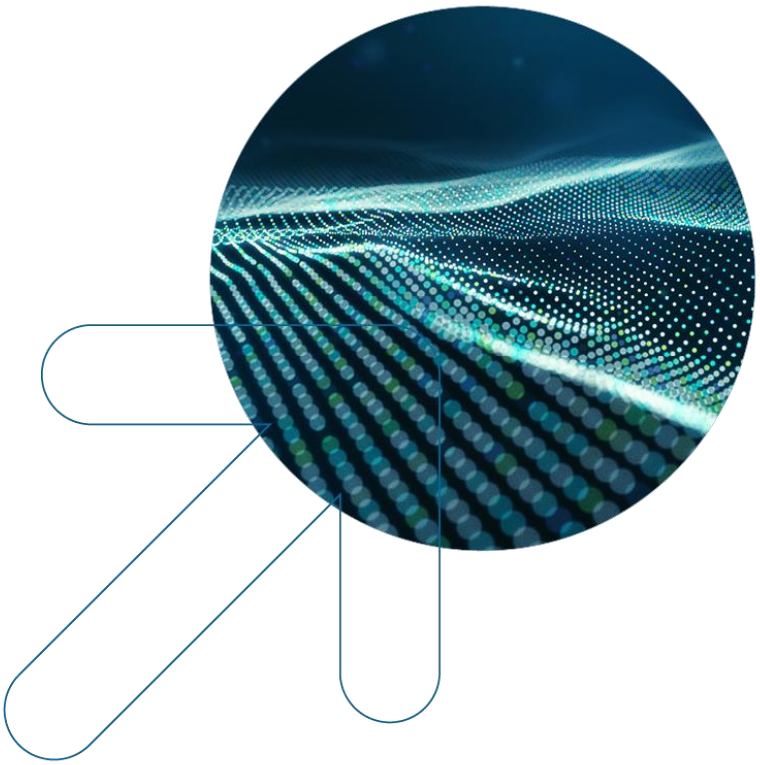
WIK • PLASMA INSIGHT



Data availability and data volume under the Digital Services Act

Authors:

Andrea Liebe
Lukas Wiewiorra

Bad Honnef, November 2023

Gefördert durch:

Bundesministerium für Digitales und Verkehr

aufgrund eines Beschlusses des Deutschen Bundestages

PLASMA

WIK
Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste

Imprint

WIK Wissenschaftliches Institut für
Infrastruktur und Kommunikationsdienste GmbH
Rhöndorfer Str. 68
53604 Bad Honnef
Germany
Phone: +49 2224 9225-0
Fax: +49 2224 9225-63
eMail: info@wik.org
www.wik.org

Person authorised to sign on behalf of the organisation

General Manager and Director	Dr Cara Schwarz-Schilling
Director Head of Department Smart Cities/Smart Regions	Alex Kalevi Dieke
Director Head of Department Networks and Costs	Dr Thomas Plückebaum
Director Head of Department Regulation and Competition	Dr Bernd Sörries
Head of Administration	Karl-Hubert Strüver
Chairperson of the Supervisory Board	Dr Thomas Solbach
Registered at	Amtsgericht Siegburg, HRB 7225
Tax No.	222/5751/0722
VAT-ID	DE 123 383 795

Date: July 2023

Table of contents

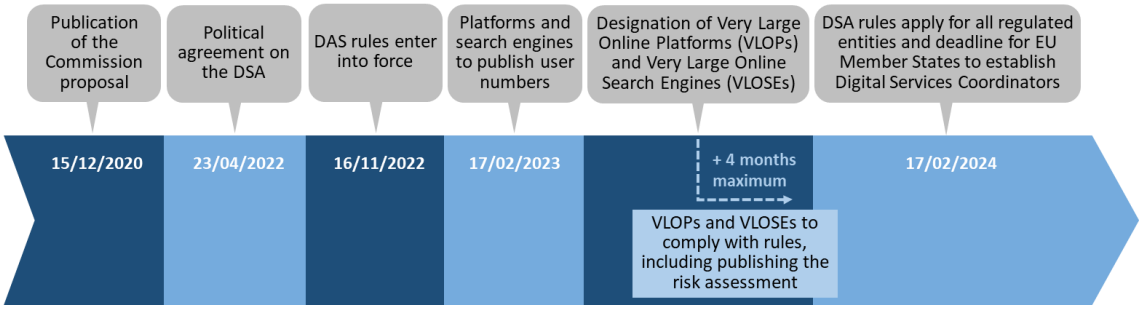
Figures

Tables

1 Introduction

The Digital Services Act (DSA) came into force in November 2022. The provisions of the DSA are already linked to specific obligations and actions for individual companies and the EU Commission. Starting in February 2024, these provisions will apply to all platforms. The DSA aims to request data from companies and services to better understand developments in the platform economy in Europe and assess systemic risks of very large online platforms (VLOP). At the same time, data-based analysis is also required as part of the duty to supervise. The chart below shows the timeline of the DSA until February 2024.

Figure 1-1: Timeline of the DSA



Source: European Commission (2023a).

The objective of this PLASMA Insight is to show what data will be available on a regular basis in the future and what the challenges, opportunities and analytical possibilities are associated with this. The analysis focuses on VLOPs and very large online search engines (VLOSEs), which are mandated by the DSA to provide data on a regular basis.

Chapter 2 provides an overview of the DSA and outlines the addressees and reporting obligations. The subsequent three chapters cover different types of reporting obligations.

Chapter 3 specifically focuses on data pools and reporting obligations, including the publication of user numbers and transparency reports. It also shows how VLOPs and VLOSEs have already reported. The chapter identifies challenges and suggests possible improvements for future reporting. It also provides an indication of the types of statements that can be expected when using the data.

Chapter 4 focuses on databases. It explains the contents of the Transparency Database, which is aimed at content moderation, and the Advertising Repositories. These repositories retrospectively document the online advertising of a VLOP or VLOSE in a given year. The chapter also highlights the challenges involved in setting up such a database, considers data quality requirements, and explores possible analytical applications.

Chapter 5 discusses data access for researchers, including the vetting process and criteria. It also identifies challenges for companies and researchers and suggests possible solutions.

Chapter 6 summarizes our findings and provides an outlook for the PLASMA project.

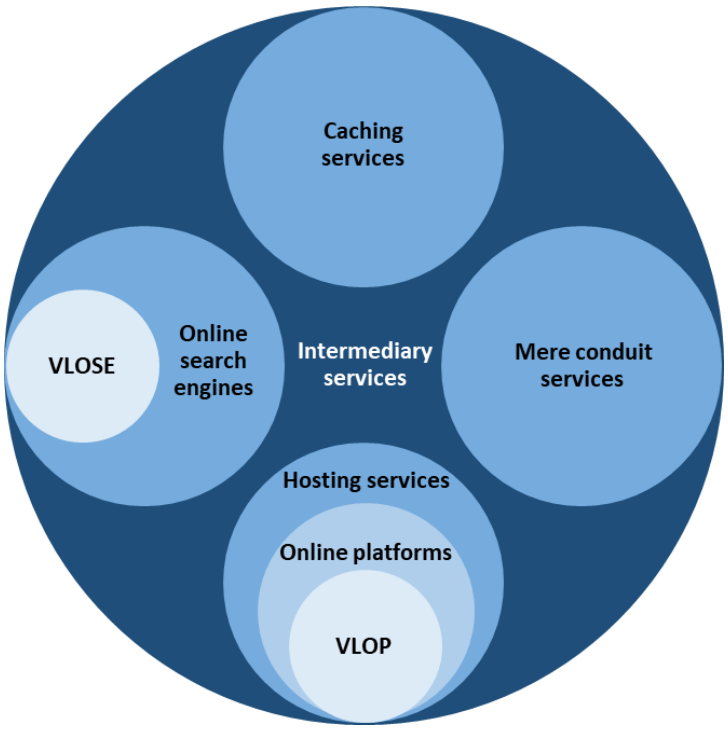
2 The DSA – Fact and data centric regulation for online platforms and search engines

2.1 Overview

In July 2022, the Council of the European Union adopted the Digital Services Act (DSA) to set accountability standards for online platforms and broader obligations for very large platform operators. The DSA aims to promote innovation, growth and competitiveness, also of smaller market participants, and is to be regarded as a kind of basic law for the internet. The law will be implemented gradually as shown before The DSA deals with harmful and illegal goods, services and online content, such as counterfeit products and hate speech. It is an update of the provisions of the E-Commerce Directive (2000) and will gradually replace it.

The addressees of the DSA are all intermediary services which are shown Figure 2-1. While services with an incentive to distribute content like online platforms (e.g. social media, online marketplaces etc.) and search engines are in focus of this paper, also intermediary services that are mere conduit services like Internet service providers, services that only temporarily store information for caching like content distribution networks and hosting services that permanently store information, but without the goal to distribute them themselves are in scope of the DSA . Moreover, the DSA applies to all providers falling into these categories operating in the EU, regardless of where their headquarters are located, meaning all services directed at EU customers are fully within the scope of the DSA.

Figure 2-1: Providers addressed by the DSA



- Intermediary services with an infrastructure network: Internet service providers, domain name registrars, including:
- Hosting services such as cloud and web hosting services, including:
- Online platforms that bring sellers and consumers together, such as online market places, app stores, collaborative economy platforms, and social media platforms.
- Very large online platforms pose risks for the spread of illegal content and harm to society. Special rules for platforms that reach more than 10% of the 450 million consumers in Europe.

Source: WIK, based on Quintais, J. P., Schwemer, S. F. (2022).

In the following, the focus is on services with an incentive to distribute content and information and in particular on very large online platforms and very large online search engines (VLOP/VLOSE). The DSA classifies platforms or search engines that have more than 45 million users per month in the EU as VLOP or VLOSE.

On 25 April 2023 the European Commission designated the following 19 platforms including two app stores, five marketplaces, eight social media services, one encyclopedia, one mapping service and two search engines, which are shown in Table 2-1. In total 13 companies are affected.

Table 2-1: Designated VLOPs and VLOSEs

		Company	Service
Social Media	**Very Large Online Platform (VLOP)**	Alphabet	YouTube
		Meta	Facebook
		Meta	Instagram
		Bytedance	TikTok
		Microsoft	LinkedIn
		Snap	Snapchat
		Pinterest	Pinterest
		X / Twitter	X / Twitter
App Stores		Alphabet	Google App Store/ Google Play
		Apple	Apple App Store
Wiki		Wikimedia	Wikipedia
Online Markets		Amazon	Amazon Marketplace
		Alphabet	Google Shopping
		Alibaba	AliExpress
		Booking.com	Booking.com
		Zalando	Zalando
Maps		Alphabet	Google Maps
Search	**Very Large Online Search Engine (VLOSE)**	Alphabet	Google Search
		Microsoft	Bing

Source: WIK, based on European Commission (2023b).

Some of the DSA's many obligations are limited to VLOPs with more than 450 million users that are central to economic transactions and have a high profile in the public debate. Very small platforms are exempt from many of the obligations.

A VLOP or VLOSE designated by the Commission has four months to comply with the DSA obligations. It must provide a contact point, report crimes, publish user-friendly terms and conditions, and be transparent about advertising, referral systems or content moderation decisions. In addition, systemic risks associated with the designated services must be identified, analyzed and assessed. Once risks have been identified and reported to the Commission for monitoring, VLOPs and VLOSEs are required to take action to

mitigate those risks. They must also establish an internal compliance function, be audited by an independent auditor, and share data with the Commission and national authorities to monitor and assess compliance with the DSA. In addition, they must allow vetted researchers access to platform data if the research contributes to the detection, identification and understanding of systemic risks in the EU. These reporting obligations are the source of the new data availability. They are discussed in detail in the following sections.

The independent national Digital Service Coordinators (DSC) will play a key role in the enforcement of the DSA and will be instrumental in the implementation of the new rules. The European legislator's aim is to ensure a single point of contact, a single complaints body and a responsible supervisory authority with an overall view. In this context, it has to be taken into account that the enforcement of the DSA will require the continuous monitoring of a large number of relevant addressees as well as the handling of a large number of individual complaints. Therefore, it is crucial for the work of the DSC to have a comprehensive understanding of the volume of data that will be generated prior to the start of DSA-based regulation.

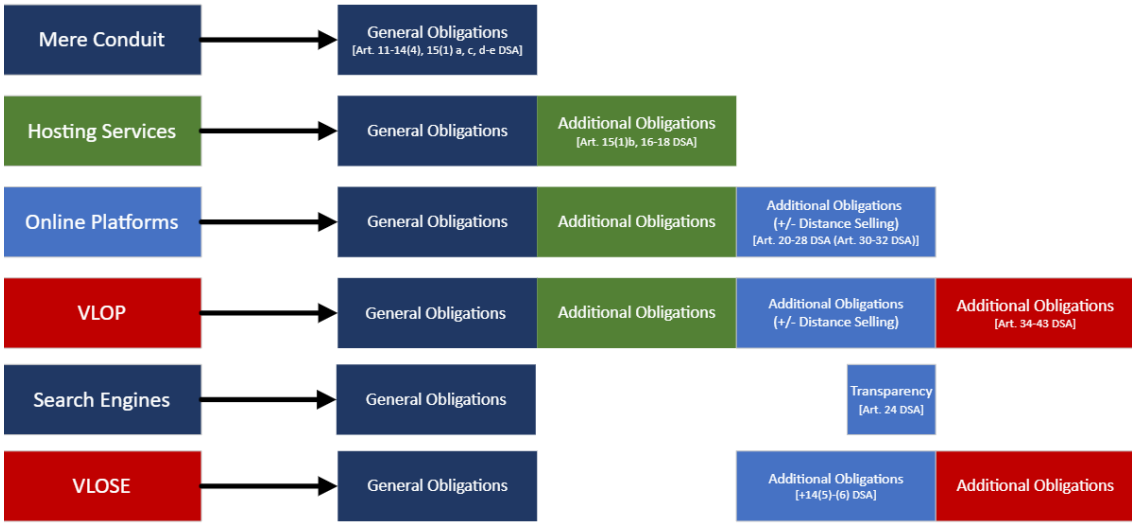
The DSA envisages that VLOPs and VLOSEs in particular must regularly provide data on various aspects of their business activities „(…) to facilitate supervision and research into emerging risks“ (DSA Art. 63). This will allow authorities and regulators to analyze and evaluate the impact and effectiveness of this legislative initiative based on this information and other available data.

Before discussing in more detail the methodological challenges related to the availability of data for the DSA, we provide a brief overview of the reporting obligations.

2.2 DSA reporting obligations

The DSA contains numerous obligations and requirements addressed to services, the Digital Service Coordinator and the EU Commission. For the services and the companies behind them, a distinction is made according to type (mere conduit, caching service, hosting service, online platform, VLOP/VLOSE). In general, two types of provisions and obligations can be distinguished. Regulations on the use of digital services like the publication and explanation of terms of use on the one hand and obligations to report in a standardized form, usually in a machine-readable format on the other hand. Examples are content moderation activities of the providers as well as reporting of user numbers. In addition, the platforms are required to comply with risk assessment and auditing regulations and to submit corresponding reports. Figure 2-1 gives an overview.

Figure 2-2: DSA rules by type and scope



Source: WIK, based on Schmid, G., Koehler, P. (2022).

The following section focuses on the reporting obligations, which, as already mentioned, are associated with a previously unknown volume of data. Table 2-2 below lists the obligations of the DSA by category. These relate to content moderation, user number reporting, VLOP/VLOSE activities and auditing, online advertising, data access as well as DSC and Commission activities. The reporting obligations vary in nature, ranging from the reporting of individual figures to transparency reports with many qualitative explanations and extensive databases that are updated in real time.

This means that a large number of reports and databases need to be created and monitored. This presents a number of challenges. Firstly, it is a matter of implementing rules for all instances that have never existed before and most of which will have to be reintroduced into their structures. Secondly, the services and platforms will have to operationalize the indicators set out in the DSA and determine them in a compliant manner. To this end, the DSA provides at various points for further legal acts and guidelines, which the EU Commission may specify in more detail if necessary. However, it is also a question of mapping the relevant information provision in internal reporting and monitoring. It will also be the responsibility of the competent authorities to monitor and analyze the various reports and draw the appropriate conclusions.

There are also challenges, particularly of a technical nature, which in turn are associated with the creation of databases and the retrieval and storage of the data provided. All of these requirements call for a wide range of capacities that need to be built up in a timely manner.

The obligations to provide data in the form of transparency reports and reports on user numbers and the data collections to be set up in the form of databases are discussed below. These areas are currently expected to generate the largest volume of primary data to be managed on a regular basis. Another area, which will also be discussed in more detail below, is data access for researchers, as this is a new instrument in the form envisaged by the DSA, which on the one hand offers many opportunities for in-depth analysis, but on the other hand whose specific design and handling is critical for success.

Table 2-2: Reporting obligations and data access under the DSA

	Type	MR	Actor	Shall/may	Target	Frequency	Article	Topic
Content Moderation	Transparency report	MR	Intermediary service	Shall	Public	Every year	15(1)	Content moderation practices
	Transparency report	MR	Online platform	Shall	Public	Every year	24(1)	Expanded content moderation practices (in addition to Art. 15 obligations)
	Report	n/a	Trusted flagger	Shall	DSC that certified them; public	At least every year	22(3)	Notices trusted flaggers submitted
	Database	MR	Commission	Shall	Public	Ongoing	22(5)	List of trusted flaggers from DSCs
	Database	MR	Commission, Online platforms	Shall	Public	Immediately	24(5)	Decisions to remove content (= Art. 17(1))
AMAR	Transparency report	MR	Online platform	Shall	DSC, Commission	Every 6 months	24(2)	AMAR reporting
	Transparency report	MR	Online platform	Shall	DSC, Commission	Upon request	24(3)	Update AMAR reporting
VLOP/VLOSE Audit	Audit report	n/a	Auditor	Shall	VLOP; later public (42(4)(c)); later Commission (75(3)), DSC of establishment (recital 93), Board (recital 93)	Every year	37(4)	Including audit opinion
	Audit implementation report	n/a	VLOP	Shall	DSC, Commission, public (42(4)(d))	if no positive audit report	37(6)	VLOPs to produce audit implementation report
	Report	n/a	VLOP	Shall	Public, DSC, Commission	Three months after each audit report	42(4)(a) + 34	Results of risk assessment
	Report	n/a	VLOP	Shall	Public, DSC of establishment, Commission	Three months after each audit report	42(4)(b) + 35(1)	Risk mitigation measures
	Report	n/a	VLOP	Shall	Public, DSC of establishment, Commission	Three months after each audit report	42(4)(c) + 37(4)	Audit report
	Report	n/a	VLOP	Shall	Public, DSC of establishment, Commission	Three months after each audit report	42(4)(e)	If applicable: Consultations by VLOP on risk assessments
Online Advertising	Database	MR	VLOP	Shall	Public	ongoing	39(1)	Online ad repositories (sarchabl, API)
Data access	Data access/analysis	n/a	VLOP	Shall	Auditor	every year	37(2)	Auditors to receive all relevant data, get questions answered
	Data access/analysis	n/a	DSC; Commission	Shall	VLOP	upon request	40(1)	DSC, Commission to receive data necessary to check compliance, API
	Data access/analysis	n/a	Vetted researcher via DSC	Shall	VLOP	n/a	40(4) + 40(8)	Vetted researcher to receive data necessary to check risk assessment and mitigation (= Art. 34, 35), API
Risks	Report	n/a	Board (in cooperation with Commission)	Shall	Public	Every year	35(2)	Assessment of most prominent systemic risks (at EU and member state level), best practices of mitigation; underlaying data to be preserved
DSC/Commission	Report	MR	DSC	Shall	Public, Commission, Board	Every year	55	Report on DSC activity
	Report	n/a	Commission	Shall	EP, Council	Every year	43(7)	Costs for supervision and fees charged

Source: WIK, inspired by Jaursch, J. (2022a).

3 Data pools and reports

The following section will focus specifically on the reporting of user numbers in accordance with Article 24(2) DSA and the transparency reports to be submitted, as well as the challenges involved.

3.1 Reporting of Average Monthly Active Recipients (AMAR)

So far, the regulations on the reporting of the number of users (AMAR Average Monthly Active Recipients) have come into force. They apply to all providers of online platforms and online search engines, except small and micro platforms. On this basis the EU Commission has designated VLOPs and VLOSEs[1] which are already shown in Table 2-1. Accordingly, the reporting of AMAR figures is of key importance both for the VLOPS and VLOSEs and for the public as recipients of the effects generated by these platforms. If the number of users exceeds 45 million, they are classified as "systemically relevant". The table below shows the AMAR figures already reported. VLOPs and VLOSEs must also report the numbers of users separated by country as part of their transparency reports which are subject of Table 3.2.

Table 3-1: Reporting of user numbers by VLOPs and VLOSEs

	Company	Service	AMAR (Million) 08/2022-01/2023	AMAR (Million) 02/2023-07/2023	Type
Social Media	Alphabet	YouTube	401+	416+	VLOP
	Meta	Facebook	255	258	VLOP
	Meta	Instagram	250	257	VLOP
	Bytedance	TikTok	125	134	VLOP
	Microsoft	LinkedIn	163+	177	VLOP
	Snap	Snapchat	96+	102	VLOP
	Pinterest	Pinterest	>45	>45	VLOP
	Twitter	Twitter	100+	112+	VLOP
App Stores	Alphabet	Google App Store /Google Play	274+	284+	VLOP
	Apple	Apple App Store	>134	123	VLOP
Wiki	Wikimedia	Wikipedia	151+	151+	VLOP
Online Markets	Amazon	Amazon Marketplace	>45	181	VLOP
	Alphabet	Google Shopping	74+	70+	VLOP
	Alibaba	AliExpress	>45	135+	VLOP
	Booking.com	Booking.com	>45	>45	VLOP
	Zalando	Zalando	83+	76+	VLOP
Maps	Alphabet	Google Maps	278+	275+	VLOP
Search	Alphabet	Google Search	332	364	VLOSE
	Microsoft	Bing	107	119	VLOSE

Source: WIK, Own compilation, based on primary data provided on the websites of the companies.[2]

1 Under the DMA, the more or less, the same providers were designated as Gatekeepers.
2 Detailed source references can be found in the table in the appendix.

It is noteworthy that in the first round, four companies did not report specific figures, but instead provided general information that they were above the 45 million threshold. In the second round, only two companies did so. Another point worth noting in relation to the publication of user numbers is that companies are still very inconsistent in what they publish on their websites, and figures can sometimes only be found with a lot of digging. Information ranges from plain raw numbers to detailed reports.

It can be assumed that there were uncertainties among the companies when the reporting obligation came into force. On the one hand, the indicator is not one of the known standard indicators for the number of users, and its determination in a legally compliant manner is therefore associated with uncertainties. Secondly, strategic considerations will also have been taken into account. It has become known that the companies would like to see the publication of a methodology for calculating the average monthly active recipients. In particular, guidance on how to distinguish between different platforms and services from the same provider, whether and how to count recipients of embedded content and search engines with auto-completed search terms, whether scrapers, bots and visitors with multiple devices should be excluded from the average monthly active recipients, etc. would be considered desirable in this context.[3] For example, Google stated that they calculate differently for different services. They are currently working on refining their process for future reports and look forward to further guidance from regulators on methodology.[4]

Article 33(3) of the DSA provides that the EU Commission may adopt delegated acts in which it specifies the method. The Commission has not yet made use of this right, but has nevertheless published a document in which it answers practical questions on the implementation of the obligation raised by companies.[5] Although the Commission clarifies a number of points in this guidance document, it also stresses that it is without prejudice to any legislation yet to be adopted.

However, the question remains whether these explanations are sufficiently concrete and whether they lead to the publication of consistent and horizontally comparable figures. In order to be able to assess this, it is necessary to compare the available figures originally calculated by the online platforms with metrics of Monthly Active Users tracked by third-party market research companies. This requires aggregated figures of Monthly Active Users across all EU countries.

As mentioned above, VLOPs and VLOSEs are required to report the number of users (AMAR) by country as part of their transparency reports. The following table contains the data for Germany in addition to the EU data reported above. In addition, the data for the

3 Ten Thije, P., van Hoboken, J. (2022).
4 https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-24_2023-1-1_2023-6-30_en_v1.pdf
5 European Commission (2023c).

common indicator of monthly active users (MAU) is also listed for Germany.[6] MAU refers to the unique customers who have actively used the service within a month, without duplication across multiple devices. The MAU figures are derived from both app and web usage. The indicators of monthly active users and monthly unique visitors (MUV) are usually considered equivalent. The data is provided and modelled by our project partner, Statista. In general these figures can be useful for individual plausibility checks and cross-checks.

Table 3-2: User numbers - Germany

Service	Time frame	GER AMAR Million (if provided: of which logged in)	GER MAU/MUV (Million) 01/2023-06/2023	Relation
YouTube	01/2023-06/2023	289.5 (82.1)	105.2	AMAR>MAU
Facebook	04/2023-09/2023	33.2	69,2	MAU>AMAR
Instagram	04/2023-09/2023	42.5	53.7	MAU<AMAR
TikTok	04/2023-09/2023	20.9	37,7	MAU>AMAR
LinkedIn	01/2023-06/2023	32 (5.7)	11.2	AMAR>MAU
Snapchat[7]	02/2023-07/2023	21.1	9.5	AMAR>MAU
Pinterest		NA	12.5	NA
Twitter / X	20/04/2023-20/10/2023	16.3 (8.9)	27.3	MAU>AMAR
Google App Store/ Google Play	01/2023-06/2023	63.8 (54.5)	33,9[8]	AMAR>MAU
Apple App Store	not specified[9]	28	19.3	AMAR>MAU
Wikipedia	not specified[10]	34.5	19.8	AMAR>MAU
Amazon Marketplace	01/2023-06/2023	60.4	85	MAU>AMAR
Google Shopping	01/2023-06/2023	27.1 (15.4)	NA	NA
AliExpress	07/2023-09/2023	11	6.8	AMAR>MAU
Booking.com	02/2023-07/2023	>8.25[11]	21.4	MAU>AMAR
Zalando		NA	14.3	NA
Google Maps	01/2023-06/2023	124.4 (54.7)	48.4	AMAR>MAU
Google Search	01/2023-06/2023	171.2 (62.8)	164.9	AMAR>MAU
Bing	01/2023-06/2023	27.3	5.2	AMAR>MAU

Source: WIK, Own compilation based on the transparency reports and PLASMA (2023).

6 Nevertheless, it is striking that the terms AMAR and MAU are mixed up and used misleadingly. For example, Amazon and Microsoft use the term MAU when breaking down their AMAR in the transparency reports.

7 Numbers are not reported in the transparency report but separately. https://values.snap.com/de-DE/privacy/transparency/european-union [Last accessed 09.11.2023].

8 App data only.

9 Report applies to 27/08/2023-27/09/2023.

10 Report applies to 25/08/2023-30/09/2023.

11 On https://www.booking.com/content/dsa.en-gb.html, booking.com provides the statement that the number of AMAR is “well above 45 million” in the reporting period in the EU. In the transparency report, only a pie chart is shown. Germany’s “slice” has an angle of roughly 66°, corresponding to 18.33% of the EU total.

A first cautious estimate, using Germany as an example, shows that AMAR and MAU figures diverge significantly for different services. This impression is mainly due to the different types of reporting.

Firstly, the figures are in some cases based on different time periods and lengths of time for which an average is calculated. A horizontal comparison of different services on this basis does not appear to be meaningful. Seasonal effects or other particularities may be over- or underestimated. It is necessary to specify the period for which corresponding figures are to be provided.

It is also notable that, for example, in the case of Google services, the figures are broken down into 'average monthly counts based on distinguishable sessions of logged-out recipients' and 'average monthly counts based on distinguishable logged-in accounts of recipients'. It seems unclear whether all other services have made this distinction. Some services reported very late, and others reported as a graph without giving specific numbers for countries.

The monthly active user figures shown here for Germany largely include app and web usage data. They therefore show the number of unique customers (not duplicated across multiple devices) who actively used the service during the month. A comparison of MAU and AMAR figures paints a mixed picture. Even the relationships between the figures are rather unclear. Of the 19 services, the data situation is currently insufficient for four services. In ten cases the AMAR figures are higher than the MAU figures and in six cases the MAU figures are higher than the AMAR figures. It is therefore not possible to identify a clear pattern. However, it is noticeable that for those alphabet services that report registered and unregistered users separately, the sum of these is higher than the MAU figures.

However, it cannot be argued that registration necessarily plays a role here, as the figures are similar for Wikipedia and exactly the opposite for TikTok, for example, where the MAU figures exceed the AMAR figures.

More guidance is needed on reporting to enable meaningful analysis. However, this also shows that monitoring from the regulatory side is also difficult and requires a certain amount of effort.

3.2 Transparency reports

The first transparency reports have already been submitted.[12] The focus is on the moderation of online content. Under Article 42, VLOPs and VLOSEs are required to submit a

12 Steiner, F. (2023).

report in machine-readable format at least every six months and for the first time by the end of October 2023, containing the following information:[13]

- Number of requests received from Member State authorities, broken down by the type of illegal content concerned, the Member State issuing the request and the time taken to respond;
- Number of complaints submitted, categorized by the type of allegedly illegal content involved, the number of complaints submitted by trusted flaggers, any action taken on the basis of the complaints, distinguishing whether the action was taken under the law or the provider's terms and conditions, the number of complaints processed by automated means, and the time taken to respond;
- Information about the content moderation engaged in at the providers' own initiative, including the use of automated tools, the measures taken to provide training and assistance to persons in charge of content moderation, the number and type of measures taken;
- Number of complaints received, the basis for those complaints, decisions taken in respect of those complaints, the median time needed for taking those decisions and the number of instances where those decisions were reversed;
- Use made of automated means for the purpose of content moderation,
- Number of disputes submitted to the out-of-court dispute settlement bodies, the outcomes of the dispute settlement, and the median time needed for completing the dispute settlement procedures, as well as the share of disputes where the provider of the online platform implemented the decisions of the body;
- Number of suspensions imposed pursuant to Article 23, distinguishing between suspensions enacted for the provision of manifestly illegal content, the submission of manifestly unfounded notices and the submission of manifestly unfounded complaints.
- Specification of the human resources devoted to content moderation by the provider of VLOPs, broken down by each of the official languages of the Member States,
- Specification of the qualifications and linguistic expertise of the persons carrying out the activities as well as the training and support given to such staff;
- Specification of the indicators of accuracy and related information referred to in Article 15(1), point (e), broken down by each official language of the Member States.

13 Intermediary platforms must report at least once a year in machine-readable form in accordance with Article 15 DSA and online platforms also once a year in machine-readable form in accordance with Article 24 DSA. The cumulative structure described above in Chapter 2.2 applies.

Transparency reports in this form are not unknown to companies, at least in Germany. The Act on the Improvement of Law Enforcement in Social Networks (Network Enforcement Act - NetzDG), which has been in force since 2017, has so far taken on the role of the DSA and also provides for transparency reporting obligations. These are not identical, but similar.

Companies are free to choose the format of their reports. The reports available so far have different layouts and vary in both length and visual presentation. Amazon, for example, produces a 25-page report, while LinkedIn produces a 17-page report. Microsoft fulfils its reporting obligation with a six-page report. It is worth noting that some VLOPs/VLOSEs publish as pdf files for download and others publish directly in the web interface.

Table 3-3: Transparency reports by October 2023

	Company	Service	Link to the report	Link to the transparency center
Social Media	Alphabet	YouTube (joint report with the other alphabet services)	https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-27_2023-8-28_2023-9-10_en_v1.pdf	https://transparencyreport.google.com/report-downloads
	Meta	Facebook	https://transparency.fb.com/sr/dsa-transparency-report-oct2023-facebook/	https://transparency.fb.com/reports/regulatory-transparency-reports/
	Meta	Instagram	https://transparency.fb.com/sr/dsa-transparency-report-oct2023-instagram/	https://transparency.fb.com/reports/regulatory-transparency-reports/
	Bytedance	TikTok	https://sf16-va.tiktokcdn.com/obj/eden-va2/fsslreh7uulsn/DSA%20Report%20October%202023/DSA%20draft%20Transparency%20report%20-%2025%20October%202023.pdf	https://www.tiktok.com/transparency/en/dsa-transparency/
	Microsoft	LinkedIn	https://content.linkedin.com/content/dam/help/linkedin/en-us/October-2023-LinkedIn-DSA-Transparency-Report10.pdf	https://about.linkedin.com/transparency
	Snap	Snapchat	https://values.snap.com/privacy/transparency?lang=en-US	https://values.snap.com/privacy/transparency?lang=en-US
	Pinterest	Pinterest	https://policy.pinterest.com/en/digital-services-act-transparency-report	https://help.pinterest.com/en/article/digital-services-act
	Twitter/X	Twitter/X	https://transparency.twitter.com/dsa-transparency-report.html	https://transparency.twitter.com/en/reports.html
App Stores	Alphabet	Google App Store /Google Play	joint report with the other alphabet services	joint report with the other alphabet services
	Apple	Apple App Store	https://www.apple.com/legal/dsa/transparency/eu/app-store/2310/	https://www.apple.com/legal/dsa/de/
Wiki	Wikimedia	Wikipedia	https://foundation.wikimedia.org/wiki/Legal:Supplemental_Transparency_Report_for_August-September_2023	https://foundation.wikimedia.org/wiki/Legal:Supplemental_Transparency_Report_for_August-September_2023#
Online Markets	Amazon	Amazon Marketplace	https://assets.aboutamazon.com/cd/28/4d02dd2e41ec8c6d1bc341e9d919/amazon-eu-store-transparency-report-jan-june-2023.pdf	https://www.aboutamazon.eu/news/policy
	Alphabet	Google Shopping (joint report with the other alphabet services)	joint report with the other alphabet services)	joint report with the other alphabet services
	Alibaba	AliExpress	https://files.alicdn.com/tpsservice/0475f95eec8798f4d6e9937a08e77c38.pdf?spm=a2g0o.tm1000005123.1782285040.1.51e96f3d1CSSYv&aecmd=true	https://campaign.aliexpress.com/wow/gcp-plus/300000844/XsB8xbyAcS?_immersiveMode=true&pha_manifest=upr&disableNav=YES&spm=undefined.undefined.undefined.0&aecmd=true
	Booking.com	Booking.com	https://r-xx.bstatic.com/data/mobile/dsa_transparency_report_bf3fdc24.pdf	https://www.booking.com/content/dsa.de.html?aid=357028&label=bin859jc-1DCBQoggJCA2RzYUgzWANoO4gBAZgBB7gBF8gBDNgBA-gBAYgCAa-gCA7gCn5aiqwbAAgHSAi-QyMzkwYjkxNC03YzJkLTQyYjgtOTJmYi0zNmQ4OWZhNTYzMjTYAg-TgAgE&sid=2582aedcc51dacd71bae383f22ba06c5&keep_landing=1&
	Zalando	Zalando	https://mosaic02.ztat.net/cnt/contentful-apps/uploads/a74cdebf-cfc7-46dd-8853-13afed1e41aa.pdf	https://mosaic02.ztat.net/cnt/contentful-apps/uploads/a74cdebf-cfc7-46dd-8853-13afed1e41aa.pdf
Maps	Alphabet	Google Maps (joint report with the other alphabet services)	joint report with the other alphabet services	joint report with the other alphabet services
Search	Alphabet	Google Search (joint report with the other alphabet services)	joint report with the other alphabet services	joint report with the other alphabet services
	Microsoft	Bing	https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1dO0h	https://www.microsoft.com/en-us/corporate-responsibility/reports-hub

Source: WIK. Own compilation..

Differences may be due to the fact that individual companies aggregate several services in one report, but there are also uncertainties regarding the implementation of reporting obligations. A dilemma can be identified. There is a dilemma here in that the more is reported in principle, the fewer (follow-up) questions the service provokes. On the other hand, the more and detailed the reporting, the more information becomes public knowledge.

The reports consist of continuous text presenting the information qualitatively and tables with concrete figures. In terms of volume, the monitoring effort appears to be manageable, but the actual plausibility and accuracy checks are another matter. It can be assumed that the first step will be to check whether the reports are available and whether they are as complete as possible. Here too, the Regulation will initially allow certain freedoms and, if necessary, gradually specify the requirements once several rounds of reports have been submitted.

So far, the focus here has been on the transparency reports of the VLOPs and VLO-SEs. It should not be forgotten that the intermediary platforms and online platforms also have to submit reports covering a large part of the indicators. At this stage, it is not really possible to estimate the extent to which monitoring will be required.

3.3 Outlook and analysis options

The challenges associated with the increasing availability of data for companies, regulation and research are numerous. To ensure comparability between services and companies and to conduct robust analyses, it is advisable to provide sufficient explanation and standardization of the principles and definitions (calculation) of the indicators. This includes avoiding subjective assessments unless clearly marked as such, using clear and concise language with a logical flow of information, adhering to conventional academic structure and formatting, employing formal register and precise word choice, avoiding bias, and ensuring grammatical correctness. Furthermore it must be ensured, that the national DSCs use the Board’s exchange to work towards harmonization and standardization of the national data collections.

Having focused on the challenges of data provision, it is natural to ask what opportunities and possibilities arise from the availability of data.

On the one hand, it is possible to monitor VLOPs and VLOSEs over time, differentiated by EU country, and to analyze possible changes. On the other hand, observations can also be made at a horizontal level for comparable platforms and competition analyzes can be carried out. This makes it possible to identify developments that require regulatory intervention and to make comparisons that can serve as a basis for exploring new competitive business models.

Particularly in the area of content moderation, it is possible to see how much harmful content is published on a platform, how it is responded to, and how this changes over time. It is possible to assess whether the methods used are fit for purpose and effective, and whether any changes to the approach are sustainable. In this way, recommendations can be made to platforms in similar segments to improve content moderation.

4 Databases

The DSA basically contains two statements on the need to set up and operate databases.[14] One is on the transparency database and the other is on advertising repositories. The corresponding regulations can be found, as already explained, in Articles 24(5) and 39(1). The fundamental challenges in setting up the two types of databases are discussed below. In both cases, the DSA provides for databases that are associated with an immense volume and whose development requires a great deal of know-how on the one hand and a great deal of information technology capacity on the other.

4.1 Transparency databases

The transparency reporting obligations in Article 24, which address the providers of online platforms, contain the provision in Section 5 that the various decisions and justifications pursuant to Article 17(1) on content moderation must be reported to the Commission. The Commission in turn feeds the information into a publicly accessible database, which is managed by the Commission and is machine-readable. Companies must ensure that no personal data is included. The database is publicly accessible via the website https://transparency.dsa.ec.europa.eu/.

Article 17 of the DSA contains the obligation to provide clear and specific reasons to all recipients concerned when removing or otherwise restricting the availability of and access to information provided by recipients. Hosting service providers must therefore inform their users of the content moderation decisions they make and explain the reasons for these decisions. The relevant indicators set out in section 3 are information on whether the decision is to remove, disable access to, downgrade or restrict the visibility of the information and

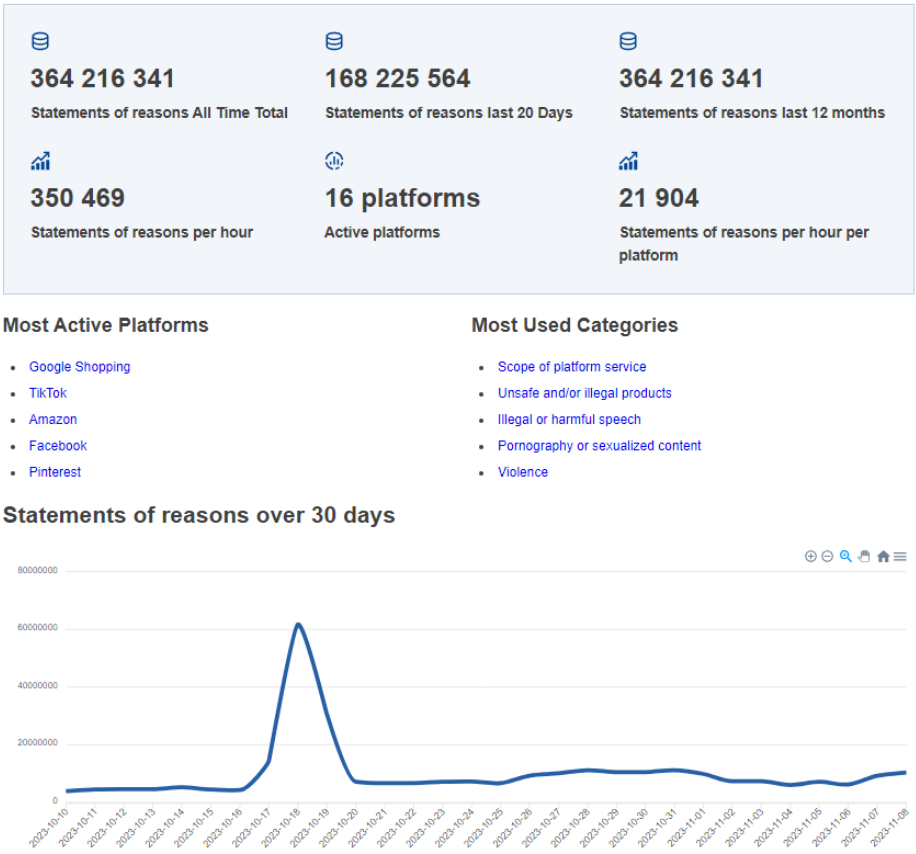
- the facts and circumstances relied on in taking the decision,
- where applicable, information on the use made of automated means in taking the decision ,
- where the decision concerns allegedly illegal content, a reference to the legal ground relied on,
- where the decision is based on the alleged incompatibility of the information with the terms and conditions of the provider of hosting services, a reference to the contractual ground relied on, and
- clear and user-friendly information on the possibilities for redress available to the recipient of the service in respect of the decision.

14 In addition, the DSA also contains the database for trusted flaggers in accordance with Article 22(5). This is to be maintained by the DSC in the form of a list and primarily provides master data on the trusted flaggers and does not contain any information that needs to be updated on an ongoing basis.

The database is updated almost in real time thanks to the fast and automated transmission of the relevant information by the providers. The Commission provides extensive information on the structure and the individual data categories it contains. Data has been included since the end of September 2023.

The source code is also publicly available. Functionality includes search, analysis and the ability to download data. In addition, summary statistics and aggregated visualizations are available, currently in a beta version of an analysis interface. A large number of other functionalities are planned and will be implemented gradually. The EU Commission is interested in interacting with users of the database. The figure below gives an overview.

Figure 4-1: Dashboard of the transparency database



Source: https://transparency.dsa.ec.europa.eu/analytics.

To give an impression of the velocity and volume of data, we highlight a few key-figures on the transparency database. For the period 25-Sep-2023 to 08-Nov-2023, 364.216 million records are available that have been provided by 16 active platforms. The data volume amounts to 365.29GB. Extrapolating from this, it can be assumed that an average storage capacity of approximately 250 GB per month will be required. In the case of the transparency database, the EU Commission has been mandated to set up and maintain it, while the databases for the ad repositories are to be set up by the major online platforms.

4.2 Advertising repositories

The second type of database covered by the DSA is the ad(vertising) repositories. Their purpose is to create transparency for the public, in particular in the area of online advertising.

Providers of VLOPs and VLOSEs that display advertising on their online interfaces shall collect the information referred to in Article 39(2) in a specific area of their online interface and make it publicly available through application programming interfaces (APIs) for the entire period during which they display an advertisement and for one year after the last display. For this purpose, a searchable and reliable tool should be provided, which can be queried according to a number of criteria, and it must be ensured that the archive does not contain any personal data of the users to whom the advertisement was or could have been displayed. It must also ensure that the information is accurate and complete. Implementation must take place by February 2024. The EU Commission may issue guidelines on the structure, organization and functioning of archives.

The online archive must contain at least the following information:

- The content of the advertisement, including the name of the product, service or brand and the subject matter of the advertisement;
- The natural or legal person on whose behalf the advertisement is presented;
- The natural or legal person who paid for the advertisement, if that person is different from the person above;
- The period during which the advertisement was presented;
- Information, whether the advertisement was intended to be presented specifically to one or more particular groups of recipients of the service and if so, the main parameters used;
- the commercial communications published on the VLOPs and identified pursuant to Article 26(2);
- the total number of recipients of the service reached and, where applicable, aggregate numbers broken down by Member State for the group or groups of recipients that the advertisement specifically targeted.

Ad repositories are not an entirely new transparency tool. Most of the VLOPs and VLOSEs targeted by the DSA have already been providing extensive data on their advertising business for several years. Google, for example, has been providing data to its ad repository since March 2018.

However, some providers also have repositories that are being built for the first time; in the case of AliExpress and Amazon, there are even no archiving facilities (yet). At first glance, not all databases meet all the information requirements imposed by the DSA. For example, some providers do not yet have APIs that can be used to retrieve data

automatically. LinkedIn, X (formerly Twitter), Snapchat and TikTok are just a few examples. Other providers do not currently provide any or only very general information about their target audiences.

It is also noticeable that the amount of information provided varies greatly. The providers that have had ad repositories for a long time and whose business model is almost exclusively based on advertising stand out due to their very extensive archives. It remains to be seen how the archives will develop in the coming months. The table below gives an impression of the current status of the archives.

Table 4-1: Overview about the current status of the ad repositories

VLOP/VLOSE	Advertising repository	Comments
AliExpress (Alibaba)	None	None
Amazon Store	None	• Court decision to suspend Amazon temporarily from creating a public ads library.[15]
Apple AppStore	Ad Repository https://adrepository.apple.com/	• Since August 2023. • Search by app, developer, or by country or region to see ads with impressions. • API link available. • Indicators: App, developer, legal name, placement, format, country or region, parameters, first impression, latest impression.
Booking.com	Booking Ads Repository https://www.booking.com/ad-repository.de.html	• Search by date, advertisement, company or person that funded the ad. • API link available. • Indicators: Last shown, first shown, total impressions, impressions per EU Member State. • Ad targets Booking.com users based on their search criteria, including destinations and dates.
Facebook, Instagram (Meta)	Meta Ad Library https://transparency.fb.com/de-de/research-tools/meta-content-library/	*Meta Ad Library (incl. API):* • Search by all advertisements that are currently placed on meta technologies and for election adverts or adverts on politically or socially relevant topics that have been placed in the last 7 years as well as for adverts that have been placed in the EU in the past year. • Indicators of the ad library: Country, language, platform, transparency information EU (targeting group (age, sex, location), coverage, information about the advertiser, information about the beneficiary). *Meta Content Library and API (Beta Version):* • Provision of unlimited access to the entire archive of public content on Facebook and Instagram. Only researchers can request access to these tools from the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan. • Indicators of the Content Library: Language, number of views, media type, content creator, differentiation by countries.
YouTube, Google Play, Google Maps, Google Shopping	Ads Transparency Centre https://adstransparency.google.com/?region=DE	• Since March 2018. • Search by country, format, time period, topic. • API link available. • Indicators: First impression, last impression, topic, format, views, targeting group, legal name.
LinkedIn	Ad Library https://www.linkedin.com/ad-library/home	• Search by company name, keyword, country, date. • Indicators: Format, publisher/advertiser, legal name, last impression, last impression, views by country, targeting criteria.
Pinterest	Ad Repository https://www.linkedin.com/ad-library/home	• Search by date, topic, country sex, age, advertiser. • Indicators: Advertiser, legal name, date, age, region, sex interests, commercial communication, key words, region, views.
Snapchat	Snaps Ads Gallery https://adsgallery.snap.com/	• Search by country, ad status, date. • Indicators: Ad publisher, brand, last impression, first impression, total impressions by country, organization charged.
TikTok	Ad Library https://library.tiktok.com/	• Since October 2022. • Search by country, type of ad, date, keywords, advertiser • Indicators: First impression, last impression, views, targeting criteria (age, sex, .etc.), views by country, Advertiser charged
X (formerly Twitter)	Ads Repository https://ads.twitter.com/ads-repository	• Search by advertiser, view, ad details such as creative, targeting information and reach. Information is compiled in a (large) csv file that can be downloaded. • No indicators listed.
Wikipedia	None	not applicable because of a non-advertising-based business model

15 https://techcrunch.com/2023/09/28/amazon-dsa-interim-measures/?guccounter=1.

VLOP/VLOSE	Advertising repository	Comments
Zalando	Ads Repository https://www.zalando.de/ads-repository/?_rfl=en	• Search by date (last year), country. • API link available. • Indicators: Advertiser, country, views, targeting group information.
Bing (Microsoft)	Ad Library https://adlibrary.ads.microsoft.com/	• Since May 2022. • Search by ad content or advertiser name, country, date, ad content • API-link available. • Indicators: Advertiser, first impression, last impression, ad link, estimated impressions/views, ad targeting criteria.
Google Search	Ads Transparency Center https://adstransparency.google.com/?region=DE	• Since May 2018, • Search by country, format, time period, topic • API-link available. • Indicators: First impression, last impression, topic, format, views, targeting group, legal name.

Source: WIK, Own compilation.

4.3 Data quality and analysis options

Ensuring high data quality is crucial for any analysis based on the data available from the DSA. This section explains the challenges in achieving data quality and presents analysis options that focus on the objectives of the DSA and identify the need for regulatory action.

Databases have inherent limitations in handling data size and complexity. Exceeding these limits can reduce the usefulness of the database. It is therefore important to be aware of these limitations and to take appropriate precautions. Increasing security or scalability often comes at a higher cost.[16]

In the context of research databases, there is often a role for a curator to review the data before it is included in the database. This review can take place in several dimensions. Examples include content and technical quality, but also legal aspects such as privacy issues. Finally, data curation is not only about inclusion in the database, but also about managing the data throughout its lifecycle. This is an essential element for searching, finding, retrieving and ensuring the quality of data.[17]

The evaluation has shown that not all providers have yet completed the development of the databases mentioned. On the one hand, they have to follow the legal requirements very closely and, on the other hand, they do not have too much time. In particular, there is a degree of freedom and uncertainty in the interpretation of the indicators, which may need to be specified in further legislation. However, the flexibility of such databases is also limited, as the database structures can only be dynamic to a limited extent.[18]

16 Ionos (2019).
17 Böker, E. (2023).
18 Ionos (2019).

As with the data pools and reports in Section 3, this section also briefly outlines the ways in which the data that are now publicly available can be used to generate insights that contribute to the objectives of the DSA.

The implementation of the DSA is currently at an early stage. As mentioned above, there are still uncertainties and a need for clarification. However, the process of monitoring and oversight is also beginning. The data provided will primarily be used by the EU Commission and the DSC for ongoing analysis. The available data will provide an opportunity to take a closer look and to learn whether the commitments are contributing to the objectives of the DSA.

In the context of content moderation, the data in the Transparency Database, which is updated daily, can be used to make relative comparisons (benchmarking) and to identify areas for improvement in content moderation practices. While the aim of the transparency reports is to provide an overall picture and assess the impact of changes, the ad repositories can be used to go into more detail. They can be used to analyze whether complaints are decreasing over time and to make relative comparisons with similar platforms. It is possible to show a decrease or increase in complaints by a certain percentage and compare it to another (comparable) platform. If the percentages differ significantly for no apparent reason, this could be a starting point for further analysis or regulatory action. Although it is impossible to avoid complaints altogether, even under ideal conditions, observations over time, trends and benchmarking are valuable tools for the regulator.

But it can also reveal bias. Automated content moderation systems are not free from bias and error. They often fail to interpret minority content correctly. Examples include filtering by demographic or ethnic characteristics and not displaying certain content. These cases are now systematically documented and can be analyzed and compared over time.

When it comes to further exploring the possibilities of data availability in the area of advertising, one of the first starting points is user protection. Based on the available data, the regulator will be able to analyze whether certain vulnerable groups are systematically exposed to inappropriate advertising. Two types of advertising can be distinguished: those that cannot be displayed due to legal restrictions and those that are targeted based on user data. Examples of vulnerable groups include children, people with chronic or addiction-related illnesses, and people with precarious financial backgrounds.

Matching with other secondary data that is also publicly available opens up further possibilities for analysis, particularly in the context of data on target audiences. The exact placement of advertisements via advertising networks is often beyond the control of publishers. Recommendations for action can be derived from this, especially if the publisher's Internet portals turn out not to be in line with the target group. Accordingly, the comparison with the target groups is informative for both advertisers and regulators.

Of course, the data available opens up the possibility of analyzing the context of advertising and the likelihood of certain ads being displayed in the context of, for example,

misinformation or hate speech, across different platforms and countries. This can help increase competitive pressure on platforms and their ad-based business models.

In addition to monitoring and regulatory bodies, companies and researchers can study the impact on welfare and society and will certainly be guided by the objectives of the DSA in the future. With this in mind, and in order to improve the databases and analysis capabilities, it is advisable to establish a feedback loop. This will allow researchers to report on additional data needs and indicators to be introduced.

5 Data Access for researchers

5.1 Process

Under the DSA, access to research data is understood as a new regulatory instrument that should contribute to greater transparency in the digital economy. Art. 40 (4) DSA creates a right of access to data against VLOPs and VLOSEs.

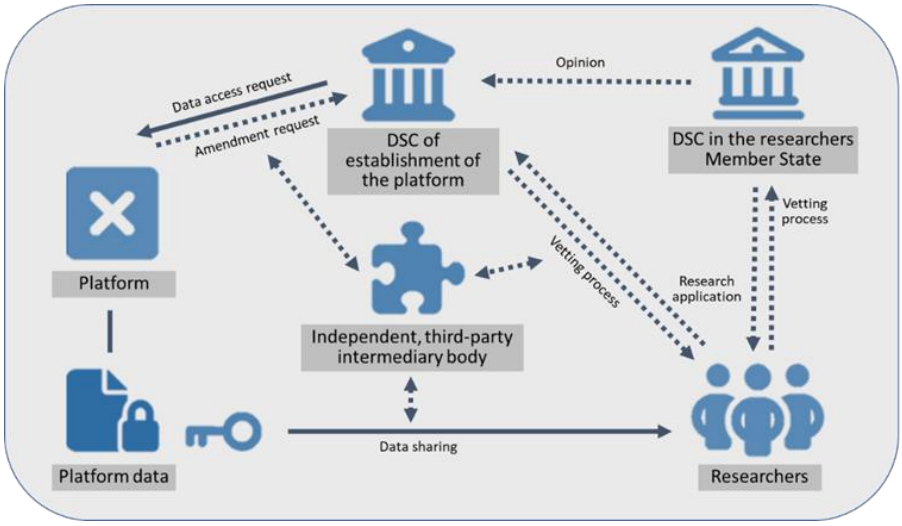
The group of entitled persons includes, on the one hand, researchers at universities and, on the other hand, at research institutions that do not pursue commercial purposes but carry out scientific research. This means that civil society organizations can also conduct research with data from online platforms and search engines, as long as they comply with protective measures on a case-by-case basis for the data provided.

The content of the right of access to data is limited. Only data that contribute to the detection, identification and understanding of systemic risks according to Art. 34 (1) DSA or to the assessment of risk mitigation measures according to Art. 35 DSA may be requested. At this point, two types of objectives become apparent. One is research funding and the other is regulatory governance. The list of systemic risks in Art. 34 (1) DSA includes socially urgent problem areas related to online platforms and search engines, but does not take into account the fact that new risks may also be identified in the future, in particular through research.

Providers of online platforms and search engines can refuse a request for access to data if it would be difficult to comply with the law on IT security or if there is a risk of revealing business secrets. Researchers are thus subject to stricter rules than accountants, who, according to Recital 92 of the DSA, must ensure confidentiality when dealing with trade secrets. This may come as a surprise, since researchers can and should also act as a possible corrective for audits and so-called audit washing. This difference in treatment shifts information asymmetries into the relationship between auditors and researchers. Furthermore, the rights of third parties, such as users of an online platform or search engine, must be respected. This applies, for example, to issues of privacy and data protection.

The extent to which obligations to make data accessible should be compensated and whether a certain level of data quality should be ensured by online platforms has not yet been regulated. The latter also concerns the question of how the data to be made accessible should be standardized or anonymized or pseudonymized. Practical experience, future legislation and possibly the European Centre for Algorithmic Transparency will provide further insights. Figure 5-1 below shows the process to be followed when a request for data access for research purposes is submitted.

Figure 5-1: Data access process



Source: WIK, based on Albert, J. (2023).

A researcher or a team of researchers intends to investigate a specific research question and needs access to data from a specific platform to do so. To obtain such access, they must prepare a detailed research proposal in accordance with Article 40 of the DSA, outlining the need for the data, the methodology and the approach to protecting data security, and commit to making the results of their research publicly available free of charge.

The next step is the so-called vetting process. Researchers must submit their application to the relevant national DSC, which will approve or reject the researcher's application after a comprehensive review process. According to Art. 40 (8) DSA, researchers must provide evidence that they meet certain criteria:

- Affiliation to a research organization,
- Independence from commercial interests,
- Explanation of funding of the research within the application,
- Fulfilment of specific data security and confidential requirements including protection of personal data,
- Demonstration of necessity and proportionality of the data access,
- Assurance that the research is for the purpose of detecting, identifying and understanding systemic risks,
- Commitment to make research results publicly available free of charge.

The purpose of this process is to ensure that researchers are working legitimately in the public interest, that they meet the required criteria, and that they have adequate technical and organizational safeguards in place to protect sensitive data once they have been granted access. The vetting process, its design and the criteria to be applied are

discussed in more detail in the following section. The details, including in particular the technical conditions under which data can be securely shared and the purposes for which it can be used still need to be clarified by the Commission in future delegated acts.

In this context, it is also worth noting the EU Code of Conduct, in which the major platforms have already committed to developing and funding a similar mechanism in June 2022.[19] To date, it is unclear whether platforms are taking their obligations under the new Code of Conduct on Disinformation seriously.

Once the research application has been approved, the DSC submits an official request for data access to the platform on behalf of the approved researchers. The platform then has 15 days to respond to the DSC's request. Either the platform grants access and makes the data available as requested, or it rejects the request. This may be on the grounds that it does not hold the data requested, or that access to the data poses a security risk or could compromise confidential information such as trade secrets. The platform must then submit a request for amendment, suggesting alternative ways of providing the data or suggesting other data that can serve the research purpose of the original request. The DSC then has a further 15 days to either confirm or reject the platform's request for amendment.

If the application or amendment is approved, the researchers must be granted access to the data. The researchers must then provide the public with an answer to their research question using the available data.

5.2 Challenges and solution approaches

There are still many technical and procedural details that need to be worked out by the European Commission before the DSA's data access regime can be fully implemented and enforced. Member States still need to establish their Digital Services Coordinators and an independent intermediary needs to be established to facilitate data access requests.

Until now, researchers have been dependent on data available due to individual national legislation (e.g. § 5 NetzDG), on commercial data products offered by the platforms or third-parties, independent data collection (e.g. web-scraping) or voluntary data access provided by the platforms. These variants pose different challenges, including insufficient scope and depth of data access and legal uncertainties in scraping. This limits adequate and systematic research.

Against this background, the data access enshrined in the DSA offers a great opportunity, provided that it is designed to meet the needs of research. Therefore, it is important to consider the following points during implementation:[20]

19 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/strengthened-eu-code-practice-disinformation_en

20 Rau, J. (2023).

- *Transparency in data availability:* Access to research data can be used more efficiently the more researchers know what data is available on the relevant platforms. A starting point could be a list of data already made available to other researchers or to the public.
- *Transparency of the methodology:* While researchers need to be very specific about what data they need from the platforms, the platforms need to be very specific and transparent about how the specific data points are generated and what the basis of the data is. Researchers need to be able to reliably understand how the data were generated in order to assess comparability with data from other platforms or secondary data. Possibility of further use of the data: The proposed deletion obligations may conflict with the rules of good scientific practice, which provide for the retention of data points for independent scientific review. A reasonable solution should be found that balances the interests involved.
- *Sustainability of research data:* Currently, some terms of use require the deletion of content from (research) data collections when the original content is deleted by the platforms or users. This is technically difficult to implement and runs counter to fundamental research interests. Against this background, it should be ensured in the future that such da-ta points can continue to be used in accordance with appropriate standards.

The obligation to provide access to research data has the potential for conflict. Researchers are concerned that a simple reference to commercial confidentiality is sufficient to deny a request for access to data. This possibility is seen as open to abuse. It is therefore important to ensure that this clause cannot be used as a pretext. From the point of view of the platforms, the review procedure and the vetting process play an important role. The procedures to be put in place and the criteria to be applied must be transparent and reliable.

As explained above, the DSA provides criteria that researchers must meet. These criteria are still relatively general at this stage and will need to be fleshed out by the DSC in order to carry out the review process. Although this is a complex task, the DSC can be guided by the established standards used to assess and evaluate research. In particular, the following three criteria are highlighted and discussed here.

- *Affiliation to a research organization:* The DSA refers to Art. 2 (1) of Directive 2019/790, which states that the primary objective of a research organization must be scientific research and/or teaching. The organization must be non-profit-making or have a mission in the public interest recognized by a Member State. It is also common practice to ask for references that demonstrate the suitability of the organization. These include publication and project track records, as well as the biographies and references of all researchers involved in the proposed project. The peer-reviewed publication process is an essential part of the scientific research community. As a rule, the more demanding and lengthy the process, the

higher the quality of the journal and the more valuable the publication as a signaling instrument. So-called journal rankings reflect the quality of a journal. These rankings are awarded by independent bodies and show the quality of publications in journals. The higher a journal is ranked and the more a researcher refers to highly ranked journals in his scientific work, the better the quality of his work will ultimately be.

- *Independence from commercial interests:* Independence from commercial interests can be demonstrated both at an organizational level and at a personal level. This applies in particular to ownership structures, project involvement in recent years, but also to existing professional and personal networks and property.
- *Fulfilment of specific data security and confidential requirements including protection of personal data:* A wide range of data must be protected by the research organization. The research organization must clearly demonstrate its ability to do so in the application process. Accordingly, an IT security concept must be submitted, describing the planned measures and existing infrastructures, in particular against loss, manipulation and illegal access.

Supporting the review process through the establishment of a commission seems sensible to demonstrate independence and transparency to all parties involved. Such organizations have been established for many years and are well known in the scientific community. Furthermore, these institutions publish their own guidelines, which provide orientation and include criteria.[21]

The indicators provided by the DSA are primarily regulatory in nature. It is necessary to take into account the different needs of regulation and research. This raises the question of whether the indicators provided reflect the information that researchers need for their analyses. Even if researchers can access other data through research data access, it remains to be seen whether this data is actually available in the form required by researchers. In the context of content moderation, for example, the transparency reports collect indicators that can be used to benchmark the actions taken and reflect whether everything that needs to be removed has been removed and everything that should not be removed has been restored. Indicators that can be used to improve quality and develop techniques and systems are also collected. These include, in particular, qualitative information and information about the personnel used in terms of their capacities and qualifications. Indicators that can be used for the further development of technologies and methods offer the public, research and industry starting points for innovation and should be specifically expanded from an innovation ecology perspective.

21 See for example Ethics Committee University of Mannheim (2023), Ethics Committee University of Tuebingen (2023).

6 Conclusion

The availability of rich data sources under the DSA provides great opportunities for researchers to shed light on systemic risks in the platform economy. However, managing these obligations is a complex task and the DSA provides a framework for monitoring compliance. Managing these tasks can be fraught with considerable effort and difficulty. On the one hand, there is the sheer volume of data and, on the other, there are information asymmetries between the regulator and the companies providing the data.

In order to meet these requirements, both the DSC and the European Commission as monitoring bodies need independence, tailor-made working methods, the development of expertise and the establishment of a self-image as a central and coordinating body for all DSA-related data analyzes from the outset.[22] For the DSC, independence means being free from directives and having sufficient staff and resources. Transparency towards the industry and the public must be ensured and appropriate contact points established. Working methods should be highly flexible in order to meet the needs of the different subject areas. This requires a close exchange between the competent authorities, for which appropriate information systems are essential. Interdisciplinary and intercultural skills are required for national and international cooperation, in addition to technical expertise, which must be available in sufficient capacity and continuously expanded through close contact with professional groups and the scientific community.

Monitoring, research and impact analysis are ideally based on a combination of primary and secondary data. Secondary data is data that is needed to fully answer specific questions and to further explore the primary data that is available. Some of this data is needed once, but some is needed over time. Indicators such as financial ratios, location data, etc. can also be counted as indicators based on an economic evaluation. An example would be determining the revenue model of a service. Tracking data can also be used to comprehensively map the user behavior of individual services.

The collection, linking and consistent storage of primary and secondary data over time is, alongside comparative analyzes, one of the core tasks of the PLASMA project.[23] PLASMA thus provides a solid scientific basis for identifying current trends, market developments and user preferences in digital markets and tracking them over the long term. The needs of both the DSA and the DMA are addressed.

However, the focus of the project is specifically on the digital economy in Germany. Therefore, the scope of the database (both in breadth and depth) and the accompanying analyzes are unique.

22 Jaursch, J. (2022b).
23 Plasma (2023).

PLASMA therefore helps to make the data generated by the new reporting requirements of the DSA usable and accessible for the German market. Building on the experience of its predecessor project, DOTT, PLASMA provides insights into the challenges of defining indicators and their consistency across different services, enterprises and geographical areas, and how to make data collection and storage processes as efficient as possible.

Literature

Albert, J. (2023): A guide to the EU's new rules for researcher access to platform data, online available https://algorithmwatch.org/en/dsa-data-access-explained/ [Last accessed 23.10.2023]

Böker, E. (223): Was ist ein Repositorium, online available https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/repositorien/ [Last accessed 21.10.2023]

Booking.com (2023): Digital Services Act, online available https://www.booking.com/content/dsa.en-gb.html [Last accessed 09..11.2023]

Ethics Committee University of Mannheim (2023): online available https://www.uni-mannheim.de/en/about/organization/bodies-and-committees/committees-and-councils/ethics-committee/ [Last accessed 11.10.2023]

Ethics Committee University of Tuebingen (2023), online available https://uni-tuebingen.de/en/faculties/faculty-of-economics-and-social-sciences/offices-resources/statutory-bodies-and-committees/ethics-committee/#c1016829 [Last accessed 11.10.2023]

European Commission (2023c): DSA: Guidance on the requirement to publish user numbers, online available https://digital-strategy.ec.europa.eu/en/library/dsa-guidance-requirement-publish-user-numbers [Last accessed 29.10.2023]

European Commission (2023b): DSA: Very large online platforms and search engines, online available https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops [Last accessed 09.10.2023]

European Commission (2023a): Digital Services Act Package, online available https://digital-strategy.ec.europa.eu/de/policies/digital-services-act-package [Last accessed 23.10.2023]

European Commission (2023): A strengthened EU Code of Practice on Disinformation, online available https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/strengthened-eu-code-practice-disinformation_en [Last accessed 09.11.2023]

European Commission (2023): DSA Transparency Database, online available https://transparency.dsa.ec.europa.eu/analytics [Last accessed 09.11.2023)

Googleapis.com (2023): Information about Monthly Active Recipients under the Digital Services Act (EU), online available https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-24_2023-1-1_2023-6-30_en_v1.pdf [Last accessed 09.11.2023]

Ionos (2019): Datenbanken: Wozu man sie braucht und welche Arten es gibt, online available https://www.ionos.de/digitalguide/hosting/hosting-technik/datenbanken/ [Last accessed 21.10.2023]

Jaursch, J. (2022a): Overview of DSA delegated acts, reports and codes of conduct, online available https://www.stiftung-nv.de/en/publication/overview-dsa-delegated-acts-reports-and-codes-conduct [Last accessed 29.10.2023]

Jaursch, J. (2022b). Wie die deutsche Plattformaufsicht aufgebaut sein sollte Empfehlungen für einen starken „Digital Services Coordinator", Stiftung Neue Verantwortung. https://www.stiftung-nv.de/sites/default/files/snv_empfehlungen_fur_einen_starken_dsc.pdf [last accessed: 13.07.2022]

PLASMA (2023) : Project internet side, online available https://plasma.wik.org/ [Last accessed 30.10.2023]

Quintais, J. P., Schwemer, S. F. (2022). The Interplay between the Digital Services Act and Sector Regulation: How Special Is Copyright? *European Journal of Risk Regulation*, *13*(2), 191-217

Rau, J. (2023): Clearing the data fog: German far right researcher's requirements for the DSA research data access, online available https://leibniz-hbi.de/de/blog/clearing-the-data-fog-german-far-right-research-s-requirements-for-the-dsa-research-data-access [last accessed 21.10.2023]

Schmid, G., Koehler, P. (2022). Digital Services Act – An Overview. https://www.taylor-wessing.com/en/interface/2022/the-eus-digital-services-act/digital-services-act---an-overview [Last accessed 09.08.2023]

Snap Inc.(2023): Transparenzbericht, online available https://values.snap.com/de-DE/privacy/transparency/european-union [Last accessed 11.10.2023]

Steiner. F. (2023): DSA:, Worauf die EU-Kommission sich beim DSA fokussiert, online available https://table.media/europe/analyze/worauf-die-eu-kommission-sich-beim-dsa-fokussiert/ [Last accessed 16.10.2023]

TechCrunch (2023): Amazon wins interim stay on delivering a public ads archive in early challenge to EU's Digital Services Act, online available https://techcrunch.com/2023/09/28/amazon-dsa-interim-measures/?guccounter=1 [Last accessed 05.10.2023]

Ten Thije, P., van Hoboken, J. (2022): Average monthly active recipients in the DSA: Definition, grey areas, and how to calculate?, online available https://dsa-observatory.eu/2022/10/10/average-monthly-active-recipients-digital-services-act-dsa-definitions-grey-areas-how-to-calculate/ [Last accessed 13.07.2023]