University of Amsterdam
Institute for Information Law

Information Law and the Digital Transformation of the University

Part II. Access to Data for Research

Information Law and the Digital Transformation of the University

Part II. Access to Data for Research

Authors: Jef Ausloos,[1] Arlette Meiring, Doris Buijs, Mireille van Eechoud, Stefanie Boss and Joanna Strycharz

Citation: Institute for Information Law (2023). Information Law and the Digital Transformation of the University. Part II. Access to Data for Research. Amsterdam: September 2023.

https://www.ivir.nl/part-ii-access-to-data-for-research/

Funding: This report was commissioned by the Executive Board of the University of Amsterdam.

The opinions expressed in this work reflect the authors' own views and not those of the commissioning organisation. The research for this report has been conducted in full compliance with the Netherlands Code for Conduct for Research Integrity (2018).

Amsterdam, 15 September 2023

Instituut voor Informatierecht
Institute for Information Law

University of Amsterdam
Institute for Information Law
P.O. Box 15514
1001 NA Amsterdam
The Netherlands
Website: https://www.ivir.nl/

[1] Jef Ausloos received funding from the Dutch Research Council (NWO), for the project *Empowering Academia through Law: Transparency Rights as Enablers for Research in a Black Box Society* (project number: VI.Veni.201R.096).

Executive summary

The digitisation of society makes it crucial to be able to observe and understand how data and digital infrastructures intermediate the world around us. Yet, whereas the amount of data being generated in our society is growing exponentially, it becomes increasingly hard for (academic) researchers to access that data and observe digital phenomena. In the context of scientific research, digital infrastructures – operated by both public and private sector actors – can be impenetrable fortresses, challenging universities' core missions as public interest-driven knowledge producers and watchdogs. While there might be various reasons for refusing researchers' access to data, important questions remain as to their validity and desirability, especially considering the ubiquity and societal impact of digital infrastructures (and the organisations behind them). In parallel to these developments, the EU legislator has recently issued a wide range of legal frameworks specifically addressing digital infrastructures, most of which comprise transparency and data access provisions. Yet, most of these provisions have not been drafted with academia's interests and mission in mind.

Against this background, this Report aims to do two things. Based on empirical insights (survey and interviews), Part A explores current practices and issues with obtaining data in and about digital infrastructures for academic research purposes. Part B maps transparency and data access provisions in 16 recent digital/data legislation at EU level, identifying the main opportunities and limitations for academic research. The report is part of a broader research project commissioned by the UvA's Executive Board aimed at clarifying digital sovereignty- and data access-claims of universities, which are faced with a rapidly evolving technological, economic, and legislative landscape.

Part A. Empirical study into the challenges to data access: findings and suggestions

Part A of the report lays out the results of an empirical study (i.e., UvA-wide survey and interviews) into (a) the reliance of UvA researchers on third-party data and (b) the obstacles UvA researchers encounter when accessing such third-party data.

- *Growing reliance on third-party data*

  According to our survey, UvA researchers are, on average, slightly reliant on third-party data. Researchers from the Faculty of Economics and Business indicated to rely on third-party data the most, while researchers from the Faculty of Law rely on such data the least. Researchers from all faculties, however, expect that their respective fields will become (much) more reliant on third-party data in the future.

- *Frequent restrictions and conditions*

  Both the survey and interviews show that UvA researchers occasionally experience issues when trying to access and subsequently use data held by third parties for research purposes. 'Legal issues,' and more specifically issues relating to data protection law and terms of service, were mentioned most frequently. Publication and sharing restrictions – for instance, not being allowed to share the data or with a limited number of parties only – were also highlighted relatively often.

- *Need for institutional support*

  Most survey respondents were neutral to unsatisfied with the institutional support they had received when confronted with data access and data use challenges. During the interviews, it came to the fore that UvA researchers and data stewards would appreciate more robust institutional support to promote researchers' access to third-party data, such as: additional legal support and/or legal

training; guidance on how to deal with specific data access and use-questions, including tensions between open science principles and data sharing restrictions/conditions imposed by third parties; better internal communications and streamlined procedures; efforts to raise more awareness among researchers of the importance of data access issues; and investments in technical resources to facilitate data access, analysis and storage.

- *Recommendations*

  Considering the findings described above, this report advises universities to:
  - Take measures to strengthen the knowledge capacity of and resources for data stewards across faculties so that they can better support researchers in dealing with data access challenges;
  - Expand the pool of legal staff and/or offer legal training to both data stewards and researchers;
  - Invest in, and promote the use of, technical resources to safely access, analyse and store third-party data; and
  - Raise awareness among researchers on data access and data use challenges.

Part B. Mapping of transparency and data access provisions in EU law: findings and suggestions

Part B of this report lays out the following key results and recommendations, resulting from a mapping exercise and analysis of transparency and data access provisions contained within 16 adopted and proposed legal frameworks under the EU's digital policy agenda dealing with digital infrastructures and data.

- *Lack of direct data access rights for researchers*

  Transparency provisions and data access rights that seem potentially useful for researchers are scattered throughout a fragmented landscape of digital/data legislation. Most of them are phrased in rather generic terms and do not address researchers in particular. Strong direct data access rights specifically aimed at researchers are thus lacking, with the notable exception of Article 40 of the Digital Services Act (though even this provision does not contain a 'general' right to third-party data for researchers).

- *Potential benefits of frameworks for researchers*

  Despite the general absence of direct access rights for researchers, the analysed legislative instruments still hold potential for researchers, for instance by providing opportunities for data donation or otherwise contributing to an enabling environment for data sharing beneficial to scientific research. For example, some of the instruments have introduced types of 'data intermediaries' that are expected to boost wider (voluntary) data sharing and the further development of the European Open Science Cloud (EOSC).

- *Balancing data access with third-party rights and interests*

  Some of the analysed provisions mandate a balancing of interests, with transparency and data access on the one hand, and the protection of third-party rights (e.g., personal data, intellectual property rights, trade secrets) on the other. However, most provisions do not specify *how* such a balancing exercise should take shape in practice. This may raise uncertainty as regards the practical value of new data access provisions and risks perpetuating existing power asymmetries.

- *Recommendations*

  Considering the findings described above, this report advises universities to:

- Invest in legal, methodological and technical capacity to enable the best use of transparency and data access provisions as laid down in EU law;
- Share knowledge, best practices and experiences with the use of transparency and data access provisions within (different departments of) the university and between relevant institutions;
- Use existing coalitions to lobby for the explicit recognition of academic interests in the adoption and implementation of transparency and data access provisions;
- Reflect on the need for, and wider implications of, data access claims in the context of academic research.

TABLE OF CONTENTS

Acronyms

AI
Artificial Intelligence

CJEU
Court of Justice of the European Union

CFREU
Charter of Fundamental Rights of the European Union

DAO
Data altruism organisation

DIS
Digital intermediation service

ECAT
European Centre for Algorithmic Transparency

ECHR
European Convention on Human Rights

ECtHR
European Court of Human Rights

EBDS
European Board for Digital Services

EDIB
European Data Innovation Board

EDMO
European Digital Media Observatory

EDPB
European Data Protection Board

EOSC
European Open Science Cloud

ERB
Ethical Review Board

IoT
Internet of Things

IP
Intellectual Property

NGO
Non-governmental organisation

VLOP
Very large online platform (service provider)

VLOSE
Very large online search engine (service provider)

Introduction

Observability constitutes an essential tenet of scientific research.[2] Indeed, knowledge production relies on the ability to observe the world around us and generate insights. In light of the growing digitisation of our society, this also means the ability to observe digital infrastructures[3] that over the past decades have nested themselves into many parts of society. Our personal and professional interactions, social and political discourse, and economic, financial, healthcare, and public safety systems all rely on digital infrastructures to function. Importantly, the urgency to observe digital infrastructures – mostly by analysing the data residing in these infrastructures – extends beyond the ability to understand the digital infrastructures themselves or their impact on humans and the environment: they are also invaluable sources to study and understand multiple *other* aspects of contemporary society, whether it is the emotional development of teens (e.g., by observing their social media usage), urban mobility (e.g., by analysing sensor data held by public transport companies), or health (e.g. by extracting data from smart watches).[4] In short, researchers need to be able to (digitally) observe, in order to study. And in order to observe, they need access to the data residing in the digital infrastructures that permeate their objects of study.[5] Yet, whereas the public's deployment of digital infrastructures and the amount of data residing in those infrastructures is growing exponentially, there have been indications that (academic) researchers experience difficulties when trying to access externally-held data for research purposes.[6]

The objective of this report is twofold. First, it aims to present insight into the current reality of data access for scientific research purposes. To this end, an empirical study was conducted into (a) the reliance of researchers affiliated with the University of Amsterdam (UvA) on third-party data, especially stemming from the private sector, and (b) the obstacles UvA researchers encountered in the past when accessing such third-party data. The results of this study are discussed in **Part A** of this report. Second, it aims to provide an overview of data access and transparency provisions throughout recent EU digital/data legislation that could potentially benefit academic researchers as a means to obtain access to certain third-party data. This overview and analysis are presented in **Part B** of this report. Together, the studies form the second part of a broader research project into the challenges associated with the digital transformation of universities that the Executive Board of the University of Amsterdam has commissioned from the Institute for Information Law (IViR). While the first part of this research discusses the key concept of "digital sovereignty" and its meaning for the university sector more broadly, the second report digs deeper into an important dimension of digital sovereignty, namely "data sovereignty". Universities and academics' data sovereignty is arguably at risk partly because academic researchers are becoming more and more dependent on third parties to obtain access to meaningful digital data for their research projects.[7] The underlying assumption of both reports is that currently, it is the owners of digital infrastructures that ultimately define the practical, technical, and epistemic terms and conditions under which knowledge is produced.

The increasing concentration and privatisation of data may have considerable effects on independent academic research. Indeed, the organisations that manage digital infrastructures – often large

---

[2] The concept of 'observability' was proposed in 2020 in the specific context of digital platforms, as a means of regulation. See Rieder and Hofmann 2020, and also Leerssen 2023a on social media regulation for observability.

[3] Cf. text box on p. 11.

[4] Compare Tromble 2021, p. 1-2: "Digital research provides important insights about the social, cultural, economic, and political phenomena that impact people's everyday lives."

[5] On the normative underpinnings of claims for access to data for research, see Part B, chapter 2 of this report.

[6] See e.g., Tromble 2021, pp. 1-8.

[7] See IViR, 'Information Law and the Digital Transformation of the University: Digital Sovereignty, Data Governance and Access to Data for Research – Part I. Digital Sovereignty', 2023, section 3.3.2.

technology companies – only rarely release data under their control for outside inquiry. Combined with their resources and reach, these organisations can *de facto* influence (certain) research agendas, that is, in research fields that (heavily) rely on third-party data.[8] While organisations may have various reasons for refusing access, questions remain as to their validity and desirability, especially considering the ubiquity and societal impact of many digital infrastructures.

Faced with legal, technical, financial and other obstacles to data access (Part A), over the years academic researchers have deployed a plethora of methods to observe digital infrastructures nonetheless. These range from concluding *ad hoc* collaborative data sharing arrangements[9] and repurposing programmatic tools for data access,[10] to more independent or adversarial approaches[11] that do not necessarily rely on the goodwill of the organisations managing the respective digital infrastructures.[12] Beneficial as they may be to some researchers, these approaches have significant drawbacks as well. For example, reliance on voluntary data access agreements could potentially lead to favouritism in that only well-funded or 'prestigious' universities receive useful data. Moreover, these arrangements cement the power of data holders in determining the scope and requirements of data sharing, as well as the ability to unilaterally retract (part of) the provision of data.[13] Other approaches may also come with technical, ethical and methodological questions, and some may even find themselves in a 'legal grey zone' such as web scraping or the repurposing of APIs for research.[14]

Considering these limitations, we suggest that *the law* could potentially serve as an additional tool to observe digital infrastructures in researchers' methodological toolboxes, as it could provide a basis for structured and uniform data access procedures and counteract strong incentives that might exist against transparency and openness. Restrictions on the accessibility of data for research have been a driver for policy debates at the EU level and, subsequently, led to the creation of new provisions in EU legislation providing for transparency and access to certain data. The role of academic research in these provisions, and how academic researchers could potentially benefit from these new provisions, are however not so clear.[15] Part B of this report therefore aims to identify and analyse relevant data access and transparency provisions throughout a number of existing and proposed instruments under the EU's digital policy agenda. Based on the mapping, legal gaps are identified which serve as a basis for recommendations to establish a clear(er) legal environment and address academic researchers' claims to observability.

[8] Ausloos and Veale 2020; Ausloos, Leerssen and Ten Thije 2020; Van Drunen and Noroozian 2023; Keller and Leerssen 2019.

[9] See e.g., the data exchange between the Mastercard Center for Inclusive Growth and the Urban Institute in the name of 'data philantrophy', <https://www.mastercardcenter.org/insights/data-philanthropy-offers-new-avenues-solving-old-problems-report-finds>; Schrage and Ginsberg 2018; Consumer Data Research Centre <https://www.cdrc.ac.uk/about-cdrc>.

[10] Ohme et al 2023, Bruns 2019, Gerlitz et al 2019.

[11] E.g. web scrapers (e.g. https://labs.polsys.net/) or plug-ins (e.g. Bodó et al 2017).

[12] For an overview of relevant approaches, see notably: Keller and Leerssen 2019; Ausloos and Veale 2020; Rieder and Hofmann 2020.

[13] E.g., Ledford 2023.

[14] See Bobrowsky 2021; Kayser-Bril 2021 (accessed 14 August 2021).

[15] As will be further explained in Part B of this report, the provisions seem to be framed in rater *generic* terms, which aim to 'broadly' stimulate the functioning of the internal market rather than scientific research specifically. See also European Commission Communication 2020b, Noto La Diega 2022 (accessed 20 March 2023).

In this report, we broadly refer to **digital infrastructures** as socio-material artefacts consisting of both technical or material components (e.g., cables, hardware devices, data centers, computing resources) and organisational or socio-technical components (e.g., settings, standards, governance structures, networks and processes that contribute to the functioning of an information system).[16] Examples of digital infrastructures are data centers and smartphones, but also the structures on top of which (commercial and non-commercial) digital platforms are built (e.g., Facebook, AirBnB, MijnOverheid and many more). Digital infrastructures must be distinguished from the data generated, collected or held by these digital infrastructures.

In this report, we define **data** as "any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording".[17] The concept of data is narrower than the concept of "information" in that it does not cover content in paper form.

[16] Ferrari 2023, p. 23; see also Henfridsson and Bygstad 2013, pp. 908-909. For a more technical/material interpretation, see: Constantinides, Henfridsson and Parker 2018, as endorsed by Van Dijck, Nieborg and Poell 2019.

[17] Article 2(1) Data Governance Act.

PART A: Data Access in Practice

An Empirical Study into the Challenges to Data Access and Data Use for Research Purposes at the University of Amsterdam

2023

Jef Ausloos, Arlette Meiring, Doris Buijs, Stefanie Boss and Joanna Strycharz
University of Amsterdam
Institute for Information Law

1. Objective and research methods

This part of the report describes the results of an empirical study into data asymmetries. More specifically, it identifies and maps (a) the reliance of scientific researchers employed at or affiliated with the University of Amsterdam (UvA) on third-party data, especially stemming from the private sector, and (b) the perceived obstacles to researchers' access to such third-party data. As such, this study contributes to a broader research project on the challenges associated with the digital transformation of universities that the Executive Board of the University of Amsterdam commissioned from the Institute for Information Law (IViR). The empirical study is two-fold and includes both an UvA-wide survey and interviews with researchers and data stewards working across all UvA faculties.

1. Survey

First, the report summarises the results of a survey conducted between 25 October 2022 and 25 November 2022 among all scientific staff – i.e., researchers and teachers – affiliated with the UvA (see chapter 2 below). The goal of the university-wide survey was to acquire a high-level understanding of the types of academic research that rely on third-party data, and moreover, of the types of challenges faced by researchers when trying to access such data for research purposes.

The survey was approved by the Law Faculty's Ethics Review Board before data collection. The survey was available in English and in Dutch and participants were recruited through an email sent by the president of the University's executive board[18] (Geert ten Dam) to all scientific personnel. If the respondent wanted to participate after receiving an email invitation, they could click on a link that would redirect them to a Qualtrics survey. After obtaining informed consent, they could proceed to the survey. Otherwise, they were exited from the study (n = 5). In total, 274 employees consented to participate and completed the survey. The median response time was 3.4 minutes.

First, respondents were asked about their professional background within the university, including questions about their faculty and department, position and time spent on research. Next, they were asked about the reliance on third-party data for their work and their field in general. Respondents received a detailed explanation about the meaning of "third-party data" in this research context. Next, respondents were asked a filter question whether they had recently engaged in a research project that relied on third-party data. Respondents who had not engaged in such a project were exited from the survey. The remaining 128 participants were asked details about this project including information about the third party they collaborated with, challenges and restrictions they faced, as well as best practices to overcome them. At the end, participants were asked if they would be willing to participate in in-depth interviews on the topic of data access.

2. Interviews

Second, the report provides an analysis of a series of 16 semi-structured interviews undertaken in December 2022 – February 2023 with the aim to better grasp the challenges identified via the survey, by collecting more detailed information on specific dependencies and challenges (see chapter 3 below). Out of the 15 interviews, 12 interviews were conducted with **scientific researchers**. For each of the seven faculties at the UvA,[19] at least one researcher was selected. For some faculties we spoke to two or three researchers,

[18] College van Bestuur.

[19] Faculties of Law; Economics and Business; Dentistry (ACTA); Humanities; Medicine; Science; and Social and Behavioural Sciences: <https://www.uva.nl/en/about-the-uva/organisation/faculties/faculties.html>.

depending on their availability and interest in being interviewed. Three researchers also responded to questions in their capacity as members of the ethics committees of their respective faculties. In addition, four interviews were conducted with **data stewards.**[20] All interviews were held online and lasted, on average, 45 minutes. They were led by two or three interviewers, i.e., Jef Ausloos (IViR[21]), Arlette Meiring (IViR) and Joanna Strycharz (ASCoR[22]). All interviewees had given their consent to an audio recording being made during the interview and to a transcription being produced afterwards. The transcripts were pseudonymised and shared with the interviewees for revision before the analysis.

The interviews had also been approved by the Law Faculty's Ethics Review Board before they took place. Prior to each interview, interviewees received an information letter which informed them of the empirical research project, the interview procedure, the identity of the interviewers and the processing of their personal data. They also received a consent form in which they could indicate whether they agreed with participation in the research, the collection and processing of their personal data, the recording and transcription of the interview, and the storage of (anonymous) research data generated through the interviews. It was also stated in the form that any responses provided by the interviewees would not be linked to their names when later cited in academic publications or reports.

During the interviews, researchers and data stewards were asked questions aimed at better understanding the various challenges researchers may run into when trying to access and use third-party data for research purposes:

Researchers were asked, in summary:

- what type(s) of data/information they typically use in their field of research and where these data are produced, generated and held, i.e., within the faculty itself or by third-party organisations such as public sector bodies, other academic institutions, research organisations, companies, NGOs, hospitals, etc.;
- whether they were (recently) involved in a specific research project for which access to externally-held data was required, and if so, what that process of data access looked like;
- whether, in that specific research project or in any other projects, they experienced any obstacles while trying to access the third-party data, and if so, what types of obstacles they ran into;
- whether the third party attached any restrictions or conditions to its permission to access and use the data;
- whether they sought (institutional) advice when bumping into obstacles, and if so, whether they had received (sufficient) support; and finally,
- whether there is anything they feel the UvA could help them with in this regard.

Data stewards were essentially asked the same questions as the researchers but with the expectation that they are able to provide more of a 'helicopter-view' on data access challenges experienced by researchers at their respective faculties. Additionally, there were asked:

- how they view their role as a data steward assisting researchers facing data access issues;
- whether their faculty pays attention to the legal aspects of data access and use and whether they advise researchers on existing legal frameworks that could potentially help them to access certain data;
- whether they feel there is sufficient guidance on the opportunities and challenges for data access; and

[20] Data stewards are university staff members who advise and support researchers and students on a daily basis on all aspects of (research) data management. Each faculty at the UvA has designated at least one, but often multiple data stewards, see UvA RDM (webpage) <https://rdm.uva.nl/en/support/support.html>. See also the text box on p. 15.

[21] Institute for Information Law (IViR).

[22] Amsterdam School of Communication Research (ASCoR).

- whether there is anything they feel the UvA could help them with so that data stewards feel better equipped to support researchers at their respective faculties.

As a disclaimer, we would like to point out that the interviewed UvA researchers and data stewards are not representative for the academic research community as a whole. Although the analysis attempts to categorise data access and data use challenges based on the anecdotes provided by the interviewees, we do not claim this categorisation to be exhaustive.

The analysis of the survey and interviews resulted in a handful of recommendations to help the UvA address the challenges flagged by its constituents (see chapter 4).

Data stewardship

Data stewards at the UvA have a key role in assisting researchers of all faculties with matters of data management (e.g., the collection, storage and sharing of research data), data protection and information security. Additionally, they are the link between researchers and the university's legal department, the faculty's ethics review board and technical support teams. The role of the data steward has increasingly been shaped and formalised over the past few years. As a result, UvA data stewards have also become more and more involved in processes of (negotiating) researchers' data access. For an overview of the (contact information of) data stewards designated for each of the UvA faculties, see UvA RDM (webpage) <https://rdm.uva.nl/en/support/support.html>.

2. Survey results

As noted above, a survey was conducted in the fall of 2022. The survey was sent out by the Executive Board of the UvA to all scientific personnel working at the university. In total, 274 people from across all UvA faculties completed the survey between 25 October 2022 and 25 November 2022.

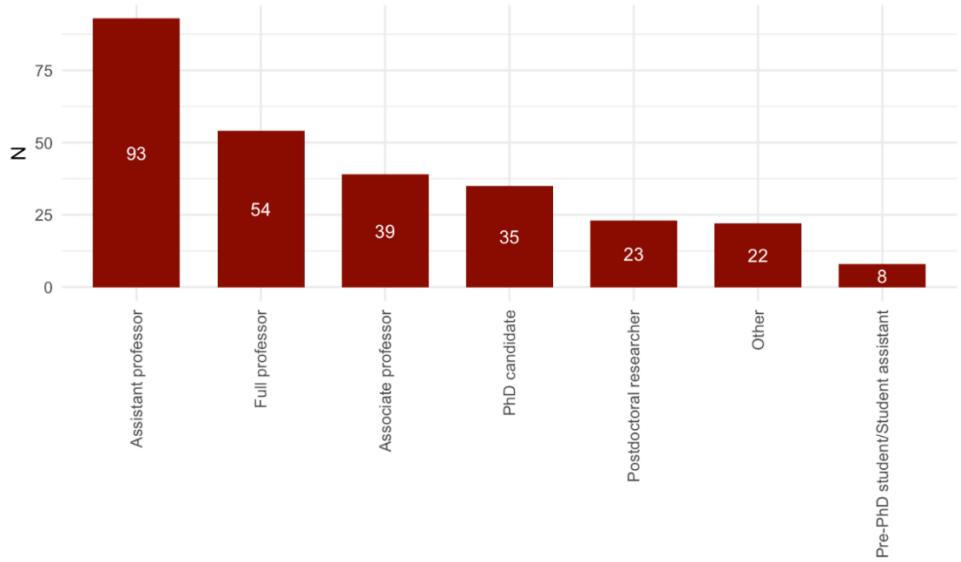

Figure 1. Position of respondents

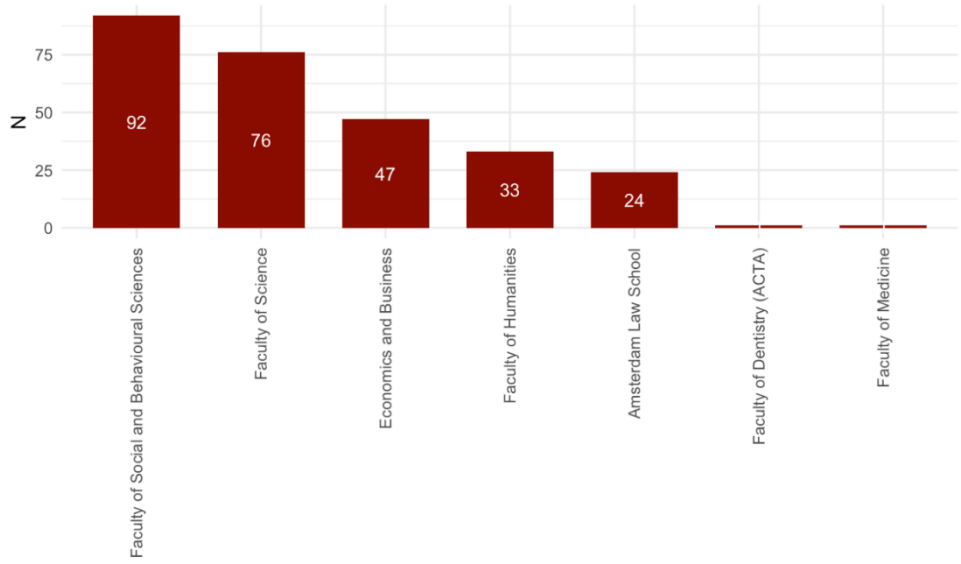

Figure 2. Faculty of respondents

On average, respondents effectively spend 52.9% of their time on research, with minor differences across different faculties. However, the amount of research time differs significantly according to the positions people hold, with only Pre-PhDs (67%), PhDs (78.6%) and Postdocs (76.2%) spending more than half of their time on research.

2.1 Reliance on third-party data

After some basic questions about their background, respondents were asked how much they rely on third-party digital data collection in their respective research. The following explanatory note was added:

> *We use **'third-party digital data collection'** to refer to situations where you collect digital data that has been collected, produced or generated by a third party in the course of their business (e.g., behavioural data collected by an online platform, or sensor data collected by a vehicle manufacturer). The data can be provided to you voluntarily (e.g., made available through an API) or obtained involuntarily (e.g., obtained by scraping the third party's infrastructure or requested on the basis of legal transparency requirements).*
>
> ***'Third-party digital data collection'** does <u>not</u> include, for example, surveys or interviews you conduct yourself, even if you do so with the help of a survey company.*

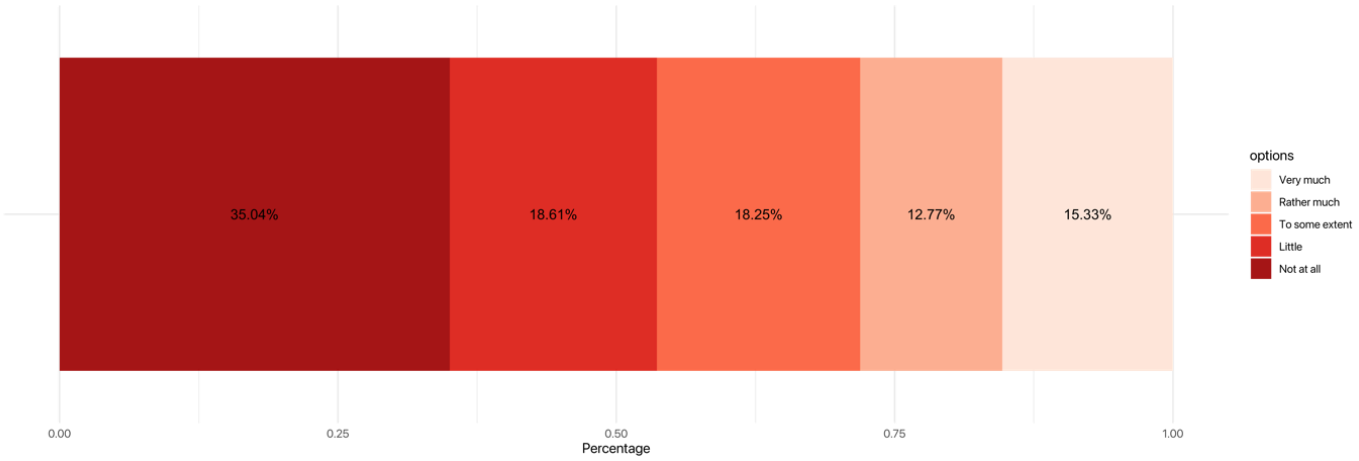

Figure 3. Reliance on third-party data

On average, UvA researchers are **slightly reliant** on third-party data (*M* = 2.54, *SD* = 1.46, 1-5 scale),[23] but this **differs strongly per faculty**, with researchers from the Amsterdam Law School (*M* = 2.17) relying on such data the least and researchers from the Faculty of Economics and Business (*M* = 3.15) and Faculty of Science (*M* = 2.58) relying on such data the most.

Respondents were also asked how often they co-design or collaborate on a joint research project with non-academic third parties to generate data in the pursuit of a common research goal (e.g., medical trials with a pharmaceutical company). Across the UvA, just over half (52%) of researchers stated they *never* do so. When further broken down per faculty, collaborations occurred least frequently at the Amsterdam Law School (*M* = 1.58, SD = 0.76) and most frequently at the Faculty of Humanities (*M* = 1.94, SD = 1.07) and Faculty of Social and Behavioural Sciences (*M* = 1.89, SD = 1.06).

Importantly, 55% researchers from all faculties indicated that they expect their respective fields to become (much) more reliant on third-party data in the future (*M* = 3.6, *SD* = 0.80, 1-5 scale).

[23] 1-5 scale = options ranged from 1 (not at all) to 5 (very much).

2.2 Use of third-party data

Nearly half of the respondents (46.4%) reported to have recently engaged in a research project that relied on digital data produced by a third party. The sources for those research projects varied significantly, ranging from private sector companies (24%) to academic institutions (15%), research organisations (13%), (international) government bodies (12%), official statistics bodies (12%), other public institutions (8%), other (7%),[24] hospitals (5%) and NGOs (5%). After clustering some of these categories, **public sector organisations** appear to be third parties **most frequently relied on** for data (32%).[25]

The majority of research projects that relied on **private sector data**, relied on the following sources: Online Platforms (21%), followed by Research (9.5%), Other (9.5%), Communications (8.6%), Finance (7.6%), the Entertainment Sector (6.7%), the Cultural Sector (4.8%), Health (4.8%), News/Journalism (4.8%), Education (3.8%), Marketing (3.8%), Agriculture (2.9%), Small and Medium Enterprises (2.9%), Energy (1.9%), Real estate (1.9%), Transportation (1.9%), Clothing (1%), Construction (1%), Hospitality (1%), and Food production (1%).

2.3 Use of private sector data

Next, respondents were asked to think about access to data held by **private sector companies** in particular (which is the focus of this report). They were explicitly asked to reflect on a **specific data-intensive research project** that they recently engaged in. As it turned out, the vast majority of private sector data is provided by companies **on a voluntary basis** (80%). For these data, approval from the private sector party was needed in 63.5% of the projects (see Figure 4 for approval rates per faculty). In 53.4% of research projects that relied on voluntarily provided data, the private sector party had a standardised process in place to grant access to the respective data.[26]

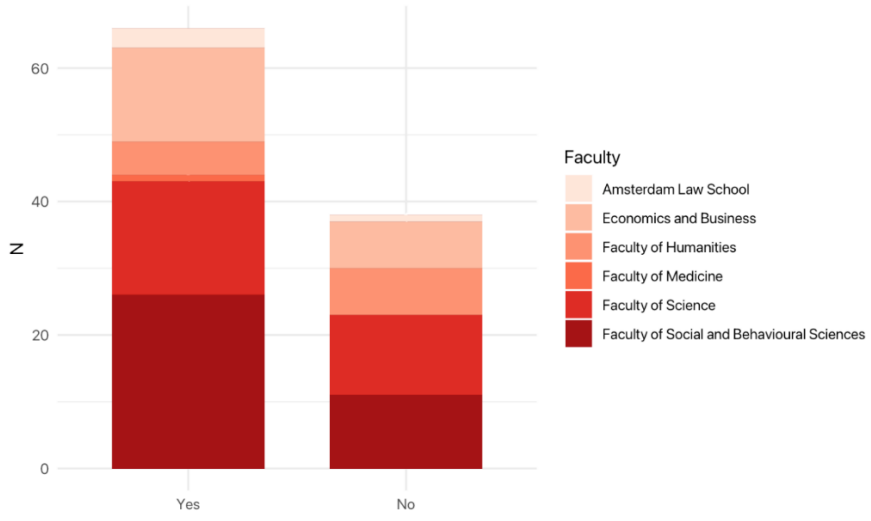

Figure 4. Approval rates for access to data per Faculty

[24] Among 'others', respondents mentioned third parties such as libraries, political parties, wikidata, news media standards developing organisations and governance bodies, schools.

[25] Combining (international) government bodies, official statistics bodies, and other public institutions.

[26] E.g., through an API, or standard terms of service.

2.4 Restrictions and conditions

In **53.1%** of research projects where third-party data was relied on, respondents stated that they **faced specific restrictions or conditions** for accessing and/or using the respective data. For example, among the researchers that faced restrictions, 21.3% had to pay a fee in order to obtain access to the desired data.

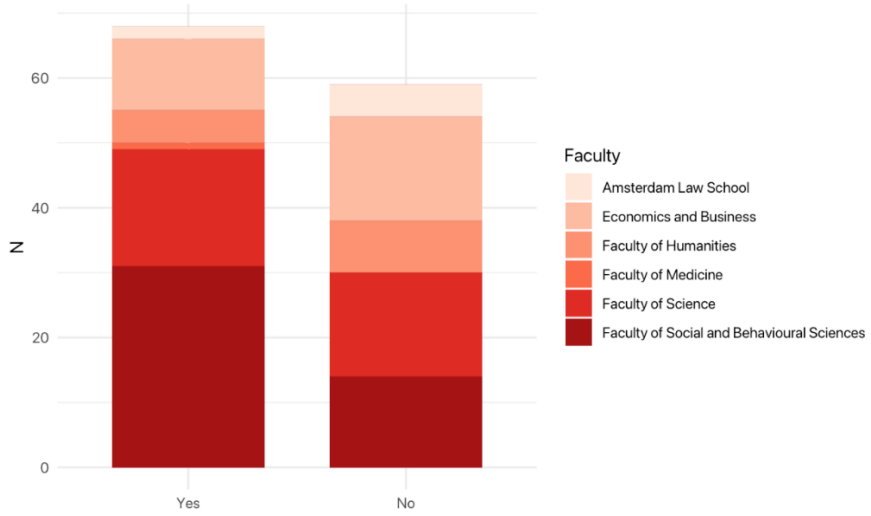

Figure 5. Did researchers face restrictions?

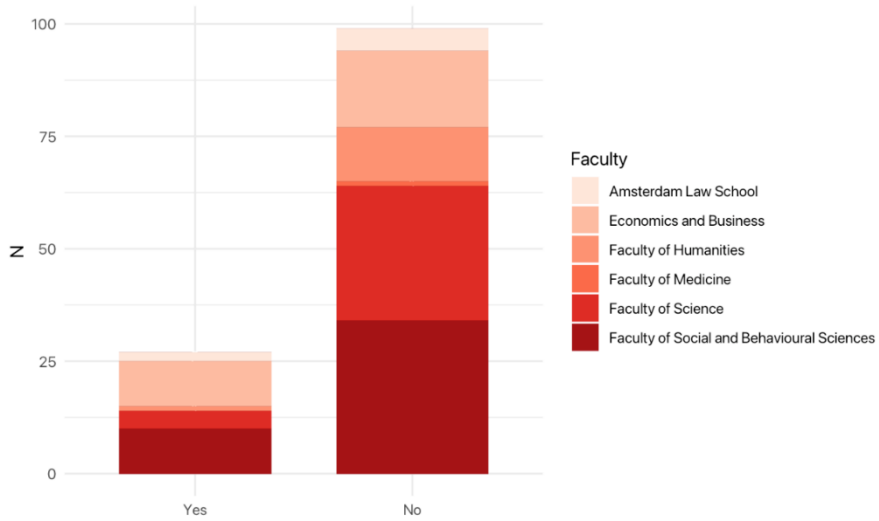

Figure 6. Percentage of researchers that had to pay a fee for accessing data

Respondents who reported to have faced certain restrictions or conditions when trying to obtain data from third parties had the opportunity to provide more details. The most frequently encountered conditions related to **how the data could be used and published** by researchers. More specifically, UvA researchers noted that the third party only provided them with the data when they would promise to keep the data confidential; to only share the data with a limited number of parties; to only use the data for academic/non-commercial purposes; or to list the data-providing third party as co-author in any publications.

One respondent flagged that in their opinion, the confidentiality requirement runs against open science standards of replicability (in Dutch):

I explained that in order to publish scientific research, I must be able to hand over the data to reviewers or other people who would want to replicate my analyses. They didn't care about that. So, in the end I had to cancel this research project because of this.[27]

Some researchers also reported legal issues, most notably **complicated data protection and/or intellectual property protection clauses in contracts** and **doubts as to the legality of web scraping**. Other issues faced by researchers relate to technical issues or requirements, such as **API constraints** or **on-site access only**, and **high financial costs**. Many of these issues are further elaborated on in the interview analysis in Chapter 3 below.

In certain cases, researchers simply **failed to obtain** (all of) the data they needed from the third party (34.7%). There are many reasons explaining why data access is partially or fully restricted. From the six predefined clusters of reasons for (partial) inaccessibility, respondents selected 'legal' reasons the most:

Legal (e.g., GDPR compliance, intellectual property)	24%
Other	18%
Organisational (e.g., issues identifying relevant person/department to share data)	15%
Technical (e.g., too complicated)	13%
Economic (e.g., too expensive)	12%
Institutional (e.g., restrictive faculty/university policies)	9%
Methodological (e.g., changing file formats and lack of replicability)	9%

Table 1. Reasons for not acquiring access to third-party data

When the experiences of respondents did not fit any of the six predefined categories (18%), they could indicate which other reason(s) for restricting access to data they were confronted with. Among these, the most commonly mentioned reason was a **lack of responsiveness** from the third party (e.g., the party stopped responding to emails). Other reasons included a **lack of personal connections** with the third-party organisation, the data being **too sensitive**, or **insufficient technical know-how and capacity of the third party** to retrieve and/or provide the data.

Apart from identifying the (clusters of) reasons that may explain the inaccessibility of research data, respondents were also asked to define whom they considered, generally speaking, to be the **main responsible actor** for not being able to obtain the respective data. Even if respondents could select multiple answers, it was still clear that the **third party** was considered the actor most responsible (64%). However, respondents sometimes also deemed responsible the faculty/university (9%), their research group (6%) or themselves (6%). Among the remaining actors deemed responsible, respondents explained very project-specific actors, for example content creators on a social media platform that removed their respective content, thus preventing them from being studied. One respondent also indicated that there was no particular actor(s) responsible, and the inaccessibility was rather a product of a system failure:

The systems. Everyone had good intentions, but linking the systems and bringing the data together meant that some data could not be obtained. Furthermore, there were various interpretations of the GDPR legislation, which complicated and delayed the process.[28]

[27] Original answer in Dutch: "*Ik heb uitgelegd dat ik voor het publiceren van wetenschappelijk onderzoek in staat moet zijn om de data te overhandigen aan reviewers of andere mensen die mijn analyses zouden willen repliceren. Daar hadden ze geen boodschap aan. Uiteindelijk heb ik dus hierdoor dit onderzoeksproject moeten afblazen.*"

[28] Original answer in Dutch: "*De systemen. Iedereen had goede intenties, maar de systemen koppelen, de data bij elkaar brengen maakten dat sommige data niet kon worden verkregen. Verder waren er diverse interpretaties van de AVG-wetgeving en die bemoeilijkten en vertraagden het proces.*"

Most researchers that faced access restrictions found these restrictions **illegitimate** or were **neutral** about them.

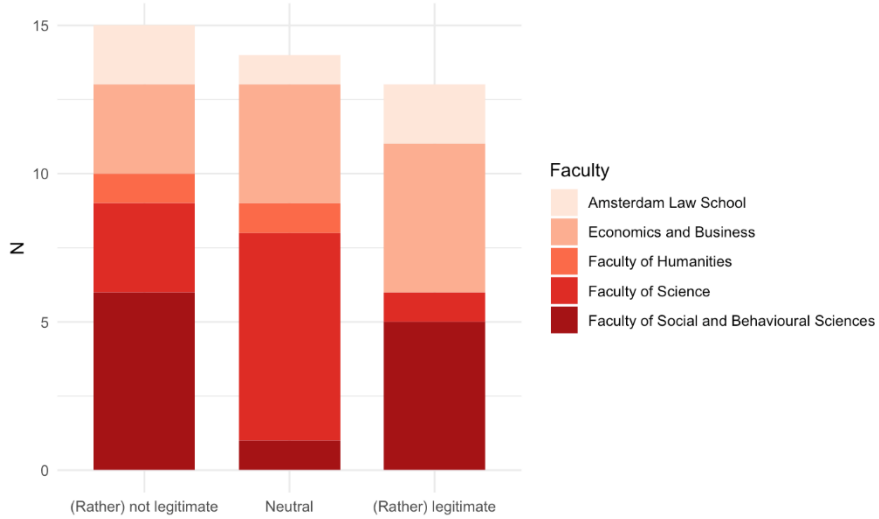

Figure 7. Perceived legitimacy of data access restrictions

Respondents considered the inaccessibility of data illegitimate for several reasons. For example, researchers that use GDPR data access rights in data donation projects, complained that third parties often only provide part of the data they are legally required to provide. Other researchers were frustrated by unwarranted non-cooperative attitudes by third parties (reflected by e.g., convoluted, unreasonable or delayed responses) or by their ERB[29] (e.g., reflected by the failure to provide constructive feedback on how to improve research proposals). As explained by one respondent who deemed the university/faculty mainly responsible for inaccessibility issues in a recent project:

> *Due to many organisational layers, the process has been taking months so far and we still do not have approval to obtain the data.*

The open answers show that many researchers are uncertain when it comes to the legal status of some data sets, notably regarding legal claims researchers may have and the duties third parties may or may not have to share the data.

Importantly, some data access-restrictions were considered (rather) **legitimate** by respondents. **Publication restrictions** were one of them, which can be explained in light of the investments and efforts a third party may have put in the collection/generation of the data, or the fact that data sets are of such nature that they should be kept confidential (e.g., for security, privacy, and strategic/commercial reasons). Some researchers also considered it legitimate that only aggregated data was made available to protect privacy.

2.5 Impact of restrictions and conditions and institutional support

The **majority of researchers** who had recently engaged in a research project which relied on third-party data were able to **obtain the data they needed (65.3%).** For **the 34.7% who did not obtain** all the data they needed, a large group reported that the inaccessibility **impacted their research** (70.7%).

[29] Ethical Review Board.

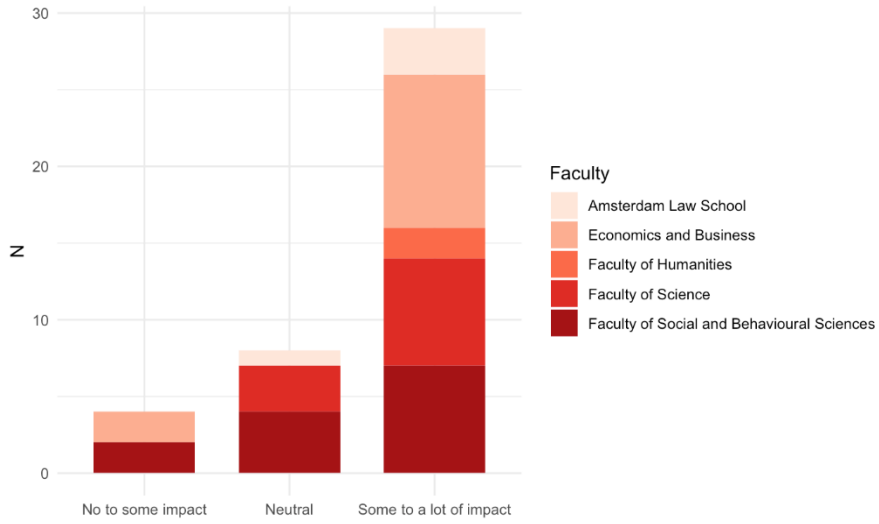

Figure 8. Impact of restrictions on research

In 30% of the cases where researchers experienced impact, they claimed that the lack of access had led them to compromise on their research goals (e.g., they had to adapt research question or analyses). In 22% of cases, the research had even become impossible altogether. In 24% of cases, researchers were able to allocate extra resources or otherwise made extra efforts to access the data after all.

I had to compromise on the research goals	30.1%
I made extra efforts and dedicated extra resources to obtain the data	24%
Research was not possible without the data	21.2%
I found a different way to obtain the same/similar data	13.7%
Other	11%

Table 2. Impact of inaccessibility of data on research projects

When asked to specify how exactly they adjusted their research project to the (partial) inaccessibility of data, researchers listed various strategies, which were often very context-specific. In some cases, an adjustment of the scope and ambitions of the research project was sufficient, but in other cases, vigorous negotiations had to be carried out (for instance with regard to co-authorship) and trust had to (re)built.

Finally, respondents were asked to what extent they considered the university's legal department, their faculty's ethics review board and/or data stewards to be useful sources of **help and advice**. Overall, respondents appeared to be either **neutral** or **rather unsatisfied** with the help they received (M = 2.73, SD = 0.92, 1-5 scale).

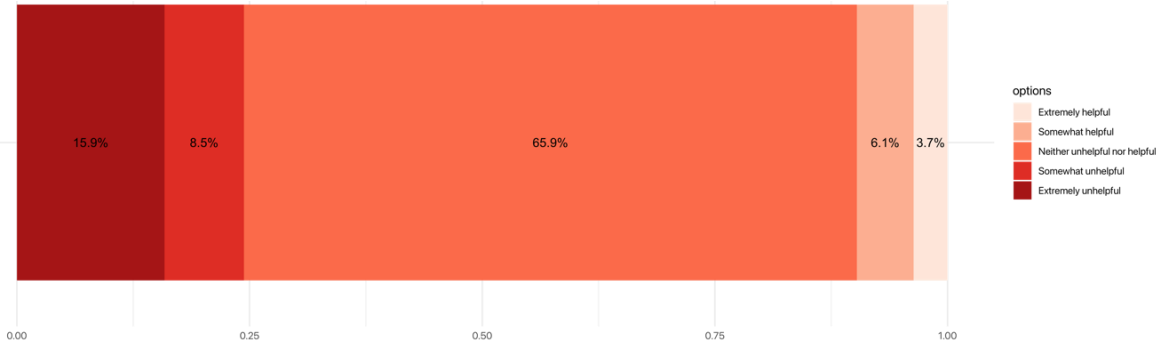

Figure 9. Perceived helpfulness of the university's legal department

3. Interview analysis

The survey was followed by a series of interviews held with researchers, including members of ethics review boards (ERBs), and data stewards working at the UvA.[30] During the interviews, it soon became clear that difficulties experienced by researchers do not only concern (the lack of) *access* to third-party data (section 3.1) but also, relatedly, the permissible *use* of third-party data when access to data is provided (section 3.2). While the interviewees also shared some positive experiences, there are certainly matters that can be improved to promote better access to, and use of, third-party data for research purposes. A lot of what is needed from the university seems to boil down to increased 'institutional support' and assistance in data management-matters. Although there are already forms of institutional support available – data stewards being one of them – it does not always prove sufficient at the moment. Researchers as well as data stewards' institutional support needs are discussed in section 3.3.

3.1 Data access challenges

Access challenges relate to the various stages of the data access phase, which includes, among other things, contacting third parties for data, negotiating terms of access, and acquiring data or a key to access the data elsewhere. In this section we distinguish between legal, social/reputational, technical, communicational, financial, and other (miscellaneous) obstacles.

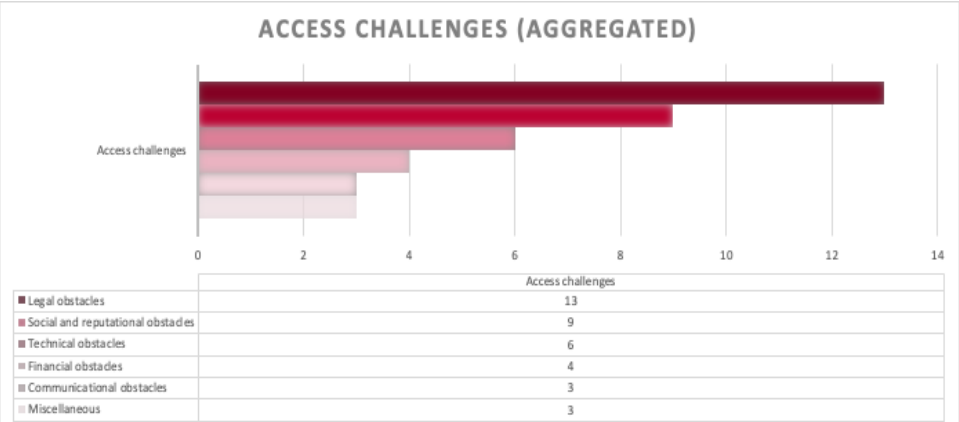

Figure 10. Overview of data access challenges ("aggregated" means the cumulative of researchers and data stewards)

3.1.1 Legal obstacles

Ten researchers and all (four) data stewards indicated that they had faced some type of 'legal' obstacle in the past when trying to access third-party data for research purposes. With 'legal', we mean that the data access challenge pertains to applicable law, for example because legislation imposes certain compliance obligations on third parties or on researchers themselves, or because regulation is considered complex or unclear. Based on the interview responses, we grouped the legal challenges that were flagged into three categories: obstacles relating to (1) data protection law, (2) terms of service, and (3) legal procedures more generally. As regards the challenges that may arise from terms of service, we particularly look at the practice of "web scraping".

[30] Please note that in this section, we use the pronouns 'they', 'them' and 'their' to refer to individual interviewees.

1. Data protection law

Data protection law, which in the EU is mainly codified in the General Data Protection Regulation (GDPR),[31] aims to ensure that the processing of 'personal data'[32] is compatible with fundamental rights, in particular the right to protection of personal data.[33] Naturally, data access challenges related to data protection law are most relevant to researchers whose research (partially) involves personal data. Examples of rules enshrined in data protection law that could affect researchers' access to data are the requirements to obtain and/or demonstrate consent from data subjects for the secondary processing of personal data ("without (additional) informed consent, the data will not be shared") and to implement appropriate security measures ("when the security of the processing cannot be guaranteed, the data will not be shared"). Sometimes, data protection law *in itself* can be an obstacle to researchers' data access in that third parties are only willing to provide researchers with access if the data are anonymised, thus rendering data protection law inapplicable. These and other 'GDPR-issues' were mentioned relatively frequently during the interviews, namely by **eight researchers and all data stewards.** Some of the issues that were mentioned, however, rather seemed to concern the *use* of data – and not so much to the process of obtaining data *access* – and will therefore be discussed in section 3.2.

One researcher described a case in which another Dutch university decided not to provide certain data to a UvA research team because the anonymity of the research subjects – a very small group of patients with very specific characteristics – could not be safeguarded. This researcher also noted that arguments of data protection and commercial reputation often go hand in hand. A commercial company that the researcher had approached in the past refused to provide them with certain client-related data, not only because the clients in question had not given their consent for such data sharing and the sharing would therefore violate the GDPR, but also because the company was afraid that clients would recognise themselves in the research paper and sue the company for breach of confidentiality/privacy law or complain about the company on social media. Another researcher brought up a research project in which the team wanted to re-purpose student data generated by the electronic learning management platform 'Canvas' to study student learning. This was deemed impossible at the time, as students had never consented to the use of their personal data for that purpose. The researcher regretted not having been granted the data needed to conduct the research, especially because it could have benefited the students. Another researcher described how a lack of security guarantees inhibited UvA-researchers' access to a video database maintained outside of the EU. The international collaboration was based on the agreement that researchers from all over the world could access the database, provided that they would also donate their own video recordings to the database. However, the entity managing the database was not able or willing to meet the high security standards required by EU data protection law. Considering the privacy-sensitivity of their data, the UvA-researchers decided not to donate their video recordings and, by implication, gave up their access to that database.

Importantly, among researchers whose access to third-party data was not restricted for reasons of privacy and data protection, some still considered GDPR compliance and the accompanying "ticking [of] boxes" by the UvA privacy officer as a "nuisance" that "slows down" the process of data access. This was also underlined by a data steward who explained that privacy lawyers "can be very strict" and ask "many questions". For some researchers this can be very uncomfortable, in particular when they have built a strong relationship of trust with a third party. At the same time, some researchers also see the good qualities of the GDPR, for example because they think anonymity could lead to access to *more* data; third-party providers

[31] General Data Protection Regulation (GDPR).

[32] Personal data are "any information relating to an identified or identifiable natural person", see Article 4(1) GDPR.

[33] See recital 2 GDPR.

(and data subjects) may be tempted to share more information when data subjects are fully shielded. One of the data stewards, however, pointed out that the process of data anonymisation can be very tricky. Interviews, for instance, may contain a lot of personal and sensitive information which can be traced back to individuals even when names are removed from the transcripts. Furthermore, one researcher described that it is simply too time-consuming to ask every patient individually for their informed consent before using their personal data for research purposes.

Notably, not all researchers may be (fully) aware of any potential data protection issues when it comes to the gathering of data for their research, for instance because they rarely work with personal data. In other departments, people are aware of the importance of GDPR-compliance when obtaining personal data, but expert legal knowledge on this topic is virtually absent. The issue of (legal) expertise is further discussed in section 3.3.1.

2. Terms of service: the case of web scraping

Apart from formal (data protection) regulation, the terms of service of (private sector) data providers may also result in data access challenges. An interesting example mentioned by **four researchers and two data stewards** is the contractual prohibition of "web scraping" laid down in the terms of service of some online platforms.[34] Web scraping refers to the automated extraction or copying of publicly available information online, a popular research strategy especially used in the social sciences.[35] Some online platforms explicitly include in their terms of service a ban on web scraping, while others allow for it under certain conditions, and again others do not have a clear policy on it. One researcher described that platforms' terms and conditions are often vague, thus making it difficult to understand what exactly is permissible. In this regard, the researcher mentioned a research project at the UvA which developed a database of Dutch-language tweets. According to the interviewed researcher, the project interpreted Twitter's terms of service in such a way "that they cannot show the tweets – the data itself – to the researcher in their tool interface, because that would constitute sharing", which "really hampers what you can do with that tool".[36] In their opinion, the project is "being overly prudent", but at the same time, it shows "that the terms, at least for the people who deal with them, are (…) not clear enough". The same researcher also stated that they did not know what the "worth" [weight, importance] of terms and conditions is, seemingly suggesting that it is unclear to them to what extent researchers should abide by the platforms' established rules. A data steward underlined that "uncertainties" about what a platform permits "can create some obstacles".

Adding to this uncertainty is also the doubtful legal status of web scraping regardless of platforms' terms of service. Last year, an appeal court in the United States decided in a case between HiQ Labs and LinkedIn that web scraping, even when conducted in violation of the terms of use of a website, cannot establish liability under the US Computer Fraud and Abuse Act, thus essentially 'legalizing' the practice.[37] In the EU, however, such clarification has not yet been provided, even though some have argued the DSA has implicitly legitimised web scraping in Article 40(12), provided that the research is privacy-compliant.[38]

---

[34] See for example the User Agreement of LinkedIn, under '8.2 Don'ts', <https://www.linkedin.com/legal/user-agreement> and the Terms of Service of Twitter, under '4. Using the Services', <https://twitter.com/en/tos>.

[35] Luscombe, Dick and Walby 2022.

[36] The researcher added that "the way the tool works, is that you make a query and then you can see some graph or something, but you cannot look at individual tweets other than maybe having a link that then links back into Twitter's interface" (following from one of the interviews).

[37] HIQ Labs vs. LinkedIn Corporation.

[38] See notably the submissions to the European Commission's call for evidence on the planned Delegated Regulation on data access provided for in the Digital Services Act. See: European Commission Delegated Act Article 40 DSA webpage <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en>. See also Pershan 2023.

One of the researchers who is also a member of their faculty's ethics review board (ERB), noted in their capacity as ERB-member that when a website's terms does explicitly not allow for scraping, the faculty's ERB "would not allow it either", since they "do not want to go against the rules of the owner of the data". If, however, there is nothing on the website indicating that scraping is *not* accepted by the platform, the ERB would likely allow the researcher to use scraping tools, unless "people want to do really weird things" such as scraping Facebook pictures and uploading them online. A data steward seemed to suggest that UvA data stewards do not necessarily discourage researchers to perform web scraping. Instead, they have been trying to create "a practical policy" that takes account of the "risks potentially involved". According to the data steward, it is for the legal department of the UvA to determine "what are the hard no[go]s".

Taking a step back, whether platform terms of service actually stop researchers from web scraping is questionable, since, in the words of a researcher, "everyone scrapes social media". At the same time, the researcher themself indicated that they deliberately did not use their self-built scraping tool in the virtual environment of a specific standard-setting body which has prohibited scraping in its terms of service, thereby stating that this "would be a violation of their terms of service" and that "of course" they will "not use that data for [their] research".

3. Legal procedures

The last category of legal access challenges mentioned during the interviews concerns legal-procedural issues in a broad sense. For example, one researcher flagged that drafting a contract on data sharing can be a **lengthy process**. When a data-providing party wants things to be specified in a certain way, and the data stewards or legal affairs department of the UvA want things specified in a different way, the access-process can be slowed down. This back-and-forth – which according to a data steward can sometimes take two or three months – can be "frustrating" for researchers, as they want to get started at some point but are hindered due to ongoing legal discussions. This is especially the case when multiple parties are involved in the drafting process. Related to this, is the **often-weak negotiating position of researchers** in relation to powerful companies. One researcher described that sometimes access to data is provided by a third party, but that the terms of access/use imposed on the research team are non-negotiable. Not signing the contract then means: no data.

> *"The thing is though, the contract is made by* [company]. *Our legal department cannot change anything, so it is 'take it or leave it'.* [The legal department] *check it, but if they want to change a sentence then* [the company] *do not allow it. They wanted to change it once, but that was not a really big thing, but they thought "Oh maybe we can re-word this", but that was not allowed. So yeah, it is checked by our legal department at least, and then signed by them and me."*

3.1.2 Social and reputational obstacles

During the interviews, **nine researchers** pointed at what we consider 'social' obstacles, i.e., obstacles that relate to how humans behave towards one another and how they want to be perceived. One researcher observed that commercial businesses seem to be more protective of their data, which, as (implicitly) suggested by another researcher, could potentially be explained by the fact that they have a commercial reputation to uphold: if a company's clients do not like the outcomes of the research that was built on their personal data, or if the clients had not given their consent to share their data for research purposes in the first place, they may perhaps decide to sue the company or complain about it (online) in public. Moreover, it was noted by another researcher that companies may fear that they will be "exposed" if they share certain

data. They mentioned the story of an American researcher who had found out that a company providing data relating to its products that are used for the selection of personnel (e.g., intelligence tests) only shared data showing that its products were of good quality, and withheld data demonstrating the opposite.

Some third parties may have a certain distrust of (academic) researchers:

"There is a lot of scepsis against researchers of any kind, especially if they're based at the law faculty, I must admit."

The researcher quoted above described how in the food sector, some people fear that academics will "steal information from farmers and sell [it] for a seed patent or something". In those cases, researchers really need to convince third-party data providers of why they need the data, which can be a very time- and energy-consuming process. At the same time, two other researchers observed that university affiliation can also *benefit* a researcher's position in terms of data access, as people might be more prone to provide access to data for academic research purposes than for commercial use.

While it could indeed be true that some people feel an aversion to academics specifically, people may in general be hesitant to share 'their' data with people they do not know. As one researcher described:

"Why would someone who does not know me give me some kind of document?"

The fact that people prefer to provide data access to researchers or research organisations they know, could potentially benefit bigger and/or more prestigious research institutions. Researchers who are not affiliated with such institutions could be in a more disadvantaged position.. For example, one researcher mentioned that Twitter did not grant a certain researcher access to a dataset because Twitter did not consider the institution the researcher had listed in their request as a "viable research institution", basically saying "no, that is not a [research] university". Another researcher emphasised the importance of contacting companies through an institutional email-address in order to sign up for an account, as these companies often wish to confirm that the researcher in question is a bona fide researcher.

Relatedly, two researchers remarked how having a network is essential to obtain access to data. Data providers may be more willing to share data with researchers they have worked with before. Plus, the process of accessing data or information may run more smoothly through existing informal networks. Younger or less-experienced researchers who are not yet part of those networks could therefore be disadvantaged. One researcher, however, was of the opinion that the level of experience or seniority of a researcher does not seem to really affect data access. Another researcher had different experiences though, stating that especially student-researchers tend to have a hard time gaining access to information from large companies.

One researcher further implied that nationality could affect data access, suggesting that Twitter seemingly favours US institutions over institutions from other countries.

Lastly, staff changes in third-party organisations could also affect data access. One researcher noted that when they asked for an updated version of a document that they had received from an employee of an organisation earlier, the colleagues of that employee, who had in the meantime left the organisation, were not willing to provide the researcher with the updated version of the document.

3.1.3 Technical obstacles

Five researchers and one data steward mentioned that they had encountered 'technical' issues during data access processes in the past. For example, one researcher explained how they had to download a data batch item-by-item from a third-party server to the local server of their research institute, which turned out to be quite a hassle. The researcher emphasised that the efficiency and simplicity of data access really

depends how a third party has stored its data. Two other researchers were particularly concerned about frequent changes of APIs.[39] When an API changes, the (structure of the) data extracted through the API also become "different" in a way. In those cases, different software may be needed to analyse the data, thus burdening researchers with the task to keep their tools for data analysis up to date. If Facebook, for instance, changes something in its functioning and researchers' tools are not adapted to that, they have a problem. Plus, when data formats change, it is much harder to compare data and draw "longitudinal conclusions". A third technical issue mentioned during the interviews is the fact that some third parties only make "unstructured data" available, which can make data analysis a lot harder. The researcher who flagged this practice considered it as "a strategy against transparency". At the same time, when third-party data are pre-processed or aggregated, it can be difficult for researchers to assess the value of the data and to sell their research to a peer-reviewed journal, since the underlying data are essentially a "black box". According to another researcher, data "always appear better than they really are" and "oftentimes fail to meet scientific standards […] for how they should be processed, […] recorded and so forth". Finally, it was noted that (platform) datasets can be so large that researchers need to run their laptops for hours, or even days, to download the data.

3.1.4 Financial obstacles

Two researchers and two data stewards mentioned that they had encountered 'financial' challenges to data access. Financial obstacles relate to costs, such as costs incurred for buying data or complying with contracts. Two researchers and one data steward flagged that they sometimes have to pay for data. While in some cases, asking for remuneration in exchange for datasets may be understandable and justifiable – considering the third-party investments sometimes needed to generate and store those data – such remuneration may also constitute a significant obstacle to academic research. One researcher described a situation in which their team either had to pay €20.000,- to a data-providing organisation *or* mention the employees of the data provider as co-author in any manuscripts produced during the research. Since the research group did not have €20.000,- available, they went with the second option, which felt somewhat "unethical" given the fact that their 'co-author' was "not really making a contribution to the academic work" other than just providing the underlying data. Another researcher noted that within their faculty and research group, external datasets are only purchased if their budget allows them to do so. One of the data stewards was of the opinion that researchers should sometimes be more careful with buying data, as some datasets could be problematic from an ethics perspective (for instance, when social media companies *think* they sell anonymous data while in fact the data are not truly anonymous).

3.1.5 Communicational obstacles

Three researchers indicated that they have faced what we call 'communicational' obstacles in the process of acquiring access to data, for instance when trying to find the right entry to data – i.e., the right person who can provide data – and when convincing contact persons to make the requested data available. As one researcher described, it had been very difficult for one of their colleagues to find and/or contact interviewees for their research. The researcher also experienced for themselves that third parties may just not respond to emails by which they request access to certain documents that are not publicly available. Another researcher specifically mentioned the effects of automated communication. In order to obtain data from Twitter, for example, academic researchers need to file a formal request. However, the requests are sometimes denied

[39] API stands for Application Programming Interface, which is, in short, a tool that ensures that two software applications can communicate with each other. Researchers sometimes use APIs to gather data from digital infrastructures.

by means of standard messages that do not contain clear explanations on why the requests have been denied. The researcher described how they felt that they were not talking to a human being and that the process was likely "outsourced to some click workers". However, even when dealing with humans, communication can be cumbersome and slow. As a researcher noted, they often need to "rattle the door a bit", "make a lot of noise" and "use [their] connections" to gain access to data held by governance bodies.

3.1.6 Miscellaneous

The following challenges do not necessarily fit within the categories as defined above and were mentioned only occasionally during the interviews, but are nevertheless worth mentioning to get a better idea of the different kinds of data access challenges researchers may see themselves confronted with:

- **Data simply unavailable** – One researcher pointed out that some data are simply not made publicly available. In some countries, for instance, not all court decisions and other court documents are published openly.
- **Time** – A researcher who (on behalf of a student) wanted to get a hold on certain clinical practice and client-related data held by a private company was told that the sharing thereof would not be in line with data protection law. Alternatively, the researcher would have to talk to the clients themselves. The company was however of the opinion that all this "would take too much time" and decided to refuse the data access request. Another researcher pointed out that collecting data from clients and contacting scholars whose papers are not publicly accessible can be very time-consuming. Lastly, a researcher pointed out that the process of data transfer can be very slow, for example when data are provided in batches rather than in one instance.

3.1.7 Differences between public sector and private sector

The number of interviews we conducted is too low to draw firm conclusions on whether certain data access challenges tend to occur more often for public sector data than for private sector data and vice-versa. Challenges are clearly very research context-dependent. One researcher commented, for example, that private governance bodies "happily publish" information, while multilateral and governmental standards bodies do not, while "you would expect that [governments] would have a higher commitment to openness". Another researcher, however, pointed out that in their experience, it is commercial businesses that seem to be more protective of their data. Lastly, a data steward noted that commercial parties often do not allow researchers to publish their datasets and generally require a payment, whereas NGOs tend to be more open towards publishing their datasets, provided that researchers use them for scientific purposes only.

3.2 Data use challenges

While data *access* – the "querying or retrieving data"[40] from an external source – is the main focus of this report, the interviewees also mentioned issues that relate to the following phase of the research lifecycle, i.e., of the subsequent *use* of the data: e.g., viewing, analysing, organising and aggregating data. Sometimes, however, data access and data use challenges are closely intertwined. A few of the use challenges that were mentioned in the interviews are therefore briefly discussed below.

[40] Based on OECD 2021.

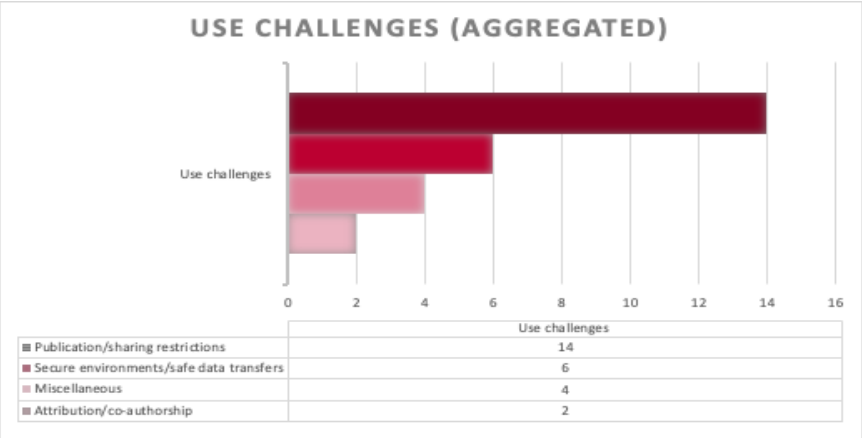

Figure 11. Overview of data use challenges ("aggregated" means the cumulative of researchers and data stewards)

3.2.1 Publication and sharing restrictions

Eleven researchers and all data stewards indicated that their past access to third-party data had been subjected to restrictions on the further sharing and/or publication of the data, or more generally, that such restrictions are common in their respective research fields.

Some researchers mentioned that they had to formally sign a non-disclosure agreement (NDA) or otherwise contractually agree to restrictions or conditions on data sharing.[41] In other cases, researchers said it was more *informally* agreed or implied that they would use the third-party data only for a specific research project and would not share them any further without prior consultation (gentlemen's agreement). The researcher who had to sign a NDA said that in their opinion, a contractual prohibition to publish the data that underpin research outputs "does not sit very well" with the principles of 'open science' and 'transparency'. The global open science-movement aims to make scientific publications and research data publicly available as much as possible;[42] restrictions on the sharing of research data thus interfere with this goal. Additionally, the researcher pointed out that a contractual prohibition on data sharing may obstruct the peer-review process, because without data, a reviewer is unable to track a researcher's analyses and therefore (in some cases) to assess the validity and value of the publication. This issue remains present when researchers are merely allowed to publish aggregated datasets.

Two interviewees flagged that some third-party data holders do allow researchers to publish the data, but only after the data provider itself has first used or published the data:

> *"They are actually working on their own papers, so they are doing their own research […]. For example, we would be working on* [research topic], *they are also working on* [research topic], *and then we ask "Oh, can we combine the two datasets?", then they sometimes say, "that's fine, but first we want to publish our own results."*

> *"They do not want to freely share that data, because there is probably a lot of information in there. So, before they publish that data, they want to investigate every possible thing that is in there, and then they share that data for others to use."*

[41] For example, the terms of service of social media companies often include restrictions on the sharing of data obtained through their APIs.

[42] See UNESCO 2021 on UNESCO's definition of 'open science'. More information on the EU's open science policy can be found on the European Commission Open Science webpage <https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en>.

Sometimes, it is rather *implied* in the researcher-third-party-relationship that the third party also benefits from the data sharing, for example because the researcher has agreed to use the data to develop policy proposals in the interest of the third party:

> *"…I will not just be processing this data […] for just publishing a paper that will help me to advance further in my career, but the point was that we* [would use] *this information as a background for the workshops that we would develop, at which they* [the data provider] *would also be present. And the idea was to kind of work towards some policy proposals together. So that was kind of the agreement, in a sense that there was something for them in it."*

3.2.2 Requirement of access and use in secure environments/safe data transfers

Five researchers and one data steward described that data providers may require researchers to access and use sensitive (personal) data stored on the third party's own servers or in an otherwise 'secured environment', which can be online or offline. One researcher noted that in the field of economics and business it is "very common" that "the data […] never leave the company's servers" and that researchers can never really download the data" and "have to work within that company". Other researchers indicated that for their specific projects, working in secured environments was not necessary but that they did have to pay special attention to the "safe transfer" of data from the third-party's storage to their own (secured) servers. One researcher explained that certain datasets they wished to access were temporarily stored on a separate server controlled by the third party, from which the researcher could then download the data to the server of their institute. While previous datasets had been shared with him via e-mail, these sensitive data – relating to cancer – were subjected to a more thorough transferring process in order "to be more safe". Another researcher explained that a safe data transfer may also imply that it is prohibited to use OneDrive, for instance, and that only SurfDrive or a similar drive is deemed safe.

3.2.3 Requirement of attribution/co-authorship

Two researchers indicated that they, or their colleagues, had been required in the past to include their third-party data provider as co-author in the scientific publication for which the data were used. As already mentioned in section 3.1.4 in the context of financial obstacles to data access, one researcher was once asked to either pay a large sum of money in exchange for the data *or* list the data provider as co-author of the publication. The researcher in question justified the requirement of co-authorship for themselves by arguing that their research could not have been conducted without the data. In this case, a compromise was found by adding a footnote at the bottom of the manuscript explaining in detail what each author's contribution to the scientific article had been, and by including the contribution by the third party, which consisted of data provision. Another researcher mentioned that a commercial party they closely work with usually requests co-authorship in cases where the data analysis is conducted by the company.

3.2.4 Miscellaneous

Some of the data use challenges mentioned during the interviews do not necessarily fit within one of the described categories above, but are nevertheless worth mentioning:

- **Other specific third-party restrictions and conditions** mentioned during the interviews were, for instance, Twitter's prohibition on the **use of sensitive personal data** such as political opinions or health data, unless the data are aggregated (inspired by data protection law), as well as Twitter and

Tumblr's requirements that acquired data must be **erased** after a certain period of time (also seemingly inspired by data protection law).[43] Notably, strict requirements on data erasure may limit the utility of the data for research purposes, as researchers have little time to properly analyse that data and compare datasets in a later stage of the research project.

- Apart from use restrictions and conditions imposed by third-party data providers, sometimes the **data themselves may be of such nature** that they can limit the successful use thereof. One researcher, for example, flagged that **language** can be an issue: many documents they find potentially interesting for their research are written in a language they do not understand, and finding data or documents on a foreign language website can be challenging.
- Finally, another researcher noted that **nationality** can be a limiting factor. The researcher in question indicated that they would very much like to access rich datasets collected and stored in Scandinavia, but that these are "hard to get if you are not from Scandinavia".

3.3 Need for institutional support

During the interviews, many researchers and data stewards indicated that they have certain needs to be met in order to overcome data access and use challenges. In this section we briefly set out those needs, which are all grouped under the broader need for 'institutional support' to deal with research projects that rely on third-party data. Importantly, several interviewees also mentioned positive experiences with existing processes, which are highlighted too.

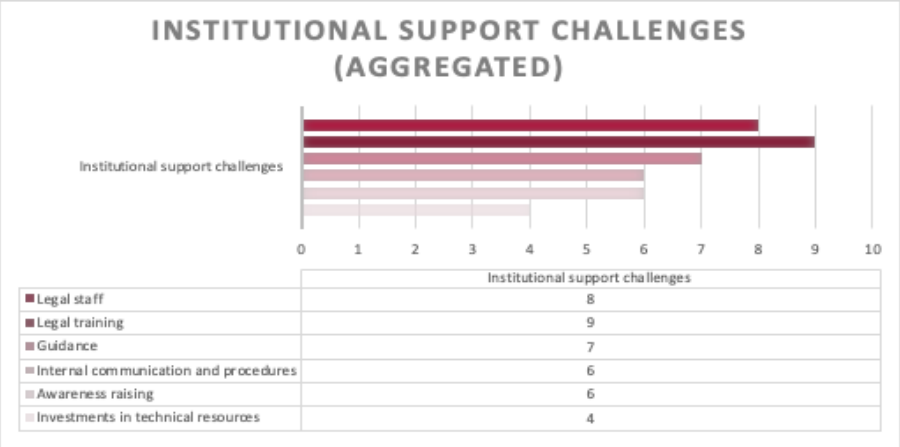

Figure 12. Overview of institutional support challenges ("aggregated" means the cumulative of researchers and data stewards)

3.3.1 Staff, knowledge, and expertise

As explained in sections 3.1 and 3.2, legal obligations and contractual requirements were frequently mentioned by both researchers and data stewards as obstacles to data access and data use. In this regard, some of the interviewees expressed the need for additional legal assistance from the UvA to help them address these challenges. Others indicated that it would be helpful if they could increase their own legal knowledge and expertise.

[43] According to the interviewed researcher, Twitter demands that data recipients regularly check Twitter's database for deleted tweets and delete these tweets in their own databases accordingly. Tumblr demands data recipients to delete the data after three days.

1. Additional legal staff

While some of the researchers and data stewards expressed that they are already quite content with the helpful staff that is available at the university (e.g., technical staff), this does not seem to be sufficient when it comes to assistance with *legal* issues.

Four researchers and three data stewards signified the need for additional legal staff at the university level as – from their perspective – the legal affairs department seems to be understaffed. One data steward observed that UvA data stewards essentially have "one lawyer" they can approach and who can check data sharing agreements, change provisions and "give a final OK on such an agreement". According to the data steward, there is "a lack of people that […] can support [them]". This was also underlined by a researcher, who stated that there are currently not enough people at the university to deal with every department's legal issues in a timely manner, and that, for example, "it takes very long for them to sign [a contract]").

Adequate legal support is important since –of course in some faculties more than in others – data protection and contractual issues may arise on a daily basis and researchers do not always possess the legal knowledge and experience to deal with them. For example, a researcher described that researchers in their faculty they "have no clue" how to prepare and assess non-disclosure agreements. Two data stewards would appreciate additional assistance from privacy lawyers that are specifically allocated to their faculties. One of them explained that in their faculty, many research projects involve the processing of personal data for which data sharing agreements must be drafted. The help of a legally trained colleague is therefore much needed. At the same time, it was emphasised by a researcher that legal questions concerning data should be addressed by someone who also knows a lot about research practice, for example a member of the faculty's Ethics Committee, who considers both the research aspect *and* the legal aspect. Another researcher suggested that the university should facilitate access to professional law firms to assist researchers with complex legal compliance issues, that is, in case the university itself is not well-equipped to do so.

2. Legal training for data stewards (and researchers)

Related to the need for additional – internal or external – legal support is the need for better legal knowledge and expertise among data stewards, and to a lesser extent, among researchers themselves. **Four researchers and three data stewards** explicitly mentioned the importance of legal knowledge among data stewards to adequately help researchers with legal issues. The problem, however, is that data stewards are not always (fully) equipped to provide the legal support researchers require, as they are often not legally trained. Indeed, one data steward reported that months into their appointment, they had not received any structured training on "very sensitive and complicated issues", including on legal issues. As an example of such legal issues, the data steward mentioned the new EU rules on text and data mining for research purposes in relation to copyright, which the data steward had to look up themselves and had to inform the faculty's researchers about, even though the data steward was not an expert on the topic. The data steward considered themselves "basically self-taught" and emphasised that looking things up on a case-by-case basis is "not an ideal way of doing things", which is the reason why they would like to see something "a little bit more systematic when it comes to support, legal support, but also just training". It thus appears that the weekly appointments with the university's lawyers – as one of the other data stewards noted – may not be sufficient and that additional (legal) training is desirable.

Similarly, according to the interviewees it would also be helpful if researchers *themselves* gain a better understanding of legal issues that could arise during their research. Of course, as a data steward observed, it is not a researcher's job to be knowledgeable about all the fine prints in terms of service, but it is important

that they come to data stewards for advice about potential problems and avoid that "things fly a bit under the radar". Another data steward described that at the UvA, GDPR workshops has been organised for data stewards, but not for researchers. A researcher suggested that it would be a good idea to organise meetings at university or faculty level aimed at making researchers aware of relevant data-legislation emerging from the European Union. Especially for researchers who are involved in large research projects that run for multiple years, it is important that they are aware of relevant legal developments. Two researchers also proposed the organisations of workshops for thesis students and PhD courses in the domain of data risk management and data handling.

3.3.2 Guidance on data access and use-questions

Four researchers and three data stewards expressed the need for institutional guidance on (legal) data-related questions. One of the researchers had the impression that the data stewards at the UvA and privacy officers working for third-party companies "all have different rules" about the "do's and don'ts [of] data collection". The researcher would thus appreciate it if these do's and don'ts were "stated somewhere", clarifying for researchers what is "really not okay" and what can be done under which conditions. Similarly, another researcher mentioned that despite the support provided by the UvA's data method centre, "some upfront guidance" would be "really helpful" so that "at the beginning (…) the basics are established" and you do not have to ask for help "every step of the way".

Other interviewees indicated they would like to have *specific* guidance. For example, one researcher said they would like to have more clarity on grey areas such as web scraping. One of the data stewards flagged that they would like the UvA to guide data stewards on the meaning of 'data ownership', which is an important but also controversial concept in research data management practice:

> *"I think in general it is… you can of course agree that the data you will buy are not your data, and the results are your results, that is fine. But sometimes it is more complicated, and the UvA does not really make a decision on what they see as data ownership."*

Another data steward signified that the data stewards at their faculty have taken the initiative together to formalise their approaches, including in regard to accessing third-party data. According to the data steward, it is important to have "some guidance that is translated to the researchers". The data stewards' goal is thus to create a document that covers the entire research data lifecycle, "from the steps of planning and starting data collection all the way to publication and potential re-use and sharing of data". Data access is part of that lifecycle. The document is envisaged to set out, for instance, which steps should be taken to use data from third parties in a way that is compliant with (inter alia) data protection law, IP law and research ethics. This means that data protection officers, IP lawyers and Ethics Committees should be involved in the drafting of the document. According to the data steward, the goal is to put a standard approach into writing so that data stewards and lawyers will not have to answer similar questions repeatedly. In addition to this effort, however, the data steward very much welcomes "any sort of harmonised guidance that has been approved at institutional level", setting out "what the university's approach [is] on this [topic]". A relevant aspect, the data steward emphasised, is what in the opinion of the university is "the accepted risk-level" when it comes to e.g., GDPR-compliance and compliance with third parties' terms of services. Relatedly, another data steward mentioned that a template for the 'informed consent' by data subjects for the sharing of personal data is currently in the making. Lastly, one of the researchers pointed out that the need for clear policy on what is allowed and what not when conducting data-driven research, becomes even more pressing when more people are getting involved in a research project and compliance with legal rules is harder to check.

3.3.3 Better internal communications and streamlined procedures

Another element that, according to some interviewees (**two researchers and all data stewards**), could potentially help tackle data access and use challenges, is better communication between researchers, data stewards and the university's legal affairs department.

One researcher pointed out that the role of 'data steward' has been relatively new at the UvA and that their faculty is "in the process of developing routines" which "is definitely something that needs improvement", because right now, "there is a lot of emailing going back and forth". According to the researcher, it would not necessarily have to be the central university to provide support in this regard, although a "standard way of organising things" would in their opinion certainly help. Indeed, a data steward acknowledged that "faster procedures" are needed but added that data stewards "do not have enough experience yet". Another data steward mentioned in this regard that their faculty is working on an 'informed consent template' to streamline the process for obtaining informed consent from data subjects. It must be noted, however, that there seem to be differences between faculties as to how the current collaboration and communication with data stewards is perceived. One researcher, for example, stated in their capacity as ERB-member that they are "very happy" with the data stewards at their faculty as well as the ways things are organised right now.

One researcher observed that it is rather the university's centralised *legal department* that is the bottleneck. While many data sharing contracts are "literally the same", it takes "a very long time" for legal affairs to sign them, which considerably slows down the data access process (and which, according to the researcher, can be attributed to a lack of legal staff, see section 3.3.1). Shorter lines are therefore preferred. Similarly, a data steward described how they had tried to reach the legal affair department three times in vain, first through the mediation of the faculty's privacy officer and Ethics Committee's secretary, and secondly (absent a response), by emailing the department directly. Because of the lack of reply, the data steward did not try to get in touch again.

Lastly, it was raised by one data steward that communication and collaboration between all the data stewards within the UvA could be improved. It became clear from the interviews that data stewards have little or no formalised contact with colleagues from other faculties, at least not as a group. Sometimes, the data stewards from one faculty join the weekly meetings held by the data stewards of another faculty to exchange knowledge and best practices. This made the respective data stewards realise, for example, that there are quite substantial differences between the faculties when it comes to data-driven research (and the types of data used). However, a data steward working at a different faculty regretted not having meetings with all data stewards on a regularly basis, because the data steward believes that they can certainly benefit from their experiences (including on data access and use challenges).

3.3.4 Awareness-raising efforts

Five interviewees expressed their satisfaction with the increased awareness on the importance of data protection and ethical considerations when conducting data-driven research among scientific staff. For instance, a data steward described how, even though not all researchers may be fully aware of the GDPR, there are nevertheless strong ethical considerations built in the data stewards' respective research institute. A researcher explained that over the last five to 10 years, researchers have become more considerate of ethical questions, for instance as regards data storage.[44]

[44] In 2016, for instance, the UvA brought together a special group who drafted a memo on data governance and on how to deal with data that researchers have about students that may be in commercial systems (following from one of the interviews).

However, **four researchers and two data stewards** also emphasised that there is nevertheless a need to increase the level of awareness among researchers on issues related to data-driven research (not just on data access, but even broader) and on the potential solutions to address them. As mentioned earlier, one researcher suggested that the central university or individual faculties should organise information sessions to update researchers on relevant legal developments. A data steward observed that in the past, GDPR-workshops had been arranged for supporting staff, but not for researchers. One of the other data stewards strongly believed that "a culture change" is necessary, in the sense that researchers must better understand what it means to involve third parties in research projects, and that they sometimes have to be careful buying datasets. A researcher who is also a member of the ethics committee at their faculty argued that "both from the management of the [research] school and from the staff" there should be "much more awareness" of the importance of law and ethics in data management practices, which the researcher thinks could be achieved by campaigning and offering proper training to staff members as well as making available special modules in bachelor's and master's programmes to students. The same researcher stated that "there is more attention to compliance and ticking boxes than actually addressing issues". Related to ethical considerations, another researcher held that researchers "need to get a much better understanding of the ethical implications of their work". Additionally, the researcher stated that the university should provide researchers with tools to engage in human rights impact assessments before they start their research process. Another researcher emphasised the need for more awareness and protocols in cases where people are hired for just a short period of time – student-assistants for example – to handle sensitive data.

3.3.5 Investments in technical resources

Finally, **four researchers** expressed the need for investments in technical infrastructures and computing and software programming expertise in order to adequately process and analyse data. One researcher stated:

> *"Just give me servers.* […] *I am continuously buying my own shit because it is so hard to get it in a uniform way. Just give me resources. Resources in terms of server space, in terms of computing power, but also in terms of like the legal support I got from* [external law firm]*."*

This was underlined by another researcher, who would like to have their "own server" rather than having to work with "Lisa", a cluster computer used by UvA faculties to perform large-scale computations, since there are typically many groups who want to work with it, which results in "waiting lists" (i.e., queues for analyses to be run). Another researcher specifically noted that the proper storage and archiving of research data is problematic in their faculty, since there is "no central storage system", or at least, "not an easy one". Again another researcher indicated that "efficient infrastructure (…) helps", but that they mostly lack "resources and expertise", and more specifically, "computing and software programming expertise".

4. Conclusions and recommendations

In this part of the report, we described the findings of empirical research into UvA researchers' (a) reliance on third-party data for doing research; and (b) the obstacles they face when doing so. The information was gathered through a survey sent to all scientific staff at the UvA as well as interviews with researchers and data stewards from different UvA faculties.

The survey allowed us to get a wider perspective across the UvA on researchers' reliance on third-party data, the types of data access and data use issues faced, and best practices. From the survey, it followed that around half of UvA researchers use third-party data relatively often: 46.4% of the respondents reported to have recently engaged in research project that relied on data produced and/or held by a third party. Researchers also indicated that private sector data is generally provided voluntarily, but often under certain conditions – resulting in data *use* challenges. Frequently mentioned in this regard were the conditions not to further share and/or publish the data and to acknowledge the data provider in any research outputs.

The interviews allowed us to learn more about specific experiences of UvA researchers and data stewards regarding data access. It was confirmed that researchers do not only experience issues when trying to *access* data, but also – when access is provided – to *use* the data in ways they consider appropriate. As regards the data access challenges, most anecdotes concerned legal, social/reputational, and technical obstacles. As regards the data use challenges, the interviewees described various restrictions and conditions imposed by third-party data providers, most notably, again, those prohibiting the publication or sharing of data with others. To better deal with all these challenges, researchers and data stewards indicated they would like to see improvements in terms of legal staff and expertise, guidance, communication and technical resources, among others.

In sum, the empirical research demonstrated that UvA researchers occasionally face issues when trying to access and use third-party data for scientific research purposes. However, it did not paint a picture of entirely 'data-deprived researchers' whose research is gravely imperilled due to unsuccessful data access attempts. Part of the reason for that could be that researchers have already internalised data access constraints by not developing certain research lines in the first place. Regardless, it still appeared that there is room to further promote and facilitate researchers' access to third-party data, especially considering that many researchers (55%) expect their research fields to become (much) more reliant on third-party data in the coming years.

The main recommendation stemming from this study is therefore for the UvA to **invest in even more robust institutional support** to facilitate researchers' access to and use of third-party data. In particular, the UvA could:

Strengthen the knowledge capacity and resources for data stewards across faculties

First, the university should *more actively advertise and enable data stewards as 'first responders'* in dealing with data access and use challenges, who can connect researchers with other support staff, including data protection officers, lawyers and ethics committees. This way, researchers may feel more inclined to involve data stewards in their research projects at an early stage. Moreover, it will foster shorter lines and a wider network of knowledge between these different actors.

Another way of improving data stewards' position and abilities to optimally fulfil their role, is for the university – for example, the university's legal department – to *draft procedural guidance on data access requests and data uses that are compliant with (inter alia) data protection law, IP law and research ethics*. Written policies on relevant aspects of data-driven research at a university- or faculty-level, kept in a centralised information/documentation pool, could provide handles for data stewards to better and more consistently assist researchers. A slightly different version of such guidance could be provided to researchers as well, tailored to their questions and needs, to prepare researchers for their interactions and discussions with data

stewards. Additionally, institutional policies and standardised forms have the potential to streamline ERB procedures and speed up internal discussions.

In particular, the policies should provide *guidance on how to solve the tension between open science/open data principles that researchers must or wish to adhere to, and the publication and sharing requirements as imposed by third parties*. Such policies can build on the Data Governance Act,[45] which specifically aims to stimulate access to and use of 'protected public data' in such a way so as to ensure that the protected nature of the data is preserved, for example by allowing access to and use data of in secure processing environments to preserve commercial business secrets.

In order to help data stewards implement data access and use policies and guidelines into practice, the university could *facilitate (regular) meetings in which data stewards exchange knowledge and best practices* as well as *workshops aimed at informing data stewards on current developments within the digital legal landscape*.

Expand the pool of expert legal staff and/or offer legal training to both data stewards and researchers

During the interviews, both researchers and data stewards emphasised the importance of legal support in matters of data access and data use, most notably as regards the interpretation of legal data access restrictions, the negotiation of data access agreements, and evaluation of data protection issues. The perceived lack of (timely) legal support could be a reason to consider expanding the pool of expert legal staff and, in addition, offering data stewards and/or researchers relevant legal training.

Invest in, and promote the use of, technical resources

To support the growing reliance on third-party data for research, the university is encouraged to invest in technical resources to safely access, analyse and store these data. Recent efforts spearheaded by UvA researchers in this regard include, for example, the virtual research environment[46] and digital data donation platform.[47]

Raise awareness among researchers on data access and data use challenges

Finally, the university may want to consider initiatives to raise more awareness among researchers on data access and data use challenges, for example through trainings, workshops and meetings, e.g., facilitated by data stewards. These initiatives should ideally also pay attention to the risk of 'data greediness' on part of researchers and encourage researchers to interrogate the necessity of third-party data for their respective research projects.

***

Anticipating the growing reliance on third-party data in the future, we consider it vital to reflect on the affordances of new legal frameworks at the EU level aimed at digital infrastructures for researchers' data access. Part B of this report will do so by mapping relevant provisions across several legal frameworks, investigating which data access opportunities they (potentially) provide for researchers.

---

[45] See Article 5 Data Governance Act.

[46] UvA 2023 (webpage) <https://www.uva.nl/en/content/news/news/2023/03/virtual-research-environment-uva-wide-available.html.>.

[47] <https://datadonation.eu/>.

Part B: Data Access in the Law

A Mapping of the Regulatory Landscape of Transparency and Data Access for Research Purposes and an Analysis of Opportunities and Gaps

2023

Jef Ausloos, Arlette Meiring, Doris Buijs and Mireille van Eechoud
University of Amsterdam
Institute for Information Law

1. Objective, scope and research methods

This Part B of the report aims to explore, analyse and evaluate how EU law could be used as a tool to tackle information asymmetries between academia and entities managing digital infrastructures.

1.1 Scope

A large part of the research in Part B is dedicated to the mapping and analysis of relevant legislative developments at the EU-level that affect access to and the use of data for scientific research purposes (see chapter 3). **In doing so, we focus on research activities performed by academics and universities, i.e., 'academic' scientific research**. This is not to say, of course, that the results of the mapping-exercise could not be relevant for other types of research, including governmental research, commercial research, and investigative research by journalists and non-governmental organisations (NGOs). We are also well aware of the wider playing field of relevant societal actors that pursue access to third-party data for other public-interest purposes, such as democratic control and economic progress, and that the examined legal provisions may serve a whole range of goals. These actors and goals, however, remain outside the scope of this report.

It should be noted that access to externally-held data can be seen as an important first step in the larger data-driven research lifecycle (see Figure 1). As evidenced in Part A, ample issues arise at this stage already. Yet, while this report mainly focuses on the role of the law as a means to tackle obstacles to data *access*, it also recognises that conditions and restrictions placed on data *use* (e.g., analysis, publication) may affect data access – for instance, if researchers do not agree with third-parties' terms they will not receive the data – and thus hinder the research workflow.[48]

Analysing data access and transparency provisions is no sinecure, considering that the last few years have been marked by the proposal and adoption of a deluge of legislation concerning data and digital infrastructures at the EU level.[49] **In this report, we exclusively focus on horizontal legislation that addresses issues related to data, digital infrastructures and the digital economy more broadly (hereinafter: "digital/data legislation")** (see Table 1). We are aware of the significant number of sector-specific frameworks that have emerged and may also contain relevant provisions on digital transparency and data access, for example in the areas of finance, transportation, agriculture, health and the environment.[50] For the purposes of this report, however, we decided to limit ourselves to an analysis of horizontal legislation. This scoping decision is in no way based on the assumption that sectoral rules are less relevant to researchers who wish to observe data and digital infrastructures (especially in those domains), but rather reflects pragmatic considerations of time and expertise as well as considerations related to the horizontal frameworks' novelty and relevance across sectors and disciplines.

As an exception to the horizontal approach, we explore the increased regulation of what has traditionally been referred to as **"the information society services sector".** The governance of online services has been a key component of the EU's recent digital policy agenda,[51] which has resulted in the adoption and proposal of various legislative acts, most notably the Digital Markets Act (DMA) and Digital Services Act (DSA). We believe that these flagship frameworks are relevant to explore in more detail, not only because of the growing ubiquity of online platforms across sectors and disciplines, but also because

[48] This also came to the fore in the empirical research (Part A) conducted for this research project.

[49] Cf. overview reports such as: Zenner 2022 (accessed 27 February 2023); Codagnone, Livia and Rodriguez De Las Heras Ballell 2022.

[50] In the context of environmental protection, see for example the European Pollutant Release and Transfer Register (E-PRTR) Regulation (Ausloos, Leerssen and Ten Thije 2020, pp. 27-52).

[51] See European Commission Communication 2020.

online platforms have become important objects of, and vehicles for, scientific research. In our UvA-wide researcher survey, online platforms were mentioned as the parties most frequently relied on for obtaining data.[52] Finally, as an exception to the analysis of *legislative* instruments, we take a look at the 2022 Strengthened Code of Practice on Disinformation, a co-regulation scheme, as it pays special attention to the topic of access to data for research purposes.

Figure 1. A simplified model of the research process, visualised as a cycle composed of smaller iterative cycles.

No.	Framework	Abbreviation
1.	E-Commerce Directive (2000/31/EC); Services Directive (2006/123/EC); Consumer Rights Directive (2011/83/EU, last amended by Directive 2019/2161)	**ECD; SD; CRD. Together: 'e-commerce and consumer law'**
2.	Access to EU Documents Regulation (1049/2001)	**EUDR**
3.	General Data Protection Regulation (2016/679)	**GDPR**
4.	Data Protection Law Enforcement Directive (2016/680)	**LED**
5.	Free Flow of Non-Personal Data Regulation (2018/1807)	**NPDR**
6.	Copyright in the Digital Single Market Directive (2019/790)	**CDSMD**
7.	Open Data Directive (2019/1024)	**ODD**
8.	Platform-to-Business Regulation (2019/1150)	**P2BR**
9	Data Governance Act (2022/868)	**DGA**
10.	Digital Markets Act (2022/1925)	**DMA**
11.	Digital Services Act (2022/2065)	**DSA**
12.	Proposed AI Act (COM(2021) 206 final)[53]	**pAIA**
13.	Proposed Political Advertising Regulation (COM(2021) 731 final)	**pPAR**
14.	Proposed Data Act (COM(2022) 68 final)[54]	**pDA**
15.	Proposed European Media Freedom Act (COM(2022) 457 final)	**pEMFA**
16	*Self-/co-regulation:* 2022 Strengthened Code of Practice on Disinformation	**2022 CoP**

Table 1. List of analysed EU (non-)legislative instruments and their abbreviations, in chronological order

[52] See section 2.1 of Part A of this report.

[53] For this report, we have used the pAIA version with amendments adopted by the European Parliament on 14 June 2023 to which we refer as 'pAIA': Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/01/06(COD) <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf>.

[54] For this report, we have used the latest proposal of the Council of the European Union to which we refer as 'pDA': Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) – Mandate for negotiations with the European Parliament, 17 March 2023, 7413/23 <https://data.consilium.europa.eu/doc/document/ST-7413-2023-INIT/en/pdf>.

1.2 Methodology

1.2.1 Categorising legal frameworks

In chapter 3, we categorised the legislative instruments listed in table 1 according to the sector(s) they aim to regulate: the public sector, the private sector, or potentially both. This categorisation has led to a structured analysis of **(1) generic frameworks, (2) frameworks regulating the public sector and (3) frameworks regulating the private sector.** The reviewed legislation has either been (a) proposed, (b) adopted or (c) already become applicable. With "proposed" frameworks, we refer to legislation that has officially been put forward by the European Commission. The content of frameworks that find themselves in this stage is not final yet and may still be subject to amendments during ongoing negotiations. Nevertheless, we decided to include the proposed frameworks in the analysis as their overall scope and substance are typically evident and changes are unlikely to greatly affect the legislation's bottom line. With "adopted" frameworks, we refer to legislation that has been adopted by the European Commission, the Council of the European Union and the European Parliament, and which has already entered into force but has not yet become applicable. This means that the text of the legislation has become final and legally exists as EU law but is not yet applicable across the Member States.[55] With "applicable" frameworks, finally, we refer to legislation that has legal effect and can be enforced. In the case of directives, applicability means that the directive applies *and* the deadline for EU Member States to transpose the directive into national law has passed.[56]

1.2.2 Analysing data access and transparency provisions

For each of the selected frameworks, we identified and addressed the following elements:

- the **type(s) of data** to which the legal framework applies (e.g., personal data, government documents, information about certain products);
- the **rationale** of the transparency or data access regime in question (aims and goals), including whether **scientific research is explicitly recognised** as a driver for transparency and data access or not (e.g., whether scientists are specifically mentioned as data recipients);
- the **most relevant transparency/data access provisions** in the framework, including corresponding recitals[57];
- whether the data specified in the provisions must be provided **proactively** by the data holder (push) or **reactively**, i.e., following a request by a data recipient or any other party (pull);
- the **key data holders** – which natural persons or organisations hold the data and have to provide the data to intended data recipients;

[55] See for instance the Digital Services Act (DSA). The DSA entered into force on 16 November 2022 (twenty days after publication in the Official Journal) but applies in its entirety from 17 February 2024, see Article 93 DSA.

[56] Directives and Regulations are two different types of EU legislation. **Regulations** are binding legislative acts that can be directly applied across the EU, meaning that the text does not have to be implemented in Member State law. **Directives** are also directly-binding, but they need to be transposed by EU member states into their own national law. The Directive only becomes applicable once the deadline for the transposition into national law has passed, and national legislation has been adopted.

[57] The articles in an act of legislation are called 'provisions', they contain the actual rules and they are legally binding and can be enforced upon. The recitals are non-binding but give more background information and explanation on the articles, they set out the reasons for the contents of the articles.

- the **intended data recipients** – which natural persons or organisations must be provided with the data, for example, the general public, regulators or specified private parties such as platform users, consumers, data subjects, vetted researchers, competitors, and so on;[58]
- **formalities/practicalities** – such as applicable fees, timeframes within which data should be provided, formats that are applied to the data, and so on; and
- whether transparency or access to data can be obtained **directly or indirectly**, and whether the acquired data are **systemic or individualised** (explained below).

1.2.3 Qualifying data access and transparency provisions

As will become apparent from chapter 3 , there is a wide variety of transparency and data access provisions throughout the explored frameworks. To help readers navigate this forest of provisions, we have qualified the relevant provisions along three additional axes, according to: (1) the required action from data holders; (2) the data access point; and (3) the type of data that can be obtained via the designated access point/entity. This translates into the following distinctions:

(1) proactive v. reactive access, i.e., data access is provided proactively by the data holder ('push') v. reactively, e.g., following an official request by a data recipient ('pull');
(2) direct v. indirect access, i.e., data access is provided directly by the data holder v. indirectly via an intermediary party; and
(3) system-level v. individual-level data, i.e., the data relate to (parts of) the digital infrastructure's system as a whole v. the data relate to specific endpoints of the digital infrastructure (people or devices).

While not absolute, these distinctions can be helpful to reflect on the relevance of data access and transparency provisions for scientific research purposes specifically. The distinctions are further explained below.

Proactive v. reactive access (or: push v. pull)
Some legal frameworks explicitly require the intended data recipient to submit a data access request with the data holder, to which the data holder must then respond (reactive). Oftentimes, however, such requests are not necessary, and data holders are expected to provide data 'proactively' in accordance with specific rules and timeframes laid down in the legal framework.

Direct v. indirect access
The second distinction is about whether datasets and information can be obtained via the initial data holder or via an intermediary party. In case of "direct" access, researchers obtain relevant data directly from the initial data holder, often the legal entity managing a digital infrastructure (or in some cases, an individual 'endpoint' within that digital infrastructure such as a platform user or smart device). "Indirect" access, on the other hand, refers to situations where data are obtained from an intermediary entity which maintains or processes data or information originating from the initial data holder, such as external auditors. The intervention of an intermediary adds a 'layer' between the original source of the data and the researcher, which could possibly affect the research utility of the data, e.g., for reasons of verifiability or accuracy. At

[58] Compare categories 2-4 of access to data ("private-party access to data", "regulator access to data", "general public access to data") in; Edelson, Graef and Lancieri 2023, pp. 23-27.

the same time, indirect access is sometimes the only option available to researchers, especially when direct access is refused by the initial data holder due to conflicting rights or interests (see section 5.3).

System-level v. individual-level data

Finally, transparency and data access provisions may cover different types of data. A broad distinction can be made between system-level and individual-level data. The type of data that can be acquired will often depend on the entity providing the data, i.e., entities managing (parts of) digital infrastructures, or endpoints within the infrastructures such as individual users. Generally, information that is obtained from a legal entity managing a digital infrastructure (either directly or via an intermediary) will amount to system-level data, enabling researchers to take a 'helicopter view' of a digital infrastructure. However, these data do often not provide detailed insights into specific aspects of the digital infrastructure, given that the data are likely to be pre-selected, pre-processed and/or aggregated. On the other hand, information that is obtained from endpoints of a digital infrastructure (either directly or via an intermediary) typically amounts to individual-level data, enabling researchers to derive more granular insights, as the data usually comprise of unprocessed information about the (use of, and interaction with) specific endpoints of the digital infrastructure. In essence, where data provided by the central data holder may provide width but lacks in depth, the opposite is typically true for data provided by endpoints of a digital infrastructure.[59]

***

The overview table (Table 2) on the next page contains examples of system-level and individual-level data that are either directly or indirectly accessible for researchers and provided either proactively (P) or reactively (R) by the data holder or intermediary entity.

[59] Ausloos and Veale 2020.

	Direct access	
System-level data	Data made publicly available by entities managing a digital infrastructure, such as: ▪ Terms of Service (**P**);[60] ▪ Privacy Policy (**P**);[61] ▪ Ad archives (**P**);[62] ▪ Transparency reports (**P**);[63]	Data provided by entities managing a digital infrastructure to researchers in their capacities of intended data recipients, such as: ▪ Personal data provided by data controllers to researchers in their capacity as data subjects, following a data access request (**R**);[64] ▪ IoT-data generated while using smart products, provided by manufacturers to researchers, on the explicit request of a product users (**R***).[65]
	Data made publicly available by other organisations than the entities managing a digital infrastructure, such as: ▪ Enforcement agencies' reports, e.g., on systemic risks of VLOPs (**P**);[66] ▪ Independent audit reports, e.g., on compliance of VLOPs (**P**);[67] ▪ Trusted flagger reports on illegal content on online platforms (**P**).[68]	Data provided via a data donation architecture:** *Personal data*[69] initially obtained by data subjects via data access requests with the digital infrastructure, and *other data* held by endpoints in a digital infrastructure (e.g., statements of reasons for content moderation actions taken by online platforms),[70] provided by these endpoints via a data donation architecture[71] for researchers to use (**R*****).
	Indirect access	

Table 2. Examples of data organised according to the distinctions of proactive/reactive data provision, direct/indirect access, and system-level/individual-level data.

[60] E.g., Article 14 Digital Services Act.

[61] Articles 13 and 14 General Data Protection Regulation.

[62] Article 39 Digital Services Act; Article 7(6) proposed Political Advertisement Regulation.

[63] E.g., Article 15 Digital Services Act.

[64] Article 15 General Data Protection Regulation.

[65] Articles 3-4 proposed Data Act.

[66] Article 35 Digital Services Act.

[67] Article 37(4) Digital Services Act.

[68] Article 22(3) Digital Services Act.

[69] Most of the data being donated will be personal data, yet it is, in theory, also possible to donate non-personal data (for instance under the Free Flow of Non-Personal Data Regulation).

[70] Article 17 Digital Services Act.

[71] See e.g., Araujo et al 2022; <datadonation.eu>.

Explanations to the overview table:

* Internet-of-Things (IoT) data generated using smart products and services can be shared with third parties, including researchers, but only on the request of a user of smart product addressed to the data holder (often the manufacturer) to share the data to a designated third party (see Article 5 of the proposed Data Act).
** In this report, we consider data access by researchers via data donation architectures as a form of indirect data access. Data donation, for our purposes here, refers to the practice where consenting participants voluntarily, actively or passively, transfer (personal) data that they are entitled to pursuant to the law, to researchers. This can be done either by sending it directly to researchers, or through a data donation platform which makes the data available to researchers.[72] We have categorised data donation as a form of indirect data access since in this case, (individual-level) data held by endpoints of a digital infrastructure are, technically speaking, not obtained from digital infrastructures directly, but from the endpoints (i.e. data subjects, users, owners of IoT devices, etc), possibly even through an institutionalised data donation platform centralising the respective data.
*** As noted above, access to individual-level data via a data donation architecture is often request-based and thus 'reactive' since researchers usually approach research subjects actively to ask them to donate their data for research purposes.

1.3 Structure

After a brief discussion on the normative underpinnings of data access claims for scientific research purposes (chapter 2), we will summarise the main features of selected data access and transparency regimes in EU digital/data legislation (chapter 3) and determine their relevance for researchers (chapters 4). Building on this analysis, we will then identify a number of relevant 'themes' recurring across the frameworks (chapter 5) and provide guidance for researchers on how to effectively use legal data access and transparency provisions to acquire useful data for research (chapter 6). We will conclude with reflections and recommendations for universities, researchers, policymakers and regulators alike on how to improve legal conditions for data access in the context of academic research (chapter 7).

[72] Compare Pereira Campos 2021, p. 18, 30.

2. Normative underpinnings of claims to data access for research

Before analysing how selected EU legal frameworks may facilitate access to data generated by and about digital infrastructures for scientific research purposes, it is useful to reflect, first, on why such (better) data access is important and justified to begin with. This section sets out the normative underpinnings of claims for access to data for purposes of **(1) public scrutiny and accountability more broadly**, and for the specific purpose of **(2) knowledge production in light of the scientific mission of universities**. The normative underpinnings are discussed in the context of fundamental rights law in Europe as a core part of a wider normative and policy framework.[73]

2.1 Access to third-party data for public scrutiny and accountability purposes

EU legal provisions on transparency and data access have traditionally been driven by considerations of public scrutiny and accountability.[74] In this regard, a distinction can be made between data held by public sector bodies and private sector entities.

Claims for access to information held by **public bodies** are not new. The world's first Freedom of Information Act (FOIA) was reportedly adopted by Sweden in 1766.[75] Today, 135 UN Member States have constitutional, statutory and/or policy measures in place to guarantee access to information held by public sector bodies.[76] It is generally assumed that the possibility to access (certain) public sector information is essential for citizens to effectuate democratic control of public administration and to exercise individual fundamental rights.[77] Although historically access laws mainly served democratic purposes, it is also increasingly recognised in the EU and elsewhere that access can have economic benefits, as public information can be re-used by the private sector for the improvement and development of (new) products and services.[78]

The right to freedom of expression as enshrined in Article 10 of the European Convention on Human Rights (ECHR) extends to access information (and by implication this is also true of Article 11 of

[73] This wider framework also includes, for instance, the EU's research policy and the framework for the European Research Area (ERA), which aims to remove barriers for researchers, scientific knowledge and technology. See Articles 179-190 Treaty on the Function of the European Union (TFEU).

[74] Even the provision in the recently adopted Digital Services Act (DSA) specifically designed to provided "vetted researchers" with access to platform data, was not created to provide researchers with access to data for scientific research purposes *as such*. Rather, it was introduced as a means to "bridg[e] information asymmetries and establish[…] a resilient system of risk mitigation" (see recital 96 DSA). The European Commission explicitly considers the new framework for researchers' access to data from very large online platforms and very large search engines "a key measure (…) to increase platforms' transparency and accountability", see <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en>.

[75] Mustonen (ed) 2006.

[76] See UNESCO (webpage) <https://www.unesco.org/en/access-information-laws>.

[77] See for example Article 15 of the Treaty on the Functioning of the European Union (TFEU) and recital 2 of the Access to EU Documents Regulation. See also Van Eechoud 2011, pp. 169-170 with reference Council of Europe 2009, par. 1 (preamble), Strasbourg, CETS No. 205.

[78] This rationale forms the basis of the 2019 EU Directive on open data and the re-use of public sector information (Open Data Directive; recast of the 2003 Directive on the re-use of public sector information). See also European Commission Communication 2015, p. 15: "[The Commission] will encourage access to public data to help drive innovation".

the Charter of Fundamental Rights of the European Union (CFREU)).[79] Article 10 ECHR does not grant individuals a 'general' right of access to official documents. However, when access to information proves to be instrumental to an individual's exercise of their right to freedom of expression – in particular the freedom to receive and impart information – the denial of access may constitute an unlawful interference with that right.[80]

Throughout its case-law, the European Court of Human Rights (ECtHR) has formulated four cumulative criteria to assess whether and to what extent the denial of access to public information constitutes an interference.[81] First, the purpose of the information request must be to enable the applicant's exercise of the freedom to receive and impart information and ideas to others. Second, the nature of the information to which access is sought must meet a public-interest test. Third, special importance is attached to the "particular role" of the seeker of information in "receiving and imparting" information to the public,[82] such as journalists[83] and NGOs whose activities are related to matters of public interest.[84] Besides these public and social "watchdogs", the Court has made clear that **"a high level of protection also extends to academic researchers"**,[85] "authors of literature on matters of public concern"[86] and potentially to "bloggers and popular users of social media".[87] What these actors have in common, is that they (may) contribute to "informing the public debate".[88] Finally, the fact that the information requested is "ready and available" is considered an important factor in the overall assessment.

While access to public sector information has long received considerable attention, claims for access to **private sector information** are not novel either.[89] Governments rely on it for the execution of public tasks, to ensure compliance with national and international laws and to inform public policies.[90] Furthermore, private societal actors such as businesses, journalists, activists, NGOs, and the general public (including, but not limited to consumers, data subjects, platform users or citizens) may also have interests in accessing information and data held by private entities. These interests can be of a commercial, personal political and/or public nature, and can range from pursuing one's own commercial self-interest

[79] According to Article 52(3) CFREU, rights in the Charter that correspond to rights guaranteed in the ECHR have the same meaning and scope as those laid down in the ECHR.

[80] Moreover, a refusal to grant access also infringes on article 10 when the public sector body fails to respect a court order mandating disclosure. See e.g., ECtHR (Fifth Section) 3 March 2020, Appl. No. 75865/11 (*Centre for Democracy and the Rule of Law v. Ukraine*), para. 60, referring to ECtHR (Grand Chamber) 8 November 2016, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 156.

[81] ECtHR (Grand Chamber) 8 November 2016, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 157-170. See also ECtHR, 'Guide on Article 10 of the Convention on Human Rights – Freedom of expression', 31 August 2022, p. 75-79.

[82] ECtHR (Grand Chamber) 8 November 2016, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 164. See also ECtHR, 'Guide on Article 10 of the Convention on Human Rights – Freedom of expression', 31 August 2022, p. 77-78.

[83] ECtHR 8 November 2016 Grand Chamber, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 165; see also ECtHR 24 June 2014, Appl. No. 27329/06 (*Rosiianu v. Romania*); and ECtHR 18 November 2021 (Third Section), Appl. No. 6106/16 (*Saure v. Germany*).

[84] ECtHR 8 November 2016 Grand Chamber, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 166; see also ECtHR 14 April 2009 (Second Section), Appl. No. 37374/05 (*Társaság a Szabadságjogokért v. Hungary*); ECtHR 25 June 2013 (Second Section), Appl. No. 48135/06 (*Youth Initiative for Human Rights v. Serbia*).

[85] ECtHR 8 November 2016 Grand Chamber, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 168; see also ECtHR 8 July 1999, Appl. Nos. 23536/94 and 24408/94 (*Baskaya and Okcuoglu v. Turkey*); ECtHR 26 May 2009, Appl. No. 31475/05 (*Kenedi v. Hungary*); and ECtHR 3 April 2012, Appl. No. 41723/06 (*Gillberg v. Sweden*).

[86] ECtHR 8 November 2016 Grand Chamber, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 168; see also ECtHR 29 June 2004 (Second Section), Appl. No. 64915/01 (*Chauvy and Others v. France*); and ECtHR 22 October 2007 (Grand Chamber), Appl. Nos. 21279/02 and 36448/02 (*London, Otchakovsky-Laurens and July v. France*).

[87] ECtHR 8 November 2016 Grand Chamber, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 168.

[88] Ibid.

[89] A recent development in access to private sector data concerns access to clinical trial data in the medical context, see: Zemla-Pacud and Lenarczyk 2023.

[90] For example, real-time data from private vehicles can be useful for the optimisation of traffic management systems, see European Commission Staff Working Document 2017, pp. 12-13.

(businesses),[91] to making certain life choices (consumers),[92] exercising or defending contractual or legal entitlements (consumers),[93] monitoring compliance and holding commercial companies publicly accountable (public interest),[94] and producing a more informed citizenry more broadly (public interest).

Individual access to public sector information for public scrutiny and accountability purposes has a strong basis in European fundamental rights law (see above) and beyond.[95] However, this is a different matter for private sector data. Under the current fundamental rights framework, there is no (conditional) right of individuals[96] to access information held and controlled by private entities for accountability purposes. As discussed elsewhere in this report (notably in Part A, sections 3.1.1 and 3.2.1, and in section 5.5 below), commercial private entities oftentimes wish to keep information confidential, on the grounds that it contains personal data, trade secrets, other types of commercially sensitive information or materials protected by (third-party) intellectual property rights. It is unclear to what extent such secrecy can be justified based on the right to freely conduct a business (enshrined in Article 16 CFREU) and the right to intellectual property (laid down in Article 17(2) CFREU). However, secrecy does not always prevail; courts, may under certain circumstances order the disclosure of certain information held by a private entity.[97] Moreover, under the General Data Protection Regulation (GDPR) and Article 8(2) CFREU data subjects have the right to access information about the processing of personal data concerning him or her.[98] Nonetheless, European fundamental rights law does not provide for a more *general* right of access to private sector data.

2.2 Access to third-party data for scientific research and knowledge production

As pointed out in the previous section, European case-law has identified academic researchers as a special category of actors that may enjoy a (conditional) right of access to public sector information under Article 10 ECHR due to their role in informing the public debate. However, it must be emphasised that academic institutions and researchers also have a societal role that goes *beyond* participation in public debate. **First and foremost, they are the key figures in the global process of knowledge production and scientific**

[91] Business users of online platforms, for instance, tend to be interested in the data collected by the platforms since they could potentially use the data for the optimisation of their internal processes, consumer relations and business decision-making as well as the improvement of their products or services, see Gineikytė, Barcevičius and Cibaitė 2020.

[92] For example, private sector information could enable citizens to take climate change mitigation measures at an individual level. There are apps on the market that inform users of their carbon footprint based on their purchasing of products or services, the accuracy of which also depends on information disclosure by private companies, see Cantillon et al 2023, p. 9-10.

[93] For example, to seek evidence to prove wrongdoing. In the *Dexia* cases in the Netherlands in 2000s, clients of financial institution Dexia collectively requested access to their complete files, including risk profiles, which were later used in class action lawsuits to prove wrongdoing by financial institutions. Dexia initially denied access to the files, but in 2007 the Dutch Supreme Court ruled that access had to be provided. See Hoge Raad 29 June 2007, ECLI:NL:HR:2007:AZ4664.

[94] For example, NGO's such as Greenpeace or Amnesty International and investigative journalism outlets such as Bellingcat, FollowTheMoney or Correctiv, often make use of corporate information published based on legal requirements, in their (background) research. See also, Hamilton 2016.

[95] Please note that transparency of the public sector has a broad normative basis that consists not only of fundamental rights law but also, among other things, the rule of law in democratic society, which requires transparency of e.g., legal procedures, law-making, elections, policy formation and public task fulfilment of the executive branch.

[96] Importantly, this report does not focus on the issue of *government* access to (commercial) private sector data (business-to-government or B2G-sharing). B2G-sharing has been a topic of recent debate during the negotiations of the proposed Data Act in the context of the proposed obligation of data holders to make data available to public sector bodies based on an 'exceptional need', e.g., in the case of natural disasters. The question in this report is rather why *individuals*, and more specifically (academic) researchers, would be entitled to access privately-held data.

[97] See e.g., the *Dexia* case described in n (93), Hoge Raad 29 June 2007, ECLI:NL:HR:2007:AZ4664.

[98] Article 15 GDPR.

advancement.[99] What sets academic researchers apart from other societal persons and entities, is their historical and societal mandate to study the world, and ultimately, to contribute to scientific and human advancement in the public interest.[100] This mission is the reason why lawmakers have traditionally endowed academics with certain privileges, most notably academic freedom and the freedom to perform research.[101] When exercising these privileges, researchers are expected to produce high-quality and efficient research that is responsive to society's needs.[102] It is also this broader mission that informs many aspects of today's open science policies.[103] While researchers' public-interest mission has remained unaltered, the world around us has. Since the development of the Internet, digital services and infrastructures have rapidly permeated people's daily lives and become objects of research as well as useful sources for research into various other aspects of the digitalising society. Contemporary research activities, thus, increasingly involve the observation, collection, processing and analysis of digital data.[104] **In short, to fulfil their traditional responsibilities of knowledge production, academic researchers need to be able to observe the (both publicly- and privately-owned) digital infrastructures penetrating modern society. Such observability hinges on the accessibility of data residing in and/or generated by these infrastructures.**[105] As we will see in chapter 3, the EU legislator has recognised the importance of access to data held by providers of certain digital infrastructures or scientific research in the recently adopted Digital Services Act, introducing dedicated rules on researchers' access to platform data.[106]

Although European fundamental rights law does not currently provide for a dedicated legal 'right' for academic researchers to access public/private sector data for the purpose of knowledge production in the public interest, it could be argued that states should stimulate such access on the basis of their positive obligation to create "an enabling and participatory environment for the development of science".[107] In addition, it is worth exploring whether the concept of a "right to research" as conceptualised in copyright literature could be applied outside the copyright-sphere and be used as a more general basis for (conditional) access to (certain) third-party data. According to copyright scholars, the right to research is a fundamental right that "lies hidden" in existing fundamental rights and can be the basis for a broader set of uses of copyrighted works for research purposes.[108] The concept has the potential to be further developed as the basis not only for researchers' use of copyrighted works, but also for researchers' access to third-party data more generally.

On a side note, one may wonder whether some kind of 'privileged' data access right tailored to academic researchers for science purposes, as opposed to 'public' access to data, is a desirable outcome. The

---

[99] On the role of universities and academic research in society, see IViR, 'Information Law and the Digital Transformation of the University: Digital Sovereignty, Data Governance and Access to Data for Research – Part I. Digital Sovereignty', 2023, section 2.1.

[100] Ibid.

[101] Laid down in Article 13 CFREU and Article 15(3) International Covenant on Economic, Social and Cultural Rights (ICESCR), and enshrined in Article 10 ECHR.

[102] Compare the objectives of the EU's open science policy: European Commission Open Science webpage <https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en#ref-8-ambitions-of-the-eus-open-science-policy>.

[103] Although not the focus of this report, it is worth mentioning that while global open science policies – aimed at, inter alia, opening up scholarly publications, accelerating new research through FAIR sharing or research data and tools, safeguarding research integrity and stimulating citizen-science – recognise the important role of data for scientific research, they mainly focus on data originating *with researchers* in publicly funded research. Wat has been missing in e.g., the EU's open science agenda, is the recognition of the importance of data *held by private entities* as inputs for research in a digitised society.

[104] See OECD 2007, p. 9.

[105] Or, as it has been put in legal literature: "the necessity to extend such [data-access] privileges [to researchers] can arise from the growing urgency at global level to support research activities indispensable for developing solutions that are necessary to realise other fundamental rights and goals of the international community", see Geiger and Jüte 2022, p. 8. See also: Rieder and Hoffman 2020.

[106] Article 40 DSA.

[107] CESCR General comment no. 25, para. 46.

[108] Geiger and Jütte 2022, p. 23.

advantage of opening up third-party data only to a select group of experts such as researchers affiliated with a university or accredited research institution, is that it may enable the sharing of more (sensitive) data, which could render the data more useful for scientific research. On the other hand, limiting data access to academic researchers would leave empty-handed other actors who share similar public-interest missions, such as investigative journalists or activists that fulfil public watchdog and/or knowledge production functions. Moreover, privileged access could reproduce problematic power dynamics and impact the independence of research.[109] On another side note, academic researchers must realise that "amassing as much digital data as possible"[110] or "hyperinformation"[111] should never be the end goal. Researchers have both legal and ethical responsibilities vis-à-vis third-party data providers and data subjects. They must therefore observe that their access to (sensitive) data for research purposes is proportionate to the infringement on third-parties' interests and, where necessary, take appropriate measures to protect e.g., privacy and confidentiality and ensure information security.[112]

That said, it remains crucial to enable the observability of digital infrastructures and safeguard high-quality, efficient and impactful research in a digitised society.[113] The next chapter will now turn to the legal reality and map out existing and proposed transparency and data access provisions in secondary EU digital/data law.

[109] See Leerssen 2023a, p. 75-76 on the pros and cons of public access.

[110] See Tromble 2021, p. 3.

[111] Obviously, researchers do not need access to *all* imaginable data residing in digital infrastructures in order to perform meaningful research and explain the world. As the philosopher Byung-Chul Han once put it: "The mass of information produces no truth. The more information is set free, the more difficult it proves to survey the world. Hyperinformation and hypercommunication bring no light into darkness", see Han 2015, pp. 37-42.

[112] Compare Edelson, Graef and Lancieri 2023, p. 23.

[113] Also see: Rieder and Hoffman 2020.

3. Selected transparency and data access provisions in EU law

This chapter summarises the main features of the transparency and data access provisions enshrined in 16 legal frameworks and one non-legal, co-regulation framework in the EU's digital/data strategy.

3.1 Generic frameworks – regulating both the public and private sector

3.1.1 General Data Protection Regulation (GDPR)[114]

The GDPR is a regulation that contains rules on the processing of personal data.[115] It aims to enhance individuals' control over their own personal data and to facilitate the free flow of personal data throughout the EU. Transparency is one of the regulation's key principles (Article 5). The principle is reflected in, for instance, the general information obligations imposed on data controllers (Articles 13 and 14, proactive provision) and the right to data access granted to data subjects (Article 15, reactive provision). Based on Article 15, data subjects have a right – upon request – to receive a copy of their personal data as well as information on processing activities,[116] which allows them to become aware of the processing of their data and to verify the lawfulness thereof.[117] In addition to the right to access, data subjects also have a right to data portability (Article 20), which has a more limited scope than the right to data access (invokable in less situations and involving less data) but can be more scalable and/or efficient to run analyses on the respective data. Portability in the context of the GDPR means that data subjects are entitled, under certain circumstances, to receive their personal data from the data controller and to transmit the data to another data controller without hindrance.[118]

While the GDPR does not contain transparency or data access provisions that are aimed at *researchers* specifically, both the data subject's right to data access and the right to data portability have the potential to be used by researchers as vehicles to collect individual-level data for scientific research purposes. Current practice shows that the data access right of Article 15 GDPR is increasingly used by researchers to facilitate 'data donation' initiatives.[119] As explained in section 1.2.3, data donation is a practice where research subjects are invited to transfer (personal) data – or have data transferred – that they are entitled to pursuant to the law, to researchers. In many current data donation architectures, research subjects must first request 'their' data held by a third party, and upon receipt, deliver these data to researchers, either directly or via a designated intermediary. A disadvantage of this practice, however, is that research subjects must *actively* take a few steps in order to actually donate their data to a specific research project. Research conducted in the

[114] Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[115] Personal data is defined in Article 4(1) GDPR as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

[116] Article 15(1) GDPR. Information on the processing includes e.g., the purpose of the processing, the recipients to whom the personal data are disclosed, the period of storage and in access the data have not been collected from the data subject, information on the source of collection.

[117] Recital 63 GDPR.

[118] This provision was created "to further strengthen the control" over one's personal data where the processing is carried out by automated means. See Article 20(1) and recital 68 GDPR.

[119] See e.g., Araujo et al 2022.

Digital Data Donation Infrastructure (D3I)[120] project has shown that many research subjects tend to drop out and do not complete the data donation process.[121] A solution to these low retention numbers could potentially be found in the right to data portability: if a social media networking site, for instance, would have an option embedded in the platform for users to directly and easily share – 'port' – a copy of the data relating to them with (intermediaries operated by) universities and/or research projects, this could perhaps reduce the number of drop-outs.[122]

3.1.2 Free Flow of Non-Personal Data Regulation (NPDR)[123]

The NPDR is a regulation that aims to ensure the free movement of non-personal data.[124] To achieve this, the NPDR lays down rules on data localisation requirements, the availability of data to competent authorities, and data porting for professional business users.[125] For the purposes of this report, the provision on the switching of data processing services providers[126] and the porting of data between different IT-systems (Article 6) are of particular relevance. The rationale behind this provision is to avoid the effects of vendor lock-ins, i.e., situations where users cannot switch between service providers because their data is 'locked' in the provider's system due to, for instance, specific data formats or contractual arrangements.[127] This is important, as data portability is a "key factor" in user choice and effective competition.[128] Article 6 states that the European Commission shall encourage the development and implementation of self-regulatory codes of conduct on the porting of data by professional users of data processing services from one IT-environment to another one, that is, either to another service provider's systems or to the user's own on-site systems.[129] In other words, if a service provider adheres to a code of conduct developed under the NPDR[130], the porting of data of a professional user has to take place in compliance with the requirements specified therein.[131]

Importantly, where the GDPR contains a special *right* for data subjects to the portability of personal data, Article 6 NPDR does not provide a right for professional users to port non-personal data. Instead, it follows a self-regulatory approach with voluntary codes of conduct for the industry.[132] An inherent downside of voluntary codes of conduct is that there is no legal obligation to adhere to them, which means that professional users will not always be able to request the porting of their data. With this in mind, portability under Article 6 NPDR does not seem to benefit scientific research in particular.[133] However, when a professional user ports data held by a service provider to its *own* IT-systems,[134] it could in theory

[120] <datadonation.eu>.

[121] Data on research subjects dropping out of the data donation process are on file with the authors of this report.

[122] Theo Araujo during the IViR Workshop on Research Data Access to Digital Infrastructures, held on 16 March 2023.

[123] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (Free Flow of Non-Personal Data Regulation).

[124] 'Data' is in Article 3(1) NPDR defined as "data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679".

[125] Article 1 NPDR.

[126] Shortened: "service provider", see Article 3(4) NPDR.

[127] European Commission, guidance NPDR p. 16-17.

[128] Recital 29 NPDR.

[129] European Commission, guidance NPDR, p. 18.

[130] For instance, the 'Switching Cloud Providers and Porting Data (SWIPO)' Codes of Conduct, see <https://swipo.eu/>.

[131] European Commission, guidance NPDR, p.19.

[132] Ibid, p. 18.

[133] This downside of voluntary codes of conduct is also acknowledged in the proposed Data Act. The proposed Data Act introduces binding obligations for service providers to enable effective switching between data processing services (recital 70 pDA). See further n (316).

[134] See Article 6(1)(b) NPDR.

further share the data with researchers. Moreover, professional users may also decide to *directly* port data to an entity providing data processing services for scientific research purposes, including data donation platforms. Article 6 NPDR thus provides a *basis* for the sharing of non-personal data with researchers.

3.1.3 Copyright in the Digital Single Market Directive (CDSMD)[135]

The CDSMD sets out rules that aim to modernise the EU copyright framework in the digital environment. One of its objectives is to ensure that research carried out with the assistance of digital technology – generally referred to as text- and datamining (TDM) – can be carried out in compliance with EU copyright law.[136] To this end, Article 3 contains a mandatory exception to copyright, related rights and sui generis database protection for reproductions and extractions made by research organisations in order to carry out TDM of copyrighted works for purposes of scientific research.[137] Notably, the exception only applies to works to which researchers already have lawful access. The exception thus merely ensures that researchers are not obstructed by copyright when conducting text- and datamining *after* lawfully obtaining access to copyrighted works.[138] In other words, while the CDSMD does not so much aim to enhance researchers' *access* to copyright protected content, we still consider the CDSMD relevant to our analysis since it explicitly aims to facilitate the subsequent *use* of lawfully accessed content for scientific research purposes.[139]

3.1.4 Data Governance Act (DGA), Chapters III-IV and VI[140]

The DGA is a broad regulation aimed at unlocking the full potential of data for the European economy and society. For this, it is deemed necessary to improve the conditions for data sharing, by creating "a harmonised framework" for data exchange and introducing basic requirements for data governance.[141] To that end, the DGA,[142] inter alia, encourages the re-use of 'protected' data held by the public sector (see section 3.2 on frameworks regulating the public sector). Moreover, it has created new notification and registration frameworks for 'data intermediation services providers' (Articles 10-15) and 'data altruism organisations' (Articles 16-25). These new data intermediaries are envisioned to facilitate the exchange of substantial amounts of data[143] and increase trust in data sharing[144] (see text boxes below). Such intermediaries are potentially helpful to researchers, as the wider sharing of data in general, most likely implies the wider sharing of relevant data for scientific research. It should be noted, however, that while in theory researchers can be the buyers or sellers of data shared via commercial data intermediation services, this is not likely to happen often in practice given the not-for-profit nature of most publicly funded research. The rules on data altruism services thus seem more relevant in this regard.

It should also be noted that the new rules for data intermediation service providers and data altruism organisations apply to services performed *in the EU*. Where an entity established outside the EU provides

---

[135] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (CDSM Directive). For the Dutch implementation of the CDSMD, the Dutch Copyright Act, the Neighbouring Rights Act and the Databases Act will be amended.

[136] This must ultimately promote the EU's competitive position as a research area, see recital 10 CDMSD.

[137] Article 3(1) CDMSD, European Commission and Senftleben 2022, pp. 36-46.

[138] European Commission and Senftleben 2022, pp. 43-45.

[139] Recitals 8-18 CDSMD. See also European Commission and Senftleben 2022, pp. 43-45.

[140] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

[141] Recital 3 DGA.

[142] In the rest of this section, where reference is made to the DGA, this is a reference to chapters III-IV and VI DGA.

[143] Recital 27 DGA.

[144] Recital 5 DGA.

data intermediation or data altruism services, it must appoint a legal representative in one of the Member States in which the services are offered for the relevant authorities in the EU to turn to (see Article 11(3) and Article 19(3)). However, the DGA also contains provisions limiting the transfer of data outside the EU and obliges data holders and intermediaries to take adequate measures to prevent data transfers that would contravene EU law (Article 31). How this will play out for e.g., data altruism organisations that are set up to support international scientific research, including collaborations with the United States, is unclear.

A **data intermediation service (DIS)** is defined in the DGA as "a service which aims to establish ***commercial*** relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data […]".[145] DIS providers are intended to function as trustworthy, neutral and independent organisers of data-sharing.[146] Some DIS providers offer their services to data subjects, while aiming to enhance data subject agency and control over personal data.[147] To increase trust in DIS providers, the EU legislator deemed it necessary to create a regulatory framework with harmonised requirements for trustworthy DSIs.[148] Articles 10 – 15 of the DGA lay down such requirements.

The Commission lists some examples of DIS providers on its website, such as DAWEX and API-ALGRO.

Data altruism refers to the ***voluntary*** sharing of data based on the consent of data subjects or permissions of the data holders without seeking or receiving a reward that goes beyond compensation related to the costs of the making available of the data for the sharing for objectives of general interest.[149] **Data altruism organisations (DAOs)** are "legal persons that seek to support objectives of general interest by making available relevant data based on data altruism at scale".[150] In other words, DAOs collect and process data made available for altruistic purposes. Whereas DISs focus on commercial data exchanges, DAOs aim to contribute to the voluntary sharing of data. For a legal person to be allowed to use the label 'data altruism organisation', it must meet the requirements as laid down in Chapter IV (Articles 16 – 25) of the DGA.

Examples of data altruism organisations mentioned by the Commission include for instance MyData Global, the Smart Citizen platform and the German Corona-Datenspende-App.[151]

Lastly, Chapter VI of the DGA is dedicated to the establishment of the European Data Innovation Board (EDIB), an expert group. The EDIB consists, inter alia, of several representatives of competent authorities for DISs and (registration of) DAOs, the European Data Protection Board (EDPB)[152] and the

[145] Article 2(11) DGA. The provision also lists four categories of services that are excluded of becoming a data intermediation service.

[146] See also recital 33 of the DGA. As such, they should comply with the GDPR (recital 35 DGA).

[147] Recital 30 of the DGA jo. Article 10(b) DGA.

[148] Recital 32 DGA.

[149] See for the full definition Article 2(16) DGA.

[150] Recital 3 DGA.

[151] European Commission Data Governance Act explained (webpage) <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>, accessed 17 May 2023.

[152] The EDPB is an official EU body established in Article 68 GDPR. It consists of delegates from supervisory authorities from all EU member states and of the European Data Protection Supervisor (or their respective representatives). The EDPB is responsible for, among other things, monitoring compliance with the GDPR.

Commission.[153] The EDIB shall also consist of three subgroups, respectively focusing on the registration of DAOs, technical discussions on standardisation, portability and interoperability and lastly on stakeholder involvement, including research and academia.[154] The EDIB shall, among other tasks[155], focus on developing a consistent practice for data altruism[156], guidelines on e.g., common European data spaces and interoperable frameworks of common standards[157], and to advise and assist the Commission with the development of the European data altruism consent form.[158] This European data altruism consent form is introduced in Article 25 DGA, and is aimed to allow for the collection of consent (in case of data altruism which is related to personal data) in a uniform format. The Commission shall adopt implementing acts for the development and establishment of this consent form.[159] This may help to further professionalise and facilitate voluntary data sharing by data subjects.

3.1.5 Proposed Artificial Intelligence Act (pAIA)[160]

The pAIA is a regulation on artificial intelligence (AI) that aims to ensure human-centric and ethical development of AI in the EU. The regulation sets out rules for AI systems and imposes obligations on the providers, importers and users thereof, based on the level of risk associated with the AI systems.[161]

The pAIA aims in particular to provide users (i.e. 'deployers')[162] of 'high-risk' AI systems with information to help them understand the systems they are using[163], make informed choices, and exercise their right to an effective remedy.[164] Article 13, for instance, requires providers of high-risk AI systems to pro-actively inform deployers of such systems by means of 'instructions for use'. These instructions should enable providers and deployers "to reasonably understand the system's functioning" and "all technical means available" are used to ensure that the system's output is "interpretable".[165] The instructions should enable to "understand and use the AI system appropriately".[166] The instructions must contain, for example,

[153] Article 29(1) DGA describes that the EDIB shall be established "in the form of an expert group, consisting of representatives of the competent authorities for data intermediation services and the competent authorities for the registration of data altruism organisations of all Member States, the European Data Protection Board, the European Data Protection Supervisor, ENISA, the Commission, the EU SME Envoy or a representative appointed by the network of SME envoys, and other representatives of relevant bodies in specific sectors as well as bodies with specific expertise. In its appointments of individual experts, the Commission shall aim to achieve gender and geographical balance among the members of the expert group."

[154] Article 29(2)(a)-(c) DGA.

[155] The tasks of the EDIB are listed in Article 30(a)-(m) DGA.

[156] Article 30(b)-(c) DGA.

[157] Article 30(h) DGA.

[158] Article 30(m) DGA.

[159] Article 25(1) DGA.

[160] For this report, we have used the pAIA version with amendments adopted by the European Parliament on 14 June 2023 to which we refer as pAIA: Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/01/06(COD) <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf>.

[161] The pAIA differentiates between prohibited AI-systems (title II), high-risk AI systems (title III) and transparency obligations (for certain AI systems) (title IV). In this study, we focus on high-risk systems, as the provisions in title III contain the most requirements that could provide access to data.

[162] Please note that the original proposal for the AI Act used the term 'user', which is defined in Article 2(4) of the original proposal as "any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity". However, the Council version from June replaced 'user' with 'deployer' in Article 2(1)(b) pAIA, maintaining the same definition.

[163] Recital 47 pAIA.

[164] Recital 47 of and paragraph 5.2.3 of the Explanatory Memorandum of the original proposal for an AI Act. Additionally, enhanced transparency on high-risk AI systems enables supervisory and enforcement authorities to fulfil their tasks, see paragraph 3.5 of the Explanatory Memorandum and Recitals 46-47 pAIA.

[165] Article 13(1) pAIA.

[166] Ibid.

information on the characteristics, capabilities, and limitations of performance of the system (e.g., the level cybersecurity), human oversight measures, and potential risks to health and safety or fundamental rights.[167] Deployers must ("where applicable") in turn use these instructions for use to comply with their obligation under the GDPR to carry out a DPIA and publish a summary of the DPIA.[168] Deployers, being best equipped to assess how the high-risk systems will be used in practice,[169] must carry out an additional fundamental rights impact assessment for high-risk AI systems.[170] All deployers carrying out a fundamental rights impact assessment are *encouraged* to make their assessment publicly available,[171] and public authorities or undertakings are required to do so.[172]

Lastly, Article 60 of the proposal states that a publicly accessible EU database for certain high-risk AI systems and foundation models shall be established. Providers of high-risk systems as referred to in Article 6(2), deployers of high-risk AI systems that are public authorities or EU institutions, gatekeepers[173] and providers of foundation models shall register their systems into the database and shall enter information on their systems in the database.[174] Foundation models are models trained on broad data at scale and are designed for generality of output and can be adapted to a wide range of tasks.[175] A specific type of foundation models are generative AI systems[176], which shall also be registered in the EU database.[177] The list of required information to be entered into the database can be found in Annex VIII and includes for instance contact details, arrangements for human oversight and a description of the data sources used in the development of the AI system.

The database may be a useful and direct source of information for many researchers from different disciplines, especially since the database will contain system-level information on high-risk AI systems in a wide range of fields such as migration, education, law enforcement and generative AI. The instructions for use for the high-risk AI system could be of particular interest for researchers in the field of AI that study the operations, methods of, and risks attached to, high-risk AI systems (system-level information), provided that these instructions will be publicly available. Unless researchers (or the research organisation they're affiliated with) act in capacity of a deployer, researchers would only be able to access those instructions indirectly, if deployers would provide them with these instructions. As the AI-Act is still a proposal, however, it is not entirely clear which types of information and data will become accessible to whom and how easy it will be for researchers to obtain such information.

3.1.6 Proposed European Media Freedom Act (pEMFA)[178]

[167] See for a full list of information that should be provided through the instructions for use Article 13(3) pAIA.

[168] Article 29(6) pAIA.

[169] Recital 58a (new) pAIA.

[170] Notably including a clear outline of the intended purpose of use, categories of people who are likely to be affected by using the AI system, the reasonably foreseeable impact on fundamental rights by using the AI system, and specific risks of harm likely to impact marginalised and vulnerable groups. The full list is set out in Article 29a(a)-(j) pAIA.

[171] Recital 58a (new) pAIA.

[172] Article 29a(5) (new) pAIA.

[173] As referred to in the Digital Markets Act.

[174] Article 51(1a) (new) pAIA and Article 28b(g) (new) pAIA. See also recital 69 pAIA.

[175] See for the full definition: Article 3(1)(1c) (new) pAIA

[176] Article 28b(4) (new) pAIA describes such models as "AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI")".

[177] Article 28b(2)(g) (new) pAIA.

[178] For this report, we have used the Council version of the pEMFA 'mandate for negotiation', to which we refer as: Proposal for a European Media Freedom Act (Council of the European Union, Interinstitutional File: 2022/0277(COD), Brussels, 21 June 2023 (OR.en) 110954/23) <https://data.consilium.europa.eu/doc/document/ST-10954-2023-INIT/en/pdf>.

The pEMFA is a proposed regulation that aims to protect media pluralism and media independence in the EU. To this end, the pEMFA contains rules on, amongst other things, transparency on media ownership, which is considered as a prerequisite for well-informed opinions and democratic participation.[179]

According to Article 6, media service providers are obliged to make information on their ownership (including shareholdings) accessible to the recipients of their services.[180] The pEMFA also introduces transparency obligations regarding the allocation of state resources so as to avoid covert subsidies and undue political influence in the media.[181] Pursuant to Article 24(2), public authorities or entities,[182] regulatory authorities, state-owned enterprises[183] must proactively make publicly available information about their state advertising expenditure allocated to media service providers.[184] This information shall be "accurate, comprehensive, intelligible, detailed and yearly".[185] Lastly, since Very Large Online Platforms (VLOPs)[186] nowadays function as 'gateways' for users in accessing media content online, the proposed EMFA also contains a provision for VLOPs to be transparent on the content moderation decisions (restrictions) they impose on media service providers.[187] VLOPs shall therefore proactively make publicly available "detailed" information on the number of restrictions or suspensions they imposed on media service providers based on their terms and conditions as well as the grounds for those restrictions, and the number of dialogues with media service providers[188] (Article 17(5)).

Although the pEMFA does not contain any provisions on direct access to information or data for researchers specifically, it does contain public transparency obligations that could also be of interest to (social) media researchers. Depending on the transparency reporting conducted by VLOPs, it remains to be seen whether the information in the reports will be more systemic (as we expect) or will also contain individual-level data.

3.2 Frameworks regulating the public sector (relationships)

3.2.1 Access to EU Documents Regulation (EUDR)[189]

The EUDR governs access to official documents[190] at the European level,[191] mirroring national access to documents regimes such as the *Wet Open Overheid* in the Netherlands.[192] It is used as an example here, because it has many elements that can be found in national access laws. The EUDR's main aim is to ensure the widest possible access to official EU documents.[193] Any EU citizen and any natural or legal person "residing

[179] Recital 19 pEMFA.

[180] We expect that this type of information will become publicly available (proactive access).

[181] Recital 49 pEMFA.

[182] Including national, federal or regional governments.

[183] Or other state-controlled entities or local governments of territories with more than 1 million inhabitants.

[184] Article 24(2) pEMFA. This information will include at least the legal names of the media service providers from which advertising services were purchased, the total annual amount spent, and the amounts spent per media service provider.

[185] Article 24(2) pEMFA.

[186] See Article 33 Digital Services Act.

[187] Recital 31 pEMFA.

[188] Cf. Article 17(4) pEMFA.

[189] Regulation (EC) No. 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission Documents (Access to EU documents Regulation).

[190] Article 3(a) EUDR defines a document as "any content whatever its medium (written on paper or stored in electronic form or as a sound, visual or audiovisual recording) concerning a matter relating to the policies, activities and decisions falling within the institution's sphere of responsibility.

[191] Held by the European Parliament, Council and the Commission.

[192] See also the Council of Europe Convention on Access to Official Documents (CETS No. 205) (*Tromsø Convention*).

[193] Article 1(a)-(c) EUDR lists the EUDR's three aims regarding access to EU documents.

or having its registered office" in the EU has the right to access EU documents (Article 2).[194] Access can be obtained directly through a register (in which the information sought has been proactively published) or following a written application (passive access).[195] Details on the application procedure are laid down in Articles 6-10; details on direct access through electronic registers are set forth in Articles 11-12. The EUDR does not mention research(ers) as a driver for transparency or access to information, although in practice (investigative) journalists and researchers constitute an important part of the user base. But access laws like the EUDR safeguard transparency for all citizens. All applicants must therefore be treated equally, no preferential access is given. The purpose of the use (e.g., research) is also irrelevant; this is another common feature of access laws. The idea is that the public interest in access is a given, that in principle need not be shown.

Like national laws, the EUDR contains various grounds on which access can be (or even must be) refused (Article 4). These include public security, the protection of judicial proceedings, legitimate commercial interest of third parties (e.g. a business that has had to provide confidential information to the EC), protection of privacy (e.g. of persons about whom documents contain details), etc. While certain categories documents are made available by default, a large amount of information is not.[196] When a request for access is denied, the requester can appeal the decision and ultimately go to the Court of Justice. These are lengthy and costly procedures. According to the European Ombudsman the current law does not meet the needs of citizens in the digital age; the EU institutions have not adapted to the realities of modern communication tools.[197] Because it is typically time-consuming (and costly) to go through access requests, and relatively little information is made publicly available pro-actively, the EUDR is of limited use to researchers. There have been attempts to modernise the EUDR, but these have so far failed. Meanwhile, at the level of Member States access laws have been updated especially where it concerns obligations of governments to proactively disclose information, and do so online and in standard formats.

3.2.2 Data Protection Law Enforcement Directive (LED)[198]

The LED governs the processing of personal data in the context of law enforcement. The rationale of its data access provisions is to enable data subjects to verify the accuracy of the data and the lawfulness of the processing, and to enable data subjects to exercise their rights conferred to them by the LED.[199]

Pursuant to Article 13, data controllers must communicate certain information to the data subject upon request, such as their identity and information on the purposes of the processing activities.[200] Based on Article 14, data controllers must also confirm to the data subject whether or not they are processing personal data. If so, the data controller must provide access to the personal data and provide information

[194] Article 2(1)-(2) EUDR. In some cases, (legal) persons outside of the EU may also be granted access (Article 2(2) EUDR).

[195] Article 2(4) EUDR. See also Article 12(2) on legislative documents specifically.

[196] An important factor in this is the fact that the EUDR enables access to 'documents', rather than information. If certain information is not captured in documents, it may not be covered by the EUDR. Because of these differences in terminology, it is not always clear what the scope of frameworks such as the EUDR are. See Rodriguez Lafuente 2022. For more information on the application of the EUDR, see 'Report from the Commission on the application in 2021 of Regulation (EC) No 1049/2001 regarding public access to European Parliament, Council and Commission documents', COM(2022) 498 final, 2022 and 'Draft twentieth annual report of the council on the implementation of Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 20 May 2001 regarding public access to European Parliament, Council and Commission documents', Council of the European Union, 8196/22, 2022.

[197] O'Reilly 2021.

[198] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing council framework decision 2008/977/JHA (Data Protection Law Enforcement Directive).

[199] Recital 43 LED.

[200] Article 13 LED.

regarding the purposes of the processing, their legal basis, what categories of data are concerned, who are the possible (other) recipients of the data, and so on. However, a "full summary" in an "intelligible form" is sufficient to comply with the right to data access.[201] Importantly, according to Article 15, Member States may adopt legislative measures that restrict data subjects' right to access.[202]

The type of data and the information that can be obtained through the LED access provisions may be interesting to researchers focusing on criminal law or criminology. However, as with the GDPR, the LED only provides *data subjects* with a right to data access, not third-party researchers. Researchers would therefore have to resort to data donation initiatives to acquire data from the data subjects. However, concerns have been raised over the potential lack of awareness (as data subjects may not always know that they are subject of a criminal investigation).[203] Moreover, empirical evidence suggests that compliance with the LED's right of access is very low in practice.[204] This, combined with the fact that researchers will be dependent on data subjects' willingness to share the obtained data – which may also not always be the case due to the sensitivity of the data – leads to the conclusion that the LED is arguably not the most useful instrument to obtain access to data for researchers.

3.2.3 Open Data Directive (ODD)[205]

The ODD is a directive that aims to stimulate the re-use of public sector information for the European economy and society.[206] Compared to its predecessor, the Public Sector Information Directive,[207] its scope has broadened. The current directive applies to (a) documents held by public sector bodies, (b) documents held by certain public undertakings,[208] and (c) research data.[209] An important limitation of the ODD, however, is that it only applies to documents access to which is not restricted or excluded under national rules on access to documents.[210] For example, the re-use obligations do not apply to documents containing e.g., commercially confidential information, which are typically excluded from access by national access

[201] Recital 43 LED.

[202] They may do so to (a) avoid obstructing legal inquiries, investigations or procedures, (b) "avoid prejudicing the prevention, detection, investigation or prosecution of criminal offences or the execution of criminal penalties", (c) protect public security, (d) protect national security and (e) protect the rights and freedoms of others, see Article 15(1) LED.

[203] Leiser and Custers 2019, pp. 374-375. See also: European Parliament, Vogiatzoglou and Marquienie 2022, p. 55 in reference to Article 29 Data Protection Working Party, *Opinion on some key issues of the Law Enforcement Directive (EU 2016/680)*, WP258, 29 November 2017, p. 18.

[204] Vogiatzoglou et al 2021.

[205] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (Open Data Directive). The Dutch implementation of the Open Data Directive will be transposed into an implementation Act, called 'Wet Implementatie Open Data Richtlijn'. For more information on the distinction between Directives and Regulations, please see Chapter II of this report. More information on the Dutch implementation of the Open Data Directive, see Kamerstukken II 2022/23, 36 382, nr. 2.

[206] Recital 4 ODD.

[207] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information (PSI Directive).

[208] Public undertakings are defined in Article 2(3) as "any undertaking active in the areas set out in point (b) of Article 1(1) over which the public sector bodies may exercise directly or indirectly a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules which govern it. A dominant influence on the part of the public sector bodies shall be presumed in any of the following cases in which those bodies, directly or indirectly: (a) hold the majority of the undertaking's subscribed capital; (b) control the majority of the votes attaching to shares issued by the undertaking; (c) can appoint more than half of the undertaking's administrative, management or supervisory body."

[209] Article 1(1) ODD.

[210] I.e., to documents that are not 'closed' for the public on grounds of e.g., security, data protection and commercial confidentiality. See Article 1(2)(a)-(l) in general and Article 1(2)(d), (e), (f), (h), Article 2(3) on national access regimes. See also Recital 23 ODD.

regimes.[211] Moreover, the directive does not cover documents protected by intellectual property rights from third parties[212]

Most of the provisions of the ODD relate to reactive access to information, specifying rules on request procedures. In principle, public sector bodies "shall make" a document available for re-use upon request (Article 4(1)), subject to certain exceptions as laid down in national access regimes and Article 1(2) of the directive. However, Article 5(2) states that documents falling within its scope should be made available "by design and by default", thus also encouraging more proactive access.

Articles 4-9 of the ODD contain (procedural) rules and conditions for re-use, e.g., with regard to the request process, formats, applicable charges and licences. Article 8 states that re-use shall in principle not be subject to conditions, unless those conditions are objective, proportionate, non-discriminatory and justified on grounds of a public interest objective.[213] To further stimulate simplified access to datasets and re-use, a 'single point of access' must be established (Article 9(2)). Article 12 stipulates that exclusive rights for re-use shall, in principle, not be granted.

A new element to the directive is the inclusion of research data into its scope. Pursuant to Article 10, Member States must adopt national policies that aim to have publicly funded research data be made openly available following the principle of 'open by default' and in accordance with the so-called FAIR[214] principles.[215] On top of that, publicly funded research data that have already been made publicly available through repositories must also be made re-usable for commercial and non-commercial purposes in accordance with the general rules of Chapters III and IV.[216]

Another novelty is the singling-out of high-value datasets[217], which are considered to provide "significant socioeconomic benefits" such as data related to the environment (e.g., on emissions), statistical data (e.g., on poverty, industrial production) and mobility data (e.g., on transport networks).[218] Article 14 prescribes that such high-value datasets should be available free of charge, machine readable, provided via APIs and as a bulk download where relevant.[219] The ODD prescribes a procedure that the Commission must follow to establish which data are to be classified as high-value datasets.[220]

Lastly, the ODD in Article 5(5) prescribes that where so-called "dynamic data" (e.g. sensor data) is made available for re-use, this must be done through suitable APIs, and where relevant, as a bulk download.[221] Real-time access to such data is important because its value for many uses depends on immediate availability. This is also true for research, although the availability of historic data is of course also of interest.

---

[211] Article 1(2)(d) ODD.

[212] Article 1(2)(c) ODD.

[213] Article 8(1) ODD. When licenses are used, 'standard licenses' should be available and encouraged (Article 8(2) ODD).

[214] FAIR stands for: Findable, Accessible, Interoperable and Re-usable, see Recital 27 ODD.

[215] Article 10(1) ODD.

[216] Legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights shall be taken into account (Article 10(2) ODD).

[217] Article 2(10) ODD defines 'high-value datasets' as "documents the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets". The thematic categories of high-value datasets are exclusively listed in Annex I of the Open Data Directive, i.e., (i) geospatial, (ii) earth observation and environment, (iii) meteorological, (iv) statistics, (v) companies and company ownership and (vi) mobility.

[218] European Commission and PwC EU Services 2023 (report), in which seven additional themes are identified as potential high-value datasets areas: Climate loss; Energy; Financial; Government and public administration; Health; Justice and Legal; Language.

[219] Article 14(1) ODD.

[220] Article 14(1) ODD. See the Commission Implementing Regulation (high-value datasets). See also a list of potential new high-value dataset themes, such as financial data, energy data, health data and climate loss data, in: European Commission and PwC EU Services 2023 (report).

[221] Dynamic data "means documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence; data generated by sensors are typically considered to be dynamic data" (Article 2(8) ODD).

The potential benefit of the ODD for researchers' access to data is that, compared to its predecessor,[222] it applies to more types of data and data holders. Researchers will therefore have more opportunities to re-use data held by the public sector.[223] However, various types of documents that might be interesting for use in research remain outside the scope of the ODD, e.g. data that is not subject to public access laws, or in which third parties hold intellectual property rights.[224] With respect to processes and practical aspects, the directive lowers barriers for re-use of public data by providing for single access points, and encouraging the use of standard licences.[225] An important financial aspect is that in principle re-use must be free of charge. For researchers, an improvement is also the regimes that provide for better access to high-value datasets and dynamic data (the latter can be high-value). Researchers and RPOs will benefit from being able to easily access and reuse more sources and larger amounts of real-time data provided via APIs.[226] Of note, datasets that are designated as high-value under the ODD will typically become publicly accessible (although the ODD does still allow for terms to be set for re-use)[227], when they may not be so now under Member States' law. Also important is that public sector bodies will (with very few temporary exceptions) not charge researchers for access and re-use.

Researchers also have an interest in accessing other researchers' data. Since Article 10 ODD obliges member states to have open access policies for research data, this should improve availability. The provision of Article 10(2) that allows re-use of research data in repositories will have an even more direct effect. In the longer term, the effectiveness of Article 10 will, inter alia, be dependent on whether these national policies adequately incentivise for research data to be published (e.g., along with research results) and whether for research data a proper balance is found at the national level between the public interest in accessible research data and safeguarding interests that necessitate the protection of know-how (e.g. through IPRs and trade secrets).[228] These derogations will be "crucial in determining the effectivity of the open access policies".[229] In sum, while the impact of the ODD on the availability of data for research is positive, it is likely that benefits may vary from Member State to Member States. A lot will depend on the details and effectiveness of national policies for open science, on the implementation of the ODD more generally and the scope of existing access regimes.

3.2.4 Data Governance Act, Chapter II (DGA)[230]

Building on the ODD, Chapter II of the DGA governs the re-use of certain categories of 'protected' data held by public sector bodies. Unlike the ODD, the DGA *does* cover documents that are protected on grounds of commercial confidentiality (including business, professional and company secrets), statistical confidentiality, the protection of third-party intellectual property rights, or the protection of personal data

[222] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, amended by Directive 2013/37/EU (Public Sector Information Directive).

[223] See also European Commission and Van Eechoud 2022, p. 14. However, still note that under Article 10 ECHR, public sector bodies are not obliged to allow re-use in all cases.

[224] In this regard, the Data Governance Act (discussed below) may be of interest as it broadens the scope of applicable documents.

[225] Standard licenses may be beneficial for researchers, as the European Commission and Van Eechoud note, because "the increased use of standardised licenses by public sector bodies will bring down transaction costs for researchers". Additionally, they note, standardised licenses may make it "easier to combinate data from different sources and share it for further (re)use", see: European Commission and Van Eechoud 2022, p. 25.

[226] European Commission and Van Eechoud 2022, p. 25.

[227] Broomfield 2023, p. 183.

[228] Gobbato 2020.

[229] Gobbato 2020, p. 153.

[230] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

(together: "protected public data").[231] In this way, Chapter II of the DGA aims to ensure the wider availability of data held by public sector bodies for re-use, in particular for research and innovation activities in the public interest.[232] Harmonised conditions for access to and use of protected public data were therefore deemed necessary.[233] However, in contrast to the ODD, the DGA does not oblige public sector bodies to make data available for re-use (a fortiori, users have no right to re-use).[234] It does lay down rules in case public sector bodies do grant access for re-use of such data.

Article 5 of the DGA sets out similar principles as the ODD, i.e. that where re-use is allowed, any terms and conditions shall be "non-discriminatory, transparent, proportionate and objectively justified with regard to the categories of data and the purposes of re-use and the nature of the data for which re-use is allowed". Public sector bodies must make public what terms and conditions they impose on re-users. Because protected data is involved (e.g. personal data), public sector bodies must ensure that the protected nature is safeguarded, and they can also impose conditions on others to ensure it. For example, a public sector body can anonymize personal data, and make the anonymized data available for re-use on the condition that re-users do not seek to de-anonymize the data. Where commercially confidential information is concerned (including trade secrets) or data protected by intellectual property rights, a public sector body could seek permission from the third party, or modify data so as to remove confidential nature .[235] The protected nature can also be safeguarded by requiring (vetted) users to access and use data in a secure processing environment.[236] Similar to the ODD, the DGA introduces a **'single information point'** through which all information concerning the conditions and fees for re-use should be made easily available (Article 8(1)), and through which re-use requests can be channelled.[237]

As noted above, the DGA complements the ODD by establishing re-use rules for protected public sector information. This could lead to the wider availability (due to the extended scope of the DGA) of data held by public sector bodies and thus, to increased opportunities for researchers who wish to obtain access to public sector datasets for re-use.

3.3 Frameworks regulating the private sector (relationships)

3.3.1 E-Commerce and consumer law: Consumer Rights Directive (CRD)[238], E-Commerce Directive (ECD),[239] Services Directive (SD)[240]

[231] Article 3(1) DGA. See also European Commission and Van Eechoud 2022, p. 26 stating that the DGA '*complements*' the ODD.

[232] Recital 6 DGA.

[233] Ibid.

[234] See also European Commission and Van Eechoud 2022, p. 26.

[235] Article 5(3)(a) DGA.

[236] Article 5(3)(b) DGA.

[237] Article 8(2) jo. Article 7 DGA (in which the rules on the competent bodies for requests for re-use are laid down).

[238] Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council Text with EEA Relevance (Consumer Rights Directive), amended by Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules (Modernisation Directive).

[239] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internet Market (E-Commerce Directive).

[240] Directive 2006/123/EC of the European Parliament and of the Council of 12 December 2006 on services in the internal market (Services Directive).

The CRD, ECD and SD provide the legal framework for the (online) offering and purchasing of goods and services in the Union. In EU e-commerce and consumer law, the provision of information is considered necessary to enable consumers to protect themselves as well as to stimulate quality improvement and competition, thereby assuming that consumers can better compare goods based on the provided information.[241]

According to Article 6 of the CRD, general information such as the trader's identity and contact details, the total price of goods or services inclusive of taxes, and information on arrangements for payment, delivery and performance must be provided[242] both in the context of distance contracts[243] and off-premises contracts.[244] In case of contracts that were concluded via electronic means, services providers must in addition provide information on e.g., the different technical steps to follow to conclude the contract and the technical means for identifying and correcting input errors prior to the placing of the order.[245]

Based on Article 22 of the SD, service providers must make certain categories of (general) information available to its recipients in a proactive manner.[246] Other, additional information must only be made available upon request.[247]

Lastly, the ECD contains specific requirements for providers of information society services. According to Article 5 of the ECD, general information must be provided to recipients of information society services.[248] In the case of commercial communications, providers must be explicit about this and disclose their identity.[249]

Researchers focusing on consumer protection and consumer behaviour might be interested in the information mentioned above. Since many provisions oblige data providers to provide the information proactively and through their public online interfaces, these researchers can easily access system-level information. Such information is likely to be generic rather than consumer-specific.

---

[241] This dual objective is something that is well-established in consumer protection scholarship, see e.g., Devenney and Kenny 2012, p. 369.

[242] Notably, it is *not* sufficient to provide the mandatory pre-contractual information merely as part of the general terms and conditions that the consumer may have to accept before moving on in the transaction process; the information must really be brought to the attention of the consumer. See Commission Commission guidance CDR, p. 22-23.

[243] A distance contract is "any contract concluded between the trader and the consumer under an organised distance sales or service-provision scheme without the simultaneous physical presence of the trader and the consumer, with the exclusive use of one or more means of distance communication up to and including the time at which the contract is concluded", see Article 2(7) CRD.

[244] An off-premises contract is "any contract between the trader and the consumer (a) concluded in the simultaneous physical presence of the trader and the consumer, in a place which is not the business premises of the trader; (b) for which an offer was made by the consumer in the same circumstances as referred to in point (a); (c) concluded on the business premises of the trader or through any means of distance communication immediately after the consumer was personally and individually addressed in a place which is not the business premises of the trader in the simultaneous physical presence of the trader and the consumer; or (d) concluded during an excursion organised by the trader with the aim or effect of promoting and selling goods or services to the consumer", see Article 2(8) CRD.

[245] Article 10(1)(a)-(d) ECD lists the information that must be provided: the different technical steps to follow to conclude the contract; whether or not the concluded contract will be filed by the service provider and whether it will be accessible; the technical means for identifying and correcting input errors prior to the placing of the order, the languages offered for the conclusion of the contract. Article 10(2) continues by stating that the provider shall also indicate whether there are any relevant codes of conduct to which the provider subscribes and how those codes can be consulted electronically; and on contract terms and general conditions – in a way that allows the recipient to store and reproduce them (Article 10(3) ECD). See also Article 8(2) CRD.

[246] Including details on the service providers' identity, general conditions and clauses, the price of the service and main features of the service, see Article 22(1) SD.

[247] This includes, e.g., information on the method uses for calculating the price, (where applicable) a reference to the professional rules of a regulated profession, and any codes of conduct to which the service provider is subject, see Article 22(3) SD.

[248] This relates merely to administrative information, such as the (e-mail) address, trade register information and tax and VAT information.

[249] Article 6 ECD.

3.3.2 Platform-to-Business Regulation (P2BR)[250]

The P2BR aims to ensure "a fair, predictable, sustainable and trusted online business environment" for smaller businesses and traders on online platforms.[251] It partly does so by imposing transparency obligations on online platforms in relation to their business users.[252] Through transparency, business users should be able to trust the platforms they are dependent on and better understand platform decisions that could significantly affect them.[253]

According to Article 4, platforms must provide a statement of reasons to a given business user in case they decide to restrict or suspend the provision of its online intermediations services to the business user concerned (individual-level data). Article 5(1) obliges platforms to provide information in their terms and conditions on their ranking mechanisms, i.e., recommender systems (system-level data). Article 11 gives business users the right to lodge a complaint through an internal complaint-handling system and receive the outcome of the complain-handling process "in an individualised manner" (individual-level data).[254] Additionally, the general public must be informed of internal complaint-handling systems, in particular on their functioning and effectiveness (system-level data, Article 11(4)).[255]

The relative importance of the P2BR for researchers seems limited, as the intended data recipients of the P2BR are mainly business users. Researchers would therefore be dependent on those (affected) business users to gather individual-level data from them. The public transparency provisions, however, might provide a gateway to potentially interesting system-level information. In sum, however, since the P2BR does not contain major transparency reporting obligations for platforms, the importance of the P2BR seems to be fairly low for researchers.

3.3.3 Digital Markets Act (DMA)[256]

The DMA is an instrument of economic regulation that aims to ensure a contestable and fair online platform economy.[257] Of particular relevance for this report, are the many transparency rules that the regulation

[250] Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (Platform to Business Regulation).

[251] Recital 7 P2BR and European Commission Platform-to-business trading practices (webpage) <https://digital-strategy.ec.europa.eu/en/policies/platform-business-trading-practices>.

[252] Pursuant to Article 2(1) of the P2BR, a business user is "any private individual acting in a commercial or professional capacity who, or any legal person which, through online intermediation services offers goods or services to consumers for purposes relating to its trade, business, craft or profession".

[253] Recitals 22-24 P2BR. It should be kept in mind that large platforms have a strong position in relation to their users, as platforms oftentimes function as important and sometimes indispensable infrastructures for business users to reach their potential clients. The fact that platforms are entitled to set the terms and conditions based on which they may impose restrictions unilaterally, helps to sustain this strong position.

[254] Article 11(2)(c) of the P2BR.

[255] This includes "the total number of complaints lodged, the main types of complaints, the average time period needed to process the complaints and aggregated information regarding the outcome of the complaints", see Article 11(4) of the P2BR.

[256] Regulation (EU) 2022/1925 of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act).

[257] Article 1(1) DMA. European Commission DMA press release 2022 (webpage) <https://ec.europa.eu/commission/presscorner/detail/en/ip_22_6423>. The DMA's rationale lays in the economic and internal-market-functioning sphere.

imposes on "core platform services"[258] provided or offered by "gatekeepers"[259] both in relation to "business users"[260] and "end users"[261] of the services.[262]

For instance, gatekeepers must provide specific business users, namely advertisers and publishers to whom they supply online advertising services, with information – upon request and free of charge – that allows both parties to understand the price paid for each of the different advertising services provided (Article 5(9)-(10)). Furthermore, Article 6(8) obliges gatekeepers to provide advertisers and publishers, upon request and free of charge, with access to the performance measuring tools of the gatekeeper and the data necessary for advertisers and publishers to carry out their own independent verification of the advertisements inventory, including aggregated and non-aggregated data.

In relation to business users more generally (or third parties authorised by a business user), gatekeepers must provide them – upon request and free of charge – with effective, high-quality and real-time access to aggregated and non-aggregated data that is provided for or generated in the context of the use of the relevant core platforms services, or the services provided by the business users together with the core platforms services. With regard to personal data, however, access to the data only has to be provided where the data are *directly connected* with the use of end users in respect of the products or services offered by the relevant business user and when the end users *opt in* to such sharing by giving their consent (Article 6(10)). Individual researchers or research institutions who wish to access the aforementioned data have the option to ask business users for an authorisation to request the data.

In relation to end users, or third parties authorised by an end user, Article 6(9) prescribes that gatekeepers must provide them – upon request and free of charge – with "effective portability" of data provided by the end user or generated through the activity of the end user in the context of the use of the core platform service. To this end, end users must be given tools to facilitate the effective exercise of data portability and be given continuous and real-time access to the data. Individual researchers or research institutions who wish to access these data can either register themselves as end users of a platform service and in that capacity request data portability (including real-time access to data), or approach other end users and ask for an authorisation to port and access the data on their behalf.

Finally, system-level information can be obtained through the Commission's publicly available annual reports with information on, inter alia, market investigations related to the DMA and findings from monitoring gatekeepers' implementation of DMA obligations (Article 35(1)-(2)). Gatekeepers must also apply transparent conditions about their ranking practices (Article 6(5)) so as to ensure that they do not engage in any form of differentiated or preferential treatment in favour of products that it offers itself.[263]

---

[258] Pursuant to Article 2(2) DMA, a core platform service "means any of the following: (a) online intermediation services; (b) online search engines; (c) online social networking services; (d) video-sharing platform services; (e) number-independent interpersonal communication services; (f) operating systems; (g) web browsers; (h) virtual assistants; (i) cloud computing services; (j) online advertising services, including any advertising networks, advertising exchanges and any other advertising intermediation services, provided by an undertaking that provides any of the cire platform services listed in points (a) to (i);".

[259] Gatekeepers are defined in Article 2(1) DMA as "undertaking providing core platform services, designated pursuant to Article 3". Article 3 states that an undertaking shall be designated as a gatekeeper if (a) it has a significant impact on the internal market; (b) it provides a core platform service which is an important gateway for business users to reach end users; and (c) it enjoys an entrenched and durable position, in its operations, or it is foreseeable that it will enjoy such a position in the near future.

[260] Business users are defined in Article 2(21) DMA as "any natural or legal person acting in commercial or professional capacity using core platform services for the purpose or in the course of providing goods or services to end users".

[261] End users are defined in Article 2(20) DMA as "any natural or legal person using core platform services other than a business user".

[262] See also: Edelson, Graef and Lancieri 2023.

[263] See also Recital 52 DMA. We assume that transparency about ranking practices refers to *public* transparency.

3.3.4 Digital Services Act (DSA)[264]

Similar to the P2BR and the DMA, the DSA is an instrument of platform regulation. The DSA aims to ensure a safe and transparent online environment, protecting people's fundamental rights online.[265] One of the strategies to achieve this, is transparency, both in regard to the functioning of online platforms and search engines (system-level data) and individuals' specific uses of the platform service (individual-level data). In order to do so, the DSA contains a large number of transparency provisions, only a selection of which will be discussed here.[266]

Examples of potentially relevant transparency provisions providing for individual-level data are Article 17, which requires hosting services to proactively provide their users with statements of reasons when they impose restrictions on content, and Article 20(5), which requires online platforms to inform users who submitted a complaint of their "reasoned decision" on the complaint (without the need of a request).[267] Anonymised versions of these statements of reasons will be included in a publicly accessible machine-readable database managed by the European Commission, pursuant to Article 24(5).

The DSA also contains provisions that force platforms to be more open towards the general public about their systems more broadly. First, there is the general Article 14(1), which urges intermediary services to include information on any restrictions they impose in respect of content provided by users in their terms and conditions.[268] Furthermore, Article 15 and Article 24(1) prescribe that intermediary services and online platforms respectively publish reports on the content moderation they engage in. Another example is Article 42(4), which requires that the largest platforms (Very Large Online Platforms or VLOPs[269]) publish reports in which they describe the results of their risk assessments as well as the underlying audit reports in which compliance with the DSA is assessed.[270] Finally, VLOPs are also obliged by Article 39 to establish a publicly available ad library, i.e., a repository in which information on advertisements that are and have been presented on the platform shall be stored. All these public transparency obligations may also help researchers to gain a broader understanding of the functioning of online platforms (system-level data).

Of special relevance for researchers, however, is Article 40(4) on direct access to data from Very Large Online Platforms (VLOPs) and Very Large Search Engines (VLOSEs) for "vetted researchers".[271] According to this provision, vetted researchers can, with the help of national authorities, request access to

[264] Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act).

[265] Recitals 3, 9 DSA.

[266] See also: Edelson, Graef and Lancieri 2023.

[267] Apart from Article 14, which is aimed at the general public, the provisions as discussed before are all result in individual-level data.

[268] Including information on their content moderation, algorithmic decision-making and human review, see Article 14(1) DSA.

[269] The "Very Large Online Platforms", with more than 45 million average monthly active recipients in the EU (Article 33(1) DSA).

[270] Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) are subject to additional obligations (as laid down in Chapter 3, Section 5 of the DSA), part of which is to assess and mitigate accordingly any systemic risks that stem from their services. See also: Leerssen 2023b.

[271] The status of "vetted researcher" is granted by the Digital Services Coordinator of establishment, upon a duly substantiated application from researchers, for the specific research referred in the application (Article 40(8) DSA). To be awarded this status, researchers must demonstrate that (a) they are affiliated to a research organisation within the meaning of the Copyright in the Digital Single Market Directive, (b) they are independent from commercial interests, (c) their application discloses the funding of the research, (d) are capable of fulfilling specific data security and confidentiality requirements and to protect personal data, (e) their access to data and the time frames requested are necessary and proportionate to the purposes of their research, (f) the planned research activities contribute to the detection, identification and understand of systemic risks, and (g) they have committed themselves to making their research results publicly available free of charges.

data,[272] provided that they conduct research on "systemic risks"[273] that the platforms' and search engines' systems and services pose to society.

In addition, Article 40(12) requires that VLOPs and VLOSEs give a broader category of researchers – i.e., those who meet the conditions as set out in the provision, including those affiliated to not for profit bodies, organisations and associations – access to data that is "publicly accessible in their online interface", again, for the sole purpose of systemic risk research.[274] In contrast to non-publicly accessible data, access to publicly accessible data is provided directly, without the intervention of a national regulatory authority. Some scholars have argued that Article 40(12) DSA should legally allow the practice of web scraping on VLOPs and VLOSEs, or that the legal status of web scraping should at least be clarified, given its importance for research.[275] Web scraping refers to the automated extraction or copying of publicly available information online, a popular research strategy especially used in the social sciences.[276]

In sum, Article 40 of the DSA has significant potential to positively affect the position of researchers in terms of access to online platform data.[277] Especially considering the fact that there are few *a priori* restrictions on what data can be asked for, apart from a number of legal carve-outs (notably for information protected under data protection and trade secrets). It should however be noted that the DSA has yet to become fully applicable, and its implementation will be further shaped by delegated acts and national authorities (so-called Digital Service Coordinators)[278], which means that the DSA's practical impact on data access for research remains to be seen.

3.3.5 Proposed Political Advertising Regulation (pPAR)[279]

The pPAR is a proposed regulation that aims to enhance transparency of (the targeting of) political advertising. According to the proposal, a high level of transparency is necessary, among others, "to support an open and fair political debate and free and fair elections or referendums and to combat disinformation

---

[272] Through appropriate interfaces as specified in the access request, including online databases or application programming interfaces, see Article 40(7) DSA.

[273] The DSA differentiates between four categories of systemic risks: (a) the dissemination of illegal content through their [VLOPs'] services; (b) any actual or foreseeable negative effects for the exercise of fundamental rights , in particular the fundamental rights to human dignity enshrined in Article 1 of the Charter, to respect for private and family life enshrined in Article 7 of the Charter, to the protection of personal data enshrined in Article 8 of the Charter, to freedom of expression and information, including the freedom and pluralism of the media, enshrined in Article 11 of the Charter, to nondiscrimination enshrined in Article 21 of the Charter, to respect for the rights of the child enshrined in Article 24 of the Charter and to a high-level of consumer protection enshrined in Article 38 of the Charter; (c) any actual or foreseeable negative effects on civic discourse and electoral processes, and public security; and (d) any actual or foreseeable negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person's physical and mental well-being (Article 34(1)(a)-(d) DSA).

[274] Article 40(12) is not restricted to 'vetted researchers' but includes researchers "affiliated to not-for-profit bodies, organisations and associations, who comply with the conditions set out in paragraph 8, points (b), (c), (d) and (e)". See for further discussion of the DSA's systemic risk approach: Eder 2023 and G'sell 2023.

[275] See the responses to the public consultation regarding the DSA: <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act>. See for instance, the responses sent in by the The Amsterdam School of Communication Research (ASCoR) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3423477_nl>; dr. E. Borra, dr. S. Peeters and dr. B. Rieder <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3423780_nl>; Academic Researcher Members of the EDMO Working Group on Platform-to-Researcher Data Access <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act/F3423820_nl >.

[276] Luscombe, Dick and Walby 2022.

[277] For an in-depth discussion of the potential of Article 40 DSA, see Leerssen 2023a. Because of the large potential of Article 40 DSA, the provision is further discussed in sections 4.4.1 and 5.1.

[278] See Articles 49-51 DSA and Table 4 in section 5.4.

[279] Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (Political Advertising Regulation), COM(2021) 731 final.

and unlawful interference from abroad".[280] The rules should further enable voters, in particular, to better understand when and by whom a political advertisement is presented, and how they have been targeted.[281]

Article 7(1) prescribes that each political advertisement shall be made available together with certain information, that is, a statement that the communication it is a political advertisement, the identity of the sponsor, and a transparency notice which enables the wider context and the aims of the advertisement to be understood.[282] In addition, Article 7(6) requires that information on political advertisements is included in the DSA's ad library mentioned above.[283]

The pPAR also specifically addresses researchers, by stressing the importance of providing specific entities such as "vetted researchers", journalists, civil society organisations and accredited election observers with information on political advertising.[284] Pursuant to Article 11, the information as mentioned in Articles 6 and 7 can be requested directly by vetted researchers[285] and other "interested parties".[286] Providers of political advertising services "shall take the appropriate measures" to transmit the information without costs[287] and "shall make best efforts to provide the requested information or its reasoned response" within one month.[288] Notably, whereas platform data under the DSA can be provided to (vetted) researchers for the sole purpose of systemic risk research, the pPAR does not contain such a purpose limitation. Researchers do not have to disclose the purpose of obtaining the information.

Lastly, Article 12 requires that data controllers[289] adopt internal policies in which they describe the use of targeting and amplification techniques, if they use such techniques and process personal data.[290] A link to this internal policy must also be included in the transparency notice as required by Article 7.[291] Moreover, controllers must include in the transparency notice any information necessary to allow the individual concerned[292] to understand the logic involved, the main parameters of the technique used, and the use of third-party data and additional analytical techniques.[293] This information, too, can be requested by interested parties such as researchers (Article 13).[294]

It should be kept in mind, however, that the pPAR does not guarantee researchers' access to data, but merely grants a right to *request* access, upon which the service provider only has to make "best efforts" to provide the data. This is in contrast with Article 40(4) of the DSA, which prescribes that VLOPs "shall" in principle provide access to data upon request, except for when they (a) do not have access to the data, or (b) giving access would lead to significant vulnerabilities in the security of their service or the protection of

---

[280] Recital 4 pPAR.

[281] Ibid.

[282] This 'transparency notice' is "to enable the wider context of the political advertisement and its aims to be understood, or a clear indication of where it can be easily retrieved", see Article 7(1)(c) pPAR. Article 7(2)(a)-(g) includes a list of additional information that shall be included in the transparency notice.

[283] Article 39 DSA.

[284] This would allow these entities to "support the performance" of their respective roles in democracies and democratic processes, see Recital 46 pPAR.

[285] Article 11(2)(a) pPAR.

[286] Pursuant to Article 6, providers of political advertisement services shall retain certain information related to its advertisements, e.g., the amounts the provider of the service invoices for its service, the value (or other benefits) it received in exchange for its service, the identity of the sponsor, including its contact details (where applicable) and communicate that to the political advertisement publisher.

[287] Article 11(1) pPAR.

[288] Article 11(3) pPAR. In this regard, Article 11(3) pPAR differs from Article 40(4) DSA, which prescribes that VLOPs "**shall**" provide access to data (as opposed to "**shall make best efforts**").

[289] Data controller as referred to in the GDPR. Controller is defined in Article 2(12) pPAR as "a controller according to Article 4(7) of Regulation (EU) 2016/679 or, where applicable, to Article 4(8) of Regulation (EU) 2018/1725".

[290] Article 12(3)(a) pPAR.

[291] Article 12(3)(a) jo. Article 12(4) pPAR.

[292] I.e., those who are presented with the advertisements and are affected by the targeting and/or amplification techniques.

[293] Article 12(3)(c) jo. Article 12(4) pPAR.

[294] Article 13(1)-(2) jo. Article 11 pPAR.

confidential information.[295] According to the pPAR, a data access request can be refused on the ground that the request is "manifestly unfounded, unclear or excessive", in particular because of a "lack of clarity".[296] Furthermore, the service provider may decide to provide the data in aggregated form or place them in a range, to the extent necessary to protect the service provider's commercial legitimate interests.[297] It remains to be seen whether the protection of "commercial legitimate interests" as listed in Article 11(4) will hinder effective access to data.[298] Although the pPAR is still a proposal, the regulation seems promising as it explicitly identifies researchers as interested parties that can request information. Additionally, the public transparency obligations may also be useful for researchers, albeit for specific (political) research questions only.

3.3.6 Proposed Data Act (pDA)[299]

The pDA is a proposed regulation that mainly aims to remove barriers to data sharing.[300] In doing so, the pDA governs a variety of topics and introduces rules for different contexts, including business-to-consumer, business-to-business and business-to-government data sharing. The pDA also has a special focus on access to, and use of, data generated by the use of so-called 'smart products', i.e., physical products that obtain, generate or collect data concerning their use or environment and that are able to communicate directly or indirectly with the Internet,[301] which leads to 'Internet-of-Things (IoT) data'.[302]

Chapter II of the pDA contains provisions on the basis of which IoT-data can be accessed. According to Articles 3 and 4, consumers should be able to easily access data generated by the use of their products.[303] In case the data cannot be directly assessed by the user, the data holder[304] must make the data generated by the product's use available to the user based on a "simple request".[305] Moreover, data holders can be requested by a product user to make data available to a third party, for instance an individual

[295] Article 40(5) DSA.

[296] Article 11(5) pPAR. It is for the service provider to prove that the request is manifestly unfounded, unclear or excessive.

[297] Article 11(4) pPAR.

[298] Reference is made to section 5.3 of this report in which the topic of protection of third-party interests (such as commercial legitimate interests) may interfere with data access rights.

[299] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM/2022/68 final. For this report, we have used the latest available version of the proposal (pDA) at the time of writing, i.e., the Council Mandate: Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) – Mandate for negotiations with the European Parliament, 17 March 2023, 7413/23 <https://data.consilium.europa.eu/doc/document/ST-7413-2023-INIT/en/pdf>.

[300] See Recital 4 pDA. According to Article 1(1a) pDA, the pDA covers the following types of data and the following contexts: (a) data concerning the performance, use and environment of products and related services; (b) private sector data that is subject to statutory data sharing obligations; (c) private sector data accessed and used on the basis of contractual agreements between businesses; (d) private sector data with a focus on non-personal data; (e) data processed by data processing services; and (f) non-personal data held in the Union by providers of data processing services.

[301] Article 2(2) and recital 14 pDA.

[302] I.e., physical products that by means of their components or operating system obtain, generate or collect data concerning their use or environment and that are able to communicate directly or indirectly via the Internet. Such smart products, or 'Internet of Things' (IoT)-products, may include vehicles, home equipment and consumer goods such as fridges and voice assistances, medical health and lifestyle equipment such as smart watches, or agricultural and industrial machinery, see Article 2(2) and Recital 14a pDA. IoT data are potentially valuable to users and to society as a whole to support innovation and the development of a digital environment.

[303] Article 3(1) pDA. Notably, before concluding a contract for the purchase of a product, the data holder must provide information to the user, inter alia on the data that are likely to be generated and how these data can be accessed by the product user, see Article 3(2) pDA.

[304] Article 2(6) defines data holders as "a legal or natural person who (i) has the right or obligation, in accordance with this Regulation, applicable Union law or national legislation implementing Union law, to make available certain data or (ii) can enable access to the data through control of the technical design or means of access, in the case of non-personal data".

[305] Article 4(1) pDA. Any agreement between the data holder and user shall not be binding "when it narrows the access rights pursuant to paragraph 1" of Article 4 pDA, see Article 4(1a) pDA.

researcher or research organisation (Article 5, indirect access).[306] The third party may only process the data, including the metadata necessary to interpret and use the data,[307] for the purposes and under the conditions agreed upon with the user.[308] In the Council version of the pDA, however, two paragraphs have been added detailing how data holders may refuse data access requests in "exceptional circumstances" when they can demonstrate "that it is highly likely to suffer serious damage from the disclosure of trade secrets, despite the technical and organisational measures taken by the user".[309]

Chapter V of the pDA contains a specific provision with regard to business-to-government data sharing which could be of interest to researchers as well. Following Article 21, public sector bodies and certain EU institutions[310] are entitled to share data they have received in the past in the context of "exceptional needs"[311] with individual researchers and research organisations.[312] Recital 68 clarifies that data may be shared with researchers in the event a public sector body cannot perform the research itself.[313] The designated researchers and research organisations may however exclusively use the data on a not-for-profit basis or in the context of a public-interest mission.[314] Indeed, researchers do not have an actionable claim to the data and are dependent on public sector bodies to actually share the information/data with them.

Chapter III of the pDA does not contain provisions carrying rights to data access as such but provides for horizontal rules on *modalities* of access to data, whenever a data holder is obliged by law to make certain data available to a data recipient (business-to-business sharing).[315]

Finally, Article 23 and 24 pDA contain rules on data portability. In particular, providers of data processing services are required to develop and implement contractual terms for, and may not raise any obstacles inhibiting customers to port (meta)data created by the customer (through the use of the service) to another provider of data processing services or to an on-premise system. The contractual clauses must also prescribe that the provider shall assist the customer and, where technically feasible, complete the porting process and ensure that throughout the porting process a high level of security is maintained.[316]

Overall, the pDA's possible benefit for researchers can mainly be found in Chapter II on IoT data, which would enable researchers to access a range of individual-level IoT data. In addition, the regulatory approach to data portability could potentially contribute to more efficient data access, i.e., via data donation

---

[306] Article 5(1) pDA.

[307] Article 5(1) pDA.

[308] Article 6(1) pDA. The obligations of the third parties receiving the data at the request of the product user are laid down in Article 6 pDA.

[309] Articles 4(3a) and 5(8a) pDA. The refusal must be duly substantiated and be provided in writing and with undue delay. The data holder must also notify the national competent authority about the refusal, see Article 4(3a) pDA.

[310] The European Commission, the European Central Bank and the Union.

[311] Article 15(a)-(c) pDA explains what qualifies as an exceptional need. It shall be deemed to exist "(a) where the data requested is necessary to respond to a public emergency; (b) where the data request is limited in time and scope and necessary to prevent a public emergency or to assist the recovery from a public emergency; (c) where the lack of available data prevents the public sector body or Union institution, agency or body from fulfilling a specific task in the public interest that has been explicitly provided by law; and (1) the public sector body or Union institution, agency or body has been unable to obtain such data by alternative means, including by purchasing the data on the market at market rates or by relying on existing obligations to make data available, and the adoption of new legislative measures cannot ensure the timely availability of the data; or (2) obtaining the data in line with the procedure laid down in this Chapter would substantively reduce the administrative burden for data holders or other enterprises".

[312] Article 21(1) pDA.

[313] Although it is not entirely clear whether this is the *only* reason data may be shared with researchers, as this is not clarified in Article 21 pDA.

[314] Article 21(2) pDA.

[315] See Chapter III and Recital 38a pDA.

[316] Please note that the rules in the pDA on data portability build on the self-regulatory approach of the Free Flow of Non-Personal Data Regulation (NPDR), which provides the basis for voluntary codes of conduct on data porting. The European Commission has established that the self-regulatory frameworks developed under the NPDR have had "limited efficacy" so far, and that "a set of minimum regulatory obligations on providers of data processing services" is deemed "necessary" to "eliminate (…) barriers to effective switching between data processing services" (recital 70 pDA). This is why the EU legislator has decided to introduce binding rules on data portability in the pDA rather than relying on the NPDR's regulatory approach.

initiatives (where customers port data to data processing services used for scientific research).[317] However, the protection of trade secrets and IPRs has gained a strengthened position under the Council version of the pDA, with a potential decrease in value of the provisions in terms of data access for researchers.[318]

3.3.7 2022 Strengthened Code of Practice on Disinformation (2022 CoP)[319]

Last but not least, the 2022 CoP is a code of conduct, i.e., a 'soft law' instrument, to which providers of online platforms can adhere with a view to combatting the dissemination of disinformation[320] on their platforms.

An important aspect of the 2022 CoP is researchers' access to data.[321] In contrast to legal frameworks, the 2022 CoP contains so-called "commitments" – rather than legal obligations – for 'signatories', i.e., actors that adhered to the 2022 CoP (semi-)voluntarily. Four of those commitments are dedicated to researchers' access to data and are laid down in Chapter VI titled 'Empowering the research community'. In a nutshell, the relevant signatories of the Code have committed themselves to provide public access to data for research purposes on disinformation (Commitment 26); to provide "vetted researchers" with access to data necessary for research on disinformation by developing an independent third-body to vet researchers and their proposals (Commitment 27); to support in good faith research into disinformation involving their services (Commitment 28); and lastly, to conduct research themselves and share datasets and research findings with relevant audiences and the wider public (Commitment 29).

The commitments dedicated to empowering the research community in the 2022 CoP are welcome and potentially useful for research. Their utility however, will largely depend on whether relevant actors will actually be or remain a signatory – adherence to (specific commitments of) the 2022 CoP is largely voluntary – and how the commitments will be implemented in practice.

[317] This is further discussed in section 4.2.

[318] This is further discussed in sections 4.3.1 and 5.3.3.

[319] The 2022 Strengthened Code of Practice on Disinformation (2022 CoP). Different than the legislation analysed in this study, this Code is not legally binding, but voluntary of nature. Stakeholders can decide to become a signatory to the Code and adhere to it. It is a form of self-regulation for non-VLOPs and non-VLOSEs yet co-regulatory of nature for VLOPs and VLOSEs that aim to tackle disinformation (in case they have become a signatory).

[320] According to the 2022 CoP, "disinformation is false or misleading content that is spread with an intention to deceive or secure economic or political gain and which may cause public harm", see the Preamble of the Code, in reference to Communication COM(2020) 790 final from the Commission on the European Democracy Action Plan (the European Democracy Action Plan).

[321] 2022 CoP <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>.

4. Relevance of transparency and data access provisions for researchers

All legal frameworks addressed in chapter 3 show potential in the broader quest for access to data for research purposes. However, the different legal frameworks may help researchers in different ways, and some provisions could be more useful than others depending on a variety of factors, such as research questions, methods or practical challenges faced. The frameworks can be categorised according to their potential utility or "relevance" as tools to access data for research purposes. In a nutshell, for our purposes here, frameworks can be relevant for research because they:

A. **Provide researchers with direct access to data**, either because the provisions grant access to researchers specifically ("researchers' access"), or to the general public ("general public access") which also includes researchers, or to specific groups of people ("specific persons") which under certain circumstances may also include researchers; and/or
B. **Enable data donation-based research projects**; and/or
C. **More generally contribute to an enabling environment for access to data for research purposes**, via
    i. the introduction of **data sharing intermediaries and facilitators**, and/or
    ii. the introduction of **other favourable conditions**, such as principles of openness, low fees, standard formats, and so forth.

Before discussing each of these categories in more detail, we would like to make three remarks. First, our analysis of the potential utility of the legal frameworks for researchers is based on the *text* of the regulations rather than empirical evidence, notably because of the relative novelty of many of the legislative instruments. Where appropriate, however, we refer to existing empirical studies. Second, some of the discussed frameworks are still in the proposal phase or are not (fully) applicable yet. Although the gist of these frameworks is unlikely to be changed fundamentally – which is why the frameworks have been included for examination in the first place – they still might undergo changes during negotiations. And third, it should be noted that legislation can be useful for multiple reasons, for instance by providing specific direct data access rights (A) *and* by creating opportunities for data donation (B).

The following criteria were used to weigh the relative relevance of the transparency and data access frameworks for researchers:

A. Direct access to data

- **Available data** – whether the framework opens up a *large* amount of data that is potentially interesting to a relatively *wide* scope of research areas (very relevant); or the framework opens up a *large* amount of data that is potentially interesting to a relatively *narrow* scope of research areas (relevant); or the framework opens up a *small* amount of data that is potentially interesting to a *wide* scope of research areas (relevant); or the framework opens up a *small* amount of data that is potentially interesting to a relatively *narrow* scope of research areas (less relevant). In addition, we considered whether the framework provides access to *detailed and specific* data (relevant) or to more *generic or processed* information (less relevant, no granular insights);
- **Complexity of the procedure** – whether the procedure to obtain access to the data is simple (relevant) or relatively complex (less relevant);

- **Costs** – whether access must be provided for free (relevant) or can be subjected to a fee (arguably less relevant);
- **Request denial** – whether data must in principle be provided upon request (relevant) or access requests can be denied rather 'easily' on the basis of broad legal grounds of refusal, e.g., to protect trade secrets (less relevant);

B. Opportunities for data donation

- **Individual access rights** – whether the framework contains access rights for specific persons within a digital infrastructure such as 'platforms users', 'consumers', or 'data subjects' which could serve as the bases for data donation initiatives (relevant), or not (less relevant);
- **Data portability and interoperability** – whether the framework stimulates data portability and/or interoperability between systems (see Figure 2), thus enabling individuals to potentially donate their data for research purposes (relevant), or not (less relevant).

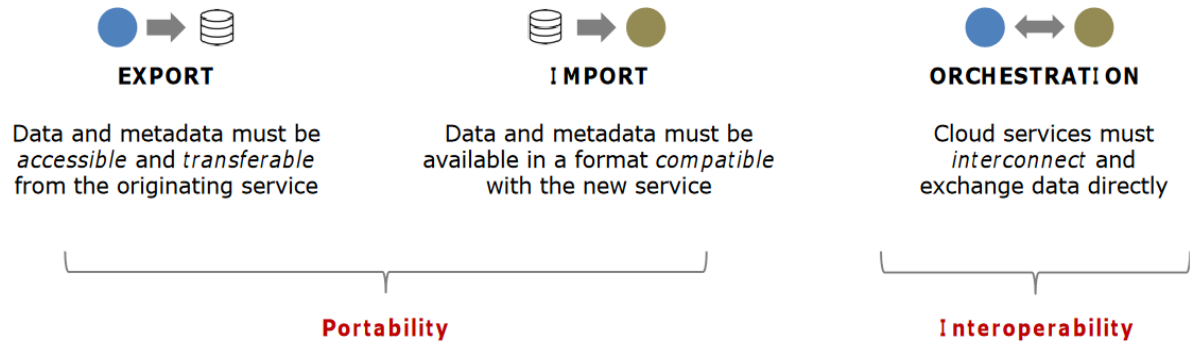

Figure 2. Data portability and interoperability[322]

C. Enabling environment

- **Data sharing intermediaries and facilitators** – whether the framework introduces rules for data sharing intermediaries and facilitators (relevant) or not (less relevant);[323]
- **Other favourable conditions** – whether the framework contains, for example, principles of transparency and openness, endorses low fees for data access, puts in place short response times to access requests, introduces standardised formats, or otherwise contributes to an enabling environment for data sharing (relevant), or not (less relevant).

[322] Source: Schnurr 2022**,** p. 11.

[323] Please note that the establishment of data sharing intermediaries and facilitators may also be useful for data donation practices because data sharing facilitators can professionalise sharing of individual-level data.

4.1 Direct access to data

4.1.1 Direct access for researchers specifically

Crucially, the only provision within the examined legal frameworks that explicitly provides researchers with a relatively strong right to access (certain) third-party data, is Article 40 of the **Digital Services Act**. Article 40(4) DSA grants "vetted researchers" a right to access data held by VLOPs and VLOSEs,[324] albeit solely for the purpose of conducting research that contributes to "the detection, identification and understanding of systemic risks in the Union" and to "the assessment of the adequacy, efficiency and impacts of […] risk mitigation measures". The potential of Article 40 DSA for researchers has been discussed in literature already.[325] Of all transparency and data access provisions reviewed for the mapping exercise, Article 40 DSA is arguably the most relevant provision in EU law to satisfy researchers' data needs, at least when it comes to online platform and search engine data.

Direct access for researchers: Article 40 DSA	
Available data	*Large amount* of online platform data that is potentially interesting to a relatively *wide* scope of research areas + *detailed and specific* data (not generic/processed)
Complexity of procedure	Complex because of the application/vetting process and the mandatory intervention by a national authority (Digital Services Coordinator)
Costs	Unclear; Delegated Act is expected to provide more clarity
Request denial	Data must in principle be provided; VLOPs and VLOSEs may only object to a request for two reasons as laid down in Article 40(5) DSA
RELEVANCE	**HIGH**

Notably, Articles 11-13 of the **proposed Political Advertising Regulation** also specifically grants (vetted) researchers a right to request access to data concerning political advertisements and targeting and amplification techniques. The difference with Article 40 DSA, however, is that there is no hard legal obligation for providers of political advertising services to provide the data: they merely have to "make best efforts" and "take the appropriate measures" to transmit the information to the researchers.

Direct access for researchers: Articles 11-13 pPAR	
Available data	*Large amount* of potentially interesting data concerning political advertising to a relatively *narrow* scope of research areas + *detailed and specific* data (not generic/processed)
Complexity of procedure	Not particularly complex
Costs	Unclear
Request denial	No hard legal obligation on providers of political advertising services to provide data: "best efforts" and "appropriate measures"
RELEVANCE	**MEDIUM**

Lastly, the **2022 Strengthened Code of Practice on Disinformation** is also relevant for researchers' direct access to data, even though it is a non-binding legal instrument. Chapter VI of the 2022 CoP is explicitly

[324] To recall, VLOPs are "very large online platforms" and VLOSEs are "very large online search engines" whose services have a number of 45 million or more average active recipients in the European Union and have been designated as such, see Article 33(1) DSA.

[325] Leerssen 2023a, Leerssen, Heldt and Ketteman 2023.

dedicated to empowering the research community and contains four specific commitments (and a set of underlying measures) to that end. Commitment 26 is of particular interest, as signatories commit to provide direct and (near) real-time access to data for research on disinformation.

Direct access for researchers: Commitment 26 of the 2022 CoP	
Available data	*Large amount* of potentially interesting data to a relatively *narrow* scope of research areas (i.e., related to disinformation) + *continuous, (near) real-time, stable* access
Complexity of procedure	To be published by the signatories, but the application process should not be overly cumbersome[326]
Costs	Unclear
Request denial	Unclear. Since the 2022 CoP is a soft law instrument, the 2022 CoP contains commitments rather than legal obligations, so signatories are not legally obliged to comply with (all the commitments and measures in) the 2022 CoP.[327]
RELEVANCE	**MEDIUM**

4.1.2 Direct access for the general public, including researchers

In addition to the very few provisions in EU law regulating data access for researchers in particular, there are a number of frameworks providing transparency and data access to the *general public*, which naturally also includes researchers.

At the EU level, the **Access to EU Documents Regulation** is important. This regulation ensures public access to documents held by the European Parliament, Council and Commission (Articles 2, 6-13), thus 'mirroring' national access laws in the EU Member States. The EUDR covers access to (legal) documents in a variety of fields (e.g., finance, environment, health care), which makes the regulation a useful tool for a lot of research areas. Some documents must be made publicly available by default, and others can be requested via a simple and low-cost application procedure. However, the EUDR contains both absolute and relative grounds of refusal. Access must for example be refused to documents concerning military matters or containing personal data if disclosure would undermine protection of the (public) interest concerned. But also (third-party) commercial interests, including intellectual property, and the protection of investigations or inspections can justify a refusal to grant access if there is no overriding public interest in disclosure (Article 4). Access for research purposes has no special status under the EUDR.[328] Moreover, concerns have been expressed about the need to update the regulation and adapt it to the communication tools of the digital age[329], but efforts to modernise the EUDR have so far failed.[330] So while the EUDR has significant potential on paper, its practical relevance has proven limited due to the lack of modernisation and wide possibilities to refuse access requests.

[326] Measure 26.2 2022 CoP.

[327] See further n (319).

[328] As opposed to some national access laws in the Member States that do treat researchers differently.

[329] O'Reilly 2021. Another example is that the EUDR, due to the lack of modernisation, does not contain any rules or conditions on *how* to provide access to documents, e.g., in a standardised format. This lack can cause a significant amount of work for researchers to be able to use the obtained information for research purposes, even when access is provided. Additionally, in case a request is needed to obtain documents, such requests can be quite time-consuming, and therefore costly. Researchers may not have the time nor resources to file such requests (including potential confirmatory application procedures). See also section 3.2.1.

[330] See Legislative Train Schedule 2023 (webpage) <https://www.europarl.europa.eu/legislative-train/theme-union-of-democratic-change/file-revision-of-the-access-to-documents-regulation>.

Direct access for the general public: EUDR as a whole	
Available data	*Large amount* of potentially interesting public sector documents to a relatively *wide* scope of research areas + *detailed and specific* information (not generic/processed)
Complexity of procedure	Relatively simple, but the procedure can take a while (and may become expensive in case legal proceedings must be initiated after refusal)
Costs	Direct access in electronic form or through the register is free of charge; the cost of producing and sending copies may be charged
Request denial	Requests can be denied on the basis of broad legal grounds of refusal, which reduces the potential for research access
RELEVANCE	**MEDIUM**

In addition to the EUDR, the **Open Data Directive** is a highly relevant framework dedicated to re-use of data held by public sector bodies. The ODD urges Member States to make as much public sector data available for re-use for both commercial and non-commercial purposes, preferably proactively (Article 5(2)) but at least upon request (Article 4).[331] These rules are further complemented by Chapter II of the Data Governance Act, which lays down certain conditions aimed at improving the general availability of data held by public sector bodies for re-use, in particular for research and innovation activities in the public interest.

Direct access for the general public: ODD as a whole	
Available data	*Wide variety* of public sector data potentially interesting to a relatively *wide* scope of research areas + *detailed and specific* information (not generic)
Complexity of procedure	Relatively simple; single point of access for requests
Costs	In principle free of charge, but marginal costs that have been incurred for the reproduction/provision/ dissemination of documents and measures taken to preserve confidentiality may be allowed
Request denial	Documents must in principle be provided; public sector bodies may however refuse re-use requests based on national access regimes and/or other provisions transposing the exceptions laid down in Article 1(2) ODD (i.e., documents that fall outside the scope of the ODD)
RELEVANCE	**HIGH**

Various public transparency provisions can also be found scattered throughout the **General Data Protected Regulation,**[332] the **proposed AI Act,**[333] **proposed European Media Freedom Act,**[334] **e-commerce and consumer law,**[335] the **Platform-to-Business Regulation,**[336] the **Digital Markets Act,**[337] the **proposed Political Advertising Regulation**[338] and the **2022 Strengthened Code of Practice on**

[331] Because the ODD is a directive, the utility of the ODD for researchers will mainly be dependent on national implementation measures by Member States (see also n (56)).

[332] E.g., Articles 13-14 GDPR (general information obligations/privacy statements).

[333] E.g., Article 60 pAIA (database).

[334] E.g., Article 24(2) pEMFA (allocation of state resources).

[335] E.g., Article 6 CRD.

[336] E.g., Article 5(1) P2BR (information in terms and conditions on ranking mechanisms).

[337] E.g., Article 35(1)-(2) DMA (Commission reports on market investigations and gatekeepers practices).

[338] E.g., Article 7 pPAR.

Disinformation[339] (in particular public reporting obligations). We consider most of these provisions of 'medium' relevance for scientific research, either because they provide access to data that may only be of interest to specific groups of researchers (e.g., pEMFA) or because they provide rather generic information which is unlikely to render granular insights to researchers (e.g., e-commerce and consumer law).

Direct access for the general public: scattered public transparency provisions in The GDPR, pAIA, pEMFA, e-commerce/consumer law, P2BR, DMA, pPAR, 2022 CoP	
Available data	Data tend to be useful for a relatively *narrow* scope of research areas only + rather *generic/processed* data
Complexity of procedure	Not complex; data should be provided proactively
Costs	Free of charge (publicly available)
Request denial	N/A; proactive access provision
RELEVANCE	**MEDIUM**

Importantly, the **Digital Services Act** contains many public transparency provisions in and of itself, similar to the frameworks mentioned above. However, we consider the provisions in the DSA to be more relevant for researchers for the reason that the provisions do not only provide for generic or processed information but also for specific and detailed information. Article 39 in particular obliges VLOPs to install publicly available advertisement repositories. These so-called 'ad archives' have already proven valuable for scientific research in practice, as they contain detailed information on advertisements presented on a platform's online interface.[340] Article 39 is a welcome provision since VLOPs that had not yet installed an ad repository are now required to do so.[341] Article 39 also describes *how* an ad repository must be made available, namely through "a searchable and reliable tool that allows multicriteria queries" and through "application programming interfaces".[342] The other public transparency provisions in the DSA relate mostly to reporting obligations by intermediary services and in particular (very large) online platforms (see e.g., Articles 15, 24 and 42). Compared to the public transparency obligations in the other frameworks mentioned above, these provisions contain relatively detailed descriptions of the information that must be made publicly available. For instance, reports on content moderation must include the number of notices submitted, categorised by the type of alleged illegal content concerned, and any actions taken pursuant to the notices.[343] Moreover, VLOPs must make a whole lot of information available that can be relevant to a wide scope of research areas.[344] In sum, we consider the public transparency provisions in the DSA to be of 'high' utility to research.

Direct access for the general public: various public transparency provisions in the DSA	
Available data	*Ad repository*

[339] E.g., the establishment of the Transparency Centre (Chapter VIII).

[340] Leerssen et al 2023. See also Leerssen et al 2019. A list of the information that a repository must include is laid down in Article 39(2) DSA. The repository must cover advertisements that are presented on the platform until one year after the advertisement was presented for the last time, see Article 39(1) DSA.

[341] A list of the information that a repository must include is laid down in Article 39(2) DSA. The repository must cover advertisements that are presented on the platform until one year after the advertisement was presented for the last time, see Article 39(1) DSA.

[342] Article 39(1) DSA.

[343] Article 15(1)(b) DSA.

[344] The risk assessment reports under Article 34 DSA, for example, may touch on several topics and research areas.

	Specific and detailed information on advertisements presented on online platforms, although maybe of interest to a relatively *narrow* scope of research areas. *Other public transparency obligations* Relatively *detailed* information (compared to the other public transparency provisions), potentially interesting to a *wide* scope of research areas.
Complexity of procedure	Data should be provided proactively + Article 39 contains several requirements to ensure that the ad repository can be easily used in practice.
Costs	Free of charge (publicly available)
Request denial	N/A; proactive access provision
RELEVANCE	**HIGH**

Lastly, the **2022 Strengthened Code of Practice on Disinformation** (2022 CoP) contains several transparency provisions[345] requiring signatories to publish information and data, inter alia through the Transparency Centre.[346] Given the large amounts of data that may become available through these (semi)voluntary commitments from platforms[347], we consider the 2022 CoP to be potentially highly relevant in in terms of public transparency. That said, the practical value of the 2022 CoP remains to be seen – as it is quite a recent framework – and its value may change over time as signatories can decide to withdraw from separate measures and commitments, as well as from the 2022 CoP as a whole.[348]

Direct access for the general public: various public transparency provisions in the 2022 CoP	
Available data	*Large amount* of potentially interesting data concerning disinformation to a relatively *narrow* scope of research areas (i.e., related to disinformation)
Complexity of procedure	Mainly proactive commitments
Costs	Free of charge (publicly available)
Request denial	N/A; proactive transparency provisions
RELEVANCE	**MEDIUM**

4.1.3 Direct access for specific persons, including researchers acting in that capacity

Seemingly less relevant, but not unimportant, are transparency and data access provisions that are aimed at specific categories of natural or legal persons, such as 'data subjects', 'platform users' and 'users of smart products' (hereinafter referred to as "individual access rights"). For example, platform users affected by platforms' content moderation decisions and policies are entitled to receive so-called statements of reasons[349]; users of IoT-products and services are entitled to receive personalised IoT-data generated by the use of the products and services[350]; data subjects have the right to receive a copy of their personal data

[345] E.g., through public policies on their approach towards political and/or issue advertising (Commitment 5), on prohibited behaviour and practices on their services (Measure 15.2), on how to limit the spread of harmful false or misleading information (Measure 18.2), or through ad repositories (commitment 10).

[346] Transparency Centre (webpage) <https://disinfocode.eu/>.

[347] See for instance the reports as published on the website of the Transparency Centre: Transparency Centre reports (webpage) <https://disinfocode.eu/reports-archive/?years=2023>.

[348] However, for VLOPs this may be slightly different as adherence to the 2022 CoP is not entirely voluntarily, but semi-voluntarily. See further n (319).

[349] Article 4 P2BR, Article 17 DSA.

[350] Articles 3-4 pDA.

processed by data controllers[351]; and individuals targeted by political advertisements based on their personal data are entitled to receive information on the techniques used to target them.[352]

Such provisions have the potential to be deployed for scientific research purposes in two ways. First, researchers may directly invoke the provisions when they are acting in the capacity as the beneficiaries targeted by the provision. For instance, when a researcher signs up to an online platform, they become a platform user and are therefore entitled to directly receive certain information in their capacity as platform user (see also section 6.1 of this report). The information provided based on these provisions typically relates to the individual interactions of the person in question (here: the researcher) with the digital infrastructure. Second, researchers could use the individual access rights to collect data at a larger scale, namely via data donation initiatives, where beneficiaries addressed by the provisions are asked to 'donate' the data they are entitled to a research project. The role of individual access provisions in data donation is further discussed in sections 4.2 and 6.1 of this report. The tables below focus on the utility of individual access rights as a gateway for researchers to *directly* access certain data.

A relevant provision in this regard is Article 6 of the **Digital Markets Act**. This provision addresses multiple categories of persons, including advertisers and publishers, end users and business users. Researchers are most likely to obtain direct access to data in their capacity as end users, or, alternatively, as third parties authorised by advertisers/publishers, end users or business users.

Direct access for specific persons: Article 6 DMA	
Available data	*Advertisers and publishers – Article 6(8) DMA* Data necessary for advertisers and publishers to carry out their own independent verification of the advertisements inventory *End users – Article 6(9) DMA* Data provided by the end user or generated through the activity of the end user in the context of the use of the relevant core platform service *Business users – Article 6(10) DMA* Data that is generated in the context of the use of the relevant core platform services, or the use of services provided together with or in support of the relevant core platforms services by business users and by end users engaging with the products or services provided by those business users In sum: a *large* amount of data is available to various specific persons and is potentially interesting to a relative *wide* scope of research areas
Complexity of procedure	Upon request but details on request procedure are unspecified
Costs	Free of charge
Request denial	Unspecified[353]
RELEVANCE	**HIGH**

As Article 6(9)-(10) of the DMA clearly illustrates, individual access rights often provide information relating to a specific individual's interactions with a digital infrastructure. In the case of a researcher, this means that the data will concern the researcher's own interactions with the digital infrastructure in their capacity as e.g., a data subject, platform user, IoT-product owner, etc. This is not to say, however, that such data are useless

[351] Article 15 GDPR, Articles 13-14 LED.

[352] Article 12 pPAR.

[353] Noto La Diega explains that Article 6 DMA "does not appear to counterbalance these obligations with IP and confidentiality in a way that would sterilise its ethos of openness" and given the obligation of 'continuous and real-time access', he notes that "the lawmaker has chosen access over IP, openness over closure", Noto La Diega 2023, pp 9-10.

by their very nature. To the contrary, research practice has shown that individual access rights are sometimes used as a first step to further research, notably through data donation. A researcher may for instance request access from a data controller to their own personal data based on Article 15 of the **General Data Protection Regulation**, just to see which information is provided and how, so as to decide whether it is worth collecting more of this information on a larger scale via data donation. Moreover, the granularity of the information is an important factor in determining its relevance for research. For example, a personal data package provided under Article 15 GDPR typically contains quite detailed and granular information and may therefore be a useful source for (certain) research in itself already, depending on the research question and research goals. Indeed, it is not always necessary to gather large amounts of data – less but detailed information may also be sufficient in some cases.

Direct access for specific persons (data subjects): Article 15 GDPR	
Available data	All personal data processed by a data controller, potentially interesting to a relatively *wide* scope of research areas + *detailed and specific* information (not generic/processed)
Complexity of procedure/formalities	Upon request but not particularly complex
Costs	In principle free of charge but reasonable fees (administrative costs) may be charged for further copies and also when requests are manifestly unfounded or excessive
Request denial	The data must in principle be provided[354]
RELEVANCE	**HIGH**

Closely linked to the GDPR, is the **Data Protection Law Enforcement Directive** (LED). Since the LED covers fewer personal data and data access requests can be denied more easily (and are denied often, as empirical research has shown), we consider it to be of less relevance than the GDPR:

Direct access for specific persons (data subjects): Articles 13-14 LED	
Available data	Only personal data processed in context of law enforcement, potentially interesting to a relatively *narrow* scope of research areas + *detailed and specific* information (not generic/processed)
Complexity of procedure/formalities	Upon request; potentially complicated due to delays, restrictions or omissions of information[355]
Costs	In principle free of charge but in case of manifestly unfounded or excessive requests reasonable fees may be charged.
Request denial	Possible on several grounds[356]
RELEVANCE	**LOW**

Furthermore, the **proposed Data Act** contains direct data access rights in Articles 3-4 for users of IoT products and services. These rights potentially cover large amounts of data which can be of interest to a wide scope of research areas, especially considering the current increase of IoT products and services in our

[354] Although Article 15(4) GDPR states that the right to obtain a copy of one's personal data shall not adversely affect "the rights and freedoms of others".

[355] Vogiatzoglou et al 2021.

[356] Articles 12(4)(b), 13(3) and 15 LED.

society. It should be noted that the utility of these rights could be decreased because of a proposed amendment allowing data access requests to be denied in order to protect data holders' trade secrets.[357]

Direct access for specific persons (users of IoT products and related services): Articles 3-4 pDA	
Available data	IoT data generated by use of a IoT service or product, i.e., a *large* amount of data potentially interesting to a *wide* scope of research areas + *detailed and specific* information (not generic/processed)
Complexity of procedure	Data should be directly accessible or otherwise available based on a 'simple request'
Costs	Free of charge
Request denial	Possible in case of exceptional circumstances, e.g., when the data holder is highly likely to suffer serious damage from the disclosure of trade secrets
RELEVANCE	**HIGH**

Last but not least, we consider the individual access rights laid down in the **Digital Services Act**, **proposed Political Advertising Regulation** and the **proposed AI Act** of 'medium' relevance as they provide data that are potentially interesting to a relatively narrow scope of research areas and/or provide rather generic information.

Direct access for specific persons (platform users): Articles 17, 20(5) DSA	
Available data	Statements of reasons addressed to platform users affected by content moderation policies + platform decisions addressed to complainants; potentially interesting data to a relatively *narrow* scope of research areas
Complexity of procedure	N/A; data should be provided proactively
Costs	Free of charge
Request denial	N/A; proactive access provision
RELEVANCE	**MEDIUM**

Direct access for specific persons (individuals targeted by targeting/amplification techniques in context of political advertising): Articles 12(3)(c)-(4) pPAR	
Available data	Information on the targeting and amplification techniques used for political advertising involving the processing of personal data, addressed to those who are presented with a political advertisement; potentially interesting data to a relatively *narrow* scope of research areas
Complexity of procedure	N/A; data should be provided proactively
Costs	Free of charge
Request denial	N/A; proactive access provision
RELEVANCE	**MEDIUM**

Direct access for specific persons (deployers): Articles 13(2) pAIA	
Available data	Information on (the output of) high-risk AI systems + instructions for use of high-risk AI systems; rather *generic and processed* information

[357] See e.g., Article 4(3a) pDA.

Complexity of procedure	N/A; data should be provided proactively
Costs	Free of charge
Request denial	N/A; proactive access provision
RELEVANCE	**MEDIUM**

4.2 Opportunities for data donation

As noted above, many of the analysed frameworks contain individual access rights. In this section, we discuss the potential of these individual access right to be **utilised by researchers for data donation initiatives**. Data donation refers to the practice where consenting participants voluntarily, actively or passively, transfer (personal) data that they are entitled to pursuant to the law, to researchers. As explained in section 3.1.1, current data donation projects often depend on research subjects to actively request 'their' data held by a third party, and upon receipt, deliver these data to researchers, either directly or via a designated intermediary (see Figure 3 below). The ideal data donation process, however, would offer research subjects the possibility to directly port their data from the third party to (an intermediary operated by) the researchers (see the dashed line in Figure 3).

For data donation to be successful, it is essential that individuals who are entitled to data on the basis of individual access rights, are technically able to transfer their data to researchers. Portability of data and interoperability of systems are therefore crucial. However, researchers have been struggling to use existing data portability rights (e.g., GDPR and NPDR), for different reasons ranging from non-compliance by data controllers/holders, to technical constraints such as volatile data formats.[358]

New provisions on data portability have recently been introduced in the **Digital Markets Act** and **proposed Data Act**. Article 6(9) DMA, for example, requires that end users of core platform services are provided with "effective portability" of the data generated through the use of a core platform service, which also implies that they have "continuous and real-time access" to the data.[359] Data portability rights hold significant potential for researchers, as they enable individual research subjects to mandate data holders to transfer data *directly* to a university or research institution without having to download the data first.[360]

An interesting provision in this regard is Article 5 pDA. Complementing the right of smart product users to access IoT-data generated while using their products, Article 5 pDA lays down an obligation for IoT data holders – often the manufacturers of the products – to also make IoT-data available to third parties upon request of the product user. In other words, researchers can agree with IoT product users that the latter submit a request with the data holder to make IoT data available to the researchers. Data holders can, in turn, be held accountable if they fail to comply with their obligation to provide the data. The potential advantage of this provision for researchers, is that it can facilitate data donation processes that do not require the active involvement of individuals. Put differently, rather than first having to ask individuals to download the data they are legally entitled to, and then have them share it with researchers, the data can be ported directly from the data holder to the data donation platform or research project.[361] Importantly, the rights

[358] See e.g., Quinn 2018, p. 3-4. Ohme and Araujo 2022; Wong and Henderson 2019; presentations of Theo Araujo (UvA researcher and project lead of www.datadonation.eu) and Lucas van der Meer (operational manager at https://odissei-data.nl) during the workshop on 'Research Access to Digital Infrastructures' (RADI) held at the Institute for Information Law (IViR) on 16 March 2023.

[359] Article 6(9) DMA.

[360] As proposed by Theo Araujo, researcher and leader of a data donation initiative at the University of Amsterdam, during the workshop on 'Research Access to Digital Infrastructures' (RADI) held at the Institute for Information Law (IViR) on 16 March 2023.

[361] Compare n (360).

and interests of the individual (here: IoT product user) should still be safeguarded throughout the process, notably through their explicit consent. While implementation (compliance and enforcement) of the new data portability and interoperability provisions remains to be seen,[362] they show potential on paper to further 'professionalise' and facilitate data donation projects.

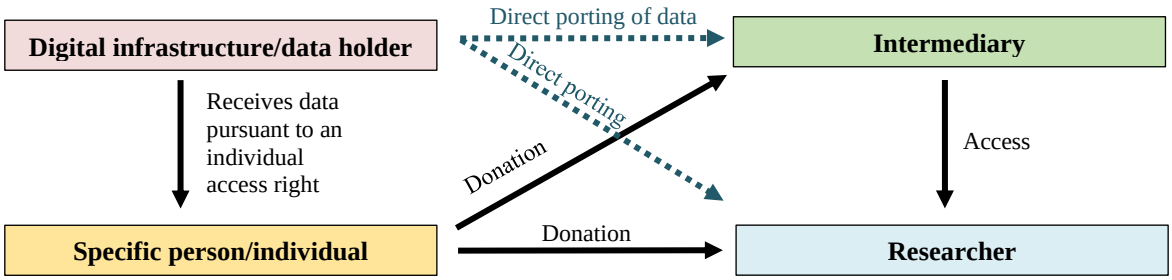

Figure 3. Various ways of data donation

4.3 Enabling environment for data-driven research

Finally, transparency and data access provisions can be relevant for researchers in that they contribute to an *enabling environment* for access to data and scientific research. Within this category, a rough distinction can be made between (i) provisions that introduce and regulate **(professional) intermediary entities** aimed at facilitating data access and sharing, and (ii) provisions that **otherwise promote data-driven research**, for example by fostering transparency, formalising data access request processes, encouraging the standardisation of formats, enhancing data portability and interoperability, preventing high fees, establishing (short) response times from the data holder, etc (to which we refer to as 'other enabling elements').

4.3.1 Data sharing intermediaries/facilitators

Some of the analysed legislative instruments have called into life new types of data sharing intermediaries. **Chapters III and IV of the DGA**, for instance, govern the registration and notification of so-called 'data intermediation services' (DISs) providers and 'data altruism organisations' (DAOs). Other types of data sharing 'facilitators' feature in for instance the ODD (single access point), Chapter II of the DGA (single information points), the 2022 CoP (independent third-party body; Transparency Centre) and the pAIA (EU database for high-risk AI systems and foundation models). The aim of these intermediaries is, in general, to support the efficient sharing of large volumes of data for both commercial and non-commercial purposes by facilitating interactions between relevant actors, including data holders, research/data subjects but potentially also researchers (see section 5.4 of this report).

4.3.2 Other enabling elements

Other enabling elements for researchers' access to data and data-driven research more broadly can be found in various legal frameworks.

[362] It should be noted that the already existing right to data portability in Article 20 GDPR (cf. section 3.1.1) has proven to be complicated to invoke and/or undercomplied with. See, e.g., Wong and Henderson 2019.

The **Open Data Directive**, for example, strongly advocates the *principle of 'openness'* in the context of public sector information, which is reflected in its provisions on, inter alia, access fees (low),[363] formats (standardised),[364] licences (standardised, open)[365] and exclusive arrangements (in principle prohibited).[366] **Chapter II of the Data Governance Act** complements the ODD in that it encourages the re-use of 'protected' data held by public sector bodies that fall outside the scope of the ODD. However, in contrast to the ODD, the DGA does *not* contain specific legal rights for re-users; it merely imposes legal obligations on public sector bodies which decide to allow the re-use of protected data.[367] Despite this lack of a hard legal obligation for public sector bodies to make 'protected' public data available for re-use, the DGA nevertheless aims to stimulate a(n) (secure) environment in which protected public data can be shared for both commercial and non-commercial purposes.[368] Taken together, the Open Data Directive and the Data Governance Act can be said to contribute to the wider availability of public sector information for scientific research purposes.

Data portability and interoperability requirements can be considered important drivers for researchers' data access as well. Indeed, even if such provisions were not called into life for scientific research in particular, their (implied and explicit) format requirements enable the sharing and re-use of data beyond the digital infrastructure that the data originate from, which may also benefit research. Data portability and interoperability requirements are enshrined in, inter alia, the **Free Flow of Non-Personal Data Regulation,**[369] **Digital Markets Act**[370], the **General Data Protection Regulation**[371] and the **proposed Data Act** (see also section 4.2).[372]

A variety of other 'enabling elements' appear in, for instance, the **General Data Protection Regulation** (*general principle of transparency*),[373] the **Digital Services Act** (*numerous public transparency provisions*),[374] the **Access to EU Documents Regulation** (*time frames for access requests*),[375] the **Data Governance Act** (see e.g., the *'European data altruism consent form'* to facilitate the collection of data based on data altruism and the power of the European Data Innovation Board to issue *standards for cross-sector data sharing*)[376] and the **Digital Markets Act** (provision of real-time data access *free of charge*).[377]

Lastly, Article 3 of the **Copyright in the Single Market Directive** is an interesting provision in this regard, not so much because it facilitates researchers' access to data *per se* but because it contributes to an environment of openness for scientific research by introducing a mandatory exception to copyright, related rights and sui generis database protection for text and data mining for scientific research purposes (in regards to copyrighted works to which researchers already have lawful access).[378]

---

[363] Article 6 ODD.

[364] Article 5 ODD.

[365] Article 8 ODD.

[366] Article 12 ODD.

[367] See also See also European Commission and Van Eechoud 2022, p. 29.

[368] An example can be found in Article 5(6) DGA: where re-use of data cannot be allowed due personal data protection issues, public sector bodies "shall make best efforts" to provide assistance to potential re-users to enable re-use after all.

[369] Article 6(1)(a) NPDR.

[370] E.g., Article 6(9)-(10) DMA.

[371] Article 20 GDPR.

[372] Chapter VI pDA.

[373] Article 5 GDPR. In this context, see also Mahieu and Ausloos 2020 on the importance of a broader culture of transparency and data access.

[374] For example, Articles 15, 22(3), 24, 42 DSA. See Edelson, Graef and Lancieri 2023, para. 1.2.3.

[375] E.g., Article 7(1) EUDR.

[376] See Article 25 DGA on the European data altruism consent form and Article 30(h) DGA on the power of the European Data Innovation Board to propose guidelines for common European data spaces, addressing inter alia standards to be used for data sharing/data use.

[377] Article 6(8)-(10) DMA.

[378] Article 3 CDSMD.

***

Based on the discussion above, Table 3 visualises the relevance of the analysed frameworks for researchers. Importantly, we would like to emphasise that this is just a rough estimation of the (potential) utility of the legislative instruments – some of which have not even been adopted yet. The table is by no means meant as a definitive categorisation or final assessment of the frameworks.

Framework	Direct access to data			Opportunities for data donation	Enabling environment	
	Researchers specifically	The general public	Specific persons		Intermediaries and facilitators	Other enabling elements
Generic frameworks						
GDPR	Low	Medium	High	High	Low	Medium
NPDR	Low	Low	Low	Low	Low	Medium
CDMSD	Low	Low	Low	Low	Low	High
DGA, Ch. III, IV, VI	Low	Low	Low	Medium	High	High
pAIA	Low	Medium	Low	Low	Medium	Low
pEMFA	Low	Medium	Low	Low	Low	Low
Frameworks regulating the public sector						
EUDR	Low	Medium	Low	Low	Low	Medium
LED	Low	Low	Medium	Medium	Low	Low
ODD	Low	High	Low	Low	Medium	High
DGA, Ch. II	Low	Medium	Low	Low	Low	High
Frameworks regulating the private sector						
CRD, SD and ECD	Low	Medium	Low	Low	Low	Low
P2BR	Low	Medium	Medium	Medium	Low	Low
DMA	Low	Medium	High	High	Low	Medium
DSA	High	High	Medium	High	Medium	Medium
pPAR	Medium	Medium	Low	Low	Low	Low
pDA	Medium	Low	High	High	Low	Medium
2022 CoP	Medium	Medium	Low	Low	Medium	Medium

Table 3. Relevance of analysed legal frameworks according to the criteria of direct access, opportunities for data donation and enabling environment

5. Recurring themes across transparency and data access provisions

Based on the regulatory mapping exercise as summarised in chapter 3, we have identified a few (interrelated) 'themes' that deserve special attention: (1) the limited recognition of scientific research and researchers in transparency and data access provisions in EU digital/data legislation; (2) abstract and generic transparency rules; (3) the tension between researchers' data access and third-party interests; (4) the introduction of institutional intermediaries that aim to facilitate data access and sharing; and (5) disparate format requirements for data sharing. Each of these themes will be discussed in more detail below.

5.1 Limited recognition of scientific research and researchers

Scientific research has been an area of interest in EU digital policymaking. The EU's Open Science policy, in particular, aims to stimulate the diffusion of knowledge through digital technologies.[379] Nevertheless, in the legislative instruments as discussed in this report, scientific research is only mentioned here and there in abstract or stand-alone ways. When looking at transparency and data access provisions specifically, **(academic) researchers are rarely recognised in explicit terms**. Indeed, only a few provisions seem to have been drafted with research and researchers in mind. The rationale for most transparency and data access provisions generally relates to the protection of individuals or internal market objectives. While this is not necessarily surprising considering the EU's policy priorities,[380] limited competences,[381] and the fact that transparency measures often serve a neoliberal regulatory agenda,[382] it may have significant consequences for the provisions' utility as tools to observe digital infrastructures for scientific research purposes.

There are several reasons why strong and explicit research data access provisions are important. First of all, the organisations managing digital infrastructures generally have significant **disincentives to share any information**, ranging from legal liability (e.g., on the basis of privacy, data protection or intellectual property law), to economic strategies (e.g., safeguarding competitive edge), technical difficulties (e.g., systems are not designed to easily retrieve and share data with researchers), and security concerns (e.g., critical information may be leaked).[383] Second, there are **significant power asymmetries** between these organisations on the one hand, and academic researchers and institutions on the other hand, making it hard for the latter to negotiate for access. Third, academic institutions are known to be **risk-averse**, which makes it unlikely that they will take active (legal) steps to compel data access in specific instances where the law is unclear. For these reasons, it is important that clear *ex ante* legal rules are established that unambiguously define the conditions for data access for (academic) research.

As has become apparent from the regulatory mapping exercise, there are some data access provisions *do* explicitly recognise the importance of academic research. Below we spotlight these provisions:

- The most prominent example in this regard, is **Article 40 of the Digital Services Act**. This provision sets out different types of data access, including by researchers.[384] In particular, Article

[379] See European Commission Open Science webpage <https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en>.

[380] See in particular: European Commission Communication 2020.

[381] Vėlyvytė 2022 (accessed 17 April 2023).

[382] Flyverbom 2019.

[383] Ausloos, Leerssen and Ten Thije 2020.

[384] Notably, Article 40(1)-(3) covers data access by regulators.

40(4) DSA explicitly requires very large online platforms (VLOPs) and very large online search engines (VLOSEs) to provide researchers that have been awarded the status of "vetted researchers" in accordance with Article 40(8), with access to any data necessary to conduct research "that contributes to the detection, identification and understanding of systemic risks in the Union."[385] A VLOP/VLOSE receiving a data access request cannot ignore or block such requests but may suggest amendments to the request where they consider that (a) they do not have access to the requested data, or (b) giving access to the requested data will lead to significant vulnerabilities in the security of their service or the protection of confidential information, in particular trade secrets. However, even this seemingly 'robust' researcher data access provision has considerable limitations. For example, researchers must demonstrate that they are affiliated to a "research organisation" as defined in the Copyright in the Digital Single Market Directive,[386] i.e., a university, research institute or other entity primarily focused on carrying out non-commercial[387] scientific research. Researchers working in for-profit research or private partners in private-public partnerships thus seem to be excluded from the DSA's data access right. Furthermore, researchers have to follow a tedious procedure – involving a regulatory body's approval[388] – and the research for which data access is requested must be specifically aimed at the detection, identification and understanding of systemic risks, or in other words: at holding VLOPs/VLOSEs accountable. As a result, broader research questions or exploratory research projects might not be able to rely on this provision.[389]

In addition to data access by vetted researchers, Article 40 DSA also regulates access by a broader group of researchers to certain platform data. Article 40(12) obliges VLOPs and VLOSEs to grant researchers who meet certain conditions, including those affiliated to not for profit bodies, organisations and associations,[390] access to data that is "publicly accessible" in their online interfaces[391] and in real-time where technically possible. For this type of data access, researchers do not have to be vetted and approved by the regulatory body first. So, whereas this provision's scope is limited to publicly available data, it is not constrained by procedural requirements. It effectively prevents VLOPs/VLOSEs from obstructing certain researchers[392] (e.g., through legal action or technical measures) to access publicly available information on their service, and arguably even comprises a *positive duty* to provide technical tools (e.g., an API) facilitating effective access.[393] Similar to Article 40(4), however, this data access right is also limited to systemic risk research.

- A second framework that is worth referring to, is the **Data Governance Act** which also explicitly recognises the importance of scientific research in its recitals and aims to stimulate the re-use of 'protected' public sector data for scientific purposes.[394] However, the DGA does not provide for

[385] Systemic risks are defined rather broadly in Article 34(1) DSA.

[386] Article 2(1) of CDMSD.

[387] I.e., on a not-for-profit basis or pursuant to a public interest mission, see Article 2(1)(a)-(b) CDSMD.

[388] A Delegated Act is expected to provide more guidance on the data access process, including the vetting process: <https://algorithmic-transparency.ec.europa.eu/news/call-evidence-delegated-regulation-data-access-provided-digital-services-act-2023-04-25_en>. See also Albert 2022.

[389] It should be noted, however, that the definition of systemic risks in the DSA seems to be very broad, which would mean that a relatively broad scope of research questions and projects could eventually fall within the scope of Article 40 DSA.

[390] See Article 40(12) in conjunction with Article 40(8) DSA.

[391] Recital 98: "for example on aggregated interactions with content from public pages, public groups, or public figures, including impression and engagement data such as the number of reactions, shares, comments'."

[392] I.e., only researchers that are independent from commercial interests; disclose their funding; can fulfil security and confidentiality requirements, and can demonstrate the necessity and proportionality of the requested data for their research purposes.

[393] Leerssen 2021.

[394] See for instance recitals 15 and 16, encouraging a more harmonised approach to make data 'easily accessible for the purposes of scientific research in the public interest'.

an *actionable right* to obtain access to data to re-use them for research purposes. That said, its provisions on re-use of public sector information may still play a positive role in enhancing a broader culture of transparency and data access. Additionally, the DGA has introduced two new actors that can benefit researchers' position in obtaining access to data and information, namely data intermediation services providers[395] and data altruism organisations[396] (see section 5.4 below).

- Thirdly, another framework that explicitly mentions researchers in relation to data use is the **Copyright in the Digital Single Market Directive**. Specifically, Article 3 CDSMD aims to remove potential copyright obstacles when researchers perform text- and data-mining techniques on large datasets that include copyrighted material. Unlike the DSA, this provision is not constrained by the purpose of the research, nor any type of actor. Importantly however, the provision does not provide researchers an actionable right to access any information, but only protects those researchers that have *already lawfully obtained access* to the data. With this in mind, the CDSM Directive is only of secondary importance for empirical researchers wishing to study digital infrastructures.

Although researchers are rarely explicitly recognised in EU digital/data legislation and do certainly not have a 'general' right to access third-party data, the examined legal frameworks can still prove useful for observing or studying digital infrastructures. As has become apparent throughout the analysis in chapters 3 and 4, the provisions in EU digital/data legislation may **still enable or facilitate data access** for researchers in different ways. A prime example of how data access provisions can be deployed in a research context is the data subject's right of access in the GDPR, which has been the basis for 'data donations' where researchers crowdsource data subjects' access requests to, for instance, obtain very detailed information about online platforms and their users.[397] Moreover, public transparency provisions that require the proactive sharing of data with the general public can render researchers with valuable information about (people's interactions with) digital infrastructures (see e.g., Article 39 DSA and Article 7(6) pPAR on ad libraries, and Article 12(2) EUDR on legislative documents).

In sum, while academia should perhaps not rely on the law as the primary tool for researchers to observe and study digital infrastructures, the growing number of data access provisions in recent (EU) digital policymaking could well **complement the methods** researchers have been developing. This is especially true considering the law's potential to compel data access in the face of significant information asymmetries and strong transparency disincentives for those controlling digital infrastructures.[398] It is important to note, however, that the law can sometimes also obstruct data access, notably through frameworks regulating the use of information such as intellectual property law (e.g., trade secrets[399]) or privacy and data protection law (e.g., sensitive data[400]). These frameworks put in place constraints on the extent to which data can be shared.

It should be pointed out that the observability of digital infrastructures and the position of academic researchers has received attention from the European Commission: its DG Research & Innovation has been taking active steps to mitigate issues through legislative and non-legislative proposals.[401] We should however remain vigilant over the framing of these new proposals. The introduction of more provisions explicitly recognising data access for research could potentially result in **problematic dependencies** between

[395] Chapter, III DGA.
[396] Chapter IV DGA.
[397] Ohme and Araujo 2022; Araujo et al 2022.
[398] Ausloos, Leerssen and Ten Thije 2020.
[399] See for instance the unclear scope of the trade secrets exemption in Article 40 DSA, Leerssen 2021.
[400] See for instance in the context of medical research: ALLEA, EASAC and FEAM 2021, European Parliament, DG EPRS 2019.
[401] European Commission 2022 <https://data.europa.eu/doi/10.2777/52110>.

academics and those managing digital infrastructures. Moreover, such provisions could also **legitimise and cement certain data collection and processing practices that may be very problematic from for instance a data protection perspective**. Policymakers, universities, researchers and transparency advocates more broadly should thus maintain a critical lens when it comes to new rules explicitly calling for data access for academic research.

5.2 Generic transparency obligations

Closely linked to the previous point, is that most data access and transparency provisions in EU digital/data legislation are rather 'generic', that is, typically designed to fulfil vaguely formulated accountability or internal market goals. Apart from research goals rarely being referred to explicitly, **the provisions also often remain fuzzy on a broader level, lacking details on the exact data points that can be requested or for which purposes, and on the procedures to be followed**. This vagueness is further exacerbated by the way in which rights and obligations are operationalised and enforced[402], thus challenging their utility for specific research needs.

As a result of this fuzziness, the information that is provided on the basis of transparency and data access obligations is often of a more general nature (e.g., privacy policies required by the GDPR, or parameters of recommender systems required by the DSA, DMA and P2BR). The fact that there is significant leeway for data holders to give shape to these provisions in practice may render it difficult for researchers to successfully make use of them. For instance, issues with nebulous and ever-changing data-formats (see section 5.5), open questions relating to how the data were generated and pre-processed, and the representativeness of datasets significantly affect their utility to researchers.

Apart from requirements to share certain data/information, there is also a growing number of obligations to publish reports on the operations of digital infrastructures. These legally ordained reports may be drafted by entities in charge of the digital infrastructures themselves (see e.g., Article 15 DSA, Article 7 pPAR and Article 17(5) pEMFA), enforcement agencies (see e.g., Article 35(2) DSA, Article 59 GDPR) or independent third parties (see e.g., Article 22(3) DSA, Article 15 DMA). Such reports may shine light on operations that otherwise cannot be rendered fully observable due to conflicting other fundamental rights, freedoms or interests, such as intellectual property, privacy and data protection rights. Yet, they do introduce an interpretational layer over the underlying data – indeed, the organisation drafting the report typically draws certain conclusions – which raises several issues possibly constraining the reports' utility for researchers. Especially self-reporting (such as the DSA's requirement on social media platforms to report on their content moderation practices[403]) raises significant questions relating to reliability, selection bias and representativeness. This is further amplified by the opacity of the underlying data that are not made public but which serve as the basis for the reports. As such, transparency through reporting duties may be less useful for quantitative scientific analyses.

The fact that most transparency and data access provisions assessed in this report are rather generic and broad, can – at least in part – be explained by the fact that the respective legislative instruments have **broad scopes** to begin with. A broad scope of application necessitates open-ended and multifunctional provisions that can be adapted to various situations. At face value, this benefits academic researchers, as they can deploy transparency and data access provisions strategically in light of their specific research agendas. A good example of how this strategic use of provisions works out in practice, are data donation projects where data subjects' access and portability rights under the GDPR are used to obtain detailed

[402] See, for example in the context of GDPR (access) rights: Ausloos, Toh and Giannopoulou 2022.

[403] Articles 15 and 42 DSA, see also section 3.3.4 of this report.

information about digital infrastructures that process personal data. Indeed, untying transparency and data access requirements from their initial objectives (e.g., accountability, stimulating competition) may support more versatile academic inquiry into the respective digital infrastructures.

At the same time, the abstract nature of transparency and data access provisions has significant downsides. Open-ended legal provisions effectively **push interpretation costs downstream**. While intended to be versatile and multifunctional, the way in which these provisions are interpreted, operationalised and enforced is generally determined by the organisations using these provisions in the first place. This gives well-resourced entities an advantage, which could **further solidify existing power asymmetries**, while weak(er) parties have fewer redlines to fall back on.[404]

In conclusion, most of the horizontal legal frameworks regulating digital infrastructures contain rather generic transparency and data access provisions. This presents both benefits (versatility) and challenges (unclarity and power asymmetries) in an academic research context. Some of the challenges could be resolved through sector-specific frameworks at different levels as well as non-legislative interpretational guidance (e.g., via code of conducts[405]).[406] An example of a (more or less) sector-specific framework is the DSA, which foresees more specific data access provisions, to be detailed even further through delegated acts[407] and independent oversight bodies. One can also think of older frameworks in other industries – from the financial sector to food production, agriculture, transportation, etc. – all of which are increasingly reliant on digital infrastructures as well.

5.3 Balancing transparency and data access with third-party interests

In the examined provisions, the interests of transparency and data access are often set against competing rights and interests, including (personal) data protection,[408] the protection of copyright and other intellectual property rights,[409] and the protection of the confidentiality of certain information such as trade secrets.[410] The balancing exercise that is required in this regard can be seen as part of the broader (political) debate on control over data versus competing societal interests (reaching beyond scientific research).

5.3.1 Personal data and privacy

Privacy and data protection are important fundamental rights enshrined in Articles 7 and 8 of the Charter of Fundamental Rights of the European Union (CFREU) and Article 8 of the European Convention on Human Rights (ECHR) respectively. These rights may be affected by transparency and data access, as illustrated by cases (see text box) below.

---

[404] In context of GDPR, see notably: Ausloos, Toh and Giannopoulou 2022.

[405] As suggested for instance in the GDPR (Article 40) and the DSA (Article 45). A concrete example of a relevant code of conduct, see EDMO 2022.

[406] Cf. Ausloos, Leerssen and Ten Thije 2020.

[407] Delegated acts are non-legislative acts of general application that can be adopted by the European Commission. They can only be adopted if there is a delegation of power in a legislative act. A delegated act may "supplement or amend certain non-essential elements of that legislative act". The Commission consults experts, or expert groups before it adopts a delegated act.

See: <https://www.consilium.europa.eu/en/council-eu/decision-making/implementing-and-delegated-acts/>.

[408] GDPR.

[409] Article 17(2) CFREU.

[410] See for instance Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Trade Secret Directive).

In the joint cases C-37/20 and C-601/20, the Court of Justice of the EU (CJEU) clarified the privacy and data protection implications of detailed information held in **public registers**. In Luxembourg, a law had been adopted in accordance with the EU Anti-Money Laundering Directive based on which information regarding the ultimate beneficial owners of registered entities must be kept in a register.[411] Some information kept in that register was publicly accessible but beneficial owners were allowed to file a request to restrict access to that information. The CJEU ruled that the provision providing the general public access to the information on beneficial ownership constituted a 'serious interference' with the fundamental rights of private and family life and the protection of personal data as respectively laid down in Articles 7 and 8 CFREU. The interference could not be considered 'strictly necessary' and was therefore deemed unlawful. In the Dutch context, these joint cases have been particularly relevant for the *UBO register*, which also contains information on beneficial ownership and is governed by the Dutch Chamber of Commerce. As a result of the CJEU ruling, extracts of information from the UBO register cannot be purchased anymore.[412]

Of all third-party interests, privacy and data protection featured most frequently throughout the reviewed frameworks as constraints on transparency and data access.[413] Article 40 of the **Digital Services Act**, for example, explicitly requires the protection of personal data from collection until publication.[414] The **Open Data Directive** is not even applicable to documents access to which is restricted under national access to documents regimes on grounds of protection of personal data.[415] Furthermore, the **Data Governance Act** states that conditions on re-use should be designed in such a way to ensure effective safeguards for the protection of personal data such as anonymisation of the data before transmission.[416] Article 1(3) DGA also prescribes that in the event of a conflict between the DGA and Union law and national law on the protection of personal data, the latter shall prevail. The same principle is laid down in the **proposed Data Act**.[417] Moreover, the pDA prescribes that any personal data generated by the use of an IoT product or service shall only be made available to users or third parties where there is a valid legal basis under the GDPR.[418]

5.3.2 Intellectual property rights

In addition to data protection, many transparency and data access provisions also take account of intellectual property rights, especially copyright. For instance, the **Open Data Directive** is not applicable to documents for which third parties hold IP rights,[419] and the directive's obligations only apply insofar as they are compatible with international agreements on the protection of IP rights.[420] Similarly, the **Data Governance Act** states that the re-use of protected public data is only allowed in compliance with IP rights[421] and that IP rights held by third parties and public sector bodies should not be affected or limited by the DGA.[422] To

[411] Judgement of the Court, 22 November 2022, WM (C-37/20) and Sovim SA (C-601/20) v. Luxembourg Business Registers, Court of Justice of the European Union, ECLI:EU:C:2022:912.

[412] See Letter of the Minister of Finances of 20 January 2023 (Kamerstukken II 2022-23, 31 477, nr. 85).

[413] See e.g., Article 4(1)(b) EUDR, Article 1(2)(h) ODD, Article 1(3)(h) and Article 5(3)(a)(i) DGA.

[414] See Article 40(8)(d) DSA. Noteworthy in this regard is the work of the European Digital Media Observatory (EDMO) in developing a detailed code of conduct on how platform data access can be operationalised in a GDPR-compliant manner, see EDMO 2022.

[415] Article 1(2)(h) ODD.

[416] Recital 15 and Article 5(3)(a) DGA.

[417] See Article 1(3) pDA.

[418] Article 4(5) and 5(6) pDA.

[419] Article 1(2)(c) ODD.

[420] Article 1(5) ODD.

[421] Article 5(7) DGA.

[422] Recital 17 DGA.

this end, the DGA lays down (technical and legal) procedural requirements that public sector bodies may impose to preserve the protected nature of the data before they are made available for re-use.[423] It also states that public sector bodies should "exercise their [own] copyright in a way that facilitates re-use".[424] Finally, the **proposed Data Act** declares that the Regulation is without prejudice to Union and national law on the protection of IP,[425] and that when data is made available, IP rights are respected.[426] Importantly, while the **Copyright in the Digital Single Market Directive** introduced a very helpful mandatory copyright exception on text and data mining for scientific research purposes,[427] this exception does not really enhance researchers' *access* to copyright-protected data, as it merely allows the use (for text and data mining purposes) of copyright-protected data that users already have *lawful access* to.

5.3.3 Trade secrets and other commercially confidential data

Trade secrets[428] and other commercially confidential data are often mentioned in the same breath as data protection and IP, and are sometimes referred to as "quasi-IP rights".[429]

The **Open Data Directive**, for instance, does not apply to documents that are excluded from access under national access to documents regimes on grounds of commercial confidentiality, including business, professional or company secrets.[430] In the specific context of research data that must be made openly available and re-usable (by research organisations), the directive also states that "legitimate commercial interests" must be taken into account, in accordance with the principles of 'as open as possible, as closed as necessary'.[431] Additionally, the **Data Governance Act** – while it aims to stimulate the re-use of, inter alia, commercially confidential data[432] held by the public sector – emphasises that Member States should provide support to public sector bodies to make use of techniques to ensure the safe re-use of commercially confidential business data for research.[433] For example, data containing commercially confidential information can be modified before transmission, to such an extent that no confidential information is disclosed.[434] The DGA also states that re-use should be without prejudice to the EU Trade Secrets Directive.[435] Moreover, according to Article 40(5)(b) of the **Digital Services Act**, VLOPs and VLOSEs may request amendments of researchers' data access requests on the grounds that access will lead to "significant vulnerabilities in the protection of confidential information, in particular trade secrets".

---

[423] Article 5 DGA.

[424] Recital 17 DGA.

[425] Article 1(4c) and recital 13 pDA.

[426] Recital 28 pDA.

[427] Article 3 CDSMD. Text and data mining is defined in Article 2(2) CDMSD as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations".

[428] Trade secrets are defined in Article 2(1) of Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Trade Secrets Directive) as "information which meets all of the following requirements: (a) it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) it has commercial value because it is secret; (c) it has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret".

[429] See e.g., Fia 2022.

[430] Article 1(2)(d)(iii) ODD.

[431] Article 10(1) and (2) and Recital 28 ODD. See also Noto La Diega, p. 4.

[432] According to recital 10 DGA, "commercially confidential data" includes "data protected by trade secrets, protected by know-how and any other information the undue disclosure of which would have an impact on the market position or financial health of the undertaking".

[433] Recital 7 DGA.

[434] Recital 15 and Article 5(3)(a)(ii) DGA.

[435] Trade Secrets Directive, recital 10 DGA.

Importantly, this provision cannot be used as a blanket refusal: VLOPs/VLOSEs have to propose alternative data access arrangements that are appropriate and sufficient for the initial request.[436] Article 42(5) DSA further states that in case a publicly available risk assessment report contains information that a VLOP deems likely to "cause significant vulnerabilities for the security of its service", this information may be removed.[437]

The growing importance of trade secrets in transparency and data access regimes is particularly underscored by the ongoing legislative process of the **proposed Data Act**. One of the main amendments proposed by the Council of the EU relates to the balancing of the protection of trade secrets and IP rights and the sharing of IoT data.[438] Under Article 4 pDA, users of IoT products and services have the right to receive IoT data from the data holders. Data holders can however require that the confidentiality of data containing trade secrets is preserved, through technical and organisational measures, before disclosure.[439] In "exceptional circumstances", a data holder may even *refuse* the request for access, despite measures already taken by the user to accommodate the data holder's interests. This is the case when the data holder can demonstrate that they are "highly likely to suffer serious damage from the disclosure of trade secrets".[440] As regards the user's right under Article 5 pDA to share IoT data with third parties (e.g., researchers) as designated by the user, Article 5(8) states that trade secrets shall only be disclosed if that is "strictly necessary to fulfil the purpose agreed between the user and the third party" and "all specific necessary measures (…) are taken by the third party to preserve the confidentiality of the trade secret". Again, under "exceptional circumstances" the data holder may refuse the access request.[441]

***

Importantly, the protection of third-party rights and interests does not have to lead to the *full* refusal of a researcher's request for data access. It is key to find **the right balance** between transparency and data access for scientific research and the protection of those rights and interests. That balance can, for instance, be achieved by providing access **under certain conditions**, such as the anonymisation of personal data,[442] the modification or aggregation of commercially confidential data,[443] or the placement of information "in a range"[444] to the extent necessary to protect commercial legitimate interests.

Although the need for researchers' access to data is increasingly acknowledged by law- and policymakers, the EU's digital/data legislation analysed for this report still seem to vacillate between the

[436] Article 40(6) DSA. Please note that the Digital Services Coordinator of establishment can decide not to amend the request to the wishes of the VLOP.

[437] Article 42(5) DSA. VLOPs and VLOSEs must however send the complete reports to the Commission and the Digital Services Coordinator of establishment, accompanied by a statement of reasons for the removal of information from the publicly available reports.

[438] Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) – Mandate for negotiations with the European Parliament, 17 March 2023, 7413/23. See also European Council 2023 (webpage) <https://www.consilium.europa.eu/en/press/press-releases/2023/03/24/data-act-member-states-agree-common-position-on-fair-access-to-and-use-of-data/>.

[439] Article 4(3a) and recital 28a pDA.

[440] Article 4(3a) pDA. "Serious damage" is understood as "damage with an adverse effect on the conduct of economic activity, when the data holder would face significant economic losses, which could, in particular, threaten its viability or pose a serious risk of bankruptcy", see recital 28a pDA.

[441] Article 5(8a) pDA.

[442] Article 5(3)(a)(i) DGA.

[443] Article 5(3)(a)(ii) DGA.

[444] Article 11(4) pPAR.

protection of third-party interests and transparency.[445] Scholars have argued that in some contexts, this balance "appears to be struck in favour of the former".[446] While the idea of balancing rights and interests may be praiseworthy, there is always a **risk that the balancing exercise ends up being decided by more powerful actors**, thus reinforcing data enclosures.[447] As also came to the fore in the empirical analysis in Part A of this report, powerful players sometimes invoke data protection arguments or trade secrets protection as an 'excuse' to deny access to sensitive, yet interesting information for both the public and the research community.[448] This could seriously obstruct important societal research into exactly these powerful players.

Finally, while many transparency and data access provisions require a balancing act on paper, they often remain unclear as to how this should work out in practice[449] since they do not provide for any rules on how to perform the balancing exercise between data access rights and the protection of IP rights and other third-party interests *in concreto*.[450] What seems clear, however, is that where the legislator has explicitly provided for data access, the protection of third-party rights and interests cannot result in an outright refusal to provide data to the data recipient.[451] In other words, the balance should tip in favour of those requesting data –respecting the provision's requirements – by default, while data holders bear the burden of proof to establish why data access cannot or only partially be accommodated.

5.4 Data sharing intermediaries/facilitators

Looking at the different legal instruments, another trend can be discerned: the creation and regulation of institutional "intermediaries" to facilitate the provision of data access for use by others. These organisations, which can be public sector bodies or private legal entities (see Table 4), are third-party agents whose common denominator is that they **connect data providers and data users**. Their services may vary from e.g., the making available of technical infrastructures[452] to non-technical services such as the issuance of data access requests on behalf of data users.[453]

Some of the novel intermediaries are envisioned as the key facilitators of what have been referred to in EU digital policymaking as "**common European data spaces**".[454] The European Commission intends to establish several domain-specific spaces in which data-sharing tools and platforms, data processing and computing capacities and data governance frameworks are provided so as to enable the easy sharing of data within and across sectors.[455] In May 2022, the European Commission issued its first proposal

[445] Indeed, recital 97 DSA notes that to ensure the DSA's objective is achieved "consideration of the commercial interests of providers should not lead to a refusal to provide access to data necessary for the specific research objective pursuant to a request" under the DSA and that "providers should ensure appropriate access for researchers, including, where necessary, by taking technical protections such as through data vaults".

[446] See e.g., Noto La Diega 2023, p. 5 on the Data Governance Act.

[447] Noto La Diega 2023, p. 18.

[448] See also Clark 2021.

[449] See for instance: Request for a preliminary ruling from the Verwaltungsgericht Wien (Austria) lodged on 16 March 2022 – CK (Case C-203/22). In this pending CJEU case, preliminary questions have been posed on how the right of access as laid down in Article 15(h) GDPR should be interpreted in relation to a potential violation of the protection of trade and business secrets (in this case, through the partial disclosure of an algorithm). The fact that such questions have been posed in a request for a preliminary ruling indicates that it is unclear how the balancing exercise must be performed.

[450] In the context of the proposed Data Act in particular, Geiregat laments the lack of any guidance on how to settle conflicts between data access and IP rights, other than providing for an exemption for IoT data from the *sui generis database* right in Article 35 pDA. See Geiregat 2022.

[451] Compare e.g., recital 63 of the GDPR.

[452] See recital 32 DGA with regard to data intermediation services providers in business-to-business and business-to-consumer contexts.

[453] See e.g., Article 40(8) DSA on the Digital Services Coordinator.

[454] See European Commission Communication 2020b.

[455] Ibid., pp. 16-17, 21-22.

for the establishment of a data space: the European Health Data Space.[456] Data altruism organisations and data intermediation services providers as defined in the DGA are likely to play an important role in the operationalisation of these data spaces. For example, the European Open Science Cloud (EOSC) – a long-term initiative which has recently been recognised as the "science, research and innovation data space"[457] (see text box below) – will largely depend on the voluntary data sharing activities facilitated by data altruism organisations. Data intermediation services providers, which aim to establish *commercial* relations between data holders and data users, could in theory also become part of the EOSC, but it seems that the Cloud will mainly take a not-for-profit approach.[458] This is not to say, however, that for-profit entities cannot be involved as customers or providers of EOSC services; they can (and are[459]), but on a not-for-profit basis.

> The **European Open Science Cloud (EOSC)** is a joint initiative of the European Commission and the European research community to develop a virtual, "federated and open multidisciplinary environment" for European researchers, innovators, companies and citizens "where they can publish, find and re-use data, tools and services for research, innovation and educational purposes".[460] Notably, the EOSC is envisioned to go beyond merely offering a technical infrastructure but also to provide for e.g., licensing models and interoperability guidelines.[461]

Other institutional intermediaries have been created with rather specific data exchanges in mind, serving as conduits between data holders and data users. Examples of these types of facilitators are the 'single point of access' under the Open Data Directive, which makes available datasets held by public sector bodies; the Digital Services Coordinator under the Digital Services Act, which issues data access requests with very large online platforms on behalf of vetted researchers; and the European Commission in its capacity as manager of a database of registered high-risk AI systems and foundation models (Table 4).

The rise of (regulatory frameworks for) data sharing intermediaries/facilitators is, in principle, a positive development from a researchers' perspective. While not always created *for* science or academia in particular,[462] researchers may certainly benefit from them. First, they might make data sharing and access **more efficient**, as researchers do not have to set up ad hoc architectures and actively gather research subjects to donate their data but instead can fall back on existing datasets and sharing infrastructures. Second, professional intermediaries can **help overcome legal obstacles** by handling data in a GDPR- and IP law-compliant manner, for instance by providing for secure analysis environments.[463] Third, and relatedly, when data holders **trust** intermediaries with their data, they may decide to **make more data available** that could be useful for scientific research.

---

[456] European Commission, Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, COM(2022) 197 final. This proposal has not been discussed in the report, as it falls outside the scope of the research being a domain-specific rather than a horizontal framework.

[457] See European Commission EOSC (webpage) <https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en>.

[458] European Commission and Van Eechoud 2022, p. 31.

[459] See e.g., <https://providers.eosc-portal.eu/stats/providers>.

[460] <https://eosc-portal.eu/about/eosc>.

[461] See e.g., the 'EOSC Interoperability Framework' (EOSC-IF), <https://eosc-portal.eu/eosc-interoperability-framework/about-eosc-interoperability-framework-governance-eosc-if>.

[462] The legal frameworks for data intermediation service providers and data altruism organisations established in the DGA, for example, have been created to improve the conditions for data sharing in the **internal market**. That said, Article 3(3) of the Treaty on European Union (TEU) explicitly recognises that the internal market "shall promote scientific and technological advance".

[463] For example, the Dutch collaborative organization for IT in research and education 'SURF' has developed, together with partners, a "secure analysis environment" (SANE), in which researchers can analyse sensitive data while the data holders retain control over the data, see <https://www.surf.nl/en/news/sane-secure-data-environment-for-social-sciences-and-humanities>.

At the same time, some questions remain unanswered as to the – so far mainly theoretical – potential of these intermediaries. A mismatch in incentives between data sharing intermediaries and researchers may render the intermediaries impractical or even useless to researchers. Moreover, there has been some unclarity as **to what extent these new legal concepts apply to entities already operating** in the data sharing field. Such organisations come in many shapes and forms, as can be seen from Table 4. Do initiatives aimed at facilitating data sharing in the context of scientific research, such as ODISSEI[464] and the Virtual Research Environment of the University of Amsterdam,[465] qualify as 'data altruism organisations' within the meaning of the Data Governance Act? And if so, where do these organisations have to be registered as such? Moreover, it is uncertain which rules are applicable to hybrid entities that facilitate both commercial and non-commercial exchanges of data.[466] AMdEX, for instance, is an exchange infrastructure that is currently being developed by the University of Amsterdam together with other private, public and scientific partners and which aims to support an open data market for all sorts of organisations and individuals ("by all and for all"[467]) where data can be shared in a secure way and under data holders' own terms and conditions. To date, the AMdEX-project has focused on four individual data markets or "use cases",[468] consisting both of commercial and non-commercial data exchanges. The question, therefore, is to whether and to what extent AMdEX fits within the new frameworks for data intermediation service providers and for data altruism organisations.

Data-sharing intermediaries/facilitators	Definition and tasks	Legal framework
"Data intermediary services provider (recognised in the Union)"	A provider of data intermediation services (see Article 2(11) DGA), which may include public sector bodies,[469] and which constitutes a separation in the data economy between data provision, intermediation and use.[470] Data intermediary services providers may only act as intermediaries in transactions and can therefore not use the data exchanged for any other purpose.[471]	DGA
"Data altruism organisation (recognised in the Union)"	A legal person that seeks to support objectives of general interest by making available relevant data based on data altruism at scale and that meet the requirements laid down in the Data Governance Act.[472]	DGA
"Single point of access"	An entity through which citizens can request the making available of datasets held by public sector bodies with regard to the documents to which the Open Data Directive applies.[473]	ODD
"(National) single information point"	A body or structure through which citizens can inquire about, or request the re-use of, protected categories of data referred to in Article 3(1) of the Data Governance Act. The single information	DGA

[464] ODISSEI is a national research infrastructure that brings social sciences researchers in the Netherlands together with data, expertise and resources provided by various partners, see <https://odissei-data.nl/en/>.

[465] The Virtual Research Environment is a cloud-based working environment for researchers to share data and analyse data, UvA 2023 (wegbpage) <https://www.uva.nl/en/content/news/news/2023/03/virtual-research-environment-uva-wide-available.html?origin=kUP%2Byx6UTZqvuJiCJKnnEQ&cb>.

[466] European Commission and Van Eechoud 2022, p. 31.

[467] See <https://amdex.eu/about/>.

[468] A description of the four 'use cases' can be found here: <https://amdex.eu/usecases/>. The use cases included markets for (i) aircraft maintenance data, (ii) data generated by smart buildings, (iii) sensor data collected in public spaces, and (iv) research data. While the research data market is mainly not-for-profit, the other markets do contain commercial data exchanges.

[469] Recital 27 DGA.

[470] Recital 32 DGA.

[471] Recital 33 DGA.

[472] Recital 3 DGA.

[473] Article 9(2) ODD.

	point makes available by electronic means a searchable asset list containing an overview of all available data resources including, where relevant, those data resources that are available at sectoral, regional or local information points, with relevant information describing the available data, including at least the data format and size and the conditions for their re-use. The single information point may be linked to sectoral, regional or local information points, and may be automated provided that the public sector body ensures adequate support.[474]	
"European single access point"	An entity that offers a searchable electronic register of data available in the national single information points (see above) and further information on how to request data via those national single information points.[475]	DGA
European Commission	The European Commission will set up and manage an EU database where providers of high-risk AI systems must register their high-risk AI system.[476]	pAIA
"Digital Services Coordinator"	An authority in a Member State appointed with the task of supervising the application and enforcement of the Digital Services Act. It has to coordinate and cooperate with other national competent authorities, and acts as the single contact point with regard to all matters related to the application of the Digital Services Act for the European Commission, the Board, the Digital Services Coordinators of other Member States and other national competent authorities.[477] The Digital Services Coordinator, amongst other things, issues reasoned requests for data access, on behalf of vetted researchers, to providers of very large online platforms or very large online search engines.[478]	DSA
"Independent third-party body"	An independent third-party body, (to be) funded and set up by the signatories of the Strengthened Code of Practice on Disinformation of 2022, which can vet researchers and research proposals and cooperates with the signatories to enable the sharing of personal data necessary to undertake research on disinformation with the vetted researchers.[479]	2022 CoP
"Common Transparency Centre website"	A publicly available, user friendly and searchable website on which the signatories of the Strengthened Code of Practice on Disinformation of 2022 publish insights and data on online disinformation.	2022 CoP

Table 4. Examples of data sharing intermediaries/facilitators throughout the legal frameworks

5.5 Format requirements and other formalities

Last but not least, we noted that the transparency and data access provisions across the examined frameworks contain many requirements as to the ways in which data should be requested and subsequently provided. Such requirements can relate to the **data or information itself** (e.g., they must be 'clear' and

[474] Article 8(1)-(2) DGA.
[475] Article 8(4) DGA.
[476] Recital 69 and Article 60 pAIA.
[477] Recital 110 DSA.
[478] Article 40(8) DSA.
[479] Commitment 27 2022 CoP.

'easily understandable'), the **formats** in which the data must be provided (e.g., in a 'commonly used' format), the **timelines** to be adhered to (e.g., 'without delay') and the **conditions** under which data may be shared (e.g., 'non-discriminatory'). The concepts describing these requirements, however, are often not clearly defined. This vagueness could potentially complicate the process of data access.

5.5.1 Concerns about unclear format requirements

Open-ended format requirements allow for flexibility and context-dependent interpretations,[480] yet also raise critical issues. To start, it is important to provide data recipients with (legal) certainty and **protect them against arbitrary outcomes** from data access request procedures – data recipients should be able to know what data they can expect from, for instance, a data controller, a platform or an IoT service provider. Furthermore, data holders may hinder scientific research by not adhering to clear and consistent formats, because **changing formats may render it difficult to conduct longitudinal and comparative research**.[481] Experiences from data donation projects based on the upload of 'data download packages' under Article 15 GDPR have shown, for example, how frequent format changes may complicate research.[482] Finally, uncertainties with regard to procedures and formats have the potential **to perpetuate certain power asymmetries**.[483] Data providers who may already be reluctant to provide data benefit from the fact that they can engineer their own online interfaces/formats and thereby complicate access to data for researchers.[484]

5.5.2 Tackling concerns about formats and other formalities

As noted above, the volatility of data formats and unclarity of procedural requirements may render it harder for researchers to observe digital infrastructure and/or re-use the respective data over time. A number of strategies and soft law instruments may help overcome this problem. One strategy is that of **standardisation**. The **Open Data Directive**, for example, encourages the use of 'standard protocols' and 'standards for datasets'.[485] Public sector bodies and public undertakings must make their documents, where possible, available for re-use in a format that complies with 'formal open standards'.[486] Similarly, Article 44 of the Digital Services Act requires the European Commission to support and promote the development and implementation of 'voluntary standards', for example in respect of the communication with recipients of intermediary services in a user-friendly manner on restrictions based on terms and conditions,[487] APIs to facilitate researchers' access to platform data,[488] and interoperability of advertisement repositories.[489] Importantly, for standardisation to serve the interests of researchers and civil society more broadly, it should be ensured that standards are the product of a democratic process – in which researchers are also represented – and be avoided that powerful actors managing digital infrastructures hijack this process.

---

[480] For instance, a format that is 'commonly used' may be very dependent on the sector in which it's used, and 'within a reasonable time' may mean something totally different in case of responding to an online personal threat than in a request for (a large amount of) governmental documents.

[481] Cf. Van Drunen and Noroozian 2023, pp. 2-3.

[482] Hase et al (forthcoming).

[483] For instance, in the context of (social media) platforms, see: Van Drunen and Noroozian 2023.

[484] Van Drunen and Noroozian 2023, pp. 2-3.

[485] Recital 32 ODD.

[486] Article 5(1) ODD. See also Article 2(15) ODD, which defines 'formal open standard' as "a standard which has been laid down in written form, detailing specifications for the requirements on how to ensure software interoperability".

[487] Article 44(1)(b) DSA.

[488] Article 44(1)(d) DSA.

[489] Artice 44(1)(f) DSA.

Additionally, the EU regulator has recognised the importance of **codes of conduct** in further clarifying format-requirements and applicable procedures for transparency and data access. Codes of conduct are self-regulatory, non-legally binding and usually sector-specific documents that contain rules to which signatories can decide to commit themselves.[490] The proposed **Political Advertising Regulation** for example, mentions codes of conduct as a means to help the proper application of transparency requirements for political advertisements.[491] Similarly, the **General Data Protection Regulation** sets out a list of requirements for codes of conduct intended to contribute to the proper application of GDPR.[492] An existing example of a code of conduct is for instance the **2022 Strengthened Code of Practice on Disinformation**, which (still) leaves much room for interpretation on key terms and requirements to its signatories. It states, for instance, that signatories must develop tools for real-time access to certain data for research purposes but fails to clearly define relevant requirements.[493] Indeed, because of their non-binding and flexible nature, codes of conduct may prove ineffective in situations with strong power asymmetries. The **Free Flow of Non-Personal Data Regulation**, for instance, introduced codes of conduct as a tool to enhance the switching between data processing services providers and the porting of data.[494] However, the European Commission is of the opinion that this self-regulatory approach has not proven effective so far, which is why a regulatory (and binding) approach has been proposed in the Data Act.[495]

Another way through which the European regulator anticipates further clarification of transparency and data access format- and procedural requirements, is by providing **regulatory guidance**. Such guidance could be drafted by enforcement agencies or advisory bodies such as the European Data Protection Board (EDPB)[496], the European Board for Digital Services (EBDS)[497] and the European Data Innovation Board (EDIB).[498] The EDPB, for instance, has published extensive guidelines on the right of access as laid down in Article 15 of the GDPR.[499] The EBDS' mission includes to "support and promote the development and implementation of European standards, guidelines, reports, templates and code of conducts".[500] The EDIB shall develop guidance on handling requests for the re-use of data, operationalisation of data altruism and more.[501]

Finally, a few legislative instruments provide that certain specifics related to transparency requirements are to be laid down in **delegated and implementing acts**.[502] Most notably, the **Digital**

---

[490] A Dutch example is the 'Code voor de journalistiek', *Nederlandse Vereniging van Journalisten* <https://www.nvj.nl/themas/journalistieke-praktijk/ethische-regels/code-journalistiek>.

[491] Article 7(7) pPAR.

[492] Articles 40-41 GDPR. One example of a code of conduct developed under these provisions is the European Digital Media Observatory's (EDMO) draft code of conduct for 'platform-to-researcher data sharing' which also aims to facilitate the operationalisation of Article 40 DSA, see EDMO 2022.

[493] 2022 CoP, p. 28 (QRE 26.2.1-3).

[494] Notably, best practices for facilitating the switching of service providers and the porting of data, minimum information requirements regarding processes, technical requirements, timeframes, etc. See Article 6(1) NPDR. See also sections 4.2 and 4.3.2 of this report.

[495] Explanatory Memorandum and recital 70 pDA.

[496] The EDPB is an official EU body established in Article 68 GDPR. It consists of delegates from supervisory authorities from all EU member states and of the European Data Protection Supervisor (or their respective representatives). The EDPB is responsible for, among other things, monitoring compliance with the GDPR.

[497] The EBDS is an independent advisory body composed of representatives of the national authorities, the Digital Services Coordinators, see Chapter IV, Section 3 of the DSA.

[498] The EDIB is an expert group, consisting of representatives of competent national authorities, the EDPB, the EDPS, ENISA, the Commission, the EU SME Envoy (or a representative), and other representatives of relevant bodies in specific sectors or bodies with specific expertise, see Article 29(1) DGA.

[499] EDPB 2023.

[500] Article 63(1)(e) DSA.

[501] Article 29-30 DGA.

[502] Delegated acts are non-legislative acts of general application that can be adopted by the European Commission. They can only be adopted if there is a delegation of power in a legislative act (in case of the DSA, Article 87). A delegated act may supplement or

Services Act anticipates the adoption of delegated acts to specify the technical and legal conditions under which VLOPs are to share data with vetted researchers,[503] as well as implementing acts to lay down templates for the form, content and other details of transparency reporting obligations.[504] The **proposed Political Advertising Regulation** also foresees a role for delegated acts to specify the form in which information in the transparency notices and information about targeting should be provided.[505] Pursuant to the **Open Data Directive**,[506] on 20 January 2023 the Commission adopted an implementing act on high-value datasets and templates for re-use, formats of data and metadata, and technical arrangements for dissemination.[507] Lastly, the **Data Governance Act** prescribes the adoption of delegated acts establishing a rulebook laying down, inter alia, appropriate information requirements regarding the use of data and recommendations on interoperability standards.[508] It also encourages the adoption of implementing acts on establishing a "European data altruism consent form", which must allow the collection of consent or permission across Member States in a uniform format.[509]

amend certain non-essential elements of that legislative act. The Commission consults experts, or expert groups before it adopts a delegated act. Implementing acts provide rules with details on the application of a basic acts, in case uniform (practical) conditions across the EU are needed. Implementing acts cannot delete, add or change anything in the basic act, they only implement the content of the act (without changing its substance). <https://www.consilium.europa.eu/en/council-eu/decision-making/implementing-and-delegated-acts/>.

[503] Article 40(13) DSA. At the time of writing, the European Commission held a public consultation for input in the drafting process of this delegated act. See: <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en>.

[504] Article 24(6) DSA.

[505] Recital 66 jo. Article 7(8) pPAR.

[506] Article 14(1) ODD.

[507] Commission Implementing Regulation (high-value datasets).

[508] Article 22(1)(a) and (d) DGA.

[509] Article 25(1) DGA. See further section 3.1.4.

6. How to use EU digital/data law to access data for research

As apparent from the legal mapping exercise and analysis, there is little clarity on the position of academic researchers and their claims to data generated and/or held by digital infrastructures. Only few provisions explicitly recognise academic researchers' interests, let alone grant them privileged transparency or data access rights. Despite the legal patchwork and general vagueness of transparency and data access provisions, **this chapter aims to offer guidance on how researchers could deploy the provisions that do exist, in particular to acquire individual-level data versus system-level data**. To do so, it is useful to briefly reiterate three ways of qualifying and categorising data access provisions in the legal frameworks discussed (cf. section 1.2.3):

- **Proactive v. reactive access.** Many legal frameworks put in place transparency duties that require data holders to make data available proactively, that is, without the need for an official request by the data recipient, for instance by publishing data on a website. We referred to this type of measures as proactive data access measures: the data should be made *readily available by the data holder* for others to observe and explore, sometimes under certain conditions (e.g., contractual). This is different from reactive data access measures, which require *data recipients to first take active steps*, often through a request (procedure), as a response to which the data holder must provide data. Oftentimes, reactive data access enables for more detailed information, as it typically concerns data that cannot be made available to the public at large due to conflicting rights and interests. In sum, data provided on the basis of proactive access provisions are typically easily accessible for researchers but comprise less detailed information, and data provided on the basis of reactive access provisions may be more cumbersome for researchers to acquire but often contain more detailed information.

- **Direct v. indirect access.** It is also useful to distinguish between provisions based on the entity providing access to the data ('access point'). This is relevant for researchers in that it determines whom they must approach and indicates whether the data may have been subject to an additional level of interpretation. Data access could either be provided by the entity that originally produced or kept the data (*directly*), or by a third-party entity (*indirectly*).

- **System-level v. individual level data**. A third and final distinction relates to the scope of the data that the provision provides access to: system-level or individual-level data. While individual-level data *relate to the specific endpoints of a digital infrastructure* (e.g., a person using the digital infrastructure, or an IoT device such as a smart fridge or car), system-level data *relate to (parts of) the infrastructure as a whole* (e.g., advertising criteria, demographics, etc.). Which type of data access provision is more relevant for researchers will very much depend on their research question.

6.1 Access to individual-level data about the endpoints of digital infrastructures

Several legal provisions analysed for this report enable access to individual-level data, i.e., information about the endpoints of digital infrastructures. These endpoints can be either **individuals** (e.g., 'data subjects' under

the GDPR), **legal persons** (e.g., 'advertisers' or 'business users' under the DSA, DMA and P2BR) or **objects** (e.g., IoT-devices under the pDA).

6.1.1 General considerations

Most provisions laying down individual-level data access requirements are *reactive* in nature, which means that the relevant data must be requested by the data recipient. Entities in control of the respective digital infrastructures may use various technical means to comply with individual-level data access provisions. For example, a growing number of large online platforms – from social networks and dating apps, to gig economy and content providers – offer "download my data"-functionalities to comply with the right of access under the GDPR and similar frameworks across the world. Although the datasets that are made available through these functionalities are rarely complete, they tend to reveal a lot of detailed information about the data subject and their interactions with the platform. It can be expected that similar tools will be developed to comply with new and forthcoming individual-level data access provisions, such as those in the proposed Data Act.

Some individual-level data access provisions, however, require *proactive* measures by digital infrastructures, meaning that information needs to be provided without a request having to be made first. For example, under the DMA, 'gatekeeper platforms' must provide real-time metrics data to publishers and advertisers about advertisements on a daily basis. This obligation is likely to be implemented through interactive dashboards or APIs.[510] Furthermore, online platforms must explain their decisions to restrict or downrank content, or to suspend or terminate certain user accounts, to the platform users affected.[511]

None of the reviewed provisions providing for access to individual-level data are explicitly aimed at researchers. Instead, they are aimed at specific types of rightholders, such as (professional) platform users, consumers and data subjects. This is not surprising, as most provisions have primarily been introduced, or at least in part, for economic reasons, namely to ensure fair competition and support internal market objectives more broadly (e.g., DSA, P2BR). Some provisions are informed by fundamental rights and freedoms and require minimum levels of transparency in that regard (e.g., GDPR, LED). The objectives of the respective data access provisions typically determine the type of rightholders: businesses for economically inspired provisions, and individuals or citizens for fundamental rights-based provisions. This, in turn, affects the utility of the individual-level data access provisions for researchers.

6.1.2 Specific opportunities and limitations

There are essentially two options for researchers to use individual-level data access provisions that are not explicitly addressed to them. **First, they can obtain access to (limited) individual-level data directly by acting in the capacity as the data recipient/rightholder addressed by the relevant provision**. For instance, a researcher can create an account with an online platform and become a platform user (DSA), buy a smart product and become a product user (pDA), or more generally be a data subject whose personal data are processed (GDPR), and in those capacities, demand their "own" individual-level data from the digital infrastructure and use it for research purposes. Unsurprisingly, this strategy might work better for certain data than for others. Indeed, individual researchers do not typically qualify as "business users" or "advertisers" (e.g., under the DMA and P2BR), which means that data access provisions tailored to these actors are less relevant to them. The biggest limitation of this approach, however, is that it only allows researchers to use their *own* data and not data of *other* individuals at a larger scale. Evidently, one can only

[510] Article 5(9) DMA.

[511] E.g., Article 4 P2BR and Article 17 DSA.

do so much research based of the data generated by themselves. Researchers may also be hesitant to use their own personal data for privacy reasons, as anonymisation techniques are of limited value when using personal data of a small group of individuals. Moreover, this approach may raise a number of economic and ethical issues since it requires researchers to act in different legal capacities and/or purchase certain services or products. It is important for researchers to consider the implications of deploying individual-level data access provisions on scientific standards.

Second, researchers can obtain access to (more) individual-level data indirectly by collecting relevant data from the rightsholders addressed by the individual-level data access provisions. Business users, data subjects or IoT-device owners may under certain conditions be open to share their individual-level data with researchers. An important advantage of this approach is that it enables the scaling up or crowdsourcing of individual-level data access, thus giving researchers access to bigger and richer datasets. In recent years, this practice has increasingly been professionalised and institutionalised via **data donation projects** (see the text box below).[512] A downside of this approach, however, is that researchers are dependent on the willingness of individuals to share their data with them.

> **Data donation projects** exist in various forms. It can range from a simple ad hoc initiative by researchers that recruit their own research subjects to donate their data, to the involvement of a professional third-party entity that facilitates the sharing of data (section 5.4). In between these two, there are initiatives such as the UvA's Digital Data Donation Infrastructure (D3I) project,[513] which offers a technical infrastructure with legal, ethical, and methodological guidelines aimed at streamlining the data donation process. This is important, considering the significant issues that may emerge in this context, varying from GDPR compliance to ethical standards applicable to research involving human research subjects. However, existing data donation projects experience very low retention numbers, with many participants dropping out because of the lengthy or complex processes of requesting their data from data holders in the first place. Looking ahead, we expect new legal categories of independent organisations – notably data intermediation services providers and data altruism organisations in the DGA[514] – to step in to play a central role in data donation projects (section 5.4).

To conclude, individual-level data access provisions offer significant opportunities to researchers, mainly to those wishing to obtain detailed information about the endpoints of digital infrastructures. Yet, these opportunities are also constrained by several drawbacks as described above. Their value will strongly depend on the specific research questions at hand, given that the provisions typically allow for granular information about specific endpoints and not the digital infrastructure as a whole. This limitation may partly be compensated by the crowdsourcing of individual-level data through data donation, bringing together a representative set of endpoints of the respective digital infrastructure, provided that researchers have the resources to scale and interpret the data in compliance with legal, ethical and methodological standards.

6.2 Access to system-level data about (aspects of) the digital infrastructure

As noted above, various legal provisions grant access to system-level data, i.e., information about (aspects of) a digital infrastructure as a whole rather than the specific endpoints. Such data include transparency reports drafted by online platforms on e.g., their content moderation practices (Article 15 DSA),

[512] See for example Araujo et al 2022.

[513] See <https://datadonation.eu/>.

[514] Recital 27 DGA.

advertisement archives (Article 39 DSA), and publicly available information on restrictions imposed on media services providers (Article 17(5) pEMFA).[515]

6.2.1 General considerations

The main advantage of system-level transparency and data access provisions is that they usually **provide a more holistic perspective on the digital infrastructure**. A 'helicopter view', if you will, enabling overarching insights into the system as a whole or at least into significant parts of it. Such a perspective may be more appropriate to respond to research questions, for instance when exploring discriminatory practices on online platforms.

In essence, system-level data access provisions require entities in control of digital infrastructures to share certain data relating to the infrastructure. The utility of the data for researchers will depend on the respective research purposes as well as the degree to which the data have been pre-processed (e.g., modified, aggregated, selected) by the respective entity. Only few of the system-level data access provisions explicitly refer to researchers as intended data recipients, and only one of the examined provisions gives them an actionable right to access data: Article 40 of the Digital Services Act. This provision enables researchers – under certain conditions – to request any data from online platforms, but due to its novelty, it has not been tested in practice yet.[516] While other provisions do not explicitly refer to researchers, they can still be valuable for research in different ways as discussed in section 6.2.2.

Many system-level transparency requirements are *proactive* in nature, meaning that those entities responsible must share or publish the respective data without anyone having to make a request. This is the case, for example, with compliance reporting obligations (e.g., Article 11 DMA), terms of service (e.g., Article 14 DSA; Articles 3-5 P2BR) and general information duties (e.g., on the parameters for ranking systems in the P2BR;[517] recommender systems in the DSA;[518] and the instructions for use for AI systems in the AI Act[519]). Furthermore, this report also identified several *reactive* system-level data access provisions, where system-level data only needs to be shared upon request, for example in the case of data intended for systemic risk research (Article 40 DSA) or certain public sector documents (Articles 6-10 EUDR).

System-level data access provisions may either target the public at large or only benefit specific data recipients, typically enforcement agencies and regulatory authorities but in some cases other entities, such as trusted flaggers. This distinction determines whether a researcher can access the data *directly* (i.e., as a member of the general public) or *indirectly* (i.e., from the designated data recipients) . Examples of system-level data access provisions aimed at the public at large are Article 60 of the proposed AI Act, which establishes a database for certain high-risk AI systems, and Article 24(2) of the proposed European Media Freedom Act, which contains an obligation for governments to publish information on the allocation of state advertising in the media sector. An example of a provision addressing specific data recipients is Article 11 of the Digital Markets Act, which imposes an obligation on gatekeepers to report to the European Commission on compliance with the Regulation.

[515] Notably, a large number of public transparency reporting obligations that produce system-level data are laid down in platform-governance frameworks such as the Digital Services Act. Other legislative instruments, such as e-commerce and consumer law frameworks, the proposed European Media Freedom Act and the proposed Political Advertising Regulation contain some public transparency obligations, but less so in formal *reporting* obligations, as imposed on platforms in e.g., the DSA and DMA.

[516] Article 40 (as most of the DSA provisions) becomes applicable from 17 February 2024, see Article 93(2) DSA.

[517] Article 5 P2BR.

[518] Articles 27 and 38 DSA.

[519] Article 13 pAIA.

6.2.2 Specific opportunities and limitations

Similar to the individual-level data access opportunities described in section 6.1.2, there are two main routes for accessing system-level data. **First, researchers can make use of system-level data that must be made publicly available based on proactive transparency obligations**. Data that must be made available to the *public at large* are obviously easily accessible. However, as mentioned above, some proactive system-level transparency obligations are only accessible to *specific data recipients*, typically a regulatory authority or other public sector body, which means that researchers can only access these data indirectly. In such cases, researchers could try to obtain access to (part of) the data by relying on open government provisions, for instance those enshrined in the Access to EU Documents Regulation, Open Data Directive and Chapter II of the Data Governance Act. Sometimes, public sector bodies themselves are legally obliged to proactively make (part of) the data received from the data holders publicly available. The European Commission must, for example, establish a publicly accessible database containing content moderation decisions taken by platforms, which the platforms must supply to the Commission.[520] A significant drawback of this indirect data access strategy is that by the time information gets to the researcher, it is often aggregated or interpreted and does not comprise the underlying source data.[521] Importantly, some proactive system-level data access provisions do not designate the general public or a regulatory authority as data recipients but rather a *third entity*, such as "trusted flaggers" (Article 22(3) DSA), the "European Centre for Algorithmic Transparency", or the "Transparency Centre" (2022 CoP) (see text boxes below). These actors are entitled to receive certain (platform) information that is not publicly available, and typically publish reports based on that information. By this route, the (platform) data become 'publicly available' anyway, albeit often in aggregated, modified or otherwise processed form.

Trusted flaggers are professional entities with a designated area of expertise in the context of illegal online content.[522] Just like other platform users, they can submit notices of allegedly illegal content to platforms (through a notice and action mechanism), with the difference that their notices must be given priority and decided upon without undue delay.[523] Trusted flaggers must also publish annual reports on the notices they have submitted through the notice and action mechanism, in which they report on the notices categorised by the identity of the platform, the type of allegedly illegal content and the action taken by the platform.[524] Those reports must be sent to the Digital Services Coordinator (national authority) and are made publicly available.[525]

The **European Centre for Algorithmic Transparency (ECAT)** is a research centre that was officially launched on 18 April 2023. The ECAT aims to improve the understanding of how algorithms work. Scientists and experts at the ECAT will "analyse transparency, assess risks, and propose new transparent approaches and

[520] Article 24(5) jo. Article 17(1) DSA.

[521] See e.g., Article 11(1)-(2) DMA, based on which gatekeepers shall provide the Commission with a report describing in a detailed and transparent way its compliance with the DMA. However, the gatekeeper shall provide the Commission with a non-confidential summary, too, which is the version that will be made publicly available by the Commission.

[522] The status of trusted flagger shall be awarded upon application by the Digital Services Coordinator to applicants who have demonstrated their particular expertise and competence, their independence and that they carry out their activities in a diligent, accurate and objective manner. In addition, trusted flaggers shall include an explanation of the procedures they have in place to ensure their independence. See Article 22(2)-(3) and Recitals 42 and 61 DSA.

[523] Article 16(1) DSA.

[524] Article 22(3) DSA.

[525] Article 22(3) DSA.

best practices".[526] With its research, the ECAT will help the European Commission monitor compliance of VLOPs and VLOSEs with their systemic risk obligations in the DSA. As such, the ECAT is part of the Commission and is hosted by the Joint Research Centre (JRC), the EU's research hub. The ECAT may publish its findings and make their research (data) publicly available, which in turn could be of use to academic researchers.

Chapter VIII of the 2022 CoP is dedicated to the establishment of the **Transparency Centre**.[527] This Transparency Centre is envisioned as a hub for information on the 2022 CoP and will host a repository of signatories' reports on their implementation of the 2022 CoP – including their commitments to empower the research community as laid down in Chapter VI. The first reports by the Transparency Centre have already been published.[528] The Transparency Centre also shows which platforms have signed up as signatory and to which commitments and measures exactly.

Second, researchers can use reactive system-level data access provisions to actively request access to such data themselves. The analysed legal frameworks regulating the public sector, and in particular the Access to EU Documents Regulation, the Open Data Directive and Chapter II of the Data Governance Act, seem most promising in this regard. Based on these instruments, public sector documents held by EU and national institutions can be requested by members of the general public and should, in principle, be provided upon such a request (subject to restrictions). These documents could potentially include information on *private sector digital infrastructures*, for instance when these infrastructures are used by the public sector bodies (e.g., productivity software). It is conceivable that system-level data which can only be obtained on request contain more detailed and specific information than data that are shared proactively and publicly by data holders, which may render them more valuable for researchers. At the same time, request procedures could also complicate matters for researchers since requests may be denied or subject to rather strict conditions (e.g., the data are only made accessible in a secure processing environment).[529]

In sum, system-level data may provide general information about the operations of digital infrastructures, the so-called helicopter view. Researchers may acquire access to system-level either using *proactive* transparency obligations both directly (i.e., from the data holder) or indirectly (i.e., from third parties such as regulatory authorities that receive data from data holders) or using *reactive* transparency obligations by requesting access to system-level data from the data holder.

[526] See European Commission 2023 (webpage) <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2186>.

[527] See Transparency Centre (webpage) <https://disinfocode.eu/>.

[528] See Transparency Centre reports (webpage) <https://disinfocode.eu/reports-archive/?years=2023>.

[529] Although the mere observation of certain parts of information being left out of a document can of course also be interesting to researchers (and something that would maybe not have become clear in case a data holder proactively publishes a (heavily edited or aggregated) report.

7. Conclusions and recommendations

As digital infrastructures are increasingly penetrating every sector and part of our society, access to data residing in and about these infrastructures is vital to observe and understand the world around us. This holds particularly true for academic research, which is driven by public interest goals, including a watchdog function and the pursuit of scientific knowledge. However, researchers experience growing difficulties in obtaining access to relevant data as it is progressively enclosed within digital infrastructures by the actors managing them. Faced with obstacles to independent observation and analysis of digital infrastructures, researchers may turn to the law to compel data access. Against this background, this report has mapped which legal provisions in the EU's digital policy agenda offer most potential in this regard and analysed their promises and constraints as tools to obtain data for scientific research. As discussed in the introduction, the working hypothesis of the study was that legal frameworks might provide more structured, uniform, and robust procedures to obtain access to third-party data and counteract strong incentives against transparency and openness, and thus, enhance data access. A summary of our findings is provided below.

7.1 Summary of findings

Normative grounding

Chapter 2 briefly discussed the normative underpinnings of claims to data access for research. We observed that researchers' access to public sector information for purposes of public scrutiny/public debate has a strong basis in European fundamental rights' law. However, researchers' access to (public- and private sector) data for their primary purpose of contributing to knowledge production has received less to no attention in fundamental rights case-law. In light of both **academia's traditional responsibility for public interest-driven knowledge production** and the **growing ubiquity of digital infrastructures** in modern society, researchers should **arguably be granted better access** to the data residing in, and generated by, these infrastructures.

Mapping and assessment of legal frameworks

Especially since the publication of its 'European Strategy for Data'[530] and Communication on 'Shaping Europe's Digital Future'[531] in 2020, the EU legislator has proposed and adopted a plethora of frameworks aimed at regulating aspects of the digital economy and targeting digital infrastructures. A considerable number of provisions within those frameworks establish transparency and data access requirements that have been identified in **Chapter 3** of this report. **Chapter 4** has shown that the provisions' effective utility for researchers differs widely. From the perspective of (academic) researchers, the relevance of the provisions can roughly be determined according to whether (a) they grant researchers direct access to relevant data; (b) the provisions can be deployed for data donation for scientific research purposes; and/or (c) the provisions contribute to an enabling environment for data access more generally.

- **Direct access to data**

  The only reviewed provision that specifically offers *researchers* an actionable claim to direct access to third-party data, is Article 40 of the Digital Services Act. On the basis of this provision, vetted researchers can request direct access to any data held by very large online platform providers and search engines (VLOPs/VLOSEs)to the extent relevant for conducting research on

[530] European Commission Communication 2020b.

[531] European Commission Communication 2020.

'systemic risks'. While Articles 11-13 of the proposed Political Advertising Regulation also specifically provide for researchers' access to data (on political advertisements), the provisions have been phrased in less strong terms, leaving leeway for the data holders not to provide the data.[532] Besides the provisions in which they are explicitly considered as data recipients, researchers may also gain direct access to third-party data based on legal provisions addressed to the **general public** (e.g., public reporting duties), or to **persons in particular capacities**, which roles researchers could potentially take on (e.g., data subjects, platform users, smart product owners). General public access provisions can be found in several frameworks, such as the General Data Protection Regulation, the Access to EU Documents Regulation, the Open Data Directive and the Digital Services Act. Provisions aimed at persons in specific capacities can be found in for example the proposed AI Act (e.g., use instructions for deployers of high-risk AI systems), the Digital Services Act (e.g., statements of reasons for platform users), the General Data Protection Regulation (e.g., information on the processing of their personal data for data subjects) and the proposed Data Act (e.g., IoT data for users of smart products).

- **Data donation**

  As mentioned above, some of the frameworks contain **access rights for specific (groups of) persons** (e.g., data subjects, platform users, owners of IoT devices). Researchers who seekcertain individual-level data may find these access rights useful for data donation purposes. In fact, the data subject's access right under Article 15 of the General Data Protection Regulation has already proven to be a valuable tool for researchers in practice.[533] The potential of individual access rights for research is further enhanced by legal provisions providing for **data portability and interoperability**, as recently enshrined in e.g., the Digital Markets Act and the proposed Data Act (in addition to existing provisions in the General Data Protection Regulation and Free Flow of Non-Personal Data Regulation). The proposed Data Act has another special feature which could facilitate data donation even further, namely the possibility for users of smart products to request IoT data holders to directly share IoT data with "third parties", such as researchers.[534] If researchers are indeed authorised by users of smart products to act as third party, researchers would no longer have to set up data donation architectures themselves and convince IoT product users to deliver their data to an intermediary. The provision in the proposed Data Act therefore has significant potential to smoothen the practice of data donation for scientific research.

- **Enabling environment**

  The third category of relevant provisions contributes more broadly to **an 'enabling environment'** for data access. Within this category, we made a rough distinction of frameworks that either introduce 'data intermediaries and facilitators', and frameworks that contain 'other enabling elements'. **Data intermediaries** are organisations functioning as third-party agents which connect data providers and data users (i.e., parties seeking data). Two prominent examples of such data intermediaries have been established in the Data Governance Act: data intermediation services providers[535] and data altruism organisations[536]. Examples of data sharing facilitators in specific data

[532] Providers or political advertising services shall make "best efforts" to provide the requested information and "take appropriate measures" to transmit the information, see Article 11(1) and 11(3) pPAR.

[533] See e.g., 'Digital Data Donation Infrastructure' (*D3I: A digital data donation infrastructure*) <https://d3i-infra.github.io/> accessed 17 April 2023.

[534] Article 5 pDA.

[535] Chapter III DGA.

[536] Chapter IV DGA.

sharing contexts are the 'single points of access' and 'information access points' as introduced in the Open Data Directive[537] and Chapter II of the Data Governance Act[538]. Examples of '**other enabling elements**' are mandatory low fees, specific time frames for requests to be processed, and arrangements for data portability and interoperability.

The legal analysis in this report has shown that the reviewed legal frameworks may contribute in different ways to data access for research purposes. Due to the diversity in scope and goals of the legal frameworks as well as the diversity of research areas, it is **impossible to rank** the legislative instruments according to their alleged relevance for academic research in general. That said, **one framework stood out in particular**: the **Digital Services Act**. The Digital Services Act is the only legal framework that contains a direct data access right specifically addressing researchers as data recipients (Article 40). Although its scope and rationale are limited to research related to (the mitigation of) systemic risks, the provision does seem to acknowledge the importance of data access for research.[539] Moreover, the Regulation also contains several public transparency provisions and opportunities for data donation. Given the growing ubiquity of (very large) online platforms, all of the DSA's data access provisions could be relevant for a wide array of research areas.

Recurring themes across access and transparency provisions

In **Chapter 5**, we observed a few recurring 'themes' across the analysed legal frameworks. In a nutshell, we noted that very few provisions in the analysed frameworks seem to have been designed with (academic) research in mind; most provisions must be viewed in light of generic internal market goals, accountability, and the protection of individuals. Although scientific research and the importance of data access for research purposes has been acknowledged here and there in recent (proposals for) legislation, researchers are still rarely mentioned as the addressees of direct data access rights. Additionally, a large number of potentially useful provisions mandate to balance transparency and data access with the protection of third-party rights and interests such as the protection of personal data, IP rights and trade secrets. While it is important to safeguard these rights and interests, the legal frameworks do not clarify how to balance them with transparency and data access *in practice*, which complicates the assessment of these provisions' potential. Lastly, many format and procedural requirements on how to access data are vaguely formulated in that they leave much room for interpretation. It is to be hoped that these unclarities will be resolved over time through soft law instruments, such as standardisation protocols, codes of conduct, regulatory guidance and delegated or implementing acts.

Strategies to use legal provisions for data access

In **Chapter 6** we finally outlined specific strategies for researchers on how to use legal provisions to acquire access to data generated and held by digital infrastructures. The two strategies focus on access to individual-level data (typically granular and detailed information) and access to system-level data (typically more generic information providing a 'helicopter view') respectively. As to the opportunities to access individual-level data, it was emphasised that researchers could: (1) take on the role the data recipient/rightholder addressed by the data access provisions – e.g., platform user, data subject, IoT product user, etc. – and in that capacity request and receive direct access to data, or (2) deploy data donation strategies to scale up individual-level data received by a (representative) group of addressees. To obtain system-level data, on the other hand, it was observed that researchers can: (1) access the data holder's information *directly* on the basis of public transparency provisions (e.g., public reporting duties), (2) access the data holder's information *indirectly* via

[537] Single point of access (Article 9(2) ODD).

[538] Single information point (Article 8(1)-(3) DGA and European single access point (Article 8(4) DGA).

[539] See also sections 4.1.1 and 5.1.

other data recipients, such as regulatory authorities or the European Centre for Algorithmic Transparency, who are legally entitled to the data holder's data.[540]

7.2 Recommendations

This report mainly aimed to map and analyse the potential of transparency and data access provisions in the EU's digital policy agenda. Based on this, we identified a number of key recommendations that may inform the development of a more robust strategy on how to operationalise these data access provisions in specific research fields. In particular, we pinpoint specific areas for action, both for universities as well as research funding organisations and law- and policymakers.

Recommendations for universities

Digital infrastructures have nested themselves deeply into our lives and environments. As such, they have become crucial objects of, and vehicles for, scientific research in general. With this in mind, there is an urgent need for universities to stress the importance of observability of digital infrastructures and take measures to improve the conditions for data access:

Invest in legal, methodological and technical capacity to make the best use of transparency and data access provisions enshrined in EU law

First and foremost, it is highly recommended that universities and university associations invest in a robust **support system** to tackle both legal, methodological and technical obstacles to effectively using transparency and data access provisions enshrined in existing and proposed EU digital/data legislation:

- **Legal and methodological capacity building**

  An important condition for realising the potential of legal transparency and data access provisions is that the research community – including individual researchers, research groups, departments and faculties – is **aware of the benefits and limitations of existing (and proposed) legal provisions**, **as well as how to operationalise them in their research fields**. Such awareness must be accompanied with the development of a robust knowledge base on the deployment of data access and transparency provisions within specific (sub)disciplines, enabling researchers, for example, to identify the right procedures and accommodate third-party interests where necessary (see below). Such legal knowledge can be imparted through a shared document pool, instruction videos and hands-on guidance documents/sessions to help people navigate through the forest of legal provisions. Additionally, it is also vital to develop methodological capacity – that is, strategies on how to integrate data access provisions into research methodologies – in a dynamic way. The potential and constraints of legal provisions for data access should be uncovered and discussed in different settings and faculties, accommodating for multifarious research questions, designs, methods, and disciplines (e.g., through workshops or discussion fora). Data stewards can play a role

[540] To indirectly access the data from regulatory authorities, researchers could invoke public sector frameworks to access government data that contain the data from the original data holder. Of course, public sector bodies that decide on an access request must take into account the legitimate interests of the data holder, including the protection of commercially confidential information (e.g., trade secrets) and IP rights.

in enabling such a platform for connecting people and expertise, as well as following relevant (legal) developments in the interest of scientific research and data access.[541]

- **Technical capacity building**

  Besides interdisciplinary knowledge, studying digital infrastructures may require significant technical resources and expertise, notably to enable safe and seamless data sharing. If data portability and interoperability requirements are to be of any value to the research community, it will be important to take proactive steps in formulating how these should be given shape in function of academic research. **It is therefore necessary that universities build, support, and maintain independent technical tools for data access, including technical systems for data donation, secure processing environments, APIs, and more**.[542] Secure processing environments, for example, are essential to enable (full) data access while respecting third-party interests such as privacy/data protection, intellectual property rights and commercial confidentiality, and arrive at a right balance between openness and closeness. It also important to invest in other techniques to preserve the protected nature of third-party data (e.g., the anonymisation, aggregation or randomisation of personal data) and to monitor their effectiveness, especially considering the growing complexity and ever-changing nature of data formats and structures. Importantly, when building technical capacity, universities should be mindful of the market power that commercial providers offering technical resources may have, and make sure they do not undermine their own "digital sovereignty" as discussed in the first report of this broader research project.[543]

For both capacity building areas, we want to emphasise the importance of cooperation and collaboration between universities, which brings us to the following recommendations below.

Share knowledge and best practices

Even when equipped with the resources to operationalise legal provisions compelling transparency and data access, individual researchers or research groups might still face challenges, especially when interacting with powerful data holders. It is therefore crucial that university departments, faculties, and academic institutions as a whole join forces and **share experiences** with the use of transparency and data access provisions in research as well as **best practices** on how to solve (recurring) issues. This could be done, for example, by creating and maintaining a knowledge repository (e.g., a wiki) and/or by organising events to share experiences.

Use existing coalitions to lobby for changes in laws and policies in favour of access to data for scientific research

As apparent from this report, scientific research is rarely taken into account when devising provisions on transparency and data access to digital infrastructures. For the interests of the research community to be more articulated in the design of future data access frameworks as well as in the implementation of existing

[541] See also Part A, section 3.3.

[542] In this regard, it is worth referring to national funding initiatives such as Platform Digitale Infrastructuur SSH <https://www.pdi-ssh.nl>.

[543] Referring back to the first report of this research project: IViR, 'Information Law and the Digital Transformation of the University: Digital Sovereignty, Data Governance and Access to Data for Research – Part I. Digital Sovereignty', 2023

legal frameworks, **it is important that universities and researchers adequately formulate and communicate their needs in this regard to law- and policymakers (lobbying).**

While over the years, universities and researchers in Europe have built several strong coalitions – both at national and European level – to make their voice heard by law- and policymakers, it is vital that these coalitions (continue to) make access to data for research an agenda item. Recent efforts by university associations have mainly focused on flagging the undesirable impacts of EU digital/data legislation and open science policy on the university sector. Issues on access to data about and residing in digital infrastructures as input for research have so far received less attention, despite the growing reliance on third-party data for research purposes.[544]

Lobbying can take place in several stages of the legislative process. The most obvious timing for lobbying is when legislation (or policy) is being drafted, or when legislation has been proposed but is still under negotiation.[545] Some of the frameworks discussed in this report find themselves in the latter stage,[546] in which it is still possible – to a certain extent – to influence any amendments to the legal text. Universities could lobby for such amendments by reaching out to law- and policymakers. After legislation has been adopted and entered into force, the operationalisation of the law in specific contexts will often still need to be developed and provisions may have to be clarified in, for instance, delegated or implementing acts or codes of conduct.[547] The development of such documents offers another occasion for the academic research community to make sure their interests are heard and incorporated into law, for instance by responding to consultations on delegated acts and by performing pilot studies and presenting the results as best practices to law- and policymakers. Lastly, after all these legal frameworks have materialised, universities can still lead by example on how certain data access provisions are interpreted and applied in practice (see the recommendation below).

Take the lead in giving shape to data access provisions

One of the ways for universities and researchers to ensure that the implementation of existing transparency and data access provisions works in their favour, is by taking the lead in giving shape to these provisions in practice rather than waiting for regulatory guidance.

For instance, where legislative instruments leave room for the development of codes of conduct, universities could **initiate such codes of conduct** themselves to facilitate easier access to third-party data.[548] And even in the absence of formal code of conducts, (associations of) universities can support or lead the development of **policies** that operationalise data access to the benefit of academic research interests. Such codes or policies could include, among others, **minimum standards** for how data should be shared with researchers for different research fields/faculties so that the data can be used most effectively (i.e., **format/formal requirements**). Additionally, they could contain **specific guidance on how to deal with third-party interests,** for example by requiring that the burden of proof is on the third party to show that its interests override the interests of scientific research.

Another area in which universities could take the lead, is the development of robust research practices and technical infrastructures for data sharing (as described above). One concrete instance of doing so could be via so-called data sharing intermediaries, an emergent category of actors in EU data legislation. If the legal mechanism to promote data altruism as established by the Data Governance Act proves to be

[544] See Part A of this report, section 2.1.

[545] See section 1.2.1 of this report for an explanation of the distinct stages of legislation.

[546] For example, the proposed AI Act, the proposed Data Act, the proposed Political Advertising Regulation and the proposed European Media Freedom Act.

[547] For a more in-depth discussion of these instruments, reference is made to section 5.5.2 of this report.

[548] Take for example, the Code of Conduct developed by the European Digital Media Observatory (EDMO) on researcher access to platform data under Article 40 of the General Data Protection Regulation (GDPR), see: EDMO 2022.

effective, researchers are likely to benefit from it since data altruism organisations (DAOs) can serve as vehicles to share data that are valuable for scientific research. However, it is unclear what these DAOs should look like exactly, i.e., in a way that they are actually useful for researchers. It is therefore advised that universities **develop a blueprint for DAOs dedicated to scientific research**, and not passively wait for DAOs to emerge, thereby risking that the data shared through these DAOs are not of much interest to researchers.[549]

Self-reflect on data access needs and impacts

In terms of responsibilities, it is also crucial that the research community critically reflects on the actual scope and impact of their data needs. Indeed, just like the technology sector has been criticised heavily for its role in surveillance/information capitalism, universities (faculties and departments) should also consider their own complicity in this. Specifically, universities and researchers should be mindful that the quest for obtaining access to third-party data does not legitimise third parties' problematic data practices and does not create or intensify dependencies. Moreover, researchers must uphold ethical and legal responsibilities governing data collection in the academic context, notably when it comes to sensitive data or exploratory research.

In short, while data access to digital infrastructures is important for academic research, it is equally important for the research community to remain critical as to their own role as well**. Universities could therefore create and reinforce tools to encourage systematic self-reflection on researcher data needs, the impact of those needs on others, as well as the broader (economic, societal, political) implications of data access**. Especially the latter is increasingly important considering the growing recognition of universities' responsibilities and exemplary role in society (e.g., with regard to fossil fuel funding). Such tools for critical self-reflection should offer space to go beyond mere procedural requirements (e.g., compliance checklists) and stimulate active discussions on these complex matters.

Recommendations for law- and policymakers

Indeed, there are limits as to what universities can do to improve the conditions for researchers' data access. Law- and policymakers are therefore encouraged to take account of the following considerations:

Recognise scientific research in digital law- and policymaking

Universities and researchers fulfil a number of missions in society which include acting as a public watchdog and participating in public debate and, most importantly, contributing to knowledge production and scientific progress. Yet, as explained throughout this report, these missions are increasingly endangered as (academic) researchers face growing obstacles in observing a world that is intermediated by digital infrastructures. While policymakers have been stepping up to constrain the power of digital infrastructures and have been imposing numerous transparency requirements, this is rarely done with (academic) research in mind.[550] When it comes to future legal frameworks, as well as the further implementation of existing frameworks, we believe it is important to consider decoupling transparency and data access provisions from sometimes narrow internal market objectives, or at least incorporating 'carve-outs' for the use of these provisions in support of public-interest driven goals, including academic research. **More generally, we encourage law- and policymakers to put scientific research and data access for research purposes**

[549] See also European Commission and Van Eechoud 2022, p. 31.

[550] Reference is made to e.g., sections 2.2, 4.1.1 and 5.1 of this report.

more prominently on their agendas. While doing this, it is advised to include research stakeholders, such as research organisations, universities, and individual researchers in (preliminary) discussions on how to shape new law and policy in ways that respect their interests.

Provide interpretational guidance

This report showed the ample vagueness of transparency and data access provisions, at least when one tries to apply them to a research context.[551] To the extent law- and policymakers wish to safeguard academia's core missions, they may **provide interpretational guidance to render these provisions more useful by (academic) researchers as well**. Such guidance may be sector-specific and should (exactly) define the data to be shared and clarify procedures, exceptions and limitations, technical aspects, quality assurance, timing, and so on. Guidance can take different forms, ranging from delegated acts to codes of conduct or other types of soft law instruments. It is essential that the frameworks and additional guidance are adequately monitored and enforced, for instance through clear liability rules, audits, sanctions, and penalties. The further development and concretisation of provisions should actively involve research communities to assess their needs when it comes to specifics in data access for research.

More legal certainty is also needed when it comes to how data can be used *after it is obtained* by researchers. Indeed, as apparent from the empirical study (Part A of this report), data providers often unduly obstruct access to data by severely limiting the possibilities of data use.[552] While legislative steps have already been taken to facilitate text- and datamining of copyright-protected works for research purposes, for other uses of copyrighted or otherwise protected data (e.g., trade secrets), it may still be unclear how much legal room researchers have. Lawmakers should therefore provide more clarity in this regard, which will also be crucial to the broader (legal-political) debate on how to reconcile *de facto* control over data (in the interest of e.g., the protection of personal data, IP and commercially confidential information) with competing societal interests such as scientific research.[553]

Invest in public technical infrastructures to facilitate data access

As mentioned in the recommendations for universities above, there is a need for robust and scalable technical tools that can be deployed by (academic) researchers to operationalise transparency and data access provisions. The European Open Science Cloud (EOSC) is an example of such a promising and European-wide technical infrastructure, which could become a useful platform for the sharing not only of research outputs but also of input data held by public and private sector entities. Law- and policymakers (and research funding organisations) should deploy their political and financial position to **stimulate the development of public digital infrastructures to facilitate the technical sharing of data for research purposes**.

***

[551] Reference is made to section 5.5 of this report.

[552] Reference is made to Part A, section 3.2 of this report, as well as Part B, section 1.1 of this report.

[553] Reference is made to section 5.3 of this report.

Bibliography

Acronyms & Introduction

Legislation

International

Soft law

UNESCO 2017
Revised Recommendation on Science and Scientific Researchers of the United Nations Educational, Scientific and Cultural Organization (UNESCO), adopted during the 39th session by the General Conference in Paris, 30 October-14 November 2017. See Records of the 39th session of the General Conference, Annex II. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000260889.page=116>.

European Union

Regulations and directives

Proposed Data Act (Council version)
Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), 17 March 2023, 7413/23.

Digital Markets Act
Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector (Digital Markets Act).

General Data Protection Regulation
Regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Literature

Ausloos and Veale 2020
J. Ausloos and M. Veale, 'Researching with Data Rights', *Technology and Regulation* 2020, pp. 136-157.

Bodó et al 2017
B. Bodó and others, 'Tackling the Algorithmic Control Crisis – the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents' (2017) 19 Yale J.L. & Tech. 133 <https://yjolt.org/sites/default/files/bodoetal_0.pdf>.

Bruns 2019
A. Bruns, 'After the "APIcalypse": Social Media Platforms and Their Fight against Critical Scholarly Research', *Information, Communication & Society* 22, No. 11 (September 2019): 1544–66.

Constantinides, Henfridsson and Parker 2018
P. Constantinides, O. Henfridsson, G.G. Parker, 'Introduction – Platforms and Infrastructures in the Digital Age', *Information Systems Research*, Vol. 29, No. 2, pp. 381-400.

Van Dijck, Nieborg and Poell 2019
J. van Dijck, D. Nieborg and T. Poell, 'Reframing platform power', *Internet Policy Review* 2019, Alexander von Humboldt Institute for Internet and Society, Vol. 8, No. 2.

Van Drunen and Noroozian 2023
M. van Drunen and A. Noroozian, 'How to Design Data Access for Researchers: A Legal and Software Development Perspective' (2023) <https://papers.ssrn.com/abstract=4330544>.

Ferrari 2023
V. Ferrari, *Money after Money: disassembling value/information infrastructures*, (diss. Amsterdam UvA) 2023.

Gerlitz et al 2019
C. Gerlitz, A. Helmond, D. Nieborg, and F. van der Vlist. 'Apps and Infrastructures – a Research Agenda'. *Computational Culture*, No. 7 (October 2019), <http://computationalculture.net/apps-and-infrastructures-a-research-agenda/>.

Henfridsson and Bygstad 2013
O. Henfridsson and B. Bygstad, 'The Generative Mechanisms of Digital Infrastructure Evolution, *MIS Quarterly* 2013, Vol. 37, No. 3, pp. 907-932.

Keller and Leerssen 2019
D. Keller and P. Leerssen, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation' in N Persily and J Tucker (eds), *Social Media and Democracy: The State of the Field and Prospects for Reform* (CUP 2019), <https://papers.ssrn.com/abstract=3504930>.

Leerssen 2023
P. Leerssen, *Seeing what others are seeing: Studies in the regulation of transparency for social media recommender systems* (diss. Amsterdam UvA) 2023.

Noto La Diega 2022
G. Noto La Diega, 'Ending Smart Data Enclosures: The European Approach to the Regulation of the Internet of Things between Access and Intellectual Property', *Cambridge Handbook on Emerging Issues at the Intersection of Commercial Law and Technology* (Cambridge University Press 2022), <https://papers.ssrn.com/abstract=4268125>.

Ohme et al 2023
J. Ohme, T. Araujo, L. Boeschoten, D. Freelon, N. Ram, B. Reeves, and T. Robinson. 'Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking'. *Communication Methods and Measures* 0, No. 0 (February 2023): 1–18.

Rieder and Hofmann 2020
Bernhard Rieder and Jeanette Hofmann, 'Towards platform observability', *Internet Policy Review* 2020, Vol. 9, No. 4.

Tromble 2021
R. Tromble, 'Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age', *Social Media + Society* 2021, Vol. 7, No. 1.

Studies and reports

Ausloos, Leerssen and Ten Thije 2020
J. Ausloos, P. Leerssen and T. ten Thije, 'Operationalizing Research Access in Platform Governance: What to Learn from Other Industries?', *Algorithm Watch* 2020. Available at: <https://dare.uva.nl/search?identifier=90e4fa77-d59a-49f1-8ccd-57d0725122bd>.

IViR 2023
IViR, 'Information Law and the Digital Transformation of the University: Digital Sovereignty, Data Governance and Access to Data for Research – Part I. Digital Sovereignty', 2023 <https://www.ivir.nl/part-i-digital-sovereignty/>.

Blogposts, webpages, videos, other

Bobrowsky 2021
Meghan Bobrowsky, 'Facebook Disables Access for NYU Research Into Political-Ad Targeting', *Wall Street Journal* 8 April 2021, <https://www.wsj.com/articles/facebook-cuts-off-access-for-nyu-research-into-political-ad-targeting-11628052204>.

Consumer Data Research Centre
Consumer Data Research Centre, 'About the CDRC' (webpage), <https://www.cdrc.ac.uk/about/>.

Kayser-Bril 2021
N. Kayer-Bril, 'AlgorithmWatch Forced to Shut down Instagram Monitoring Project after Threats from Facebook', *AlgorithmWatch*, <https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/>.

Ledford 2023
H. Ledford, 'Researchers scramble as Twitter plans to end free data access', *Nature* 14 February 2023, <https://www.nature.com/articles/d41586-023-00460-z>.

Schrage and Ginsberg 2018
E. Schrage and D. Ginsberg, 'Facebook Launches New Initiative to Help Scholars Assess Social Media's Impact on Elections', *Meta Newsroom* 9 April 2018, <https://newsroom.fb.com/news/2018/04/new-elections-initiative>.

The Politics of Systems
The Politics of Systems, LabsPolsys (webpage), <https://labs.polsys.net/>.

PART A

Legislation

International

Soft law

OECD 2021
Organisation for Economic Cooperation and Development (OECD), Recommendation of the Council on Enhancing Access to and Sharing of Data, nr. 0463, 6 October 2021.

UNESCO 2021
Recommendation on Open Science of the United Nations Educational, Scientific and Cultural Organization (UNESCO), adopted during the 41st session by the General Conference in Paris, 9 – 24 November 2021 <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.

European Union

Directives and regulations

General Data Protection Regulation
Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

Data Governance Act
Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724.

Literature

Luscombe, Dick and Walby 2022
A. Luscombe, K. Dick and K. Walby, 'Algorithmic thinking in the public interest: navigating technical, legal, and ethical hurdles to web scraping in the social sciences', *Quality & Quantity* 2022, No. 56, pp. 1023-1044.

Case law

HIQ Labs vs. LinkedIn Corporation
United States Court of Appeals for the Ninth Circuit 18 April 2022, No. 17-16783 (*HIQ Labs v.. LinkedIn Corporat*ion), <https://cdn.ca9.uscourts.gov/datastore/opinions/2022/04/18/17-16783.pdf>.

Blogposts, webpages, videos, other

European Commission Delegated Act Article 40 DSA webpage
European Commission, Have your say > Published initiatives > Delegated Regulation on data access provided for in the Digital Services Act (webpage), <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en>

European Commission Open Science webpage
European Commission, Strategy on research and innovation > Strategy 2020-2024 > Our digital future > Open Science (webpage), <https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en>.

LinkedIn User Agreement webpage
LinkedIn User Agreement, '8. Dos and Don'ts', <https://www.linkedin.com/legal/user-agreement>.

Mastercard Center for Inclusive Growth data exchange
Mastercard Center for Inclusive Growth, 'Data Philanthropy Offers New Avenues for Solving Old Problems', 2 August 2018, <https://www.mastercardcenter.org/insights/data-philanthropy-offers-new-avenues-solving-old-problems-report-finds>.

Pershan 2023
C. Pershan, 'The DSA must ensure public data for public interest research', *The Mozilla Foundation* 27 June 2023, <https://foundation.mozilla.org/en/blog/the-digital-services-act-must-ensure-public-data-for-public-interest-research/>.

Twitter Terms of Service webpage
Twitter Terms of Service, '4. Using the Services', <https://twitter.com/en/tos>.

UvA RDM (webpage)
UvA Research Data Management, 'Contact & support', <https://rdm.uva.nl/en/support/support.html>

UvA 2023 (webpage)
UvA, 'Virtual Research Environment UvA wide available', 7 March 2023 <https://www.uva.nl/en/content/news/news/2023/03/virtual-research-environment-uva-wide-available.html?origin=kUP%2Byx6UTZqvuJiCJKnnEQ&cb>.

PART B

Legislation

International

Conventions

ICESCR
United Nations (General Assembly), International Covenant on Economic, Social and Cultural Rights (ICESCR), Treaty Series, 16 December 1966, Treaty Series, Vol. 993.
Soft law

CESCR General comment no. 25
Committee on Economic, Social and Cultural Rights (CESCR), General comment No. 25 (2020) on science and economic, social and cultural rights (article 15(1)(b), (2), (3) and (4) of the International Covenant of Economic, Social and Cultural Rights

OECD 2007
Organisation for Economic Cooperation and Development, Principles and Guidelines for Access to Research Data from Public Funding (2007).

Council of Europe

Conventions

ECHR
Council of Europe, European Convention on Human Rights (ECHR), as amended by Protocols Nos. 11, 14 and 15 and supplemented by Protocols Nos. 1, 4, 6, 7, 12, 13 and 16.

Tromsø Convention
Council of Europe Convention on Access to Official Documents (CETS No. 205) (Tromsø Convention).

Soft law

Council of Europe 2009
Explanatory Report to the 2008 Council of Europe Convention on Access to Official Documents 2009, par. 1 (preamble), Strasbourg, CETS No. 205.

European Union

Primary EU law

TFEU
Treaty on the Functioning of the European Union (TFEU), consolidated version, *OJ* EU C 326/47.

CFREU
Charter of Fundamental Rights of the European Union (CFREU), *OJ* EU 2012, C 326/391.

Directives and regulations

E-Commerce Directive
Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internet Market (E-Commerce Directive).

Access to EU Documents Regulation
Regulation (EC) No. 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission Documents (Access to EU documents Regulation).

Services Directive
Directive 2006/123/EC of the European Parliament and of the Council of 12 December 2006 on services in the internal market (Services Directive).

PSI Directive
Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information (PSI Directive).

General Data Protection Regulation
Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Data Protection Law Enforcement Directive
Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing council framework decision 2008/977/JHA (Data Protection Law Enforcement Directive).

Trade Secrets Directive
Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Trade Secrets Directive).

Free Flow of Non-Personal Data Regulation
Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (Free Flow of Non-Personal Data Regulation).

Copyright in the Digital Single Market Directive

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (CDSM Directive)

Open Data Directive
Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (Open Data Directive).

Commission Implementing Regulation (high-value datasets)
Commission Implementing regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use (Commission Implementing Regulation)

Platform to Business Regulation
Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (Platform to Business Regulation).

Consumer Rights Directive
Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council Text with EEA Relevance (Consumer Rights Directive) amended by Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules (Modernisation Directive).

Data Governance Act
Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

Digital Markets Act
Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector (Digital Markets Act).

Digital Services Act
Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act).

Proposals

Proposed Political Advertising Regulation
Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (Political Advertising Regulation), COM(2021) 731 final.

Proposed Data Act
Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM/2022/68 final.

- Latest version: Council of the European Union, Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) – Mandate for negotiations with the European Parliament, 17 March 2023, 7413/23.

Proposed AI Act
Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial

Intelligence act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/01/06(COD) <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pd>.

Proposed European Media Freedom Act
Proposal for a European Media Freedom Act (Council of the European Union, Interinstitutional File: 2022/0277(COD), Brussels, 21 June 2023 (OR.en) 110954/23) <https://data.consilium.europa.eu/doc/document/ST-10954-2023-INIT/en/pdf>.

Proposal for a European Health Data Space Regulation
European Commission, Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, 3 May 2022, COM(2022) 197 final.

Soft law

European Commission, guidance NPDR
European Commission, Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union, COM(2019) 250 final

European Commission Communication 2020
Communication from the European Commission, 'Shaping Europe's digital future', 19 February 2020, COM(2020) 67 final.

European Commission Communication 2015
European Commission, Communication on 'a Digital Single Market Strategy for Europe', 6 May 2015, COM(2015) 192 final.

European Commission Staff Working Document 2015
European Commission, Commission Staff Working Document 'A Digital Single Market Strategy for Europe – Analysis and Evidence', 6 May 2015, SWD(2015) 100 final.

European Commission Communication 2017
Communication from the European Commission, 'Building a European Data Economy', 10 January 2017, COM(2017) 9 final.

European Commission Communication 2020b
Communication from the European Commission, 'A European strategy for data', 19 February 2020, COM(2020) 66 final.

European Commission guidance CDR
Commission notice Guidance on the interpretation and application of Directive 2011/83/EU of the European Parliament and of the Council on consumer rights, 2021/C 525/01 ("Commission Guidance").

European Commission Staff Working Document 2017
Commission Staff Working Document on the Free Flow of Data and Emerging Issues of the European Data Economy, Accompanying the Document Communication Building a European Data Economy, SWD(2017) 2 final, January 2017.

2022 CoP
The 2022 Strengthened Code of Practice on Disinformation.

EDPB 2023

European Data Protection Board, *Guidelines 01/2022 on data subjects right – Right of access. Version 2.0*, <https://edpb.europa.eu/system/files/2023-04/edpb_guidelines_202201_data_subject_rights_access_v2_en.pdf>

The Netherlands

Proposed

Kamerstukken II 2022/23, 36 382, nr. 2
Voorstel – Wijziging van de Wet hergebruik van overheidsinformatie en enkele andere wetten in verband met de implementatie van richtlijn nr. 2019/1024/EU van het Europees Parlement en de Raad van 20 juni 2019 inzake open data en het hergebruik van overheidsinformatie (Wet implementatie Open data richtlijn), *Kamerstukken II* 2022/23, 36 382, nr. 2 (Dutch proposal for implementation of the Open Data Directive): 'Wet implementatie Open Data Richtlijn'

Soft law

Letter of the Minister of Finances of 20 January 2023 (Kamerstukken II 2022-23, 31 477, nr. 85).
Letter of the Minister of Finances of 20 January 2023 (Kamerstukken II 2022-23, 31 477, nr. 85), < <https://zoek.officielebekendmakingen.nl/kst-31477-85.pdf>.

Caselaw

Court of Justice of the European Union

Final

Judgement of the Court, 22 November 2022, WM (C-37/20) and Sovim SA (C-601/20) v. Luxembourg Business Registers, Court of Justice of the European Union, ECLI:EU:C:2022:912.

Pending

Request for a preliminary ruling from the Verwaltungsgericht Wien (Austria) lodged on 16 March 2022 – CK (Case C-203/22).

European Court of Human Rights

ECtHR (Fifth Section) 3 March 2020, Appl. No. 75865/11 (*Centre for Democracy and the Rule of Law v. Ukraine*)
ECtHR (Grand Chamber) 8 November 2016, Appl. No. 18030/11 (*Magyar Helsinki Bizottság v. Hungary*), para. 156
ECtHR, 'Guide on Article 10 of the Convention on Human Rights – Freedom of expression', 31 August 2022
ECtHR 24 June 2014, Appl. No. 27329/06 (*Rosiianu v. Romania*)
ECtHR 18 November 2021 (Third Section), Appl. No. 6106/16 (*Saure v. Germany*).
ECtHR 14 April 2009 (Second Section), Appl. No. 37374/05 (*Társaság a Szabadságjogokért v. Hungary*)
ECtHR 25 June 2013 (Second Section), Appl. No. 48135/06 (*Youth Initiative for Human Rights v. Serbia*).
ECtHR 8 July 1999, Appl. Nos. 23536/94 and 24408/94 (*Baskaya and Okcuoglu v. Turkey*)
ECtHR 26 May 2009, Appl. No. 31475/05 (*Kenedi v. Hungary*);
ECtHR 3 April 2012, Appl. No. 41723/06 (*Gillberg v. Sweden*)
ECtHR 29 June 2004 (Second Section), Appl. No. 64915/01 (*Chauvy and Others v. France*)
ECtHR 22 October 2007 (Grand Chamber), Appl. Nos. 21279/02 and 36448/02 (*London, Otchakovsky-Laurens and July v. France*).

- See also: European Court of Human Rights, 'Guide on Article 10 of the European Convention on Human Rights – Freedom of expression' (updated on 31 August 2022).

The Netherlands

Hoge Raad 29 June 2007, ECLI:NL:HR:2007:AZ4664.

Literature

Ausloos and Veale 2020
J. Ausloos and M. Veale, 'Researching with Data Rights, *Technology and Regulation* 2020, pp. 136-157.

Araujo et al 2022
T. Araujo et al, 'OSD2F: An Open-Source Data Donation Framework', *Computational Communication Research* 2022, vol. 4(2), pp. 372-387.

Broomfield 2023
H. Broomfield, 'Where is open data in the Open Data Directive?', *Information Polity* 2023, vol. 28(2), pp. 175-188.

Devenney and Kenny 2012
J. Devenney and M. Kenny, *European Consumer protection: theory and practice*, Cambridge University Press 2012.

Van Drunen and Noroozian 2023
M. van Drunen and A. Noroozian, 'How to Design Data Access for Researchers: A Legal and Software Development Perspective' (2023) <https://papers.ssrn.com/abstract=4330544>.

Eder 2023
N. Eder, 'Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation' (26 June 2023) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4491365>.

Van Eechoud 2011
M. Van Eechoud, 'Friends or Foes? Creative Commons, Freedom of Information law and the European Union Framework for Reuse of Public Sector Information', in: L. Guibault and C. Angelopoulos, *Open Content Licensing: From Theory to Practice*, Amsterdam University Press 2011.

Fia 2022
T. Fia, 'Resisting IP Overexpansion: The Case of Trade Secret Protection of Non-Personal Data', *International review of intellectual property and competition law (IIC)* 2022, vol. 53, pp. 917-949.

Flyverbom 2019
Mikkel Flyverbom, *The Digital Prism: Transparency and Managed Visibility in a Datafied World*, Cambridge University Press 2019.

Geiger and Jütte 2022
C. Geiger and B.J. Jütte, 'Conceptualizing a 'Right to Research' and Its Implications for Copyright Law: An International and European Perspective' 2022, *Joint PIJIP/TLS Research Paper Series*.

Geiregat 2022
S. Geiregat, 'The Data Act: Start of a New Era for Data Ownership?', <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4214704>.

Gobbato 2020
S. Gobbato, 'Open Science and the reuse of publicly funded research data in the new Directive (EU) 2019/1024', *Journal of Ethics and Legal Technologies* 2020, vol. 2(2), pp. 146-161.

G'sell 2023
F. G'sell, 'The Digital Services Act: a General Assessment', in: A. von Ungern-Sternberg (ed.), *Content Regulation in the European Union – The Digital Services Act, Schriften des Irdt – Trier Studies on Digital Law* 2023, vol. 1, Verein für Recht und Digitalisierung e.V., Institute for Digital Law (IRDT) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4403433>.

Hamilton 2016
J. Hamilton, *Democracy's Detectives: The Economics of Investigative Journalism,* Harvard University Press 2016.

Han 2015
B. Han, *The Transparency Society*, Redwood City: Stanford University Press 2015.

Hase et al (forthcoming)
Hase, et al., Fulfilling their data access obligations: Platforms need to increase their compliance to enable data donation studies, *Internet Policy Review* (Forthcoming).

Leerssen et al 2019
P. Leerssen et al., 'Platform ad archives: promises and pitfalls', *Internet Policy Review* 2018, vol. 8, no. 4 (accessed 20 July 2023)

Leerssen 2023a
P. Leerssen, *Seeing what others are seeing: Studies in the regulation of transparency for social media recommender systems* (diss. Amsterdam UvA) 2023.

Leerssen et al 2023
P. Leerssen et al., 'News from the ad archive: how journalists use the Facebook Ad Library to hold online advertising accountable', *Information, Communication & Society* 2023, vol. 26, no. 7, pp. 1381-1300.

Leerssen, Heldt and Ketteman 2023
P. Leerssen, A. Heldt and C.M. Ketteman, 'Scraping by? Europe's law and policy on social media research access', in: C. Strippel, S. Paasch-Colberg, M. Emmer & J. Trebbe (eds.), *Challenges and perspectives of hate speech research* (pp. 405-425).

Leiser and Custers 2019
M. Leiser and B. Custers, 'The Law Enforcement Directive: Conceptual Challenges of EU Directive 2016/680', *European Data Protection Law Review* 2019, vol. 5(3), pp. 367-378.

Luscombe, Dick and Walby 2022
A. Luscombe, K. Dick and K. Walby, 'Algorithmic thinking in the public interest: navigating technical, legal, and ethical hurdles to web scraping in the social sciences', *Quality & Quantity* 2022, no. 56, pp. 1023-1044.

Mahieu and Ausloos 2020
R. Mahieu and J. Ausloos, 'Harnessing the Collective Potential of GDPR Access Rights: Towards an Ecology of Transparency', *Internet Policy Review* 2020.

Mustonen (ed) 2006
J. Mustonen (ed), *The World's First Freedom of Information Act*, Anders Chydenuius Foundation 2006.

Ohme and Araujo 2022

J. Ohme and T. Araujo, 'Digital Data Donations: A Quest for Best Practices', *Patterns* 2022, vol. 3, no. 4, 100467.

Pereira Campos 2021
J.F. Pereira Campos, *The Dynamics of Data Donation: Privacy Risks, Mobility Data, and the Smart City* (diss. St. Andrews, Scotland) 2021.

Quinn 2018
P. Quinn, 'Is the GDPR and Its Right to Data Portability a Major Enabler of Citizen Science?', *Global Jurist* 2018, vol. 18, no. 2.

Rieder and Hofmann 2020
Bernhard Rieder and Jeanette Hofmann, 'Towards platform observability', *Internet Policy Review* 2020, vol. 9, no. 4.

Tromble 2021
R. Tromble, 'Where Have All the Data Gone? A Critical Reflection on Academic Digital Research in the Post-API Age', *Social Media + Society* 2021, vol. 7, no. 1.

Vogiatzoglou et al 2021
P. Vogiatzoglou et al, 'From Theory To Practice: Exercising The Right Of Access Under The Law Enforcement And PNR Directives', *Jipitec* 2021, vol. 11, no. 3.

Wong and Henderson 2019
J. Wong and T. Henderson, 'The Right to Data Portability in Practice: Exploring the Implications of the Technologically Neutral GDPR', *International Data Privacy Law* 2019, vol. 9, no. 3, pp. 173-191.

Zemła-Pacud and Lenarczyk 2023
Ż. Zemla-Pacud, and G. Lenarczyk. 'Clinical Trial Data Transparency in the EU: Is the New Clinical Trials Regulation a Game-Changer?' *IIC* 2023, vol. 54, no. 5, pp. 732–763.

Reports

ALLEA, EASAC and FEAM 2021
ALLEA, EASAC and FEAM, *International Sharing of Personal Health Data for Research*, 2021<https://www.feam.eu/wp-content/uploads/International-Health-Data-Transfer_2021_web.pdf>.

Ausloos, Leerssen and Ten Thije 2020
J. Ausloos, P. Leerssen and T. ten Thije, *Operationalizing Research Access in Platform Governance: What to Learn from Other Industries?* (Algorithm Watch) 2020 <https://dare.uva.nl/search?identifier=90e4fa77-d59a-49f1-8ccd-57d0725122bd>.

Cantillon et al 2023
E. Cantillon et al, *Mobilizing private sector data for climate action* (Solvay Public Policy House) 2023.

Codagnone, Livia and Rodriguez De Las Heras Ballell 2022
C. Codagnone, G. Livia and T. Rodriguez De Las Heras Ballell, *Identification and Assessment of Existing and Draft EU Legislation in the Digital Field* (European Parliament) 2022 <https://www.europarl.europa.eu/RegData/etudes/STUD/2022/703345/IPOL_STU(2022)703345_EN.pdf>.

Edelson, Graef and Lancieri 2023
L. Edelson, I. Graef and F. Lancieri, *Access to Data and Algorithms: For an Effective and DMA and DSA Implementation* (Centre on Regulation in Europe) 2023 <https://cerre.eu/wp-content/uploads/2023/03/CERRE_Access-to-Data-and-Algorithms-DMA-DSA.pdf>.

European Commission 2022
European Commission (Directorate General for Research and Innovation*), European Research Area Policy Agenda: Overview of Actions for the Period 2022 – 2024* (Publications Office 2022) <https://op.europa.eu/en/publication-detail/-/publication/490ee6ca-aa58-11ec-83e1-01aa75ed71a1/language-en> (accessed 1 March 2023).

European Commission and Senftleben 2022
European Commission and M. Senftleben, *Study on EU copyright and related rights and access to and reuse of data* (European Commission) 2022 <https://www.ivir.nl/publicaties/download/KI0822205ENN.en_.pdf>

European Commission and Van Eechoud 2022
European Commission and M. Van Eechoud, *Study on the Open Data Directive, Data Governance Act and Data Act and their possible impact on research* (European Commission) 2022 <https://www.ivir.nl/publicaties/download/KI0822204ENN.en_.pdf>

European Commission and PwC EU Services 2023
European Commission and PwC EU Services, Directorate-General for Communications Networks, Content and Technology, *Identification of data themes for the extensions of public sector High-Value Datasets: final study*, (European Commission) 2023 <https://data.europa.eu/doi/10.2759/739414>

European Ombudsman 2021
European Ombudsman, *Results of the European Ombudsman's ad hoc survey of stakeholders' experiences requesting access EU documents* (European Ombudsman) 2021 <https://www.ombudsman.europa.eu/en/document/en/149496 >.

European Parliament, DG EPRS 2019
Health Ethics and Policy Lab, ETH Zurich, E. Vayena and others, *How the General Data Protection Regulation changes the rules for scientific research,* (European Parliament) 2019 <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf>.

European Parliament, Vogiatzoglou and Marquenie 2022
European Parliament, P. Vogiatzoglou and T. Marquenie, *Assessment of the implementation of the Law Enforcement Directive* (European Parliament) 2022 <https://op.europa.eu/en/publication-detail/-/publication/215de10f-75ee-11ed-9887-01aa75ed71a1>.

EDMO 2022
European Digital Media Observatory, *Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access* (Institute for Data, Democracy & Politics, The George Washington University) 2022 <https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf?ref=static.internetfreedom.in>

Gineikytė, Barcevičius and Cibaitė 2020
V. Gineikytė, E. Barcevičius and G. Cibaitė, *Analytical paper 5: Business user and third-party access to online platform data* (Observatory on the Online Platform Economy) 2020 <https://platformobservatory.eu/app/uploads/2020/09/Analytical-Paper-5-Business-user-and-third-party-access-to-data_final.pdf>

IViR 2023
IViR, 'Information Law and the Digital Transformation of the University: Digital Sovereignty, Data Governance and Access to Data for Research – Part I. Digital Sovereignty', 2023 <https://www.ivir.nl/part-i-digital-sovereignty/>.

Schnurr 2022

D. Schnurr, *Switching and Interoperability between Data Processing Services in the Proposed Data Act* (Centre on Regulation in Europe) 2022 <https://cerre.eu/wp-content/uploads/2022/12/Data_Act_Cloud_Switching.pdf>

Blogposts, webpages, videos, other

Albert 2022
J. Albert, 'A Guide to the EU's New Rules for Researcher Access to Platform Data', *AlgorithmWatch* 7 December 2022) <https://algorithmwatch.org/en/dsa-data-access-explained/> (accessed 12 December 2022).

AMdEX (webpage)
AMdEX > About AMdEX <https://amdex.eu/about/>

Ausloos, Toh and Giannopoulou 2022
J. Ausloos, J. Toh and A. Giannopoulou, 'How the GDPR Can Exacerbate Power Asymmetries and Collective Data Harms. Exploring How Power Asymmetries Operate across the Law and Collective Harms' *Ada Lovelace Institute,* 29 November 2022 <https://www.adalovelaceinstitute.org/blog/gdpr-power-asymmetries-collective-data-harms/> (accessed 1 March 2023).

Clark 2021
M. Clark, 'Research Cannot Be the Justification for Compromising People's Privacy', *Meta Newsroom* 3 August 2021, <https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/>.

European Commission 2023 (webpage)
European Commission, 'DSA Enforcement: Commission launches European Centre for Algorithmic Transparency', 17 April 2023 <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_2186>.

European Commission Data Governance Act explained (webpage)
European Commission, Shaping Europe's digital future > Policies > Data Governance Act explained <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>

European Commission Delegated Act Article 40 DSA webpage
European Commission, Have your say > Published initiatives > Delegated Regulation on data access provided for in the Digital Services Act (webpage) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en>

European Commission DMA press release 2022 (webpage)
European Commission > Press corner > Digital Markets Act 'Digital Markets Act: rules for digital gatekeepers to ensure open markets enter into force' 31 October 2022 <https://ec.europa.eu/commission/presscorner/detail/en/ip_22_6423>

European Commission EOSC (webpage)
European Commission > Strategy on research and innovation > Strategy 2020-2024 > Our digital future > Open Science > European Open Science Cloud <https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en>.

European Commission Open Science (webpage)
European Commission, Strategy on research and innovation > Strategy 2020-2024 > Our digital future > Open Science (webpage), <https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en>.

European Council 2023 (webpage)

European Council, 'Data act: member states agree common position on fair access to and use of data', 24 March 2023 <https://www.consilium.europa.eu/en/press/press-releases/2023/03/24/data-act-member-states-agree-common-position-on-fair-access-to-and-use-of-data/>.

European Commission Platform-to-business trading practices (webpage)
European Commission, Shaping Europe's digital future > Policies > Platform-to-business trading practices <https://digital-strategy.ec.europa.eu/en/policies/platform-business-trading-practices>

European Open Science Cloud (webpage)
European Open Science Cloud > Providers Statistics <https://providers.eosc-portal.eu/stats/providers>

Legislative Train Schedule 2023 (webpage)
Legislative Train Schedule, Revision of the access to documents Regulation 2023 <https://www.europarl.europa.eu/legislative-train/theme-union-of-democratic-change/file-revision-of-the-access-to-documents-regulation>

Leerssen 2021
P. Leerssen, 'Platform research access in Article 31 of the Digital Services Act: Sword without a shield?' *Verfassungsblog*, 7 September 2021 <https://verfassungsblog.de/power-dsa-dma-14/> (accessed 9 February 2023).

Leerssen 2023b
P. Leerssen, 'Counting the days: what to expect from risk assessments and audits under the DSA – and when?', *DSA Observatory,* 30 January 2023 <https://dsa-observatory.eu/2023/01/30/counting-the-days-what-to-expect-from-risk-assessments-and-audits-under-the-dsa-and-when/>.

NVJ (webpage)
Nederlandse Vereniging van Journalisten > Thema's > Journalistieke praktijk > Ethische Regels > Code voor de journalistiek <https://www.nvj.nl/themas/journalistieke-praktijk/ethische-regels/code-journalistiek>

ODISSEI (webpage)
Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI) <https://odissei-data.nl/en/>

O'Reilly 2021
E. O'Reilly, '20 years of public access to EU documents: time for a makeover?', *EUobserver,* 15 November 2021 <https://euobserver.com/opinion/153414>.

Rodriguez Lafuente 2022
O. Rodriguez Lafuente, 'It's time to update the EU's Regulation on Access to EU documents', *Shaping Europe,* 17 March 2022 <https://shapingeurope.eu/en/its-time-to-update-the-eus-regulation-on-access-to-eu-documents/>

SWIPO
Switching Cloud Providers and Porting Data (SWIPO)' Codes of Conduct <https://swipo.eu/>

Transparency Centre (webpage)
Transparency Centre (webpage) <https://disinfocode.eu/>.

Transparency Centre reports (webpage)
Transparency Centre > Reports <https://disinfocode.eu/reports-archive/?years=2023>

UNESCO (webpage)
UNESCO, Access to Information Laws (webpage) <https://www.unesco.org/en/access-information-laws>

UvA 2023 (webpage)
UvA, 'Virtual Research Environment UvA wide available', 7 March 2023 <https://www.uva.nl/en/content/news/news/2023/03/virtual-research-environment-uva-wide-available.html?origin=kUP%2Byx6UTZqvuJiCJKnnEQ&cb>

Vėlyvytė 2022
V. Vėlyvytė, 'Does the Court of Justice of the European Union Respect the Limits of EU Competence?', *EU Law Analysis*, 18 December 2022) < https://eulawanalysis.blogspot.com/2022/12/does-court-of-justice-of-european-union.html> accessed 17 April 2023.

Zenner 2022
Kai Zenner, 'Digital Factsheets #3: The Upcoming Digital Files and Their Place in the Existing DSM Framework', *Digitizing Europe*, 24 October 2022 <https://www.kaizenner.eu/post/digital-factsheets3>.