying an absolute number of disputes received, broken down by the outcomes, and the average time for a dispute handling. They should also report any shortcomings that they have encountered. The relevant DSCs then must issue biannual reports summarising the data for all certified ODS bodies, including best practices and recommendations for how to improve the system.

Given the limited reporting obligations, one of the possible improvements surely can be the transparency of the ODS bodies, specifically the process of reporting decisions and publication of reports concerning their own activities. The DSCs can arguably require increased transparency as part of fair procedural rules. As explained in Chapter 11.5, the DSCs are tasked with certifying and supervising such bodies across the EU.

Finally, the DSA does not explicitly deal with the question of transparency by *public authorities*.[48] However, if such authorities operate as trusted flaggers, they have obligations under Article 22(3). If they issue orders, they might have obligations based on national or European law, such as Article 8 TCR. If authorities send simple notifications and thus are not acting as trusted flaggers, the only possible sources of transparency are national freedom of information laws or annual activity reports.

16.4 Conclusion

The DSA creates a set of default safeguards applicable to most of the situations of delegated enforcement. By putting gatekeepers of digital services under scrutiny, it will help to reveal the extent of content moderation undertaken and some aspects of its quality. However, as much as the conventional method of annual transparency reporting serves its purpose, it ultimately only acts as a rear mirror to what has already happened. Therefore, the DSA has supplemented it with an exciting new measure: the involvement of researchers in transparency endeavours. Indeed, if we wish to better examine what is happening now in the digital ecosystem, and potentially what will happen in the future, data access by researchers presents a novel way forward. Here, too, the DSA institutionalises data access as a general safeguard that can step in to provide a look at the actions of regulators and companies alike. Because the researchers' mandate is

[48] For instance, Europol's pre-DSA reports: European Union Agency for Law Enforcement Cooperation., *EU Internet Referral Unit Transparency Report 2021: Terrorist Propaganda Monitoring and Analysis, Referrals, and Public Private Partnerships*. (EU Publications Office 2022) <https://data.europa.eu/doi/10.2813/341047> accessed 6 September 2023.

worded very broadly, the sky is the limit for its potential application. This access can put even some of the most controversial policies of the EU, such as Article 17 CDSM Directive, on a corrective course even after they were already adopted. However, for this to happen, there will be numerous practical obstacles to access that first have to be cleared.