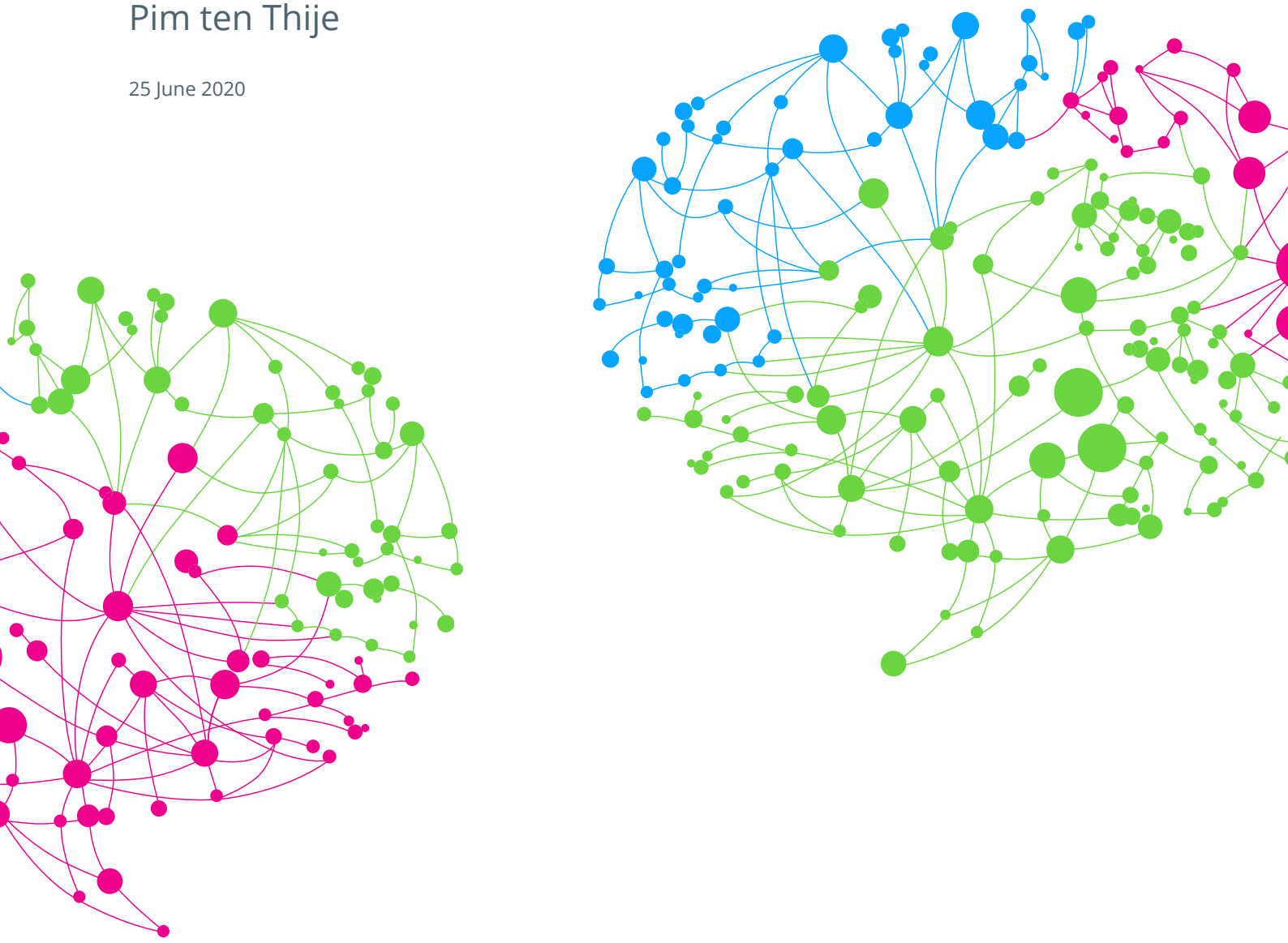
GOVERNING PLATFORMS

Operationalizing Research Access in Platform Governance

What to learn from other industries?

Jef Ausloos
Paddy Leerssen
Pim ten Thije

25 June 2020



Published as part of Governing Platforms, a research project by

in partnership with

with support by

About the authors

This report has been prepared by Dr. Jef Ausloos, Paddy Leerssen and Pim ten Thije.

It was written under the academic guidance of Prof. Dr. Natali Helberger, Prof Dr. Claes de Vreese and Dr. Kristina Irion.

The authors are part of the Institute for Information Law (IViR) and the Information, Communication & the Data Society (ICDS) Initiative at the University of Amsterdam (https://www.ivir.nl; https://www.uva-icds.net/).

Preface

This report has been commissioned by AlgorithmWatch.

The findings and views expressed in this report are solely those of the authors and should not be attributed to any of the other aforementioned entities.

Comments and suggestions are welcome at <operational-transparency@ivir.nl>.

Executive Summary

This Report identifies best practices for research access regimes in the platform governance context, by learning from existing legal frameworks in other domains. **Meaningful research access is a precondition for informed and effective platform governance**. Platforms play a central and ever-expanding role in modern society. Due to their influence, scale, and complexity, a wide range of expert research is necessary to understand their impact on society, and hold them accountable where necessary. Yet, the data access needed to perform this research is sorely lacking. Ensuring adequate research access should therefore be a paramount priority in upcoming transparency reforms.

This Report contributes to the existing debates on data access by taking a step back from platform governance per se, and learning from other (regulatory) transparency frameworks in existence already. Specifically, **the Report examines how key challenges have been tackled in other sectors, and formulates a number of best practices for a clear and effective research access framework in platform governance.**

The best practices for research access in this Report are drawn from two case studies of data access frameworks in two different sectors: environmental protection, and medical research. In environmental law, we consider the European Pollutant Release and Transfer Regime. This case study is instructive for platforms for what we term the **incentive problem:** the regulated party has strong incentives to oppose, avoid and obstruct transparency demands. In medical research, we consider the Findata program, a recent Finnish legal initiative for enabling research access to health data in a data protection-compliant manner. This case study is instructive for what we term **data protection concerns**: the reticence of regulated companies to share information that might be traced back to identifiable data subjects.

The case studies reveal the following best practices for data access regulation:

Overcoming the incentive problem

- **Specific disclosure rules**
  Clearly delineate what data should be included
- **Standardized methods for data generation**
  Standardized methodologies for generating reported data
- **Liability for data quality**
  Adequate safeguards ensuring the quality (completeness, consistency and credibility) of the reported data
- **Size-based regulation**
  A well-considered threshold below which no data should be reported
- **Transparency by default**
  All pre-defined data should be transparent by default and can only be kept confidential exceptionally, subject to strict requirements
- **Public transparency by default**
  The default should be that all pre-defined data is publicly accessible, and privileged access is the exception

- **Tiered oversight structure**
  Different layers of oversight ensuring accountability at multiple levels
- **Sanctions / penalties**
  Genuine threat of penalties in case of non-compliance
- **Mandatory proactive support**
  Legal requirement for public bodies to raise awareness and encourage civil society to engage with the data access framework
- **Strict timing**
  Clearly defined time-frames within which data has to be reported

Overcoming Data Protection Concerns

- **Request-based, adaptive regime**
  A comprehensive, request-based access regime, where data access is collected and provided in silos, on a case-by-case basis
- **Mandatory data sharing**
  Data sources are legally required to comply with access requests
- **Iterative regulation**
  Including a pilot-phase to test key aspects of the data access regime in practice, before drafting the actual legal framework
- **Pre-processing**
  Intermediary institution collects, combines and pseudonymizes data (in a transparent manner), before making it accessible to researchers/applicants
- **Different forms of data access**
  Depending on the purposes and risks involved, access can take different forms, from aggregated statistics to granular datasets
- **GDPR Compliance**
  An ambitious and comprehensive access regime involving potentially very sensitive personal data *can* be GDPR compliant

Overall, these case studies also suggest a number of more general, cross-cutting best practices, regarding the overall governance structure for research access:

Cross-Cutting Best Practices Overcoming the incentive problem & data protection concerns

- **Binding rules**
  A meaningful research access regime in platform governance requires a robust legal framework
- **Independent institutions**
  Given the complexity and many challenges faced, it is advisable to have an independent institution intermediate the research access regime
- **Tiered regulation**
  Even if an independent institution is called into life, multiple levels of oversight are required in order to ensure accountability of, and trust in the access regime overall
- **Proactive support for researchers**
  In order for it to be meaningful, a platform research access regime should also ensure that researchers have the resources and tools necessary to actually conduct research
- **(Public) transparency by default**
  Barriers to gaining access to data should be minimized
- **Verification and pre-processing**
  The intermediary role of independent institutions not only serve as assurance that data is suitable for disclosure, they can also maintain relevant access infrastructures, such as public databases, virtual machines, and discussion fora

Table of Contents

1 Introduction

Transparency requirements are a common denominator in scholarly and policy debates on platform governance. Generally seen as a precondition for corporate accountability, transparency is a recommendation that almost all policy proposals, legal initiatives and stakeholder proposals share. Yet when it comes to *how* exactly such transparency should be given shape, there is much less consensus or clarity. **This Report aims to provide some guidance on how to operationalize transparency measures for platform governance, by taking inspiration from established transparency-frameworks.** In particular, it focuses on enabling data access for public interest research, and issues regarding data protection compliance and companies' strong incentives against sharing data.

1.1 Problem statement

Society is increasingly datafied and intermediated through digital infrastructures. This is true for the way we interact with our environment (e.g. smart city/home), how we move around (mobility), work, date, exercise, learn, entertain ourselves, and much more. Crucially, the pivotal players in these emerging and expanding ecosystems are private companies, establishing themselves at central nodes as ineluctable 'platforms'. **Platforms serve as an increasingly central infrastructure for modern society**, placing them in a position to collect vast amounts of data about individuals, and to shape how they interact with each other and their environment.

As discussed in two earlier Reports commissioned by AlgorithmWatch,[1] access to data has become a precondition for evidence-based law-making and accountability mechanisms for platform governance.[2] This follows from the inherently digital, 'data-based' infrastructure platform companies have constructed. Indeed, Stark et al. clarify that content moderation issues (among others) 'should be tackled with reason and based on empirical evidence' and 'scientific evidence is needed on both thematic complexes in order to investigate the extent of the phenomena and their consequences in more detail, so that evidence-based measures can be developed.'[3] Cornils et al. explain that 'decisions to regulate communications (both on and offline) should be grounded in empirical experience' and '[t]ransparency obligations are therefore the entry level of any imperative regulation.'[4] This evidence base can not only serve to assist regulators, but also other actors in the platform governance ecosystem such as users, journalists, academics and NGOs.

1 Birgit Stark and others, 'Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse' (AlgorithmWatch 2020) <https://algorithmwatch.org/en/governingplatforms/communications-study-stark-may-2020>; Matthias Cornils, 'Designing Platform Governance: A Normative Perspective on Regulatory Needs, Strategies, and Tools to Enhance the Information Function of Intermediaries' (AlgorithmWatch 2020) <https://algorithmwatch.org/en/governingplatforms/legal-study-cornils-may-2020>.

2 See also (the many references in): Robert Gorwa and Timothy Garton Ash, 'Democratic Transparency in the Platform Society' in Nate Persily and Josh Tucker (eds), *Social Media and Democracy: The State of the Field* (Cambridge University Press 2020) 2.

3 Stark and others (n 1).

4 Cornils (n 1).

A system 'needs to be understood to be governed,'[5] and transparency – data access in particular – is often the first step required towards such understanding.[6] Yet, **whereas the amount of data being generated is growing exponentially, it has become increasingly difficult for civil society to access that data**. This can be explained partly by the growing complexity and privatization of data ecosystems.[7] Specifically, the ways in which data is captured and subsequently used are constantly developing and require vast resources. Moreover, global interconnectivity and the rise of so-called surveillance/informational capitalism[8] also precipitate constantly reinforcing data ecosystems, with a tendency of concentrating data in the hands of central nodes.[9] Both these central nodes, as well as the surrounding ecosystems more broadly are predominantly in private hands, rendering the vast majority of data being captured and generated proprietary – kept exclusive through combinations of legal and technical restrictions.

This *concentration* and *privatization* of data has a deep impact on independent investigations and research by civil society actors of all stripes, including academics, journalists, NGOs and policymakers. Indeed, large technology companies only rarely release data under their control for independent outside inquiry. Combined with their unrivalled capacity to capture and process vast amounts of data, these **companies become the *de facto* gatekeepers of research and reporting agendas.** In this role as gatekeeper, strategic secrecy can prevent robust accountability mechanisms from being established, for instance when it comes to scrutinizing companies' algorithmic practices. While there might be various (legal, economic, technical) reasons for refusing access to data under their control, it is clear that platforms' interests in maintaining exclusivity may not always align with public interests in transparency – and that their arguments for maintaining secrecy may not always be valid or in good faith.[10]

European Commission White Paper on Artificial Intelligence[11]

'The specific characteristics of many AI technologies, including opacity ('black box-effect'), complexity, unpredictability and partially autonomous behavior, may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights. Enforcement authorities and affected persons might lack the means to verify how a given decision made with the involvement of AI was taken and, therefore, whether the relevant rules were respected. Individuals and legal entities may face difficulties with effective access to justice in situations where such decisions may negatively affect them.'

5 Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973, 982–83.

6 Gorwa and Ash (n 2) 20.

7 Jef Ausloos, 'GDPR Transparency as a Research Method' (Institute for Information Law (IViR), University of Amsterdam 2019) Draft Paper <https://papers.ssrn.com/abstract=3465680> accessed 17 October 2019. See also more generally: Archon Fung, 'Infotopia: Unleashing the Democratic Power of Transparency*' (2013) 41 Politics & Society 183, 187–88.

8 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019); Julie E Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019).

9 Seda Gürses and Joris van Hoboken, 'Privacy after the Agile Turn' in Jules Polonetsky and others (eds), *Cambridge Handbook of Consumer Privacy* (Cambridge University Press 2018).

10 See in this regard also: European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (2020) https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf.

11 European Commission, 'White Paper on Artificial Intelligence – A European approach to excellence and trust 2020' (White Paper) COM (2020) 65 final.

Council of Europe on the Manipulative Capabilities of Algorithmic Processes[12]

Put briefly, growing information asymmetries challenge scrutinizing platform's algorithmic practices; investigating emerging data ecosystems; their impact on society and individuals; and 'datafied' social phenomena more broadly. Fine grained, sub-conscious and personalized levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions. These effects remain underexplored but cannot be underestimated. Not only may they weaken the exercise and enjoyment of individual human rights, but they may lead to the corrosion of the very foundation of the Council of Europe. In light of this, the Committee of Ministers also stresses the societal role of academia in producing independent, evidence-based and interdisciplinary research and advice for decision-makers regarding the capacity of algorithmic tools to enhance or interfere with the cognitive sovereignty of individuals.

As will be explained further in the following chapter, civil society has tried to overcome this 'transparency crisis' through a variety of methods. Perhaps the most visible strategies are the self-regulatory initiatives championed by platforms themselves (e.g. data grants;[13] ad archives;[14] and Social Science One[15]). These initiatives have been widely criticized for being incomplete, ineffective, unmethodical, and unreliable. In addition, civil society actors have objected to associating with private entities as a precondition for doing research, based on real or perceived threats to research independence that may result from, for example, an obligatory sign-off procedure on produced findings.[16] For these reasons, **there is a clear need for a more robust data access framework that is legally enforceable**.[17]

1.2 Objective of the report

While rhetorically useful as a high-level concept, **transparency needs to be further specified in order to gain practical meaning**. Zooming into the growing calls for accountability in platform governance, an important subset of 'transparency' relates to 'data access';[18] more specifically, **granular access to the data feeding into (and coming out of) the computational infrastructures governing content/ information flows**. This comprises, for example,

12 Council of Europe Council of Ministers, Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies).

13 Axel Bruns, 'After the "APIcalypse": Social Media Platforms and Their Fight against Critical Scholarly Research' (2019) 22 Information, Communication & Society 1544, 1551.

14 Paddy Leerssen and others, 'Platform Ad Archives: Promises and Pitfalls' (2019) 8 Internet Policy Review <https://policyreview.info/articles/analysis/platform-ad-archives-promises-and-pitfalls> accessed 7 February 2020.

15 Gary King and Nathaniel Persily, 'Unprecedented Facebook URLs Dataset Now Available for Academic Research through Social Science One' (Social Science One, 13 February 2020) <https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one> accessed 4 March 2020.

16 Bruns (n 13) 1553; Jef Ausloos and Michael Veale, 'Researching Through Data Rights' (2020) (forthcoming); Stark and others (n 1). See generally the open letter regarding corporate support of research into technology and justice by Lina Dencik and others, 'Funding Matters – a Statement about the Corporate Funding of Academic Conferences' (Funding Matters, 2018) <https://fundingmatters.tech/> accessed 9 June 2020.

17 See more generally: Fung (n 7) 190.

18 Other manifestations of transparency, for example, could relate to information on internal procedures or governance structures.

access to takedown decisions,[19] advertisements,[20] law enforcement requests,[21] and so on. The Stark and Cornils Reports[22] – alongside many others[23] – already stressed the normative and societal urgency of robust data access regimes in the platform governance debate. These arguments will be reviewed in Chapter 2, but this Report aims to take a step further and explore how calls for data access can be operationalized. More specifically, its objective is to draw lessons from existing data access frameworks in other sectors of industry, so as to guide policy-makers and stakeholders more broadly, in developing a robust platform transparency framework.

The added value of this approach lies in the fact that platform governance debates often remain inward-looking, despite the fact that data access regulation has many precedents in other areas of regulation. To prevent us from reinventing the wheel, this report takes a step back from platform governance debates to explore the wealth of (regulatory) transparency frameworks in existence already. Few studies appear to have taken such a comparative approach until now, or at least done the heavy lifting in systematically comparing other data access regimes with the transparency requirements in platform governance. That is what this Report aims to do: to **identify the key challenges raised in relation to platform data access; examine how similar issues have been tackled in other sectors (successfully or unsuccessfully); and draw lessons for a clear and effective data access framework for platforms**. Importantly, the Report particularly focuses on the requirements and limitations of frameworks enabling data access to public interest researchers.

1.3 Approach and structure of the report

This Report can be situated within the broader 'platform governance' debate, with a particular focus on dominant social media platforms (e.g. Facebook, Twitter and YouTube).[24] Within that focus, it looks in particular at **best practices for operationalizing data access for public interest researchers** (e.g. academia and investigative journalists), summarized as 'research access'. That said, many of the findings in this Report (notably in Chapter 5) can easily be extrapolated to other 'platforms' and/or be valuable for operationalizing data access not just to public interest researchers but also to other stakeholders, such as government regulators or commercial entities.

This Report is composed of 5 chapters, progressively building up to recommendations for a robust research access framework in the context of platform governance. The first substantial chapter (Chapter 2) describes the current 'research access crisis' giving rise to calls for data access reform, and sets the stage for the rest of the Report. Two key obstacles are identified, which inform the selection of case studies in Chapters 3 and 4: (a) establishing a data access framework aimed at promoting platform accountability in the face of conflicting incentives; and (b) how to do so in a data protection compliant manner. The

19 Berkman Klein Center for Internet & Society at Harvard University, 'Lumen' (Lumen Database) <https://lumendatabase.org/> accessed 9 June 2020.

20 Ad Archives, cf. Leerssen and others. 'Platform Ad Archives' (n 14).

21 See e.g. Google Inc., 'Requests for User Information – Google Transparency Report' (Global requests for user information) <https://transparencyreport.google.com/user-data/overview> accessed 9 June 2020.

22 Stark and others (n 1); Cornils (n 1).

23 See (the many references in): Daphne Keller and Paddy 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* (Cambridge University Press 2019); Gorwa and Ash (n 2).

24 Cf. the definitions and scoping in Cornils (n 1) ch A.III; Stark and others (n 1) ch 2.1; See more generally: Martin Moore and Damian Tambini (eds), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (Oxford University Press 2018).

case study in Chapter 3 examines environmental law – in particular, the European Pollutant Release and Transfer Register[25] – a context with strong disincentives to disclose information. Chapter 4 examines medical research – in particular the recently established Finnish medical data sharing platform Findata – a sector with strong concerns around the disclosure of (sensitive) personal data. Chapter 5, finally, builds on the insights gained in the case study chapters and connects it back to the platform governance debate in particular. It describes six concrete lessons that can be drawn from the respective data access regimes and considers how they might apply with regard to online platforms. The Chapter ends by listing a number of elements the case studies do not give an answer to, but which require careful attention when designing a research access regime in the platform governance context.

1.4 Limitations and scope

This Report is limited in scope. It concentrates specifically on research access regimes for accountability in platform governance, with a focus on social media platforms and how they mediate content. The Report aims to contribute to the development of such regimes by offering insights from two existing data access frameworks. With this in mind, the analyses are legal and policy-oriented in nature (rather than technical or economic for instance), specifically focusing on the two central issues as represented by the two case studies: 'the incentive problem' and 'data protection concerns'. Questions regarding regulatory design and legislative competences are answered in the Cornils Report.[26]

Furthermore, the data access regimes discussed in this Report are primarily (though not exclusively) aimed at a more expert audience. In light of their intended goals (generating accountability) and subject matter (complex algorithmic infrastructures), the intended target audience will primarily be 'experts' with the resources and incentives to understand and act upon such data. For the purposes of this Report, this 'expert audience' can be understood as policymakers, oversight bodies and civil society interpreted widely, including academia, NGOs and journalists. These can be considered the actors with an interest and ability to independently scrutinize and understand platform practices, something which necessarily precedes evidence-based and accountable lawmaking.[27]

It should also be acknowledged that despite their intuitive appeal, transparency and data/research access are by no means silver bullets for producing platform accountability.[28] As Gorwa and Garton Ash explain, the 'major reason for the widespread popularity of transparency as a form of accountability in democratic governance is its flexibility and ambiguity. [...] Transparency in practice is deeply political, contested, and oftentimes problematic; and yet, it remains an important — albeit imperfect — tool which, in certain policy domains, has the potential to remedy unjust outcomes, increase the public accountability of powerful actors, and improve governance more generally.'[29] As is widely remarked, transparency need not be seen as an alternative to other forms of regulation, but rather as an important complement or precondition.

25 In particular: Regulation (EC) No 166/2006 of the European Parliament and of the Council of 18 January 2006 concerning the establishment of a European Pollutant Release and Transfer Register and amending Council Directives 91/689/EEC and 96/61/EC (Text with EEA relevance) [2006] OJ L33/1 (E-PRTR Regulation).

26 Cornils (n 1).

27 Stark and others (n 1); Cornils (n 1); Keller and Leerssen (n 23).

28 Ananny and Crawford (n 5).

29 Gorwa and Ash (n 2) 3.

2 The research access crisis in platform governance

2.1 What is at stake: why research access matters for platform governance

Public interest research from academics, journalists and other civil society actors plays a vital role in platform governance. This section outlines some of the key functions of public interest research:

- Diagnosing new harms and threats in online ecosystems, and detecting wrongdoing by platforms and/or users.
- Assisting governments in developing and enforcing evidence-based policies and standards.
- Helping to raise awareness and mobilize other forms of social and political accountability from users, commercial actors, opinion-makers or politicians.
- Holding governments accountable for their actions online, and monitoring the protection of fundamental rights.
  Leveraging highly valuable platform data for investigating (e.g. social, economic or political) phenomena more broadly.

The backdrop for this Report is an online ecosystem that is increasingly reliant on a handful of dominant platforms.[30] These range from service platforms such as Airbnb and Uber to e-commerce platforms such as Amazon and, at the center of this present Report, social media platforms such as YouTube, Facebook, Twitter and Instagram. Thanks to their dominant positions, these platforms wield significant influence over increasingly large aspects of our society and economy. As their influence grows, so do calls for governments and policymakers to hold these services accountable towards public interests and values.[31]

As European societies scramble to formulate new governance structures for dominant platforms, an overarching concern is a lack of transparency. At present, it is difficult for outside stakeholders to observe what occurs on platforms – let alone to hold them accountable. Recent policymaking in Europe has therefore taken 'transparency' as a core value in platform governance, with a growing body of standards and legislative proposals dedicated to imposing new disclosure obligations on platforms (see below).

Different transparency measures are needed for different stakeholders. A majority of recent policymaking has focused on transparency for platform users, including end-users in business-to-consumer

30 Martin Moore and Damian Tambini (eds), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (Oxford University Press 2018).

31 José van Dijck, Thomas Poell and Martijn de Waal, *The Platform Society: Public Values in a Connective World* (Oxford University Press 2018).

relationships as well as commercial users in business-to-business relationships, and to a lesser extent investigative powers for regulatory agencies.[32] Whilst these forms of transparency are certainly worthwhile, this report focuses specifically on data access for independent public interest researchers. This topic has until now received relatively less attention, and has only recently started to gain traction in European policymaking.

The importance of data access for researchers has been acknowledged by a variety of European government institutions. The Council of Europe's Committee of Ministers has called on states to encourage social media platforms to develop *open, independent, transparent and participatory initiatives* that bring together social media services providers not only with regulators but also with 'media actors… civil society, academia and other relevant stakeholders'.[33] In the European Commission's Code of Practice on Disinformation, a central point of attention is 'empowering the research community' and creating public disclosures about political microtargeting.[34] The Commission also recently announced plans to develop a European Digital Media Observatory, and data access for researchers also recurs as a point of attention in their recent AI White Paper, Digital Strategy, and the European Strategy for Data.[35] Recent committee reports from the EU parliament regarding the Digital Service Act, including the IMCO, LIBE, and JURI committees, also underscore the importance of transparency in general, and/or research access in particular.[36] A Preliminary Opinion on Data Protection and Scientific Research from the European Data Protection Supervisor (EDPS) also emphasizes the importance of independent scientific research into platform services.[37]

These principles have also been raised in the fight against Covid-19, and the related challenges of public health communication and the combatting of misinformation. A Commission Communication on published in June 2020, promotes platform research access as a central part of Europe's strategy for tackling Covid-19 disinformation. Amongst other transparency measures, it recommends that platforms 'agree with [the European Digital Media Observatory] upon a framework providing academic researchers privacy-protected access to relevant platforms' data to enhance

32 Such measures can be seen, for instance, in data protection, consumer protection, e-privacy, and competition law. A prominent example is the recent 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 (Text with EEA relevance) [2019] OJ L186/57 (P2B Regulation).

33 Council of Europe Committee of Ministers to Member States, Recommendation CM/Rec (2018)1 on Media Pluralism and Transparency of Media Ownership (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies), para 2.5. [emphasis added]

34 European Commission, 'EU Code of Practice on Disinformation' <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation> accessed 9 June 2020.

35 European Commission, 'Commission Launches Call to Create the European Digital Media Observatory' (Shaping Europe's digital future – European Commission, 7 October 2019) <https://ec.europa.eu/digital-single-market/en/news/commission-launches-call-create-european-digital-media-observatory> accessed 9 June 2020.

36 The JURI report, for instance, recommends public disclosures about targeted advertising. The LIBE report demands public reporting about content removal by platforms and national authorities, as well as calls for 'accountability- and evidence-based policy' which requires 'robust data'. The IMCO report also requires public reporting about notice-and-action procedures, as well as demands for 'transparency' (although these are focused primarily on consumers rather than researchers). European Parliament Committee on Legal Affairs, 'Draft report with recommendations to the Commission on a Digital Services Act: adapting commercial and civil law rules for commercial entities operating online', PE650.529v01-00, 22 April 2020 (JURI report), ch 2.2.4; European Parliament Committee on Civil Liberties, Justice and Home Affairs, 'Draft report on the Digital Services Act and fundamental rights issues posed', PE650.509v01-00, 24 April 2020 (LIBE report),; European Parliament Committee on the Internal Market and Consumer Protection, 'Draft report with recommendations to the Commission on Digital Services Act: Improving the functioning of the Single Market', PE648.474v02-00, 24 April 2020 (IMCO report).

37 European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (2020) https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf.

the detection and analysis of disinformation'.[38] This recent recommendation is yet to be implemented or operationalized, underlining yet again the need for more detailed guidance on the design and governance of research access frameworks.

What are the policy grounds motivating these calls for public interest research into platforms? There are several.

- Independent research is essential in diagnosing new harms in online ecosystems. At present, policy discussions about platforms suffer from a thin evidence base, and many high-profile concerns are poorly understood even by experts in the field.[39] Indeed, many argue that allegations such as 'filter bubbles' and 'foreign interference' are exaggerated or even unfounded.[40] At the same time, the structural risk remains that platforms may abuse their algorithmic agenda-setting power for their own goals, without a clear record of their decisions. More research is needed to understand these problems, which may be difficult or even impossible to perform without access to platform data.

- Public interest research contributes to developing and enforcing evidence-based regulatory policies. Diagnosing harms is of course a first step towards evidence-based policy, since one cannot respond to unknown harms. Without this evidence, regulation risks being ineffective, or even entirely misguided. Researchers can also play an important role in assessing the effectiveness of existing policies, in order to gain a better understanding of how complex online phenomena such as harmful content can best be tackled. Such research can also assist in enforcement challenges, e.g. by identifying particular instances of wrongdoing but also by helping to prioritize investigative efforts with greater accuracy.

- Relying on regulators to perform this all of this research is not advisable, since regulators are capacity-constrained and often lack much of the essential expertise needed to oversee this vast and highly technical field. By mobilizing academics, media, civil society or other independent researchers, policymakers can bring a wealth of expertise and research capacity to bear on urgent regulatory issues – a wealth that no reasonable amount of regulatory funding can match.

- Public interest research not only serves regulators and policymakers, but can also help to mobilize other forms of social and political accountability from users, commercial actors, opinion-makers or politicians. By diagnosing online harms and drawing attention to them, independent researchers can play an essential role in raising awareness about these online governance issues, and stimulating public debate, critique, and social action. As the Cornils report recognizes, 'public pressure on the practice of social networks is a very important element of platform governance.'[41] Under certain circumstances, platforms have shown themselves

38 European Commission and High Representative of the Union for Foreign Affairs and Security Policy, 'Joint Communication to the European Parliament, The European Council, The Council, the European Economic and Social Committee and the Committee of the Regions: Tackling COVID-19 disinformation – Getting the facts right' (Joint Communication) JOIN (2020) 8 final, 10 June 2020 (Joint Communication on Disinformation).

39 Cf. Birgit Stark and others, 'Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse' (AlgorithmWatch 2020) <https://algorithmwatch.org/en/governingplatforms/communications-study-stark-may-2020>.

40 E.g. Frederik J Zuiderveen Borgesius and others, 'Online Political Microtargeting: Promises and Threats for Democracy' (2018) 14 Utrecht Law Review 82; Yochai Benkler, Robert Faris and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford University Press 2018); Axel Bruns, 'Filter Bubble' (2019) 8 Internet Policy Review <https://policyreview.info/concepts/filter-bubble> accessed 9 June 2020.

41 Matthias Cornils, 'Designing Platform Governance: A Normative Perspective on Regulatory Needs, Strategies, and Tools to Enhance the Information Function of Intermediaries' (AlgorithmWatch 2020) <https://algorithmwatch.org/en/governingplatforms/legal-study-cornils-may-2020>.

responsive to public criticism and reputational threats. And if platforms are *not* responsive to public pressure, then uncovering wrongdoing may provide an additional impetus for government action.

- It is worth noting that public interest research can also serve to hold other stakeholders accountable – in the first place, users. In many cases, it may be more effective to address wrongdoing by users directly, rather than relying solely on enforcement by platforms. Popular social media 'influencers', for instance, can be an important source of harms and, accordingly, also an important point of attention for both binding regulation and more informal forms of public pressure.[42] Public interest research can help to chart online communities and spheres of influence, and hold relevant speakers accountable.[43]

- Relatedly, and crucially, platform data can also be instrumental in holding *governments* accountable for their regulation of social media. After all, social media regulation is highly sensitive from a fundamental rights perspective, raising urgent concerns relating to e.g. the freedom of expression, non-discrimination, data protection and privacy. If regulators base their interventions on confidential platform data, it may be difficult for citizens to hold them accountable for these fundamental rights concerns. In this light, data access for independent researchers can also serve a democratic function in contributing to a system of checks and balances for government action regarding online media. As Gorwa and Garton Ash argue, '[t]he challenge is to articulate a version of transparency—principles and proposals—that can produce an informational environment that is just and democratic in that it enables individuals to protect their interests and, collectively, to control the organizations that affect their lives.' [44]

- Whilst this report focuses on data access for purposes of platform governance, it should be emphasized that platform data can also be highly valuable for other forms of research that are not expressly focused on accountability. For instance, social media data is in high demand in countless areas of social science, related to e.g. psychological, sociological, economic and legal research, even if they are not directly aimed at an urgent societal harm or governance issue. To give one example, economists have expressed great interest 'in the remarkable richness of data the generated by Uber', which can help in the micro-economic study of pricing effects.[45] But within the realm of content curation, social scientists and investigative journalists have also been struggling to chart and examine the political microtargeting ecosystem,[46] or how social media are affecting children's wellbeing.[47] As more of our society

42 Some of the most visible controversies surrounding social media governance in recent memory concerned the treatment of prominent influencers, including conspiracy theorist Alex Jones, alt-right pundit Milo Yiannopoulos and youth entertainer. Alex Hern, 'Facebook, Apple, YouTube and Spotify Ban Infowars' Alex Jones' *The Guardian* (6 August 2018) <https://www.theguardian.com/technology/2018/aug/06/apple-removes-podcasts-infowars-alex-jones> accessed 9 June 2020; Kari Paul and Jim Waterson, 'Facebook Bans Alex Jones, Milo Yiannopoulos and Other Far-Right Figures' *The Guardian* (2 May 2019) <https://www.theguardian.com/technology/2019/may/02/facebook-ban-alex-jones-milo-yiannopoulos> accessed 9 June 2020; 'YouTube Punishes Star over Suicide Video' (*BBC News*, 11 January 2018) <https://www.bbc.com/news/world-asia-42644321> accessed 9 June 2020.

43 E.g. Rebecca Lewis, 'Alternative Influence: Broadcasting the Reactionary Right on YouTube' (Data & Society Research Institute 2020) <https://datasociety.net/library/alternative-influence/> accessed 9 June 2020.

44 Archon Fung, 'Infotopia: Unleashing the Democratic Power of Transparency' (2013) 41 Politics & Society 183, 184.

45 E.g. Peter Cohen and others, 'Using Big Data to Estimate Consumer Surplus: The Case of Uber' (National Bureau of Economic Research 2016) Working Paper 22627 <http://www.nber.org/papers/w22627> accessed 9 June 2020 ('In this paper we exploit the remarkable richness of the data generated by Uber, and in particular its low-cost product UberX, to generate consumer surplus estimates that require less restrictive identifying assumptions than any other prior research that we are aware of.').

46 Zuiderveen Borgesius and others (n 40).

47 Project AWeSome, 'For Researchers' <https://www.project-awesome.nl/for-researchers> accessed 9 June 2020.

moves online and onto platform services, so too have research agendas and, correspondingly, the need for access to relevant data. In the longer term, data access regimes as discussed in this paper have the potential to make vital contributions to scientific advancement.

2.2 How we got here: barriers to research access in platform governance

At present, technical and legal circumstances make public interest research into platform services difficult if not impossible. Although a range of transparency initiatives exist, mostly in the form of self-regulation and occasionally through binding laws, their contribution to public interest research is limited.[48] Below we discuss several high-profile transparency measures and their shortcomings for purposes of public interest research.

2.2.1 Public APIs (and how platforms restricted them)

Perhaps the most valuable resource that platforms have offered to third party researchers are their public APIs. These tools, often developed in commercial contexts, allow third parties to request machine-readable data in bulk, on a range of relevant topics. From the outset, APIs have always had important restrictions; certain features and data remained unavailable, or only accessible at a premium. Nonetheless, APIs have in certain circumstances and for certain platforms served as an important tool for independent research. Twitter and YouTube, for instance, offer relatively generous data access through APIs -- although these too are by no means immune from criticism -- whereas Facebook and Instagram offer comparatively lower levels of access.[49]

Unfortunately, the capacities of research APIs have in recent years actually regressed rather than expanded. Platforms originally devised these APIs as tools to generate publicity and relevant know-how, but as Bruns observes, the relationships with academic and research communities gradually 'soured' as critical research about platforms increased.[50] Starting in 2014-2015, important functionalities started to be restricted. A key turning point was the Cambridge Analytica scandal,[51] which prompted platforms to drastically curtail their APIs or, in the case of Instagram, to shutter them entirely. This development has been described by researchers as 'the APIcalypse' or the move to a 'post-API age'.[52]

---

48 It can be argued that user-facing transparency measures such as the General Data Protection Regulation (GDPR) and Platform-to-Business Regulation (P2B) can also be leveraged for purposes of public interest research, but it is clear that these instruments are not designed with this purpose in mind. E.g. Jef Ausloos, 'GDPR Transparency as a Research Method' (Institute for Information Law (IViR), University of Amsterdam 2019) Draft Paper <https://papers.ssrn.com/abstract=3465680> accessed 17 October 2019.

49 Kevin Munger and Joseph Phillips, 'A Supply and Demand Framework for Youtube Politics' (Department of Political Science, Pennsylvania State University 2019) Draft Paper <https://osf.io/73jys/download> ('The disproportionate (to its influence among the general population) amount of research using Twitter data has been well-noted, and is often ascribed to their open API from which researchers can scrape tweets [.] YouTube, however, also has an open API, 2 which is in some ways even more generous than Twitter's. Researchers can easily query search results from the first day that YouTube went live, and scrape the entirety of a given user's history... In contrast to Facebook, which does not, and which has been restricting access to data collection that was once opt-in in the wake of the misuse of that data access.'); On the overrepresentation of Twitter in academic research due to its relatively generous API, see: Zeynep Tufekci, 'Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls' (2014) <http://arxiv.org/abs/1403.7400> accessed 9 June 2020.

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

51 The Guardian, 'The Cambridge Analytical Files' (The Cambridge Analytical Files) <https://www.theguardian.com/news/series/cambridge-analytica-files> accessed 16 June 2020.

52 Deen Freelon, 'Computational Research in the Post-API Age' (2018) 35 Political Communication 665; Bruns, 'After the 'APIcalypse" (n 50).

Whilst the Cambridge Analytica scandal highlighted the privacy risks of data sharing and the need for privacy-complaint design, researchers have argued that the subsequent shutdown of APIs by platforms has been excessive, and fails to recognize the important public interests served by public APIs.[53] This is highly plausible given that platforms have few immediate incentives to offer research access (a topic returned to below). In the words of Justin Littman:

> 'Research by academic institutions is clearly perceived as a liability post-Cambridge Analytica. ... While there's clearly a huge societal benefit to this research, it's not necessarily research that benefits social media companies directly. It's easier to say no than to figure out how to handle it properly.' [54]

Another concern when it comes to APIs is the risk of sudden changes to these systems, which can render past or ongoing research methods obsolete. Developing the skills and tools to work with APIs can require major investments from researchers, which in turn can be undermined when platforms update and alter these systems. This often occurs without prior consultation or notice towards relevant researchers, leading to significant disruptions of their research. Researchers have no guarantee of consistency, or any right to contest platform decisions, leaving them in a situation which is fundamentally precarious. As communications professor Deen Freelon puts it:

> 'When companies can restrict or eliminate API access at any time, for any reason, and without any recourse, computational researchers and students need to seriously consider how to proceed. We find ourselves in a situation where heavy investment in teaching and learning platform-specific methods can be rendered useless overnight.' [55]

In light of these developments, a growing cohort of communications scientists are taking a 'data activist' stance, and calling on governments to regulate and ensure data access.[56] Others try to work with independent research tools, but these face their own crucial limitations and restrictions.

2.2.2 Independent auditing tools (and how platforms prohibit them)

Independent researchers have tried to understand online ecosystems by capturing and observing user-facing data from platforms. This practice, which is sometimes referred to as 'auditing' or 'scraping', allows for large-scale data gathering performed with the help of volunteers or automated bots.[57] In this way, researchers can obtain large-scale datasets about platform operations. This method has a key advantage over APIs, namely that researchers are not reliant on platforms to determine the validity and completeness of information obtained. However,

53 Bruns, 'After the 'APIcalypse'' (n 50).

54 ibid.

55 Freelon (n 52).

56 Bruns, 'After the 'APIcalypse'' (n 50).

57 Christian Sandvig and others, 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms' (2014) <https://www.semanticscholar.org/paper/Auditing-Algorithms-%3A-Research-Methods-for-on-Sandvig-Hamilton/b7227cbd34766655dea10d0437ab10df3a127396> accessed 15 June 2020.

without the cooperation of platforms, the capacities of this research are also limited in important ways.[58] As the following section explains, the legal and technical design of platforms is increasingly restrictive for independent research methods.

The first and most fundamental limitation in this type of web-scraping research is that it is limited to user-facing data. This is a major restriction in the possible scope of research. Many of the most salient datasets that platforms, such as training sets for machine learning systems, collect are never shared with users in the first place.

Another problem for independent research tools is that they are often restricted technically and legally by platform operators.[59] The majority of platforms prohibit this type of research in their terms of service, and they have also taken steps to enforce them. One prominent example comes from the investigative research group ProPublica, which had developed a web-scraping tool to monitor political advertisements in US elections.[60] In August 2018, they received a notice from Facebook demanding that they discontinue their work due to violations of their Terms of Service. After ProPublica refused, Facebook implemented technical measures that effectively blocked ProPublica's tool, alongside several other comparable tools.[61]

These contractual and technical restrictions on independent research have been criticized widely in civil society. Following the ProPublica incident, for instance, a group of civil rights lawyers associated with the Knight First Amendment Institute published an open letter proposing that Facebook 'amend its terms of service to create a safe harbor for certain journalism and research on its platform'.[62] It is also worth noting that the EU Code of Practice on Disinformation, which Facebook has signed, requires not only that they 'support good faith independent efforts to track disinformation and understand its impact' but also more specifically that they 'commit not to prohibit or discourage good faith research' into this topic.[63] The actual legal status of web scraping remains unclear and disputed; recent rulings in the United States have underscored the importance of web scraping for freedom of expression, and declined to apply criminal sanctions to this practice.[64] However, the status of web scraping under European law, including private-law doctrines such as contract and (intellectual) property, remains an area of legal uncertainty.

The hesitance of courts to grant *carte blanche* to web scraping is entirely understandable in light of the fact that web scraping can also be abused for harmful purposes, and therefore requires certain restrictions.

---

58 E.g. Tobias D Krafft and others, 'Filterblase geplatzt? Kaum Raum für Personalisierung bei Google-Suchen zur Bundestagswahl 2017' (AlgorithmWatch, 8 September 2017) <https://algorithmwatch.org/filterblase-geplatzt-kaum-raum-fuer-personalisierung-bei-google-suchen-zur-bundestagswahl-2017/> accessed 9 June 2020; 'SCHUFA, a Black Box: OpenSCHUFA Results Published' (*AlgorithmWatch*, 29 November 2018) <https://algorithmwatch.org/en/schufa-a-black-box-openschufa-results-published/> accessed 9 June 2020 (Technically speaking, 'scraping' refers to the act of downloading website information and storing it locally. 'Auditing' refers in general to attempts to interact with services in order to study their behaviour. Regarding platforms, these methods are often combined in what Christian Sandvig et al have termed the 'scraping audit'. When user accounts are impersonated in this process, this is referred to as a 'sock puppet audit'. When real users participate as volunteers, as a 'crowdsourced audit' or 'collaborative audit'.).

59 Bodo et al. 'Tackling the Algorithmic Control Crisis: The Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents', *Yale Journal of Law and Technology* (2018) 19.

60 Jeremy B Merrill and Ariana Tobin, 'Facebook Moves to Block Ad Transparency Tools — Including Ours' (*ProPublica*, 28 January 2019) <https://www.propublica.org/article/facebook-blocks-ad-transparency-tools> accessed 9 June 2020.

61 ibid.

62 Letter from Jameel Jaffer and others to Mark Zuckerberg (6 August 2018) <https://knightcolumbia.org/sites/default/files/content/Facebook_Letter.pdf> accessed 9 June 2020.

63 European Commission, 'EU Code of Practice on Disinformation' (n 34), recital 13.

64 Jamie Williams and Naomi Gilens, 'Federal Judge Rules It Is Not a Crime to Violate a Website's Terms of Service' (*Electronic Frontier Foundation*, 6 April 2020) <https://www.eff.org/deeplinks/2020/04/federal-judge-rules-it-not-crime-violate-websites-terms-service> accessed 9 June 2020.

Although web scraping can serve important public interests, bad actors can use data scraping in ways that undermine service integrity or user privacy. Perhaps the most telling example is ClearView AI, a US-based technology company which garnered widespread notoriety for developing powerful, privacy-invasive facial recognition software.[65] The 'backbone' of this service, according to the New York Times, 'is a database of more than three billion images that Clearview claims to have scraped from Facebook, YouTube, Venmo and millions of other website'.[66]

Such potential harms from data scraping can in theory be addressed through legal tools such as data protection, unfair commercial practices and intellectual property law – but effective enforcement against these bad actors is not always possible. These circumstances arguably lend support for technical restrictions on web scraping, and it remains to be seen whether and how a 'safe harbor' for public interest research, as envisaged by the Knight First Amendment Institute, can be implemented.

2.2.3 Data access grants and partnerships (and how platforms have failed to deliver)

As an alternative to public APIs, a number of alternative data access regimes have emerged. Most prominent among them is Facebook's Social Science One project – a partnership led by US academics intended to offer privacy- and data protection-compliant access for academic researchers. In its original design, this project was intended to provide a structure for academics to request research data from Facebook. However, the project faced many delays and organizational difficulties, drawing criticism not only from third parties but ultimately even from the project funders and its European Advisory Board.[67] As the cause for these delays, the project cites legal constraints related to e.g. US privacy law and EU data protection law. As discussed further in Section 2.3.2, the ultimate merit of these legal claims may be questionable, but it is undeniable that this is an area of significant complexity and uncertainty.

Ultimately, these legal considerations seem to have prompted a course adjustment for Social Science One; instead of granting privileged access to trusted researchers, as was initially envisaged, the platform has aimed to develop more restrictive, privacy-compliant APIs and datasets which can be shared with a broader range of academic researchers with minimal privacy concerns. The program released its first dataset in early 2020. Only seventeen groups of researchers have thus far been granted access, but the program claims it will accelerate researcher accreditation in the coming period.[68]

In addition to its many delays, Social Science One has also faced more fundamental criticisms about its aims and methods. In particular, the project has been criticized for insufficiently safeguarding the independence, inclusiveness and diversity of potential researchers and topics.[69] A recent report from the

65 Kashmir Hill, 'The Secretive Company That Might End Privacy as We Know It' *The New York Times* (18 January 2020) <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> accessed 9 June 2020.

66 ibid.

67 European Advisory Committee Social Science One, 'Public Statement from the Co-Chairs and European Advisory Committee of Social Science One' (*Social Science One*, 11 December 2019) <https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one> accessed 9 June 2020; Craig Silverman, 'Funders Are Ready To Pull Out Of Facebook's Academic Data Sharing Project' (*BuzzFeed News*, 27 August 2019) <https://www.buzzfeednews.com/article/craigsilverman/funders-are-ready-to-pull-out-of-facebooks-academic-data> accessed 9 June 2020.

68 Gary King and Nathaniel Persily, 'Unprecedented Facebook URLs Dataset Now Available for Academic Research through Social Science One' (*Social Science One*, 13 February 2020) <https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one> accessed 4 March 2020.

69 Bruns, 'After the 'APIcalypse'' (n 50); European Advisory Committee Social Science One (n 67).

European Regulators of Audio-visual Media Services (ERGA) praised Social Science One for the release of their URL dataset, but also criticized the project, highlighting *inter alia* the limited access to, and limited utility of, the program's tools, the lack of dialogue with relevant stakeholders and the recurring delays in the program's rollout.[70]

Besides Social Science One, other relevant data access regimes to emerge in recent years include Stanford University's Internet Observatory (headed by former Facebook executive Alex Stamos). Microsoft has also entered launched a number of research access programs, including Microsoft Research Open Data program.[71] Platforms may also enter into ad-hoc data-sharing arrangements with local researchers or governments, or issue so-called 'data grants', also known as 'data philanthropy'.[72]

A common objection to each of these new data-access regimes is that platforms may use these types of discretionary data grants in a self-interested and opportunistic fashion, which distorts the research agenda and fails to deliver true accountability. Platforms exercise influence in the first instance by determining the topics and materials available for disclosure. Secondly, platforms exercise influence by determining which researchers they partner with, depending on, for instance, their research discipline, research agenda, and prior publications. Indeed, the prospect of data access can also have on chilling effect on researchers who might otherwise pursue critical lines of research. So long as transparency is designed on the platform's own terms, and access depends on their continuing goodwill, the resulting research are unlikely to deliver true accountability. Overall, these types of influence related to data access can create a chilling effect on research that discourages the most critical research and researchers from stepping forward. As Bruns writes, 'the proposed price of such harmony [...] is to limit scholarly inquiry to issues and topics that are unlikely to put pressure on the platform providing the data'.[73]

2.2.4 Public reporting about content moderation and targeted advertising (and its lack of meaningful detail)

On select issues, platforms have also developed public reporting practices. In particular, public reporting is a common practice regarding content moderation activities such as removing or delisting certain content, or banning certain users. More recently, public reporting has also been deployed in the context of (political) microtargeted advertising. These features may have some limited research utility but they are widely criticized for lacking sufficient detail to offer meaningful insights.

Most major platforms issue regular Transparency Reports documenting their content moderation activities. This practice was spurred by digital rights activism from groups including AccessNow and Ranking Digital Rights, and by self-regulatory compacts such as the Global Network Initiative (GNI).[74] These transparency reports document, with aggregate data, the number of content removals on various issues, such as copyright or hate speech. Their adoption has grown over the past years, as well as the level of detail

70 European Regulators Group for Audiovisual Media Services, 'ERGA Report on Disinformation: Assessment of the Implementation of the Code of Practice' (2020) <http://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf>.

71 Jennifer Yokoyama, 'Closing the Data Divide: The Need for Open Data' (*Microsoft on the Issues*, 21 April 2020) <https://blogs.microsoft.com/on-the-issues/2020/04/21/open-data-campaign-divide/> accessed 24 April 2020.

72 On the concept of 'data philanthropy', see Bruns, 'After the 'APIcalypse'' (n 50).

73 Bruns, 'After the 'APIcalypse'' (n 50).

74 Daphne Keller and Paddy 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* (Cambridge University Press 2019).

in their disclosures. Germany's NetzDG is the first law to regulate this type of disclosure by law, requiring that all removals under the NetzDG are disclosed in semi-annual reports.[75]

A recurring criticism of these aggregated reports – both in self-regulation and under the NetzDG – is that they provide no little to no insight into the nature of the content affected.[76] Generic figures about the number of takedowns and appeals, removed from their particular case contexts, say little about the merits of such content moderation and its actual impact in practice. Yet platforms refuse to disclose the underlying content on the grounds that this content is presumed to be harmful and/or unlawful, and therefore not suitable for publication. Accordingly, experts now argue that public reporting mechanisms such as those in the NetzDG should be supplemented with privileged access for researchers and regulators, who would be allowed to study illegal and/or harmful content in a safe environment.

Another well-known program is Project Lumen, a project operated by the Berkman Klein Center for Internet & Society at Harvard University. Project Lumen maintains a public database of takedown requests submitted to various online companies including Google, Twitter, YouTube, Wikipedia, Medium and Vimeo.[77] Researchers are able to examine individual takedown notices, which is a marked advantage compared to the aggregate reporting described above. Lumen has therefore been a relatively common source of academic research, compared to platform takedown reports.[78] However, Lumen still does not guarantee access to the data itself: when content is removed from the original URL, researchers are also unable to access it.[79]

One area where public reporting has generally failed to take place, is the algorithmic curation of content in recommender systems and other ranking algorithms. Platforms routinely intervene in these systems to reward certain forms of content and to punish others, but the transparency of these gatekeeping decisions is limited. Under the P2B regulation, platforms are required to offer high-level descriptions of relevant weighting criteria and to give notice to commercial users affected by downranking, but systematic, platform-wide reporting about downranking actions is largely absent, and public interest research into these systems remains a challenge.[80]

In one specific area of algorithmic content curation, however, public disclosures have in fact started to emerge: targeted (political) advertising. Specifically, the major platforms Google, Facebook and Twitter have each developed public archives documenting their political ads, which are accessible through browser-based search interfaces as well as through automated APIs.[81] Increasingly, this archiving is also being proposed as a mandatory requirement in legislation, such as Canada's Election Reform Act and the US Honest Ads Act, and is also a key feature of the European Commission's Code of Practice on

75 ibid; Heidi Tworek and Paddy Leerssen, 'An Analysis of Germany's NetzDG Law' (Institute for Information Law (IViR), University of Amsterdam 2019) <https://hdl.handle.net/11245.1/3dc07e3e-a988-4f61-bb8c-388d903504a7> accessed 9 June 2020.

76 Julia Powles, 'The Case That Won't Be Forgotten' (2015) 47 Loyola University Chicago Law Journal 583; Keller and Leerssen (n 74); Tworek and Leerssen (n 75).

77 Berkman Klein Center for Internet & Society at Harvard University, 'Lumen – About' (*About Us*) <https://www.lumendatabase.org/pages/about> accessed 9 June 2020.

78 For an overview, see Berkman Klein Center for Internet & Society at Harvard University, 'Lumen – Research' (*Research*) <https://www.lumendatabase.org/pages/research> accessed 9 June 2020.

79 Keller & Leerssen (n 74).

80 Paddy Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems' (Institute for Information Law (IViR), University of Amsterdam 2020) Preprint Paper <https://papers.ssrn.com/abstract=3544009> accessed 9 June 2020.

81 Paddy Leerssen and others, 'Platform Ad Archives: Promises and Pitfalls' (2019) 8 Internet Policy Review <https://policyreview.info/articles/analysis/platform-ad-archives-promises-and-pitfalls> accessed 7 February 2020.

Disinformation. However, major platforms including Facebook and Google have typically resisted these binding measures and continue to present their archives as primarily self-regulatory efforts.[82]

As research tools, platform ad archives have been shown to be deficient in numerous ways.[83] For instance, their selection of 'political ads' has been shown to contain numerous false positives as well as false negatives. Furthermore, in the case of Facebook's ad archive, names listed for ad buyers have been proven false, since Facebook failed to verify the identity submitted by buyers. The Facebook archive also lacks crucial information on ad targeting, which is essential to understanding why particular ads reach a particular audience. More fundamentally, the API through which this data was offered was so riddled with bugs as to be effectively unusable. Most catastrophically, Facebook's entire library suffered a major outage only days before the UK election which it was supposed to help cover.[84] Whilst the majority of this criticism has been directed towards Facebook, it is worth noting that Twitter and Google's implementations offer even less detailed data.[85] In most cases, these crucial limitations and pitfalls do not appear to be based on concrete legal or ethical concerns. Instead, the more plausible explanation is that platforms simply lack the necessary incentives to invest in in meaningful transparency and accountability for their targeted advertising services.

2.3 Discussion: towards binding regulation of research access?

The above shows that platforms have tended to over-promise and under-deliver when it comes to the self-regulation of public interest research access. Although there is increasing public pressure on platforms to deliver transparency, in practice research access has broadly *diminished* as the result of new restrictions on important resources such as APIs and web scraping tools. Proactive disclosures from platforms, in the form of data access partnerships or public reporting, have broadly failed to compensate. This explains the growing consensus that research access should be regulated by law.

However, regulating research access is not straightforward. Below we outline two of the key challenges in this space. Firstly, the *incentive problem*: platforms have several strong incentives to oppose meaningful transparency, and are unlikely to comply in earnest unless the law imposes clear and enforceable duties on them. Secondly, the *security problem*: platform datasets can be highly sensitive to abuse, and creating adequate safeguards is essential both legally and ethically.

82 ibid.

83 ibid.

84 CNN Business and Hadas Gold, 'Facebook Promised Transparency on Political Ads. Its System Crashed Days before the UK Election' (*CNN Business*) <https://www.cnn.com/2019/12/11/tech/facebook-political-ads-uk-election-ge19/index.html> accessed 16 June 2020.

85 Leerssen and others. 'Platform Ad Archives' (n 81).

2.3.1 The incentive problem: imposing transparency on unwilling platforms

Platforms' repeated failure to meet researchers' data access demands need not be surprising if we consider the many (economic, technical, and legal and political) incentives that run counter to research access.

The opposing incentives to research access are various. First, platforms typically offer data access as a commercial service, and this business model gives them an interest in maintaining the *exclusivity* of this asset.[86] This clashes with the ideal of offering the same data for free to the public and/or to public interest researchers. Second, developing data access frameworks involves important costs and risks. Indeed, platform data may be *sensitive to abuse* by bad actors, creating risks to, for instance, user privacy or service integrity (see section 2.3.2 below). Third, however, platforms may also be averse to greater scrutiny of their services by third parties, due to the reputational, political and legal risks this may generate if wrongdoing is exposed. In other words, as profit-driven companies, platforms may simply have an interest in avoiding accountability.

Binding access regulation can play a crucial role in overcoming platform opposition to public research access. Achieving this in practice, however, is not straightforward. Platforms may not always comply with timely, accurate and complete disclosures. Indeed, they may appeal in bad faith to the supposed sensitivity of this data, in order to avoid scrutiny. In this light, effective regulation will require a framework to verify platform disclosures, and sanction them for non-compliance.[87] In doing so, regulators must be able to distinguish legitimate claims to confidentiality based on grounds of data sensitivity, from illegitimate claims to secrecy based solely on platforms' commercial self-interest. Given the scale, complexity and heterogeneity of platform services, this is a significant regulatory challenge that requires a high degree of technical expertise.

However, these challenges are not wholly unique to platforms. Corporations in many other industries have been forced to disclose information even though they have strong incentives to avoid such transparency. To ensure such compliance, governments have developed robust systems of monitoring, verification, and enforcement.[88] Platform governance need not reinvent the wheel, but can look to such examples for inspiration. Along those lines, Chapter 3 will explore how the European Union has regulated public interest data disclosures in the context of industrial facilities and their pollutant data.

---

86 See generally: Bruns, 'After the 'APIcalypse'' (n 50) (Most major platforms monetize audience data for marketing purposes in various ways, including premium, enterprise APIs. Examples include Facebook's Crowdtangle program, and Google's Analytics program, which offers free and premium access levels.).

87 Lubos Kuklis and Ben Wagner, 'Disinformation, Data Verification and Social Media' (*Media@LSE*, 7 January 2020) <https://blogs.lse.ac.uk/medialse/2020/01/07/disinformation-data-verification-and-social-media/> accessed 10 June 2020 (Argued in the context of government access but also applicable to research access); Cornils (n 41).

88 E.g. food safety (Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety [2002] OJ L31/1), fisheries (Council Regulation (EC) No 1224/2009 of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy, amending Regulations (EC) No 847/96, (EC) No 2371/2002, (EC) No 811/2004, (EC) No 768/2005, (EC) No 2115/2005, (EC) No 2166/2005, (EC) No 388/2006, (EC) No 509/2007, (EC) No 676/2007, (EC) No 1098/2007, (EC) No 1300/2008, (EC) No 1342/2008 and repealing Regulations (EEC) No 2847/93, (EC) No 1627/94 and (EC) No 1966/2006 [2009] OJ L343/1; European Commission, 'The EU's Fisheries Control System' (*Fisheries – European Commission*, 16 September 2016) <https://ec.europa.eu/fisheries/cfp/control_en> accessed 10 June 2020.) and finance (Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (Text with EEA relevance) [2014] OJ L173/349).

2.3.2 Data protection concerns: safeguards against data harms and abuse

Even if platforms are made to cooperate fully, research access poses important design questions related to the safeguarding of sensitive data. The challenge is to create frameworks that enable public interest research to the broadest extent possible, whilst preventing abuse of the data involved.

Platform data can be sensitive for various reasons. First and foremost, the disclosure of personal data about platform users can infringe their privacy and data protection rights. Secondly, platforms may object to certain data disclosures based on economic/proprietary grounds such as trade secrecy regarding the design and operation of their services.[89] Thirdly, the disclosures may disseminate content otherwise considered harmful or illegal, such as hate speech, child sexual imagery, copyright-infringing material, and so forth. Such considerations have hindered research into content removal programs, such as Google's implementation of the Right to Be Forgotten,[90] as discussed in Section 2.2.4 above. However, privacy and data protection are arguably the more cross-cutting concern, as it may implicate virtually all aspects of social media insofar as they relate to the activities of individual end-users.

Various safeguards are possible to facilitate public interest research despite the presence of such sensitive data. A first set of safeguards revolves around omitting sensitive data from relevant datasets. In the case of personal data, for instance, anonymization can help to mitigate privacy and data protection risks. This being said, anonymization is rarely fool proof, and reidentification of users often remains possible, particularly in rich social media datasets.[91] These interventions can also reduce the research utility of the dataset. To put it simply: some research questions can only be answered by studying sensitive data.

If the use of sensitive data cannot be avoided, then organizational safeguards may help to ensure their proper handling and prevent abuse. Data access can be limited to trusted researchers, who can be held to legal and ethical research standards under the threat of sanction. In addition, their access conditions can be restricted technically in various ways, to prevent them from using data for prohibited purposes (see Section 4).

The challenges related to personal data are not merely ethical but also legal: a key challenge for research access frameworks will be to comply with EU data protection law, including the General Data Protection Regulation (GDPR).[92] Although the GDPR applies a relatively light touch to journalism and to scientific research, it nonetheless applies and imposes important restrictions that must be taken into consideration.[93] Relevant considerations include the need for a processing ground (such as consent, legitimate interest of the controller);[94] the need to minimize the storage of personal data and to

89 For nuancing of this line of arguments, see: Nicholas Diakopoulos, 'Accountability in Algorithmic Decision Making' (2016) 2 *Communication of the ACM* 56.

90 Powles (n 76); Theo Bertram and others, 'Three Years of the Right to Be Forgotten' (Google Inc 2018) <https://pdfs.semanticscholar.org/13f5/e3cd0e8e522238f5df2ce279e6188664165e.pdf>.

91 See (the many references in): Jef Ausloos, Réne Mahieu and Michael Veale, 'Getting Data Subject Rights Right A Submission to the European Data Protection Board from International Data Rights Academics, to Inform Regulatory Guidance' (2020) 10 JIPITEC 294, 294-96.

92 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) (Text with EEA relevance) [2016] OJ L119/1 (GDPR).

93 *Inter alia* GDPR, arts 85 & 89.

94 GDPR, art 6.

introduce adequate security measures;[95] to comply with data breach obligations;[96] to respect the rights of individuals e.g. to request access to, or erasure of, the data involved;[97] and additional restrictions on the handling of special categories of 'sensitive data', such as those related to health or racial origin.[98] A useful starting point in this area is the European Data Protection Supervisor's recent Preliminary Opinion on Data Protection Guidelines, which proposes an array of best practices including enhanced engagement between of Data Protection Authorities and Ethical Review Boards, and the creation of EU Codes of Conduct for Research Integrity, including a specialized code for social networks research.[99] In addition to the scientific research exemptions present in the GDPR, another important safeguard for public interest research is Article 85 GDPR's protection of freedom of expression, which expressly includes 'processing for journalistic purposes and the purposes of academic, artistic or literary expression'.[100] Of course, if and when data access frameworks can manage to avoid disclosing personal data altogether, then the GDPR remains out of scope.

Again, the challenges discussed here are not unique to platforms. Governments have already developed access frameworks for sensitive data in other sectors, most notably in (public) health research. As a second case study, Chapter 4 will review Finland's cutting edge Findata access regime, which enables accredited researchers to access sensitive health data under secure conditions.

95 GDPR, arts 1(c), 1(f) & 32.

96 GDPR, art 33 & 34.

97 GDPR, ch III.

98 GDPR, art 9 (Here, scientific research benefits from a tailored regime under article 9(1).).

99 European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (2020) <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>.

100 GDPR, Article 85 & recital 153; David Erdos, *European Data Protection Regulation, Journalism, and Traditional Publishers: Balancing on a Tightrope?* (Oxford University Press 2019) (Who argues that article 85 GDPR risks being understudied and undervalued relative to the Article 89 regime for scientific research.).

3 Research access and the 'incentive problem' – Learning from environmental protection law

The European Pollutant Release and Transfer Register ('E-PRTR', 'the register', or 'the access regime') is a European Union-wide register where operators of ca. 35.000 industrial facilities located in Europe self-report the amount of pollutants they release into (waste) water, air and land or transfer to other locations every year. Created by the European Commission, the register is facilitated by national competent authorities of Member States and managed by the European Environmental Agency (EEA). The data are freely available to the public via a dedicated web-platform and as a standalone dataset.[101] Through this form of transparency, the E-PRTR aims to impose accountability on operators of industrial facilities in Europe towards to the public, NGOs, scientists, politicians, governments and supervisory authorities.

The selection of case studies for this report started with a preliminary mapping exercise of EU legislation imposing transparency obligations for accountability purposes in three sectors: finance, environment and food safety. It mapped access regimes that provide for accountability by making company-held data available to scientific researchers and the public. This preliminary research yielded several legislative measures that provide for access regimes and accountability measures in the environmental, food and financial spheres. Examples of such measures in the environmental realm include: the Emissions Trading System, which creates a market for emissions rights for greenhouse gasses[102] and the EU's Environmental Liability Directive, which implements the 'polluter pays' principle for heavy industry.[103] In the financial sector: the anti-money laundering measures creating an EU-wide system proving the ownership of bank accounts,[104] and legislation ensuring that firms publish more information on trading in listed stocks.[105] Finally, in the food industry: the Rapid Alert

101 The web-platform is located at: European Environment Agency, 'E-PRTR' (E-PRTR) <https://prtr.eea.europa.eu/#/home> accessed 9 June 2020; thError! Hyperlink reference not valid.e full dataset can be downloaded at: European Environment Agency, 'The European Pollutant Release and Transfer Register (E-PRTR), Member States Reporting under Article 7 of Regulation (EC) No 166/2006' (European Environment Agency, 6 February 2020) <https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-23> accessed 15 May 2020.

102 Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC (Text with EEA relevance) [2003] OJ L275/32.

103 Directive 2004/35/CE of the European Parliament and of the Council of 21 April 2004 on environmental liability with regard to the prevention and remedying of environmental damage. [2004] OJ L143/56.

104 Regulation (EU) 2015/847 of the European Parliament and of the Council of 20 May 2015 on information accompanying transfers of funds and repealing Regulation (EC) No 1781/2006 (Text with EEA relevance) [2015] OJ L 141/1.

105 Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (Text with EEA relevance) [2014] OJ L 173/349; Regulation (EU) No 600/2014 of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Regulation (EU) No 648/2012 (Text with EEA relevance) [2014] OJ L173/84.

System for Food and Feed creates consumer and industry facing web portals and notification systems for defects in food products in the EU;[106] and an Electronic Recording and Reporting System for Fishery on the catching of fish, landing, sales and transshipment of fishing vessels in the EU.[107]

After this initial mapping exercise, we selected the European Pollutant Release and Transfer Register ('E-PRTR Regulation')[108] as the main focus of the present case study. The choice for this particular data access regime was made for several reasons. Firstly, important parallels can be drawn between environmental protection and intermediary governance frameworks. Both relate to the governance of complex and cross-border ecosystems, safeguarding against more 'collective' harms, and curtailing negative externalities (mainly from industry).[109] Secondly, pollutant registries constitute a well-established transparency measure within cross-border environmental protection frameworks. While the E-PRTR was established in the 2000s, regulatory debate about information provision on pollution can be traced back well into the 1990s.[110] Thirdly, and following from the previous point, a lot of information on the E-PRTR is readily available, notably via the EEA's forums. Moreover, the system has been subject to several official evaluations already.

This Chapter is composed of five sections. The first section provides essential background on the E-PRTR: focusing on its legislative history, grounding in international and European law and providing some examples of how E-PRTR data has been used so far. The second section looks more closely at how the E-PRTR has been implemented, i.e. what data is made available and how exactly? The third section, zooms in on the E-PRTR governance structure: the role of EC, EEA and national environment agencies as participants and supervisors in this access regime and its arrangements on sanctions, liability and funding. Finally, the Chapter discusses how the E-PRTR stimulates accountability of industrial facilities in Europe and what key components of this regime contribute to producing such accountability.

3.1 Background

The 1992 Rio Declaration on Environment and Development gave an initial impulse to create national systems that provided citizens with 'information on hazardous materials and activities in their communities'.[111] In the European Union, this idea was first implemented in the 1996 Directive on Integrated Pollution Prevention and Control (IPPC),[112] and in

106 Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety [2002] OJ L31/1.

107 Council Regulation (EC) No 1224/2009 of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy, amending Regulations (EC) No 847/96, (EC) No 2371/2002, (EC) No 811/2004, (EC) No 768/2005, (EC) No 2115/2005, (EC) No 2166/2005, (EC) No 388/2006, (EC) No 509/2007, (EC) No 676/2007, (EC) No 1098/2007, (EC) No 1300/2008, (EC) No 1342/2008 and repealing Regulations (EEC) No 2847/93, (EC) No 1627/94 and (EC) No 1966/2006 [2009] OJ L343/1.

108 Regulation (EC) No 166/2006 Of the European Parliament and of the Council of 18 January 2006 concerning the establishment of a European Pollutant Release and Transfer Register and amending Council Directives 91/689/EEC and 96/61/EC (Text with EEA relevance) [2006] OJ L33/1 (E-PRTR Regulation).

109 Many have made this parallel before. See notably the recent seminal books: Cohen (n 8),; Zuboff (n 8); In the field of privacy and data protection particularly, see: e.g. Dennis D Hirsch, 'Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law' (2006) 41 Georgia Law Review 1; Dennis D Hirsch and Jonathan H King, 'Big Data Sustainability: An Environmental Management Systems Analogy' (2016) 72 Washington and Lee Law Review Online 409.

110 Article 10 of the 1992 UN Rio Declaration on Environment and Development "[c]alls upon States to ensure that each individual has access to information, public participation in decision-making and justice in environmental matters. … it nevertheless represents a trail blazer, laying down for the first time, at a global level, a concept that is critical both to effective environmental management and democratic governance. 'Declaration of the United Nations Conference on the Human Environment – Main Page' <https://legal.un.org/avl/ha/dunche/dunche.html> accessed 30 April 2020.

111 UNGA 'Report of the United Nations conference on environment and development' (12 August 1992) A/CONF.151/26 (Vol. I), art 10.

112 Council Directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control [1996] OJ L257/26.

2000 the European Commission's Decision on the European Pollutant Emission Register (EPER).[113] This Directive and Decision led the Commission to publish the results of its inventory of principal emissions and their responsible sources every three years.[114] The triennial report of EPER can be considered a predecessor to the E-PRTR.

Facilitating public participation in environmental decision-making

Prevention and reduction of pollution

Figure 1 – Goals of the E-PRTR

In a parallel process, the European Union signed the 1998 UN Aarhus Convention.[115] This convention granted the public rights to access environmental information, stating that: 'Each Party shall ensure that (...) public authorities, in response to a request for environmental information, make such information available to the public'.[116] Under this convention a protocol on Pollutant Release and Transfer Registers ('PRTR Protocol') was adopted in 2003.[117] The implementation of the PRTR Protocol into EU-law was achieved by the 2006 adoption of the E-PRTR Regulation.[118] In 2019 the E-PRTR Regulation was slightly amended by omnibus Regulation 2019/1010 to bring the reporting obligations of the E-PRTR in line with those in other EU environmental legislative texts.[119]

The E-PRTR Regulation has two main goals. It aims to 'facilitate public participation in environmental decision-making, as well as contributing to the prevention and reduction of pollution of the environment'[120] (cf. Figure 1). Public participation in environmental decision-making is facilitated first and foremost by making pollution data of industrial facilities accessible to the public at large, including key stakeholders such as governments, competent authorities, policymakers, NGOs, journalists and scientists.[121] Together, these parties can use the E-PRTR data to inspect facilities and create a complete picture of complicated problems on multiple levels.

Article 15

Awareness raising

The Commission and the Member States shall promote awareness of the public of the European PRTR and shall ensure that assistance is provided in accessing the European PRTR and in understanding and using the information contained in it.

Figure 2 – Art.15 in the E-PRTR Regulation

There are numerous examples that show how different parties use the E-PRTR to facilitate public

113 Commission Decision 2000/479/EC of 17 July 2000 on the implementation of a European pollutant emission register (EPER) according to Article 15 of Council Directive 96/61/EC concerning integrated pollution prevention and control (IPPC) (Text with EEA relevance) [2000] OJ L192/36.

114 ibid, preamble 2.

115 UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (adopted 25 June 1998, entered into force 30 October 2001) 2161 (UNTS) 447 (Aarhus Convention).

116 Aarhus Convention, art 4(1).

117 Protocol on Pollutant Release and Transfer Registers to the Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters (adopted 21 May 2003, entered into force 8 October 2009) 2629 (UNTS) 119.

118 E-PRTR Regulation.

119 Regulation (EU) 2019/1010 of 5 June 2019 on the alignment of reporting obligations in the field of legislation related to the environment, and amending Regulations (EC) No 166/2006 and (EU) No 995/2010 of the European Parliament and of the Council, Directives 2002/49/EC, 2004/35/EC, 2007/2/EC, 2009/147/EC and 2010/63/EU of the European Parliament and of the Council, Council Regulations (EC) No 338/97 and (EC) No 2173/2005, and Council Directive 86/278/EEC (Text with EEA relevance) [2019] OJ L170/115 (Environmental Omnibus Regulation).

120 E-PRTR Regulation, art 1.

121 European Environment Agency, 'E-PRTR FAQ' (*Frequently Asked Questions*) <https://prtr.eea.europa.eu/#/faq> accessed 27 March 2020.

participation in environmental decision-making at different levels. For example, the EEA can report on pollution of mercury (a heavy metal) in the whole European Economic Area,[122] while an NGO might solely focus on the mercury production by German coal plants.[123] Both use the E-PRTR data as an important source of information. Additionally, the E-PRTR Regulation also requires both the Commission and Member States to raise awareness and assist interested parties in accessing the data in the E-PRTR.[124] Sections 3.2.3 and 3.3.2 provide examples of how the EEA and Member States choose to fulfil this obligation and thereby contribute to the E-PRTR's goal of facilitating public participation in environmental decision-making.

Prevention and Reduction of Pollution – The E-PRTR Regulation's second goal, prevention and reduction of pollution, is to be achieved indirectly. The E-PRTR provides operators of industrial facilities the opportunity to compare their pollution levels with similar facilities and see how they perform. The possibility of public scrutiny could also incentivize operators to lower their emissions levels, because they do not want to be perceived as a heavy polluter by the public. As the Kyiv Protocol states: '[PRTR's are] expected to contribute [to] promoting a downward trend of pollution, as no company will want to be identified as among the biggest polluters.'[125] The latter is exactly what some NGOs use the E-PRTR data for, for example, ARNIKA's report on the top ten biggest environmental polluters in Bosnia and Herzegovina.[126]

NGOs are not the only ones using E-PRTR data, public authorities do as well. The European Environment Agency maintains 'industrial pollution profiles' of all 33 EEA-members: showing air, water and waste pollution and the trends therein through time per country,[127] and of the EEA as a whole.[128] It has also published an overview of learnings from ten years of pollution reporting,[129] and many other reports using E-PRTR data.[130] The European Commission's DG Environment uses E-PRTR data for policy evaluation, for example, of the Industrial Emissions Directive,[131] and the creation of industry sector specific recommendations for available techniques to lower pollution.[132]

---

122 European Environment Agency, 'Mercury in Europe's Environment' (2018) Publication 11/2018 <https://www.eea.europa.eu/publications/mercury-in-europe-s-environment> accessed 1 May 2020.

123 European Environmental Bureau, 'Mercury Emissions from Coal Power Plants in Germany' (2017) <https://eeb.org/library/mercury-emissions-from-coal-power-plants-in-germany-de/> accessed 30 April 2020.

124 E-PRTR Regulation, art 15.

125 European Commission DG Environment, 'The European Pollutant Release and Transfer Register (E-PRTR) – Environment – European Commission' (*The European Pollutant Release and Transfer Register (E-PRTR)*, 30 January 2020) <https://ec.europa.eu/environment/industry/stationary/e-prtr/legislation.htm> accessed 10 June 2020.

126 Arnika and Eko forum Zenica, 'Top Ten of Biggest Environmental Polluters According to Data of Integrated Pollutant Release and Transfer Register (PRTR) of Bosnia and Herzegovina – Report for the Year 2016' (2016) <https://issuu.com/arnika.org/docs/grafy-bosna-en1>.

127 European Environment Agency, '2019 Industrial Pollution Country Profiles' (*2019 Industrial pollution country profiles*, 2 December 2019) <https://www.eea.europa.eu/themes/industry/industrial-pollution/2019-industrial-pollution-country-profiles> accessed 1 May 2020.

128 European Environment Agency, 'EEA-33 – Industrial Pollution Profile 2019' (*EEA-33 – Industrial pollution profile 2019*, 2 December 2019) <https://www.eea.europa.eu/themes/industry/industrial-pollution/industrial-pollution-country-profiles-2019/eea33> accessed 1 May 2020.

129 European Environment Agency, 'A Decade of Industrial Pollution Data' (2019) Briefing 4/2019 <https://www.eea.europa.eu/ds_resolveuid/b8208000593e49d3aabaa8500b31b087> accessed 1 May 2020.

130 For a full list, see: European Environment Agency, 'Publications' (*European Environment Agency*) <https://www.eea.europa.eu/themes/industry/publications/publications_topic> accessed 16 June 2020.

131 AMEC Environment & Infrastructure UK Limited, 'Contribution of Industry to Pollutant Emissions to Air and Water' (2014) 32790–01 FR 13298i5 <https://circabc.europa.eu/ui/group/06f33a94-9829-4eee-b187-21bb783a0fbf/library/c4bb7fee-46df-4f96-b015-977f1c-ca2093/details> accessed 1 May 2020.

132 These Best Available Techniques (BATs) Reference Documents (BREFS) are available for many different industrial sectors, usually Chapter 1 of the BREFS uses E-PRTR data to assess the pollution of an industrial sector. See an overview of BREF's here: European IPPC Bureau, 'Reference Documents | Eippcb' <https://eippcb.jrc.ec.europa.eu/reference> accessed 1 May 2020.

Investigative researchers (in academia, NGOs and journalism) have been capitalizing on the data rendered accessible through E-PRTR data as well. Academic researchers use E-PRTR data for a diverse range of purposes: to investigate the pollution of ship traffic[133] and the impact of landfills on public health[134] to measuring the impact from industrial wastewater treatment facilities on water quality and drinking water sources.[135] NGOs have used the E-PRTR to analyze heavy metal emissions from coal plants;[136] research how those plants contribute to air pollution,[137] and even introduce a 'death ticker': showing the amount of deaths and chronic diseases that could be avoided by faster implementation of environmental performance standards.[138] Finally, journalists have benefitted from the E-PRTR both directly and indirectly. Firstly, they use the access regime to conduct their own investigations, for example on the pollution allowed by the EU's agricultural policy.[139] Secondly, they report on the outcomes of investigations by others (notably NGOs, authorities and scientists using the E-PRTR), for example, on pollution of a local manufacturing facility.[140]

This diverse set of examples of E-PRTR data being used by NGOs, journalists, scientific researchers, oversight bodies and policy-makers, illustrates the diverse use cases the E-PRTR makes possible. These examples also show how the E-PRTR access regime can facilitate public participation in environmental decision making by providing information on pollution in many different forms and how it contributes to the reduction and prevention of pollution by holding industrial facilities and policy makers accountable for their actions.

E-PRTR Shortcomings – The E-PRTR has received criticism for not entirely achieving its objectives. The European Environmental Bureau (an NGO), for example, argued that a number of important data points are missing from the E-PRTR: i.e. inspection and compliance reports of facilities; quick access to their national permits; continuous monitoring data, as is already available in the U.S. through comparable regimes; and explanations for when and why derogations of reporting standards have been granted to operators.[141]

133 MA Russo and others, 'Shipping Emissions over Europe: A State-of-the-Art and Comparative Analysis' (2018) 177 Atmospheric Environment 187.

134 G Shaddick and others, 'Towards an Assessment of the Health Impact of Industrially Contaminated Sites: Waste Landfills in Europe' (2019) 3 Environmental Epidemiology 324.

135 Annemarie P van Wezel and others, 'Impact of Industrial Waste Water Treatment Plants on Dutch Surface Waters and Drinking Water Sources' (2018) 640–641 Science of The Total Environment 1489.

136 European Environmental Bureau (n 123).

137 This report also refers further to other examples of work by NGOs using or on the functioning of the E-PRTR, e.g.: Christian Schaible and others, 'Lifting Europe's Dark Clouds – How Cutting Coal Saves Lives' (European Environmental Bureau (EEB), Sandbag, Climate Action Network (CAN) Europe, Health and Environment Alliance (HEAL), WWF European Policy Office 2016) <https://eeb.org/lifting-europes-dark-cloud-how-cutting-coal-saves-lives/>.

138 Anton Lazarus, 'Explaining the Death Ticker' (European Environmental Bureau) <https://eeb.org/publications/61/industrial-production/1070/explaining-the-death-ticker.pdf>.

139 Mark Lee Hunter and others, 'Special Investigation: How the Common Agricultural Policy Promotes Pollution' (*The Ecologist*, 23 March 2018) <https://theecologist.org/2018/may/23/special-investigation-how-common-agricultural-policy-promotes-pollution> accessed 1 May 2020.

140 Craig Smith, 'Diageo Defiant despite Distillery Listed as One of Europe's Worst Polluters' *The Courier* (12 July 2017) <https://www.thecourier.co.uk/fp/news/local/fife/466983/diageo-defiant-despite-distillery-listed-as-one-of-europes-worst-polluters/> accessed 1 May 2020.

141 Christian Schaible, Pedro Ogando and Anton Lazarus, 'Burning the Evidence: A Case Study on Large Combustion Plants' (European Environmental Bureau 2017), 35 <https://eeb.org/library/burning-the-evidence-a-case-study-on-large-combustion-plants/> accessed 30 April 2020.

The 2016 REFIT Evaluation[142] of the E-PRTR Regulation included a stakeholder questionnaire with ca. 40 respondents, including Member State's competent authorities, industry operators and others.[143] It showed that 50% of respondents somewhat or fully disagreed with the statement that 'Data presented in the E-PRTR are complete'.[144] However, 66% of respondents asserted that users trust the E-PRTR 'very much',[145] while 68% found the E-PRTR's data *quantity* and 93% found data *quality* moderately to fully suitable.[146] Bünger provides a highly detailed overview and description of the E-PRTR's strengths and weaknesses (comparing it to the U.S. Toxic Release Inventory).[147] Further critical notes on different aspects of the E-PRTR are included throughout this chapter. These include the lack of reporting on production outputs;[148] the distorted perception of the biggest polluters due to its reporting thresholds[149] and the general delay in reporting.[150] No system is perfect, and neither is the E-PRTR. Still, valuable lessons can be drawn from the E-PRTR for the platform governance debate on how an access regime can facilitate public participation by providing information on actors holding commercially and politically sensitive data, and how it can contribute to the accountability of these actors.

3.2 Implementation

This section describes how the E-PRTR Regulation is implemented in practice. Specifically, it delineates the three different types of data the E-PRTR provides access to; how that data is generated by operators of industrial facilities; gathered by Member States and the EEA and under what conditions it can be kept confidential. Finally, it discusses how the data can be accessed by interested parties.

3.2.1 Data available in the E-PRTR

Data to report on pollutant emissions passing the thresholds

- Amount of pollutant emission into air, water and land;
- Amount of pollutant released by accident;
- Off-site transfers into waste water destined for waste-water treatment outside the facility;
- Off-site transfers of waste for recovery or disposal;
- The method of gathering the pollution data;

Figure 3 – data to be to be gathered on pollutant emissions passing the threshold

142 Additional comments highlight several critiques of respondents: inconsistencies of emissions reported for similar activities in different countries; mistakes in reported emissions; differences in reporting methods and the lack of information needed to compare environmental performance within the same activity. Amec Foster Wheeler Environment & Infrastructure UK and IEEP, 'Supporting the Evaluation of Regulation (EC) No 166/2006 Concerning the Establishment of a European Pollutant Release and Transfer Register and Its Triennial Review: Final Report.' (Publications Office of the European Union 2016) <http://op.europa.eu/en/publication-detail/-/publication/5b347a4a-9ae6-11e6-868c-01aa75ed71a1> accessed 2 May 2020 (REFIT Evaluation).

143 REFIT Evaluation, 28.

144 REFIT Evaluation, 234-35.

145 REFIT Evaluation, 270.

146 REFIT Evaluation, 235.

147 It must be noted that this book dates from 2012, and the scope of the E-PRTR's implementation has become wider since. Dirk Bünger, *Deficits in EU and US Mandatory Environmental Information Disclosure: Legal, Comparative Legal and Economic Facets of Pollutant Release Inventories* (Springer-Verlag 2012).

148 Section 3.2.1.

149 Section 3.2.1.

150 Section 3.2.3.

The E-PRTR Regulation requires operators of industrial facilities partaking in 65 specific types of economic activities[151] to report the emission of 91 different pollutants when their emission surpasses a pre-defined threshold.[152] The threshold for Carbon dioxide (CO2) release to air, for example, is 100 million kg per year. Therefore, if a facility emits more than 100 million kg CO2 into the air in a year it must report data on these emissions, for inclusion in the E-PRTR. Facilities only have to report emissions if they partake in one of the 65 designated economic activities, which are divided into 9 groups. Both the precise list of pollutants and their thresholds have been determined by the European Commission in Annexes to the E-PRTR Regulation. The EC has thus determined which data is available via the E-PRTR.

Under the E-PRTR, operators of industrial facilities are obliged to report: (a) data on pollutant emissions of their facilities,[153] (b) identifying information data on themselves;[154] and they are free to voluntarily report (c) additional data on their facility and/or operations.[155] First, the data on pollutants includes the total and accidental emissions to air, water, land; different types of transfers of the pollutant to other locations and the method for how the data was gathered (see Figure 3). Secondly, the identifying information the operator has to report on the emitting facility includes: its name, parent company, full address and geographical location and some additional information (see Figure 4). Finally, apart from these two mandatory types of information, operators may voluntarily report additional data on their facility and operations via the same reporting mechanism as the mandatory data.[156] This can be information such as their production volume, number of employees, operating hours per year or a website address.

Identifying information on industrial facilities

- Name of facility & parent company;
- (national) Identification number;
- Street address: town/village, postal code, country, geographical coordinates;
- River basin district the facility is located in;
- 4-digit NACE-code, classifying the economic activity of the facility;

Figure 4 – Identifying data of operators in the E-PRTR

Two critical notes with respect to data that does not need to be reported to the E-PRTR: Firstly, operators do not need to report their production output (e.g. the amount of energy, poultry or paper they have produced). As the E-PRTR does not require the reporting of the production output of facilities, comparing two or more facilities with similar activities is difficult.[157] If one does not know the production output: 'a single large plant may appear to have a higher release level

151 These economic activities are defined in Annex 1 of the E-PRTR Regulation. The activities are divided into nine different categories: (1) energy, (2) production and processing of metals, (3) mineral industry, (4) chemical industry, (5) waste and waste water management, (6) paper and wood production and processing, (7) intensive livestock production and aquaculture, (8) animal and vegetable products from the food and beverage sector, and (9) other activities.

152 These thresholds per pollutant are listed in Annex 2 of the E-PRTR Regulation.

153 E-PRTR Regulation, arts 5(1)-(3); For a full list of data to be reported cf. Figure 3.

154 Environmental Omnibus Regulation, art 7(1).

155 Commission Implementing Decision (EU) 2019/1741 of 23 September 2019 establishing the format and frequency of data to be made available by the Member States for the purposes of reporting under Regulation (EC) No 166/2006 of the European Parliament and of the Council concerning the establishment of a European Pollutant Release and Transfer Register and amending Council Directives 91/689/EEC and 96/61/EC (Text with EEA Relevance) [2019] OJ L267/3, Annex I.

156 ibid; European Commission, 'Guidance Document for the Implementation of the European PRTR' <https://ec.europa.eu/environment/industry/stationary/e-prtr/implementation.htm> accessed 10 June 2020 (E-PRTR Guidance Document).

157 Mahelet Getachew Fikru, 'Does the European Pollutant Release and Transfer Register Enable Us to Understand the Environmental Performance of Firms?' (2011) 21 Environmental Policy and Governance 199, 202.

than a series of smaller plants producing more pollution per unit of output.'[158] Although production output can be considered 'competition sensitive' data, its inclusion would have made the E-PRTR more useful for research and accountability purposes. Secondly, the reporting thresholds for the E-PRTR mean that small facilities have no reporting duties. This exclusion of small(er) facilities from the system 'creates an erroneous assumption that large industrial polluters are solely or mostly responsible for toxic risks faced by the public.'[159] The E-PRTR's thresholds can be criticized for facilitating 'death by a thousand cuts': many facilities with lower emissions levels can still create a harmful total amount of pollutants of which users of the E-PRTR will not be aware due to its reporting thresholds. In sum, these concerns demonstrate the importance of trade-offs between economic interests (e.g. protection of competition sensitive data) and regulatory objectives (improved comparisons of polluting facilities).

3.2.2 Methods for generating data

Not all types of pollutant emissions data are easy to produce. This is especially true, for example, with regard to accidental releases of pollutants. Therefore, operators may have different methods for producing the required pollutant emissions data. Operators of industrial facilities are obliged to use the 'best available information'[160] and follow the Commission's guidelines on quality assurance[161] when reporting their emissions data. They should use 'internationally approved methodologies' for their reporting, listed by the Commission in its E-PRTR guidance document.[162] Operators must also report which of the three types of reporting methods they used, and in some cases the specific reporting method used.[163]

Operators may collect emissions data of their facilities using three different types of methods: measurement (M), calculation (C) or estimation (E).[164] When more than one method is used to collect emissions data of a facility, the method which provides the highest amount of emissions must be reported.[165] The three methods are further described below.

- **Measurement Method** – Measurement is used when the emissions of a facility are derived from 'direct[ly] monitoring results for specific processes at the facility, based on actual continuous or discontinuous measurements of pollutant concentrations,' or from 'short term and spot measurements'.[166] In this case the operators thus report data directly based on the output of measurement equipment in their facility.[167] The European Commission provides a list of approved measurement methods.[168]

- **Calculation Method** – 'Calculation' means that yearly pollutant emissions levels are created by means of input data, which are put into an existing calculation model to compute the yearly emissions levels of a facility. For example, activity data

158 Bünger (n 147) 430.
159 ibid.
160 E-PRTR Regulation, art 5(4).
161 E-PRTR Regulation, art 9(4).
162 E-PRTR Guidance Document, 35.
163 E-PRTR Regulation, art 5(4).
164 E-PRTR Regulation, art 5(1); E-PRTR Guidance Document, 33-43.
165 E-PRTR Guidance Document, 33.
166 E-PRTR Guidance Document, 33.
167 Some calculations can be used to calculate yearly pollution totals, but the origin of the data must lay in actual measurement of pollution emissions. E-PRTR Guidance Document, 33.
168 E-PRTR Guidance Document, 36; 103-112.

of the facility '(fuel used, production rate, etc.)'[169] are used in combination with standard pollutant emission values for a specific industrial process to calculate the total yearly emissions of a facility. Again, the European Commission provides a list of approved calculation methods.[170]

- **Estimation Method** – Finally, an operator can base emissions data on non-standardized estimations. This method requires that data 'are determined by best assumptions or expert guesses that are not based on publicly available references or in case of absence of recognized emissions estimation methodologies or good practice guidelines.'[171] Thus, when no other means are available to an operator, it can ask experts to estimate emissions levels based on their best assumptions. Estimation methods are especially relevant when operators want to report accidental releases of pollutants, since accurate data on these events are not immediately available to the operator.[172]

3.2.3 Data gathering process: from facility to EEA

The gathering of the pollutant emissions data that ends up in the E-PRTR system passes through three different levels. It is collected by the operator at the facility level, then pooled on the country level by 'national competent authorities' that report the data to the EEA and EC for inclusion in the E-PRTR. Figure 5 shows a schematic high-level overview of this process. More information on the checking of data and other governance relations *between* these parties can be found in section 3.3 on the governance of the E-PRTR access regime.

- **Role of Operators of Industrial Facilities** – Operators of industrial facilities have to report the three types of data to the E-PRTR (cf. section 3.2.1) on a yearly basis. In the 2016 REFIT evaluation, nine Member States reported operators missing deadlines for reporting required data to national competent authorities. This was mainly due to operators'

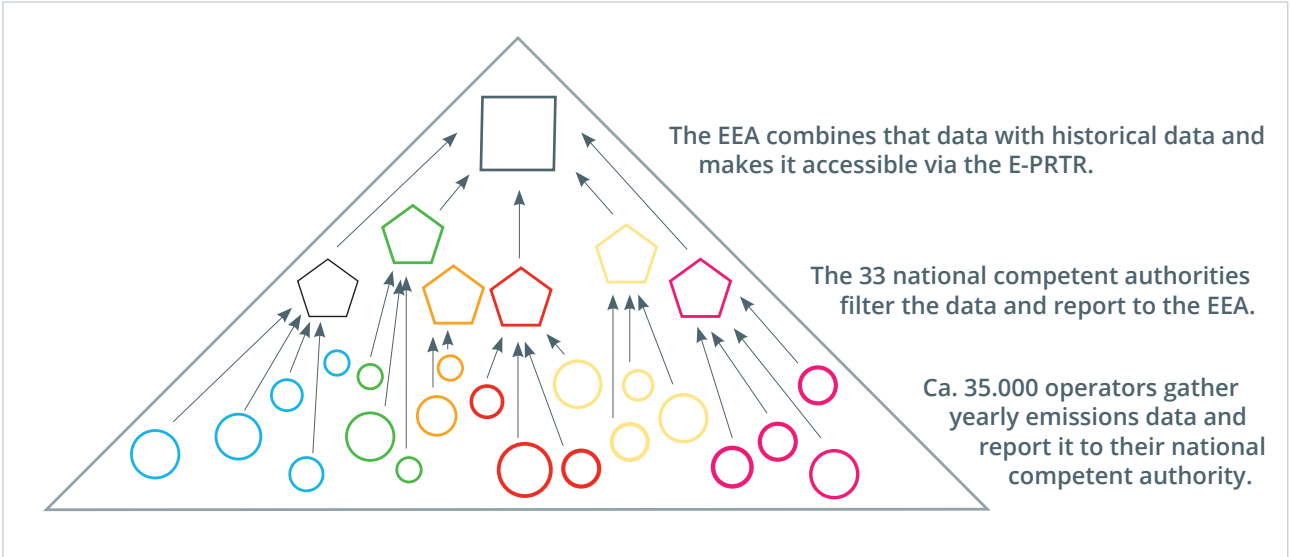


Figure 5 – Data gathering process for the E-PRTR

169 E-PRTR Guidance Document, 33.
170 E-PRTR Guidance Document, 36; 103-112.
171 E-PRTR Guidance Document, 33.
172 E-PRTR Guidance Document, 16.

uncertainties on how to report emissions data of their facilities or technical issues.[173] In almost all Member States, operators can report the data to the E-PRTR using an electronic reporting tool of the national competent authority.[174] Once reported to the competent authorities, operators must save their reporting data for five years.[175]

- **Role of National Competent Authorities** – The National Competent Authorities of the members to the E-PRTR gather the data from all operators of industrial facilities in their respective countries. Depending on the Member State, there might be one competent authority or several working together. In most Member States (23) the responsibility for the implementation of the E-PRTR Regulation is shared between several national authorities. In 15 Member States, the ministry responsible for environmental policy is co-responsible for this implementation, in 14 Member States one or more (regional) environmental agencies are. The competent authorities combine the data of all industrial facilities in their countries into one dataset. In doing so they can also claim some data to be confidential.[176] Competent authorities have to report the dataset with country-wide data to the EEA in a format and by a date to be established by the Commission by means of implementing acts, but not later than 11 months after the end of the reporting year.[177]

- **Role Of The EEA** – Finally, the EEA takes in all the data from the datasets reported by the national competent authorities. These datasets are combined with the historical data of earlier years and integrated into the E-PRTR dataset. This processing has to be completed within one month after the completion of reporting by the Member States.[178] The Commission (through the EEA) also provides a data validation tool to the Member states, which they may use to check the quality of the data they report to the EEA. The EEA has also created and now maintains the E-PRTR website and registers to provide access to E-PRTR data to the public.[179]

- **Validation Tool** – The use of the validation tool created by the EEA is not mandatory for Member States. It only assists them in quality assurance and control, which they *are* obliged to carry out.[180] The validation software (including a user manual)[181] is freely available for download.[182] The software checks the 86 variables that operators can report using the E-PRTR (of which 45 are mandatory to report) for fourteen mandatory requirements; four additional requirements and fourteen complementary requirements.[183] Examples of errors the validation tool can detect are 'incorrect co-ordinates, wholly incorrect figures, pollutants reported twice and facilities with no reported releases.'[184]

---

173 REFIT Evaluation, 27; 101.

174 REFIT Evaluation, 27 (In 2016, Denmark, Finland and Sweden did not allow paper reporting. Competent authorities in Slovenia and Greece did not have an electronic reporting tool and data was to be reported in hardcopy (paper)).

175 E-PRTR Regulation, art 5(5).

176 Section 3.2.5.

177 Environmental Omnibus Regulation, art 7(2).

178 Environmental Omnibus Regulation, art 7(2) (In other words, the EEA has one month to process the reporting data it receives from Member States).

179 E-PRTR Regulation, art 10(1).

180 E-PRTR Regulation, art 9(2).

181 Atkins Danmark, GIS & IT and Tripledev, 'E-PRTR Validation Tool – User Manual Version 3.0' <https://www.eionet.europa.eu/schemas/eprtr/EPRTRUserManual.pdf> (E-PRTR Validation Tool Manual).

182 European Commission, 'European Commission E-PRTR Validation Tool' <https://www.eionet.europa.eu/schemas/eprtr/validation-tool> accessed 7 May 2020.

183 A detailed description of all the checks the validation took carries out can be found in the E-PRTR Validation Tool Manual.

184 E-PRTR Guidance Document, 53.

The reporting deadlines set by the E-PRTR Regulation are cause for important concern, as data can be quite outdated by the time they are finally reported. Pollutant emissions taking place on e.g. the 1st of January 2020 may only be appear in the E-PRTR on the 1st of June of 2022. It can thus take more than two years before relevant stakeholders can access the relevant data. The European Environmental Bureau (an NGO), has suggested that operators should be able to submit pollution data measured by continuous monitoring devices at a much higher frequency, maybe even on a daily basis.[185]

3.2.4 Two types of access to E-PRTR data

The EEA makes the E-PRTR data accessible both via a website and as a downloadable dataset.[186] Both of these types of access are available online to any interested party. There are no requirements to register or file specific access requests in order to use the website, nor to download the entire dataset. The EEA has also made available a summary dataset with the data they consider most important. The website offers the opportunity to find facilities via a map interface or to filter data based on self-selected criteria such as emissions per geographical area, industrial activity or type of pollution.

Additionally, the EEA Enquiry Service provides assistance in accessing E-PRTR data[187] by answering questions about the E-PRTR data on its EEA forums.[188] For example, to clarify whether specific data is included in the register,[189] or whether and how E-PRTR data can be linked to other datasets.[190] The EEA answers questions as soon as possible, but no later than 15 working days.[191] While posts are public by default, a user can opt to ask questions to the EEA privately (without publication) via the forum, which makes the overall use of the forum difficult to assess.

3.2.5 Claiming confidentiality

The national competent authorities can decide to grant confidentiality for both data on pollutant emissions and identifying information of an operator itself. While national authorities make the final decision on confidentiality, operators must provide information in order to process a confidentiality claim (e.g. the legal grounds for confidentiality) when they report their pollutant emissions data.[192] Importantly, the operator must always share complete emissions data with the national authorities, even if this data is later decided to be kept confidential. In that case, the data is then not passed on to the EEA and E-PRTR.

---

185 Schaible, Ogando and Lazarus (n 141).

186 The web-platform is located at: European Environment Agency, 'E-PRTR' (*E-PRTR*) <https://prtr.eea.europa.eu/#/home> accessed 9 June 2020; the full dataset can be viewed and downloaded at: European Environment Agency, 'The European Pollutant Release and Transfer Register (E-PRTR), Member States Reporting under Article 7 of Regulation (EC) No 166/2006' (*European Environment Agency*, 6 February 2020) <https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-23> accessed 15 May 2020.

187 As it is obliged by the E-PRTR Regulation, arts 10(2) & 15.

188 European Environment Agency, 'EEA FORUM' (*EEA FORUM*, 15 May 2020) <https://community.eea.europa.eu/search?SearchableText=e-prtr&x=0&y=0> accessed 15 May 2020.

189 vvilla, 'Hi All. Animal by-Products and Derived Products Not Intended for Human — EEA FORUM' <https://community.eea.europa.eu/home/environmental-topics/air-emissions/hi-all.-animal-by-products-and-derived-products-not-intended-for-human/?searchterm=E-PRTR> accessed 2 May 2020.

190 Robin Sogalla, 'Dear EEA Team,

I Would like to Analyze the Emissions of Air — EEA FORUM' <https://community.eea.europa.eu/home/environmental-topics/air-emissions/dear-eea-team-br-br-i-would-like-to-analyze-the-emissions-of-air/view?searchterm=E-PRTR#1571320782> accessed 2 May 2020.

191 European Environment Agency, 'EEA Forum Quick User Guide' <https://community.eea.europa.eu/home/Forum_manual_by_EEA.pdf>.

192 E-PRTR Guidance Document, 55.

A confidentiality claim must be based on one of eight reasons considered legitimate by the E-PRTR Regulation.[193] These reasons include i.a.: public security, due course of justice, protection of intellectual property rights and the confidentiality of personal data (see Figure 6 for the full list). In general, all eight grounds for confidentiality can be invoked to withhold any type of information reported by operators. There is an exception for actual data on emission/releases (amount and type of pollutant emitted/released), these can only be withheld for reasons of public security, the due course of justice or intellectual property rights (grounds listed in italics in Figure 6).[194] A granted claim for confidentiality for emissions data does not automatically mean that the respective fields are left blank in the eventual Member State report to the EEA. If possible, the data is reported in a more generalized manner: for example, the group name of a pollutant is reported instead of the specific name of the pollutant itself. A Member State must always report the specific legal ground which made it decide to keep data confidential.[195]

Eight Member States have claimed confidentiality for pollutant data over the period 2010-2013. Belgium and Germany claimed confidentiality the most (BE: for 128 facilities in 2012, DE for 32 facilities in 2010), while Luxembourg claimed confidentiality the least (twice in 2012).[196] Most confidentiality claims are made for data on offsite transfer of waste or waste water.[197] The most cited legal ground for confidentiality was the commercial or industrial sensitivity of data,[198] although Belgium also claimed confidentiality for reasons of personal data protection for ninety poultry farms in Flanders.[199] The REFIT Evaluation provides an explanation for all confidentiality claims in the period 2010-2012 can under its 'Member State summary overview'.[200] Altogether the total number of confidentiality claims (320, 1,07%) is rather low in light of the total amount of facilities included in the E-PRTR in this period (30.000).

The EEA and EC leave it to national competent authorities to check the validity of individual confidentiality claims.[201] The EEA does monitor the total amount of confidentiality claims made in the yearly submissions by national competent authorities. If confidentiality is claimed for more than 10% of its reported data, the DG Environment of the Commission will enter into a dialogue with a Member State to discuss possible overuse of confidentiality claims. Ideally, confidentiality claims are not made for more than 5% of the data a national competent authority submits.[202]

193 E-PRTR Regulation, art 11; Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information [2003] OJ L41/26, art 4(2) (Public Access to Environmental Information Directive).

194 E-PRTR Guidance Document, 55.

195 E-PRTR Guidance Document, 55-56.

196 REFIT Evaluation, 109-110 (All Member States claiming confidentiality (with total number of claims in period 2010-2012): Belgium (128x), Bulgaria (19x), Denmark (15x), Germany (115x), Ireland (6x), Luxembourg (2x), Romania (6x) and the UK (29x)).

197 REFIT Evaluation, 109-110 (Figure 5.6: Number of facilities affected by confidentiality claims during reporting period).

198 REFIT Evaluation, 116 ('Belgium, Member State Summary').

199 REFIT Evaluation, 121 ('Summary of Member State Response'); E-PRTR Guidance Document, 121 (These confidentiality claims for protection of personal data probably relate to the fact that these facilities are located at the home address of the company owners (farmers living on their farm). Sometimes companies carry family names, which would mean that the individual owners' name and home address would be published in the E-PRTR. This can be avoided through a confidentiality claim.).

200 REFIT Evaluation,112-151 ('Summary of Member State Response').

201 E-PRTR Guidance Document, 55.

202 These thresholds are laid down in a quality assurance guidance documents for the new EU Registry on Industrial Sites, which uses E-PRTR data among other sources. European Environment Agency, 'Quality Assurance Logic EU Registry on Industrial Sites – Document for Users – Version 5.0' (European Topic Centre for Air pollution, Transport, Noise and Industrial Pollution (ETC/ATNI) 2020) <https://cdr.eionet.europa.eu/help/euregistry/Documents/QAQC%20Master%20Document_CID_V5_January2020.pdf>.

Legitimate grounds for a confidentiality claim

- The confidentiality of the proceedings of public authorities, where such confidentiality is provided for by law;
- *International relations, public security or national defense;*
- *The course of justice, the ability of any person to receive a fair trial or the ability of a public authority to conduct an enquiry of a criminal or disciplinary nature;*
- The confidentiality of commercial or industrial information where such confidentiality is provided for by national or Community law to protect a legitimate economic interest, including the public interest in maintaining statistical confidentiality and tax secrecy;
- *Intellectual property rights;*
- The confidentiality of personal data and/or files relating to a natural person where that person has not consented to the disclosure of the information to the public, where such confidentiality is provided for by national or Community law;
- The interests or protection of any person who supplied the information requested on a voluntary basis without being under, or capable of being put under, a legal obligation to do so, unless that person has consented to the release of the information concerned.
- The protection of the environment to which such information relates, such as the location of rare species. oth the E-PRTR website and the E-PRTR downloadable dataset provide insight into how data has been affected by confidentiality claims. The E-PRTR website shows a yellow notification bar for every page that has been affected by a confidentiality claim. The E-PRTR database includes three variables that indicate confidentiality and display the code and name for the grounds for confidentiality As such, it is clear to users of E-PRTR services which information is incomplete due to confidentiality.

Figure 6 – Confidentiality grounds (only the grounds listed in italics can be relied on to keep emission/release data confidential)

Both the E-PRTR website and the E-PRTR downloadable dataset provide insight into how data has been affected by confidentiality claims. The E-PRTR website shows a yellow notification bar for every page that has been affected by a confidentiality claim. The E-PRTR database includes three variables that indicate confidentiality and display the code and name for the grounds for confidentiality.[203] As such, it is clear to users of E-PRTR services which information is incomplete due to confidentiality.

203 European Environment Agency, 'Reported Information under Regulation (EC) No 166/2006 on the Establishment of a European Pollutant Release and Transfer Register Information on the Database Structure and Use', 7 <https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-23/database-structure-and-use-information/eprtr_database_metadata_v11.pdf-1/at_download/file> accessed 2 May 2020.

3.3 Governance

This section describes the governance structure of the E-PRTR. For this purpose, five main parties can be discerned: (a) the operators of industrial facilities, (b) Member State's national competent authorities, (c) the EEA, (d) the European Commission and (e) the European Parliament and Council. Figure 7 provides an overview of the governance relations between these parties. The following section first discusses the responsibilities and obligations of these parties; followed by the sanctions & enforcement, liability and funding arrangements.

3.3.1 Operators of industrial facilities

Operators of industrial facilities have two main obligations under the E-PRTR. First, operators are obliged to report their pollutant emissions on a yearly basis.[204] Secondly, operators are obliged to assure the quality – i.e. completeness, consistency and credibility – of the pollutant data they report.[205] *Completeness* denotes that all releases and off-site transfers of pollutants and wastes that exceed their thresholds and all information to identify a facility (name, address etc.) are reported without omissions. *Consistency* is attained through the use of the same definition and methodologies in reporting over several

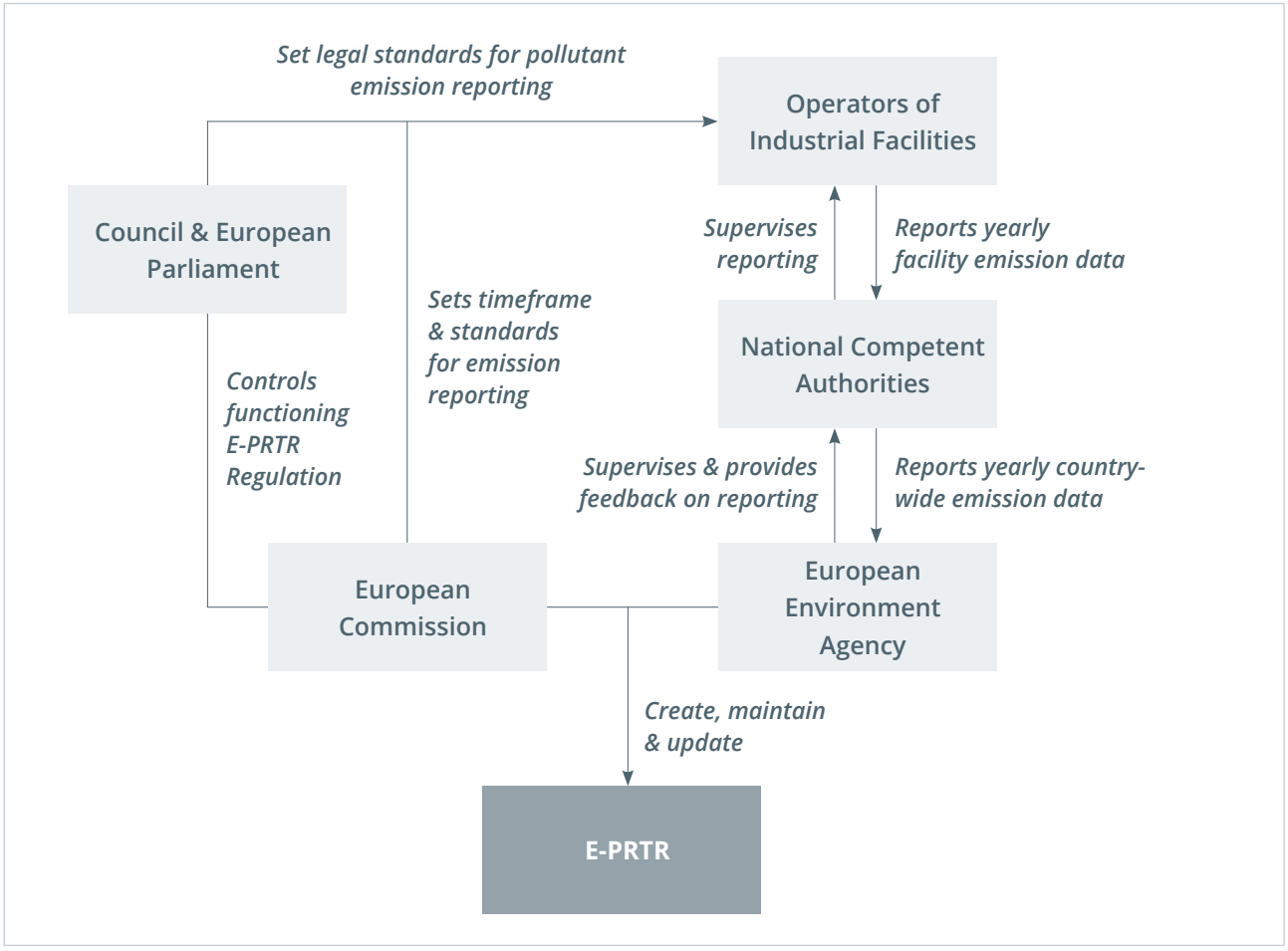


Figure 7 – E-PRTR Governance structure

204 Section 3.2.1 and 3.2.2.
205 E-PRTR Regulation, art 9.

years. Finally, *credibility* is achieved by assuring that the reported data is authentic, transparent and reliable and comparable, because of consistent reporting.[206] Operators are advised to use the best available reporting techniques to achieve these standards of completeness, consistency and credibility.[207]

3.3.2 National competent authorities

National competent authorities have three main obligations. They are obliged to (a) gather and combine the reported data for their jurisdiction on a yearly basis; (b) assure its data quality and (c) report this data to the EEA. Competent authorities may issue sanctions and penalties to operators to ensure compliance with their reporting obligations both in the quality of reported data and timeliness of reporting.[208] The data quality must be assured by competent authorities through checks for the completeness, consistency and credibility of data reported by industrial facilities.[209] How competent authorities achieve this quality assurance differs per Member State, for example, through verification by experts, facility visits by environmental supervisory authorities, or through comparison of reported data with similar facilities. Member States often have a slightly different method to assure data quality.[210] Thirdly, competent authorities are obliged to report the data of the facilities in their country to the EEA within at least 11 months after the end of the reporting year.[211]

Apart from the obligations directly related to data disclosure, Member States have two additional duties under the E-PRTR Regulation. First, they have to raise awareness for the E-PRTR on the national level.[212] Member States do so in different ways. Some only link to the E-PRTR from their national PRTR's website. On the other side of the spectrum, the Irish Environmental Protection Agency has created an 'Environmental Queries Unit' which answers the public's questions about the environment.[213] Secondly, Member States' competent authorities check whether the confidentiality claims of operators are legitimate.[214] Before 2019,[215] Member States were also obliged to create a triennial report for the European Commission on the implementation of the E-PRTR in their country, which provided a survey of several topics covered by the E-PRTR regulation.[216] Although two of these reports were delivered in 2013 and 2017,[217] they were 'considered of limited value and/or [did] not meet policy needs' and therefore were removed by the 2019 amendments to

206 E-PRTR Guidance Document, 47-48.
207 E-PRTR Guidance Document, 47.
208 Section 3.3.6.
209 E-PRTR Regulation, art 9(2).
210 REFIT Evaluation, 104-108 (Table 5.8: Processes for verification of completeness, consistency and credibility of data reported by operators to competent authorities).
211 Environmental Omnibus Regulation. art 7(2).
212 E-PRTR Regulation, art 15.
213 REFIT Evaluation. 112-151 (Summaries of Member State responses, subsections on 'public awareness')
214 Section 3.2.5.
215 In 2019 the E-PRTR Regulation was amended by the Environmental Omnibus Regulation, removing Article 16 from the E-PRTR Regulation.
216 This report had to include information on (a) the reporting of facilities; (b) quality assurance and assessment; c) access to information; (d) awareness raising activities; (e) confidentiality of information, and (f) penalties and experience with their application; E-PRTR Regulation, art 16(1).
217 European Commission, 'Report from the Commission to the European Parliament and the Council on progress in implementing Regulation (EC) 166/2006 concerning the establishment of a European Pollutant Release and Transfer Register (E-PRTR)' COM(2013) 111 final; European Commission, 'Report from the Commission to the European Parliament and the Council on progress in implementing Regulation (EC) 166/2006 concerning the establishment of a European Pollutant Release and Transfer Register (E-PRTR)' COM(2017) 810 final.

the E-PRTR Regulation to avoid 'excessive administrative burden' on the Member States.[218]

3.3.3 European Environment Agency (EEA)

The European Environment Agency has two main obligations: (a) it assists the EC in checking the data reported by national competent authorities and (b) it publishes this data in the E-PRTR. The EEA assists the EC in checking the E-PRTR data that national authorities report through yearly automated and manual checks by experts.[219] For an informal review of E-PRTR data covering 2007–2009, the EEA involved three of its European Topic Centers, which are consortia of organizations with expertise in specific environmental areas contracted by the EEA to support its work.[220] These three Topic Centers provide detailed feedback to national authorities on the quality of the E-PRTR data they reported. This feedback included an evaluation of the number of facilities and release reports, amounts of releases and transfers reported, confidentiality claims, accidental releases and more. The main errors and gaps in the data reported by national authorities were identified and published by the EEA.[221] The EEA also provides a validation tool for national authorities to pre-check the data they are reporting.[222] This is how the EEA assists the EC in checking the quality of E-PRTR data. Secondly, the EEA maintains the E-PRTR website and E-PRTR database. The data of national competent authorities is stored and processed at the EEA's ReportNet site; combined with legacy data and finally published in its entirety on the E-PRTR website.[223]

3.3.4 European Commission

The European Commission has two main obligations under the E-PRTR Regulation. First, it sets the legal standards for pollutant reporting and therefore determines the scope of the E-PRTR. It does so through the annexes of the E-PRTR Regulation. Annex I lists the economic activities for which operators of industrial facilities must report the pollutant emissions. Annex II sets the maximum thresholds for these pollutant emissions. Importantly, the Commission sets the European Economic Area's *minimal* reporting standards. This means that Member States are free to lower thresholds and require operators to report on additional industrial activities, which means that operators also have to report lower levels of emissions.[224]

Secondly, the Commission had to provide a guidance document for both operators of industrial facilities and Member State's competent national authorities on the concrete implementation of the E-PRTR Regulation.[225] This document was issued in 2006 and is

218 European Commission, 'Proposal for a regulation of the European Parliament and the Council on the alignment of reporting obligations in the field of environment policy and thereby amending Directives 86/278/EEC, 2002/49/EC, 2004/35/EC, 2007/2/EC, 2009/147/EC and 2010/63/EU, Regulations (EC) No 166/2006 and (EU) No 995/2010, and Council Regulations (EC) No 338/97 and (EC) No 2173/2005' COM(2018) 381 final, 11.

219 Eva Krtková and others, 'E-PRTR Data Review Methodology – Update 2019' (European Environment Agency – European Topic Centre on Air pollution, transport, noise and industrial pollution) Eionet Report ETC/ATNI 2019/5 <https://www.eionet.europa.eu/etcs/etc-atni/products/etc-atni-reports/etc-atni-report-5-2019-e-prtr-data-review-methodology-update-2019>.

220 European Environment information and Observation Network (Eionet), 'European Topic Centres' (European Topic Centres) <https://www.eionet.europa.eu/etcs> accessed 12 June 2020 (As of January 1st 2019, there are seven European Topic Centres).

221 See, for example: E-PRTR, 'E-PRTR Data Completeness and Errors' <https://prtr.eea.europa.eu/docs/Errors%20and%20emissions%20disclaimer%20Oct2011.pdf> (Listing the errors made by Germany in its 2011 reporting).

222 Section 3.2.3.

223 E-PRTR Guidance Document, 63.

224 Recital 21, E-PRTR Regulation ("[T]he provisions of this Regulation should not affect the right of the Member States to maintain or introduce a more extensive or more publicly accessible pollutant release and transfer register than required under the [Kyiv] Protocol,").

225 E-PRTR Regulation, art 14,

still in use today.[226] In 2018 the EC opened a tender for the review and updating of this guidance document.[227]

Until 2019,[228] the EC was also obliged to perform an additional review for the pollutant data provided by the competent national authorities, and publish a report on that review every three years, within six months after online publication of the E-PRTR data.[229] It was assisted by the EEA to comply with this obligation. In the same report, the EC had to assess the operation of the entire E-PRTR access regime and report on that assessment to the European Council and European Parliament.[230] This obligation was repealed by an amendment in 2019.[231]

3.3.5 European Parliament and Council

The European Parliament and the Council have a limited role in the E-PRTR access regime. Before 2019 they received the triennial report of the EC on the functioning of the system.[232] Nowadays, they can still evaluate the E-PRTR's functioning, the EC's implementing decisions on it, and propose changes or additions to the E-PRTR Regulation according to the regular legislative procedures. Other than this monitoring role, the Parliament and Council are more like outside parties that use E-PRTR data for (evaluating) policy making.

3.3.6 Sanctions and enforcement

The E-PRTR Regulation requires Member States to lay down rules on penalties for non-compliance with the E-PRTR access regime in their national laws.[233] Sanctions and their enforcement are thus a national matter. Member States do have to notify the Commission of the rules they implement in this area.[234] This subsection provides examples of offences that can be sanctioned, and discusses the broad range of administrative fines and criminal penalties adopted by the Member States. Then, it discusses the sanctions that have been imposed in the reporting period 2010-2013 as well as the lack of E-PRTR specific options for sanctioning Member States for non-compliance with their obligations under the E-PRTR.

Member States have adopted different types of sanctions for non-compliant operators under the E-PRTR. Many Member State simply list 'non-compliance' with regulations as the ground for sanctions. Others list more specific offences, such as: missing or non-reporting (France and Italy), incomplete or inaccurate reporting (idem), failure to report accidents or reporting of false information (Lithuania) or late submission of reporting (Sweden).[235]

Abovementioned offences are sanctioned differently within Member States. For example, Lithuania has the lowest fines (€29 – €58) for a failure to report information, while Italy fines €5000-€52.000 for the same offence. Some countries have different fines for

226 European Commission, E-PRTR Guidance Document (n 156).

227 European Commission, 'SERVICE REQUEST – ANNEX "Specific Terms of Reference": Review of E-PRTR Implementation and Related Guidance' <https://ec.europa.eu/environment/industry/stationary/e-prtr/pdf/terms_of_reference_external_use.pdf>.

228 The Environmental Omnibus Regulation amended the E-PRTR Regulation and repealed article 17.

229 E-PRTR Regulation, art 17(1).

230 E-PRTR Regulation, art 17(2).

231 Environmental Omnibus Regulation, art 7(4); European Commission, 'Proposal on the alignment of reporting obligations in the field of environment policy' (n 218).

232 E-PRTR Regulation, art 17.

233 E-PRTR Regulation, art 20(1).

234 E-PRTR Regulation, art 20(2).

235 REFIT Evaluation, 97-98 (Table 5.6 Level of fine reported for non-compliance),

different categories of facilities (e.g. Belgium, Greece), while others fine differently for air, water or land pollution (e.g. Cyprus, Estonia, Hungary). The fines range from the €29 of Lithuania to a maximum of €250.000 in Belgium. Some countries can also impose imprisonment (from min. eight days in Luxembourg to 2 years in the U.K.).[236] All in all, both the type and severity of penalties differ widely between Member States.[237]

The number of sanctions that Member States imposed in the period 2010-2013, also differs widely. Sweden (211), Poland (127) and Belgium (111) issued the most fines for non-compliance. Austria (1) and France ('very few') punished the least, together with eight Member States that have not imposed any sanctions.[238] Ireland and the Netherlands reported that they initiated enforcement procedures for non-compliance, but that these threats of penalties were often enough for operators to comply with their reporting obligations.[239] This amount of sanctions is low compared to the total number of facilities (30.523 – 31.677) and the total number of reports to be filed during the reporting period of four years (ca. 124.000).

The available sanctions and enforcement measures for non-compliance with the E-PRTR Regulation (as well as how frequently they are actually imposed) differ wildly per Member State. This is understandable from the perspective of subsidiarity: it gives Member States the opportunity to pass sanctions and penalties that are adequate their national context. However, environmental pollution is a negative externality that can not only be felt by a single Member State, but by many. Therefore, some form of minimal sanctions, or at least a prescribed list of offences to be sanctioned on the Member State level (e.g. non-reporting, insufficient reporting, false reporting) would make enforcement of E-PRTR compliance more uniform throughout the EU.

The E-PRTR Regulation does not lay down specific sanctions or enforcement measures against Member States' national competent authorities that do not comply with their reporting obligations. Even more so, the E-PRTR Regulation does not mention non-compliance by Member States at all. The Commission notes in its 2017 triennial review of the E-PRTR that: '[It] had to take follow-up action to prompt certain Member States to submit their [yearly submissions to the E-PRTR]. However, these isolated cases were dealt with rapidly and the Commission has not pursued formal infringement proceedings.'[240] Since there is no formal procedure for these 'follow-ups' it is unclear which specific Member States took more time to comply with E-PRTR obligations.[241] This quote suggests that the only formal sanction for non-compliance with the E-PRTR by a Member State is an official infringement procedure.[242]

---

236 ibid (Countries that have criminal punishment for non-compliance with the E-PRTR are: Belgium, Cyprus, Germany, Luxembourg, the Netherlands and the UK).

237 For a complete overview, see: REFIT Evaluation, 97-98 (Table 5.6 Level of fine reported for non-compliance).

238 REFIT Evaluation, 99 (These Member States are Denmark, Finland, Hungary, Italy, Luxembourg, Malta, Romania and Spain).

239 REFIT Evaluation, 99.

240 Commission (EC), 'Report from the Commission to the European Parliament and the Council on progress in implementing Regulation (EC) 166/2006 concerning the establishment of a European Pollutant Release and Transfer Register (E-PRTR)' COM (2017) 810 final.

241 Nor the REFIT Evaluation, nor the triennial reports of the Commission clarify which Member States this were.

242 Possible reasons to start an infringement procedure against a Member State for non-compliance with the E-PRTR Regulation could be: (a) an excessive use of confidentiality claims to keep pollutant emissions data secret; (b) consistently reporting country datasets too late, after the 15 month deadline, or (c) national authorities consistently reporting false pollutant data. Although currently not applicable, these scenarios are not unthinkable and therefore it is interesting to note that the E-PRTR Regulation does not to take Member State non-compliance into account. For an introductory explanation of the Commission's formal infringement proceedings against Member States for non-compliance with EU-law, see: 'Infringement Procedure' (European Commission) <https://ec.europa.eu/info/law/law-making-process/applying-eu-law/infringement-procedure_en> accessed 15 May 2020.

3.3.7 Liability

The data gathering process combined with the sanctions and enforcement measures described above suggest that the E-PRTR contains relatively high-quality data. However, the E-PRTR makes no such claims and only refers to the European Union legal notice. This notice states that 'the Commission accepts no responsibility or liability whatsoever with regard to the information on this site.'[243] Therefore, the EC seems to disclaim all liability for the E-PRTR data once published.

However, the notice also claims that 'This disclaimer is not intended to limit the liability of the Commission in contravention of any requirements laid down in applicable national law nor to exclude its liability for matters which may not be excluded under that law.'[244] The E-PRTR website thus falls under the general liability regime used for the publication of the EC's information on the internet.

3.3.8 Funding

The funding for the E-PRTR can be split up between the funding for the gathering and quality assessment by operators and competent authorities; and the funding for the quality assessment, processing and publication of the data by the EEA and EC. Industrial facilities incur personnel costs to contribute emissions data to the national PRTR's and E-PRTR: the REFIT Evaluation provides an estimate of 0.015 FTE (220 hours per year) based on a public consultation of operators. The Netherlands reports an aggregate cost of €12 million for all its operators to provide the necessary data.[245] It must be noted that the E-PRTR's thresholds exclude many small and medium size enterprises from reporting, which could incur higher costs to keep up with reporting requirements as they can rely less on existing reporting systems.[246]

The costs for national authorities to gather data, assure data quality and report to the EEA fall to the Member States. Some Member States reported start-up costs for national authorities to range between €130.000 to €1-€2 million.[247] Costs to maintain the national systems differ: Spain estimates costs at €150.000 – €170.000 to maintain both its national and the European PRTR, while the Netherlands spends €970.000 on the system and €1.2 million on the national competent authorities. These figures provide a slightly distorted image, as the costs reported by the Dutch cover their entire country, while the Spanish only covers the federal expenses and not those incurred by authorities in the regions.[248] However, Member States would have incurred these set-up and maintenance costs also without the E-PRTR, as they are obliged under the Kiev Protocol to maintain a national PRTR.[249]

Lastly, the EEA and EC incur some E-PRTR specific costs. The E-PRTR website and the preparation of its data are funded by the EEA, but how much the EEA

243 European Commission, 'Legal Notice' (European Union, 16 June 2016) <https://europa.eu/european-union/abouteuropa/legal_notices_en> accessed 7 May 2020.

244 ibid.

245 REFIT Evaluation, 247.

246 REFIT Evaluation, 46 ('For example, a large chemical company with 230 installations has an integrated environmental management system, so the actual cost of data management for PRTR for one facility is low. Other companies might have more disaggregated systems, which would lead to higher costs.').

247 REFIT Evaluation, 55; 247-48.

248 REFIT Evaluation, 45.

249 Section 3.1; REFIT Evaluation, 45 (The Kiev Protocol to the Aarhus Convention stipulates the introduction of Pollutant Release and Transfer Registers for all parties to the protocol.).

spends on the E-PRTR is unknown.[250] The European Commission says that its reporting tasks related to the E-PRTR are carried out by 1 FTE staff per year, which costs circa, €150,000.[251]

3.4 Accountability

The notion of accountability is notoriously vague. Indeed, accountability's '[e]vocative powers make it also a very elusive concept because it can mean many different things to different people.' [252] To provide some clarity, this section first provides a concise definition of accountability, discerns two types of accountability in particular (procedural and substantive accountability), and directly applies these concepts to the E-PRTR.

Accountability is 'the obligation to explain and justify conduct.' [253] In the context of the E-PRTR access regime and this Report more broadly, accountability can be defined as the ability to check the compliance of operators (of industrial facilities) with national and European (environmental) laws and obligations applicable to them. The forum to which industrial facilities are accountable to through the E-PRTR consists of a diverse set of actors: national supervisory authorities; local, national and supranational policy makers, and in the end the general public. The E-PRTR facilitates the accountability of industrial facilities through the provision of their pollutant emissions data to journalists, NGOs, scientific researchers and environmental agencies to inform the public, ask questions and pass judgement on the conduct of the industrial facilities in their neighborhood and country.[254]

For the purposes here, it is useful to further distinguish between *substantive* and *procedural* accountability. In simple terms, substantive accountability entails measures answering *what* questions, whereas procedural accountability relates to *how* questions. Measures that provide *substantive accountability* answer the question: 'What was the result?' They allow verifying whether a law or obligation has been complied with or not. For example: did an industrial facility report its yearly emissions of pollutants according to EU regulations yes or no? *Procedural accountability*, on the other hand, is provided by measures that enable answering the question: 'How has this result been achieved?' They explain and justify the process to reach compliance. For example, an industrial facility reports its methodology to measure its yearly pollutant emissions. This report looks at how accountability is operationalized in the E-PRTR context by determining which type of accountability its different components provide: substantive accountability, procedural accountability, or both.

The E-PRTR mainly provides accountability to an external forum, notably including policy makers, NGOs, journalists and national supervisory authorities, as well as the general public more broadly. This *external* accountability, enables *a priori* any stakeholder outside an industrial operator's company structure to hold that company accountable with regard to its pollution emissions (e.g. in light of environmental protection legislation). The E-PRTR provides an additional dimension of accountability, which can be called 'tiered-accountability'. This entails the fact that every party in their respective tier (or level) in the E-PRTR's reporting structure (operators,

250 The EEA does not specify how much it spends on the data gathering and management of the E-PRTR system in its yearly budgets. European Environment Agency, 'EEA Budgets' (*European Environment Agency*) <https://www.eea.europa.eu/ds_resolveuid/6675272a4c594cc0912114008f35dd17> accessed 11 June 2020.

251 REFIT Evaluation, 45.

252 Mark Bovens, 'Analysing and Assessing Accountability: A Conceptual Framework' (2007) 13 European Law Journal 447, 448.

253 ibid 450; Another definition can be found in: Jerry Mashaw, 'Accountability and Institutional Design: Some Thoughts on the Grammar of Governance' in Michael Dowdle (ed), *Public Accountability: Designs, Dilemmas and Experiences* (Cambridge University Press 2006)

254 Bovens (n 252) 447 ('Accountability is a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences.').

national competent authorities, EEA, EC), is accountable to the next tier. This creates a form of hierarchical quality control, as every 'watcher' is being 'watched' by the next tier in the hierarchical reporting chain.[255] This tiered-accountability of the E-PRTR is further discussed in section 3.4.4.

In the context of the E-PRTR access regime, accountability can be defined as the obligation of operators of industrial facilities to explain and justify their compliance with national and European environmental laws and obligations applicable to them. Operators are accountable both on a substantive and procedural level to external actors. With this in mind, we can now evaluate the individual elements of the E-PRTR access regime that enable this accountability.

For the purposes here, we focus on three main characteristics of the E-PRTR access regime that create the accountability of operators of industrial facilities covered by it. These three characteristics are: (1) the accessibility of the E-PRTR data; (2) its large scope, and (3) the reporting of confidentiality claims. Finally, the E-PRTR's extra level of tiered-accountability it discussed.

3.4.1 Public accessibility of E-PRTR data

The data within the E-PRTR is publicly accessible to anyone that wishes to access it. All pollutant emissions data is freely accessible via the EEA's website and can easily be viewed and parsed through a dedicated interface. Expert users can also download the complete dataset via the EEA website and analyze it using any software they want. No specific registration procedures or access requests are necessary.[256]

This accessibility of E-PRTR data creates both procedural and substantive accountability. Any interested party can check how much pollution an industrial facility has emitted in previous years and how that data was collected. It also opens up pollution data to a wide array of parties, such as policy makers, NGOs, journalists to see whether an industrial facility complies with laws, obligations, or promises it has made itself.

3.4.2 Large scope of the E-PRTR

The E-PRTR has a large scope: it covers a wide variety of industrial facilities and provides detailed (granular) data per industrial facility. The E-PRTR's data covers all 27 Member States of the EU, the members of the European Free Trade Association (Iceland, Liechtenstein, Norway, Switzerland) and the United Kingdom and Serbia. Since the E-PRTR is partly based on a Protocol to an international convention,[257] there are even more similar systems worldwide. This coverage enables external substantive accountability of industrial facilities, as it provides the parties using the E-PRTR with many extra opportunities to compare and benchmark the pollutant emissions of industrial facilities. They can do so not only locally, but also nationally, and demand explanations and justifications for higher pollution levels.

E-PRTR data is quite granular as well, gathered at the facility level and hence enabling substantive accountability as well. If data would be gathered only by national competent authorities on a national level or only by the EEA on the European level through occasional audits of industrial facilities, the E-PRTR would not provide accountability for the individual operators of industrial facilities. The high granularity of the E-PRTR data means that even comparisons between facilities of the same operator are possible. This provides accountability to external parties, such

255 See also Figure 5 in Section 3.2.3, which shows this hierarchical, tiered structure.
256 Section 3.2.3.
257 Section 3.1.

as the public, who can better pinpoint the source of pollution.

3.4.3 Obligation to explain confidentiality claims

The obligation for national authorities to explain and substantiate their confidentiality claims for data of industrial facilities is aimed at ensuring external procedural accountability. This element of the E-PRTR shows external parties *why* certain data is left out of the access regime and provides them a chance to inquire why this happened and take action if they deem it appropriate. This way of reporting ensures that data analysis on E-PRTR data is affected as little as possible by the confidentiality of some data points and that the reason for confidentiality is always clear.[258]

3.4.4 Tiered accountability

The three elements of the E-PRTR highlighted above provide accountability for operators of industrial facilities through an obligatory reporting scheme to national authorities, the European Environment Agency and the European Commission. For every such system the question should be asked: what ensures the accountability of the accountability system, or: *who watches the watchers?* The way in which the E-PRTR regime aims to ensure accountability of the access regime overall, can be referred to as 'tiered accountability'.

The four-tiered reporting structure of the E-PRTR creates the access regime's tiered accountability. Operators ensure the quality of their data (tier 1); then they report to their national authority who checks the data again (tier 2), which in turn reports to the EEA, who tests it another time (tier 3) and finally the E-PRTR's operation is triennially checked by the EC (tier 4). These four tiers entail that normally, the E-PRTR's data and general operation is (partly) audited at four different levels. Two examples of elements of the E-PRTR that further facilitate tiered accountability are: the EC's validation tool for data of national competent authorities, which weeds out errors before submission;[259] and the structure in place to monitor confidentiality claims, which prevents abuse of this exception.[260] Together, this structure ensures that the accountability of every party in the E-PRTR's reporting structure (facilities, national competent authorities, EEA and EC) is checked at least once every reporting cycle and therefore further strengthens the accountability of the industrial facilities in the access regime.

Altogether, the E-PRTR is an intricate system that ensures the accountability of operators of industrial facilities and obliges them to explain and justify their conduct.

258 Section 3.2.5.
259 Section 3.2.3.
260 Section 3.2.5.

3.5 Lessons Learned

Specific disclosure rules	A key feature of the E-PRTR access regime is that it clearly defines the exact types of data that need to be included. Not just that, it also specifies in detail which 'economic activities' are subject to reporting duties. This granular specification of what data is (not) included in the access regime provides legal certainty. Moreover, it helps to standardize the quality of data and increase its research utility. This is a particularly useful lesson for research access in the platform context. It will be vital to clearly list what exact data should (not) be included in such an access regime. Given the complexity and scale of platform services, and the variety of public policy concerns they raise, it may be advisable to specify this in delegated regulation rather than legislation (see Section 5.2.3 for further discussion).
Standardized methods for data generation	The E-PRTR regime addresses the question of how data is generated. The Regulation specifies that the respective industrial facilities need to indicate whether their reported data is based on *measurement*, *calculation* or *estimation*. Internationally accepted methodologies for each of these three approaches are laid down in further regulatory guidance. This approach ensures that reported data is comparable (among different operators as well as over time); and provides methodological assurances to scientific researchers (and other dataset users). Given the complexity of platform data ecosystems, clarity as to how reported data came to be will be crucial. Policymakers are recommended to consult with experts for developing – and at least outlining the minimum requirements for – such methods in the platform governance context.
Liability for data quality (completeness, consistency and credibility of disclosures)	Industrial facilities subject to the E-PRTR Regulation are liable to assure the quality of reported data. The competent authorities to which they report assess the data quality, 'in particular as to their *completeness*, *consistency* and *credibility*' (Art. 9). The European Commission has developed further guidance on how these three terms should be interpreted. Similar to the previous point, making platforms liable for the quality of reported data appears vital in making sure a new research access regime will have any meaningful role to play (both in enabling research and as an accountability mechanism). Methods for assessing data quality should be developed and made publicly available. The E-PRTR's standards of completeness, consistency and credibility may serve as valuable inspiration when doing so in the platform governance context.

Size-based regulation	The E-PRTR sets a threshold below which industrial facilities are not required to report data. It should be said that this threshold has been criticized for being too high, resulting in a large amount of pollution data produced by small(er) industrial facilities *not* being included in the database. Policy makers may wish to consider a *de minimis* rule for platform data access regimes as well (cf. existing size-based platform rules, e.g. Germany's NetzDG and France's Law on Hateful Content Online). That said, it will be important to carefully and explicitly make the trade-off between lowering the threshold and the regulatory burden on smaller players.
Transparency by default	All of the data listed in the E-PRTR Regulation has to be made fully transparent by default. By way of exception, the data source can request some data to be kept confidential, subject to certain minimum criteria. This is subject to a case-by-case assessment by the competent authority. A similar approach certainly seems appropriate for platform research access. Whilst social media data may often be more sensitive than environmental data, given the privacy/data protection implications, the basic starting point may still be that the burden of proof lies on the company to demonstrate that secrecy is required; not on the public to demonstrate that transparency is required. As mentioned before, policy makers will have to carefully consider the potential trade-offs between competing rights, freedoms and interests for each data category listed. Additionally, they should define an appropriate procedure with clear criteria for applying for confidentiality.
Public transparency by default	All of the data rendered accessible through the E-PRTR regime is available to *anyone* with an internet connection. There are no criteria to be fulfilled to gain access to E-PRTR data. Relevant authorities are legally required to promote awareness and facilitate access to the public at large. This has led the European Environmental Agency to *inter alia* develop an interface to navigate through the data, apart from giving access to the entire 'raw' data-set. Platform research access regimes would also benefit from being publicly available by default. This appears especially important if (part of) the aim of such a regime is to increase the accountability of platform operators. Availability to the public at large enables wider scrutiny and prevents concerns over preferential treatment and institutional bias or capture. Such *public* transparency should also come with a responsibility to create the enabling environment that makes publicly available data meaningfully accessible (e.g. through interfaces, APIs, etc.). Of course, given the privacy-sensitive nature of social media data, these public access regimes may also be supplemented by more limited access regimes (as explored further in Chapter 4 in the context of medical research).

Tiered oversight structure	The E-PRTR regime foresees different tiers of oversight, with each entity being held accountable by another one. Specifically, while the regime is aimed at enabling accountability of polluting industrial facilities, it also ensures that the national authorities overseeing the reporting are held accountable themselves by the European Environmental Agency, which is in turn held accountable by the European Commission. This tiered structure might also be necessitated by the large amount of industrial facilities falling within the regime's scope (> 33.000), requiring a delegation of oversight bodies. Because the number of actors involved in a potential platform research access regime will presumably be far lower, a tiered structure might not be as essential. That said, it is important that necessary safeguards are put in place so that the independent institution(s) at the heart of the access regime are held accountable as well. Delegation to Member-State level authorities might also be relevant in order to create the enabling environment referred to in the previous points: ensuring easy accessibility of platform access regimes to a wide range of stakeholders.
Sanctions / penalties	Related to the previous point, the E-PRTR regime mandates the relevant authorities to sanction non-compliance with the data access framework. Because of the lack of harmonized (minimum) amounts for penalties, enforcement differs considerably among Member States. A genuine threat of sanctions and penalties is particularly important in the platform context, which involves a handful of powerful, well-funded actors with strong disincentives to enable the type of data access regime envisaged here. A more harmonized approach (than the E-PRTR) might be more appropriate here.
Proactive support for researchers	The E-PRTR regime explicitly calls for the Commission and Member States to 'promote awareness of the public of the European PRTR', as well as requiring them to 'ensure that assistance is provided in accessing the European PRTR and in understanding and using the information contained in it.' (Art. 15) A common criticism of transparency and data access initiatives (notably in the platform governance context) is that they are meaningless without an active civil society making use of the available data. One step towards meeting these important objections is to legally require governments (e.g. independent institutions) to invest in relevant research initiatives that engage with the available data, as well as awareness raising.

Strict timing	The E-PRTR Regulation sets maximum timeframes within which data has to be reported, and made publicly accessible. Originally, these timeframes were set by the Regulation itself and criticized for being too long. A 2019 amendment gave the Commission powers to change them via implementing regulation, limiting them to a maximum period of 11 months. This may serve both as a lesson learned *and* a cautionary tale for the platform governance context. Clearly defined time limits for data reporting are important. Equally important is for those time limits to be as short as possible in light of platforms' high-paced nature (e.g. 'without undue delay and no later than 72 hours after the event they relate to has occurred'). Ideally, a platform access regime should enable near-to-real-time data access. The existence of commercial APIs enabling real-time access to large amounts of data indicates that this is a feasible exercise.

4 Research access and data protection concerns – Learning from medical research with Findata

Findata is an independent unit within the Finnish Institute for Health and Welfare of the Ministry of Social Affairs and Health,[261] which in January 2020 launched an access regime for the secondary use of health and social care data in Finland. This access regime provides access to both aggregated statistics as well as data on individual patients collected during the provision of health and social care.[262] The data originates from the Finnish electronic patient dossiers and 10 other data registers. As an administrative authority, Findata takes in requests from those wishing to use data, gathers it from the respective sources, combines it into one dataset and provides access to it in a secure way to researchers, care providers and authorities in and outside of Finland.

The sharing of health-related data for the purpose of scientific research has been under discussion for a long time. One of the main dilemmas in this debate revolves around the apparent trade-off between public benefit and progress of medical science, and privacy and data protection interests of research subjects. We see this trade-off also reappear in current discussions on contact-tracing apps and other initiatives to fight the COVID-19 pandemic. This case study demonstrates that the alleged trade-off between medical research and privacy and data protection can be overcome.

As mentioned in Chapter 2, the second case study is aimed at learning from an established data sharing framework where (sensitive) personal data is shared. With this in mind, the selection-process for this case study focused on access regimes covering health data. Sharing of health data is a well-established practice in the medical sector and, because of its highly sensitive nature, the sharing of such health data is generally explicitly regulated. Both of these aspects of health data access regimes can offer valuable lessons for drafting data sharing policies and regulation in the platform governance context.

First, the sharing of health data is already a long-established practice between public institutions, private companies and researchers. For example, in the use of clinical trials for the development of new

261 Findata is referred to in the Act on Secondary Use of Health and Social Data (nr. 552/2019) (Laki sosiaali- ja terveystietojen toissijaisesta käytöstä) (Fi), section 4 (ASU) as 'the Data Permit Authority'. Findata is an independent unit of the National Institute for Health and Welfare of the Finnish Ministry of Social Affairs and Health. For Findata's website, see: Findata, 'Findata – Health and Social Data Permit Authority | Tervetuloa!' (Findata) <https://www.findata.fi/en/> accessed 11 June 2020. Findata is not part of the Data Protection Authority (DPA) of Finland. For more information on the Finnish DPA, see: Office of the Data Protection Ombudsman, 'Office of the Data Protection Ombudsman' *(Tietosuojavaltuutetun toimisto)* <https://tietosuoja.fi/en/office-of-the-data-protection-ombudsman> accessed 11 June 2020.Error! Hyperlink reference not valid.

262 ASU, sections 3(2)-(3) (Health and social data' refer to data collected during the provision of primary health care and/or social care. Social care includes, for example, the provision of pensions, welfare or social insurance. It is important to note that the term 'social data' does not refer to 'social media data'.).

medicine;[263] the use of human tissue in biobanks;[264] for the improvement of hospital treatments or indeed, to fight a global pandemic.[265] In recent years, many initiatives have sprung up to facilitate the sharing of health data 'digitally'. One such initiative is the subject of this case study: Findata, the Finnish Health and Social Data Permit Authority (*hereafter: Findata*). Close examination of this regime can provide lessons on how a similar regime could be created between governments, internet platforms and scientific researchers.

Because of the sensitive nature of health data,[266] access regimes in the medical sector generally have to be explicitly provided for by law in order to be lawful under the GDPR.[267] This 'forced' explicitness makes a health data access regime valuable to learn from, as it lays bare the legislator's reasoning in enabling the sharing of sensitive categories of personal data with the necessary safeguards built in.[268] The legal infrastructures allowing – or even obliging – the sharing of health data, can inform the framework for sharing personal data controlled by internet platforms, much of which arguably constitutes 'sensitive' personal data as well.

In light of the above, Findata – mentioned in the European Commission's data strategy[269] – was selected as the concrete case study for this Report. Preliminary research indicated Findata to be well suited for the purposes of this Report for several reasons. Firstly, participation with Findata's access regime is mandatory for both public and private controllers of health data under a dedicated legislative framework. Secondly, Findata has a comprehensive reach, covering the entire process from applications for, gathering, making available and removal of data. Finally, considerable information on Findata's access regime is easily available (in English), enabling a more thorough analysis.

Firstly, the Findata regime is based on a law specifically created to set up and regulate this access regime.[270] The *Act on Secondary Use of Health and*

263 The Clinical Trials Directive (2001) and the Clinical Trials Regulation (since 2014) regulate the testing of medicine and medicinal products in the European Union. Regulation EU No 536/2014 of the European Parliament and of the Council on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC (the "Clinical Trials Regulation") [2014] OJ L158/1, arts 1 & 2(1); Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use (the "Clinical Trials Directive") [2001] OJ L121/34.

264 Ciara Staunton, Santa Slokenberga and Deborah Mascalzoni, 'The GDPR and the Research Exemption: Considerations on the Necessary Safeguards for Research Biobanks' (2019) 27 European Journal of Human Genetics 1159.

265 South Korea, for example, has created an access regime which allows mediated access to pseudonymised data on the prior health insurance history of COVID-19 patients. Researchers can create analysis code on a sample dataset, which is then performed on the real data by the Korean Ministry of Health. The data is available via: South Korean Ministry of Health and Welfare and Health Insurance Review & Assessment Service, '#opendata4covid19' (*#opendata4covid19*) <https://hira-covid19.net/> accessed 11 June 2020; requirements for access can be found at: Sang Woo Park, 'Sang Woo Park on Twitter: "For More Information: Https://T.Co/4gtJce6RIK"' <https://twitter.com/sang_woo_park/status/1247313805752885248> accessed 11 June 2020.

266 Data gathered during the provision of health care is qualified as a 'special category of personal data' under GDPR, art 9(1). Therefore, its processing is in principle prohibited, unless one of the ten exceptions is satisfied listed under GDPR, arts 9(2)(a)-(j).

267 GDPR, arts 9(h)-(j) (Three of the exceptions allowing processing of special categories of personal data in the GDPR, refer to health data or research with it: the provision of preventive or occupational medicine; public interest in the area of public health and scientific of historical research purposes.).

268 A recent example of how a legislator specifies legal grounds for the sharing of data relating to health, is the European Data Protection Board's opinion on the use of personal data to speed up the mitigation of COVID-19: European Data Protection Board, 'Guidelines 04/2020 on the Use of Location Data and Contact Tracing Tools in the Context of the COVID-19 Outbreak' (2020) Text <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-042020-use-location-data-and-contact-tracing_en> accessed 23 April 2020.

269 European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee of the Regions and the Committee of the Regions: A European strategy for data' (Communication) COM (2020) 66 final, 12.

270 An English translation of the ASU by the Ministry of Social Affairs and Health is available here: Ministry of Social Affairs and Health (Sosiaali- ja terveysministeriön), 'Secondary Use of Health and Social Data' (*Sosiaali- ja terveysministeriö*) <https://stm.fi/en/secondary-use-of-health-and-social-data> accessed 23 April 2020).

Social Data (hereinafter also: 'Act on Secondary Use' or 'ASU') makes participation in the access regime mandatory for both public and private health care providers in Finland.[271] This stands in contrast to other access regimes (e.g. in France, Germany, the Netherlands and at the European level) which offer less extensive access to data,[272] or are under review for compliance with data protection laws.[273] The fact that the Findata access regime also requires private parties (e.g. private health care clinics) to participate in the data sharing framework created by the government, renders the case study quite relevant for platform governance discussions. Indeed, as mentioned in Chapter 2, there are a growing number of legislative initiatives putting in place varying degrees of data sharing obligations on internet platforms.

Secondly, the scope of Findata's access regime is very broad. The law regulating Findata covers the entire process of access to health data: from the types of health data to be rendered accessible and which parties can apply to access them, to detailed descriptions of the IT-systems for requesting, gathering, accessing and analyzing the data. Additionally, all these stages of the data sharing process are overseen by a single entity, which focusses only on health and social care data. Finally, despite its novelty, there is a lot of (English-language) information available on Findata already and the organization was also approachable for several interviews. This makes Findata an interesting and useful case to get a comprehensive picture of an operational access regime involving considerable amounts of (sensitive) personal data.[274]

It must be acknowledged that this case study has two important limitations. First, Findata's access regime is country-specific and therefore only includes data of people with a Finnish personal identity code.[275] Therefore, it is situated within the legal context of Finnish national laws. Secondly, at the time of writing, the Findata regime has only been operational for a few months (accepting data requests since January 2020, data permits since April 2020) and therefore there is no conclusive evidence of its workings in practice. Yet, valuable lessons from Findata's access regime can still be drawn for platform governance, because of its mandatory participation for private parties; the broad range of health data included and its adherence to EU data protection law, such as the GDPR. Additionally, challenges for cross-border access regimes are covered by the E-PRTR case study also in this report.[276]

---

271 ASU, section 1.

272 The German Forschungsdatenzentrum of the German federal statistical authority- also mentioned in the EC's Data strategy – provides access to only one health related dataset (Diagnosis-Related Groups Statistic): 'Research Data Centre' <https://www.forschungsdatenzentrum.de/en#understand-rdc> accessed 15 May 2020; Eurostat, the statistical authority of the European Union, also provides access to only one health related microdata set (the European Health Interview Survey): 'Overview – Eurostat' <https://ec.europa.eu/eurostat/web/microdata> accessed 15 May 2020; The extensive scope of Findata's access and its inclusion of data from both public and private sources were chosen over these alternatives.

273 While Statistics Netherlands (the Dutch national statistical authority) offers access to many different types of health data on a case-by-case level ('Gezondheid En Welzijn' <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata/gezondheid-en-welzijn#id=zorgzvtab-personen-die-zorg-zonder-verblijf-hebben-ontvangen--voorheen-cakzzv---vervangen-vanaf-2009-door-gebzzvtab--0> accessed 15 May 2020), it is currently conducting a legal review of its access regime to determine whether it is in line with the GDPR: Meindert Kappe, 'CBS: Inquiry into Risks of Data Access' (*Centraal Bureau voor de Statistiek*, 20 December 2019) <https://www.cbs.nl/en-gb/corporate/2019/49/cbs-inquiry-into-risks-of-data-access> accessed 15 May 2020. A case study of this regime therefore risked to be superseded by changes to the regime soon after completion.

274 In contrast, documentation on the French Health Data Hub (Health Data Hub, 'Health Data Hub | Plateforme Des Données De Santé | France' (*Healthdatahub*) <https://www.health-data-hub.fr?lang=en> accessed 20 March 2020.) – also mentioned in the EC's Data strategy – was not readily available at the time of this case study. Findata's English-language website, translation of regulations and a swift response to inquiries for an interview expedited the choice for Findata as a case study.

275 Digital and Population Data Services Agency, 'Personal Identity Code' (*The personal identity code*) <https://dvv.fi/en/personal-identity-code> accessed 14 May 2020.

276 While the E-PRTR case study does not cover personal data, it does operate across the borders of 30+ countries, involves as many supervisory authorities and is based on a supra-national legal framework.

This chapter first continues explaining the background of Findata: its history, legal ground and main goals. Subsequently, it will discuss the implementation of Findata's access regime: the exact data it makes available and the process from a data request to research results. After that, the governance structure of this access regime will be analyzed, looking at the different supervisory authorities involved, the arrangements for enforcement and sanctions and liability and funding. The following section describes how the law establishing Findata interfaces with the GDPR and its Finnish implementation. The chapter ends with listing the main lessons learned from Findata's access regime for the platform governance debate.

4.1 Background

Finland has a long history of collecting health and social care data in registers. The first Finnish register with aggregated 'vital statistics' including births, deaths and marriages was introduced in 1749. The first nation-wide computerized register was the Cancer Register, established in 1952. In the period 1952-1994, ca. 20 registers were established.[277] In the period 2003-2014 several government bodies provided funding for the Finnish Information Centre for Register Research ('ReTki').[278] ReTki aimed to promote the use of national registers for research in the health and social sciences and provided an alphabetical list of all available registers and their controllers.[279] In the last two years of its operation, its limited funding only allowed for small-scale operations.[280]

Although access to health and social data via these registers was possible for researchers, two aspects made access cumbersome and inefficient. First, there were no clear rules on the requirements for data protection. The responsibility for data protection and security fell completely to individual researchers who obtained access to data. Data was sometimes shared via hard drives and USB flash drives, creating considerable risks and responsibilities.[281]

Secondly, obtaining access to different registers could take a long time as researchers had to make multiple applications for access to different registers. These registers could also use different IT-environments and application procedures, as the legal grounds for sharing data were scattered throughout sector-specific laws and regulations. The necessary combining and linking of data from different sources was the responsibility of individual researchers, which again was not beneficial for data protection and security.[282]

The process eventually leading to the creation of Findata started in November 2015. Ministries, business interest groups, hospital districts and research groups were involved in eight pilot projects funded by The Finnish Innovation Fund (Sitra) under the title 'Isaacus – Digital Health Hub'.[283] These projects were meant to test several aspects of an access regime for health and social care data: a permit and information portal, common metadata descriptions and collection

277 For a historical overview of health and social welfare registers in Finland, see: Jari Haukka and Mika Gissler, 'Finnish Health and Social Welfare Registers in Epidemiological Research' (2004) 14 Norsk epidemiologi 113.

278 Finnish Information Centre for Register Research, 'ReTki Info' (*Finnish Information Centre for Register Research*, 29 April 2012) <https://rekisteritutkimusen.wordpress.com/retki-info/> accessed 14 May 2020.

279 Finnish Information Centre for Register Research, 'Register i alfabetisk ordning' (*Informationscentret för registerforskning – ReTki*, 28 April 2012) <https://rekisteritutkimussv.wordpress.com/register/register-i-alfabetisk-ordning/> accessed 14 May 2020.

280 Pim ten Thije, Interview with Antti Piirainen, Head of Communications, Findata (via Zoom videoconferencing, 26 March 2020).

281 ibid.

282 ibid.

283 For a detailed description of all eight projects and the parties involved, see: Heli Parikka, 'One-Stop Shop for Well-Being Data – Isaacus Laid the Foundations for the Future' (*Sitra*, 9 November 2018) <https://www.sitra.fi/en/articles/one-stop-shop-well-data-isaacus-laid-foundations-future/> accessed 17 April 2020.

and handling and remote user environments for data.[284] Most pilot projects were finished by the summer of 2018 and the project was taken over from Sitra by the Ministry of Social Affairs and Health. The ministry combined the practical experience from the Digital Health Hub projects with its efforts to develop legislation for a new data permit authority.

Simultaneously with the Isaacus project, between October 2015 and December 2017, a governmental working committee prepared the new Act on the Secondary Use of Health and Social Data.[285] This Act was subsequently introduced (October 2017), improved upon (October 2017–October 2018) and approved by parliament (March 2019) and then introduced into Finnish law on May 1st, 2019.[286] It regulates both the practical and legislative aspects of the access regime for Finnish health and social data and will be referred to throughout this case study. Additionally, the Finnish Data Protection Act – which implements the GDPR into Finnish law – applies to the access regime.[287] Findata has started accepting data requests since the 1st of January 2020, and applications for data permits since the 1st of April 2020. At the time of writing, Findata has received 27 data requests and 83 applications for data permits.[288]

Figure 8 – Combined goals of the Findata access regime

The legally binding objectives of the Act on Secondary Use are threefold: (1) to enable efficient and secure processing of health and social care data; (2) to allow health and social care data of patients to be combined with health and social care data in national registers, and (3) to secure the legitimate expectations, rights and freedoms of individuals when processing personal data.[289] The ASU thus focusses on the efficient and secure processing of health and social care data from different sources, while ensuring citizens' rights and managing their expectations.

Findata's self-stated goals give further shape to the objectives and requirements stated in the ASU. Findata has formulated four main goals for itself: (1) to improve data security and the data protection of individuals; (2) to speed up and streamline the utilization of social welfare and health care data; (3) to decrease the duplication of work to get access to data, and (4) to improve the quality of meta-data descriptions of data registers together with the controllers.[290] Findata

284 Ibid.

285 Act on Secondary Use of Health and Social Data (nr. 552/2019) (Laki sosiaali- ja terveystietojen toissijaisesta käytöstä) (Fi).

286 Heli Parikka and others, 'A Finnish Model For the Secure and Effective Use of Data – Innovating and Promoting the Secondary Use of Social and Health Data' (2019) 153, 11 <https://www.sitra.fi/en/publications/a-finnish-model-for-the-secure-and-effective-use-of-data/> accessed 23 April 2020 (Figure 1: "The Process Of Developing The New One-Stop Shop Body In Finland" shows this process in detail.).

287 Data Protection Act (nr. 1050/2018) (Tietosuojalaki Dataskyddslag) (Fi) (An English translation is available via 'FINLEX ® – Translations of Finnish Acts and Decrees: 1050/2018 English' <https://www.finlex.fi/en/laki/kaannokset/2018/en20181050?search%5Btype%5D=pika&search%5Bkieli%5D%5B0%5D=en&search%5Bpika%5D=Data%20Protection> accessed 16 June 2020) (Data Protection Act).

288 Findata, 'Data Requests' (*Findata*, 15 June 2020) <https://www.findata.fi/en/services/data-requests/> accessed 16 June 2020.

289 ASU, section 1.

290 Findata, 'About Us' (*About Us*) <https://www.findata.fi/en/about-us/> accessed 29 April 2020.

thus tries to improve data register quality, data security and data protection of individuals, while making the process to get access to health and social care data more efficient.

Taken together, the ASU and Findata jointly pursue four objectives for the Finnish access regime for health and social care data: (a) to enable the effective sharing and combining of health and social care data; (b) to improve data register quality; (c) to do so in a secure and efficient manner, and (d) while respecting individual's rights and expectations (cf. Figure 8).

4.2 Implementation

Looking at how the Findata framework has been implemented on the ground, several elements are worth highlighting:

- the types of data sources;
- the different types of access to data;
- what parties can request access;
- the overall process, from application to publication of results

4.2.1 Data sources

Findata provides access to the personal health and social care data of persons with a Finnish identity code. This data originates from two different types of sources: (a) the nationwide system in Finland for electronic patient dossiers, 'Kanta Services', and (b) the health and social care in the data registers of circa 10 governmental bodies and supervisory authorities.[291] The former includes data collected during the provision of primary care in public and private hospitals like medical diagnoses, laboratory test results and welfare allowances. The latter consists of the data in national registers like the cancer and infectious disease registers and occupational illnesses and social benefits registers.[292]

Not included in the Findata access regime is data gathered for statistical purposes of the Finnish national statistical authority, Statistic Finland and the National Institute for Health and Welfare. These organizations have to maintain their own access regimes for this type of data.[293] In the end, the ASU does not specify what exact data should be accessible via the Findata access regime. This is due to Findata's demand-driven and purpose-based nature: it is up to Findata's employees to determine on a case-by-case basis what exact health data from pre-defined data sources (which *are* laid down in the ASU) can be made available to fulfil the specific purpose of a research application. The ASU does not further specify the wide range of available data than the two broad categories described above.

---

291 What follows is a list of authorities and organisations providing access to data registers of health and social care data. The access to registers of the parties marked with an * is limited by some condition(s). For the full list, see: ASU, section 6. Parties with health and social data registers that can be accessed are: the Finnish Ministry of Social Affairs and Health, the Finnish Institute for Health and Welfare, the Social Insurance Institution of Finland*, the National Supervisory Authority for Welfare and Health Valvira, the Regional State Administrative Agencies*, the Finnish Institute for Occupational Health*, the Finnish Medicine Agency Fimea, public and private service organisers of social and health care, Statistics Finland*, the Finnish Centre for Pensions* and the Population Register Centre*.

292 ASU, sections 6 & 51(4),

293 ASU, section 7; Disclosure of this data needs to happen in line with the Finnish Statistics Act (nr 280/2004) (Tilastolaki) (Fi) (for an English translation see: Tilastokeskus, 'Statistics Act (280/2004)' <https://www.stat.fi/meta/lait/statistics-act-2802004_en.html> accessed 16 June 2020) and the EU Regulation on Community Statistics (Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (Text with relevance for the EEA and for Switzerland) [2009] OJ L87/164).

4.2.2 Types of data access

Findata provides for two different types of access to the respective health and social data through what it calls 'data permits' and 'data requests'. A *data permit* provides access to a dataset with pseudonymized data of individual patients.[294] A *data request* provides access to a dataset with aggregated, anonymized statistics. Such a dataset could, for example, contain an inventory of the average time Finnish patients with different diseases spend in different hospitals. Proportionality is important here: if Findata comes to the conclusion that the purpose of a data permit application (of pseudonymized data) can also be achieved through a data request (including only anonymized data), it must propose a change of the application to the applicant.[295] This practice ensures that personal data is only disclosed when necessary. Finally, Findata employs a specific system where Finnish citizens can assert their individual data subject rights of access, rectification and erasure of personal data in Findata's systems.[296]

Grounds to apply for data permit or request at Findata

- scientific research
- statistics
- development and innovation activities
- education
- knowledge-based management
- steering and supervision of social and health care by authorities
- planning and reporting duties of an authority

Figure 9 – Seven ASU approved purposes for accessing data

The distinction between a data permit and a data request can be further clarified through a (fictitious) example relating to a dataset of Finnish cancer patients. A data permit could provide access to a dataset which specifies the progression of treatment per patient: time between treatments, total duration of treatment, results of different medical examinations, etc. A data request would only provide aggregated data: average scores for different groups of patients of a certain size.[297] For example, average time between treatments, average total duration of treatment per age group of patients or per group with a specific form of the disease.

294 ASU, sections 3(19), 14 & 51(1); For more information see Section 4.2.4.

295 ASU, section 43(5).

296 See Section 4.4.

297 Daniel L Oberski and Frauke Kreuter, 'Differential Privacy and Social Science: An Urgent Puzzle' 2 Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/g9o4z8au/release/2> accessed 29 April 2020 (Such a group needs to be of a certain size, because if it is too small there is still a risk that individual patients can be identified through a combination of use of other data sources and the characteristics of the group.).

As Findata is a 'data permit authority' under Finnish administrative law,[298] it has the administrative power to oblige data sources to provide it with all data it requests from the data sources on behalf of its applicants, even if subject to secrecy obligations and other restrictions on the use of data by these sources.[299] The ASU does not specify any conditions allowing the data source to withhold data from Findata. Therefore, the only possibility for withholding data would be when a data source can show that its data does not fall within Findata's mandate as specified in the ASU.[300]

4.2.3 Eligible parties for data access

There are no a priori constraints as to which parties can hand-in an application for a data permit or data request.[301] This means that in principle, anyone who wants to obtain access to health or social care data via Findata is free to apply for it, as long as they have an approved purpose to access the data they request. The ASU specifies seven approved purposes to access data (see Figure 9), some of which further constrain the types of data that can be obtained.[302] For example, for the purpose of 'development and innovation' applicants can only obtain a data request (not a data permit).[303] For 'knowledge based management' purposes, a health care provider can only obtain a data request when it wants to compare its data to that of others; it does not need a data request or permit to use data it already has.[304] Section 4.4.3 on data protection law provides furthers explanation of the lawful grounds and limitations for all seven purposes.

4.2.4 Process: from application to publication of results

Key elements to be included in a data utilisation plan

- Description of the data requested (controller, register, time period);
- Intended purpose of the data in the application;
- Controller and processor of the data and people involved;
- Legal ground for processing;
- Data security and protection measures taken throughout the data lifecycle, including storage, erasure or archiving;
- If intent to provide own data for combination: detailed information on legal basis, ethics statements for own data;

Figure 10 – List of elements to be included in data utilisation plan

Application – To apply for either a data permit or data request, a party has to hand in a data utilization plan via Findata's data request management system. The data request management system is an online environment where the progress of the application can easily be followed. The data utilization plan should include a research plan, project plan or similar document stating the intended purpose of the data requested; the controller and processor(s) of the data; the legal ground for processing, and the 'essential factors' of data security and protection measures taken throughout the data's lifecycle, including

298 See also: Section 4.3.
299 ASU, section 36.
300 This mandate contains the sources specified in ASU, section 6 under the conditions described in ASU, section 36.
301 The cost of access is higher for parties based outside the European Economic Area, because the process to obtain compliance with the GDPR is more complicated; See also Section 4.3.5.
302 All seven purposes are listed in ASU, section 2 and further explained and detailed in ASU, sections 37-42.
303 ASU, section 37.
304 ASU, sections 41(1)-(3).

storage, erasure or archiving.[305] Findata also requests information on, for example, the funding of a project, billing information, the extraction method of data (random or stratified samples), and possible control groups. The entire list of requirements is available via Findata's website.[306] If a data utilization plan needs a 'statutory ethical preliminary assessment', this assessment by a special committee of the Institute for Health and Welfare can be requested via the data request management system as well.[307] Figure 10, gives a list of the most important requirements for a data utilization plan.[308] Prior to handing in their application, interested parties can contact Findata's free helpdesk to get advice on their applications, for example on available data sources fitting their research aims and a price estimate to obtain them.[309]

Assessing Applications – Subsequently, Findata grants or denies the application for data. It has to do so within 3 months after the applicant hands in a complete application for a data permit.[310] Findata's employees might contact the applicant to request additional information for completing the application.[311] For data requests, no deadline is specified. As the Finnish data permit authority, Findata is the sole party which decides on applications when the data applied for (a) originates from two or more different sources; (b) originates from *private* social welfare and health service providers or (c) originates from the electronic patient dossiers in Kanta Services.[312] If Findata considers it necessary, it may consult the Data Protection Ombudsman (the Finnish Data Protection Authority) on the merits of an application and halt it until it has received a response.[313]

Gathering Data – Once an application has been granted, Findata's employees start gathering data from the original sources. Data sources are required to disclose the necessary data within 30 days, but this period can be extended if an application is complicated.[314] The data is then saved into a secure hosting environment. This is an IT-system maintained by a contractor, where parties can disclose and receive data.[315] The integrity and origin of data must be verified by technological means, as reliable as an electronic signature of a natural person.[316] Both access and use of the data can be restricted.[317] Findata employees can also pre-process the data from within this secure hosting service.

Pre-processing Data – During pre-processing, Findata employees link data entries from different

305 These elements are required according to ASU, section 3(17).

306 For a full list of requirements of a data permit application, see: Findata, 'Data Permits' (*Data Permits*) <https://www.findata.fi/en/services/data-permits/> accessed 29 April 2020.

307 This statutory preliminary ethical assessment is *only* mentioned in ASU, section 16(3); ten Thije, 'Interview 1' (n 280) (The committee doing this assessment mainly check whether the consent given for data collection, covers the research that is proposed by the applicant.).

308 Findata, 'Data Permits' (n 306) (For a full list of requirements of a data utilisation plan and data permit application.).

309 For the costs of data access via Findata see Section 4.3.5.

310 ASU, section 47 (A data permit application must be considered for decision 'without delay', and at least within 3 months of the submission of the application. "If the processing of the application and the associated data utilisation plan require unusually extensive processing of data from several different controllers or a particularly challenging consideration process", this period can be extended by another 3 months.).

311 Pim ten Thije, Interview with Antti Piirainen, Head of Communications, Findata (via Zoom videoconferencing, 13 May 2020).

312 ASU, section 44(1).

313 ASU, section 44(4).

314 Section 48, ASU ('The Data Permit Authority may extend the deadline for disclosure of the data if the intended use of the data requires unusually extensive processing of data from several different controllers or a particularly challenging combination process.' The permit holder must be informed of the extension, justification and a new deadline.).

315 For more information of the responsibilities of different parties involved, see Sections 4.3.2 on external supervision and 4.3.3 on liability.

316 ASU, section 17(2).

317 ASU, section 3(10).

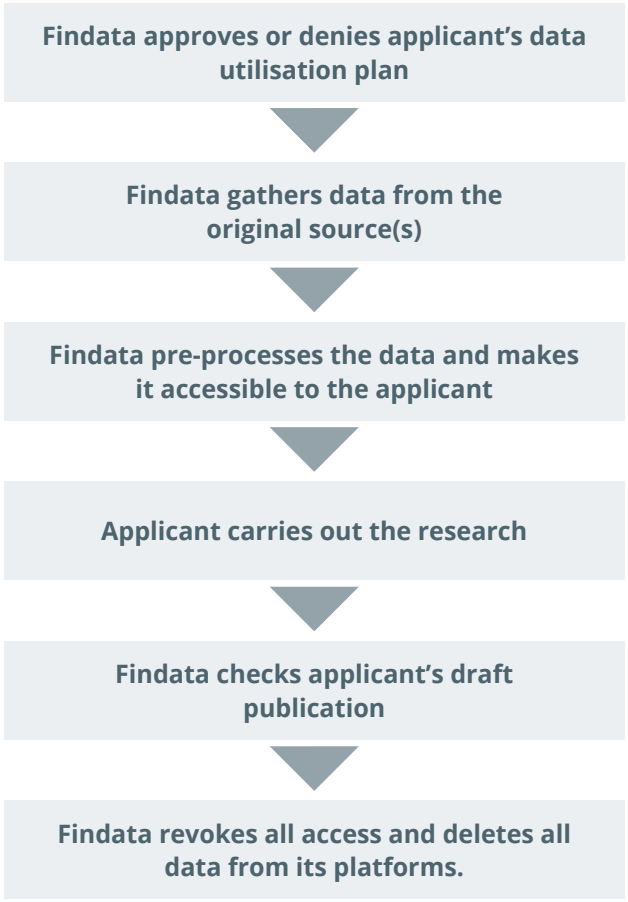


Figure 11 – Process at Findata from application for a data permit to research results

datasets that relate to the same individuals in the secure hosting environment, which only they can access.[318] This linking is achieved through the Finnish personal identity code, which is an individual ID-number that every resident of Finland obtains at birth, naturalization or when his or her residency in Finland is registered in the registration in the Population Information System. This ID number is i.e. used to register health and social data.[319] Once this linking of datasets is complete, the data is either pseudonymized (for data permits) or aggregated and anonymized (for data requests).[320]

Pseudonymization – In the context of Findata's access regime, pseudonymization works as follows: Every personal identification code in the dataset is replaced by a randomly created pseudo-identifier (called 'research number'). The combination of the two is kept by Findata until the data is deleted from the secure hosting environment. Subsequently, Findata's employees delete all directly identifying personal information (e.g. identity number, name, address etc.) from the dataset. If any additional data must be deleted, because it might lead to identification (e.g. in the case only five people have a rare disease unrelated to the application and inclusion of this data can lead to their unwanted identification) extra pseudonymization measures like these will always be communicated to the applicant.[321] Additionally, every application will receive unique pseudo-identifiers, even if exactly the same data is linked or requested by a different party.[322]

Making Data Available to the Applicant – Once the data is fully pre-processed, it is either directly made available for download (for data requests) or transferred from the secure hosting environment to the secure operating environment, accessible to the applicant (for data permits). Since the data in a data request is fully anonymized, it can simply be shared with the applicant, who can use it for the purpose stated in their data utilization plan. For data requests, the process ends here.

With a data permit, the applicant can now access and analyze the pseudonymized data via the secure operating environment. This is a virtual machine running

318 ASU, section 3(19).
319 Digital and Population Data Services Agency (n 275).
320 ASU, section 14 ('collection combination and pre-processing service for data'),
321 ten Thije, 'Interview 3' (n 311).
322 ASU, section 14(4).

statistical analysis software such as SPSS and R,[323] and is not freely accessible through the internet. Only pre-approved individuals listed in the data utilization plan can get access to this environment and all actions in it are logged. Findata monitors these logs and can revoke access to the environment at all times.[324]

End of the Process – Once the applicant has finished its project or the time period of the data permit expires, Findata revokes the applicant's access to the secure operating environment. The applicant is also obliged to let Findata check any publications of its results, to make sure these are fully anonymized and contain no personal data.[325] After this has happened, all data is deleted from all Findata's IT-platforms, which marks the end of the process.

4.2.5 Possible consequences of results

There are two possible results of projects in the Findata access regime worth discussing here. The first is the creation of intellectual property. Suppose a new treatment or medicine has been developed through data obtained via the access regime. Findata sees itself as the facilitator of data sharing between parties and therefore does not claim any potential intellectual property derived from the results obtained with the data it makes available to applicants.[326]

A second implication that may occur is when the results of a project create a clinically significant finding that enables 'the prevention of a risk to a certain patient's health or significant improvement to the quality of care'.[327] In this case, a data permit holder has the right to notify a contact person at Findata. The ASU lays down a procedure to further deal with such a situation: to identify the patient, its health care providers and inform the patients of the findings if he or she allows for this.[328]

4.3 Governance structure

The Findata data sharing framework is governed by a number of parties, both internally and externally. Internally, Findata as an authority falls under the Ministry of Social Affairs & Health. This Ministry also appoints the members of both Findata's Steering Committee as well as its high-level expert group. Externally, Findata and all other parties within the access regime (applicants, data sources and supporting parties, such as IT-contractors) are supervised by three national authorities: (a) Valvira (the Finnish National Supervisory Authority for Welfare and Health); (b) the Finnish Data Protection Authority (Data Protection Ombudsman) and (c) the Finnish National Cyber Security Centre (NSSC-FI).[329] Figure 12 shows the relations between Findata and these parties and they are further explained below.

4.3.1 Internal supervision

Findata is internally guided and monitored by a steering committee and advised by a high-level expert group. The Steering Committee consists of nine members appointed by the Ministry of Social Affairs

323 The applicant can request Findata to install other applications in the secure operating environment for an extra renumeration (see also Section 4.3.5).

324 ASU, section 5.

325 ASU, section 52.

326 ten Thije, 'Interview 1' (n 280).

327 ASU, section 55.

328 ASU, section 55 (describes the whole process in such a situation).

329 The tasks and powers of the supervisory authorities are clustered throughout the ASU: in general (ASU, section 56), for Valvira (ASU, sections 30-34); NSSC-FI (ASU, sections 26-29) and the Ombudsman (ASU, section 44(4); Data Protection Act, section 14).

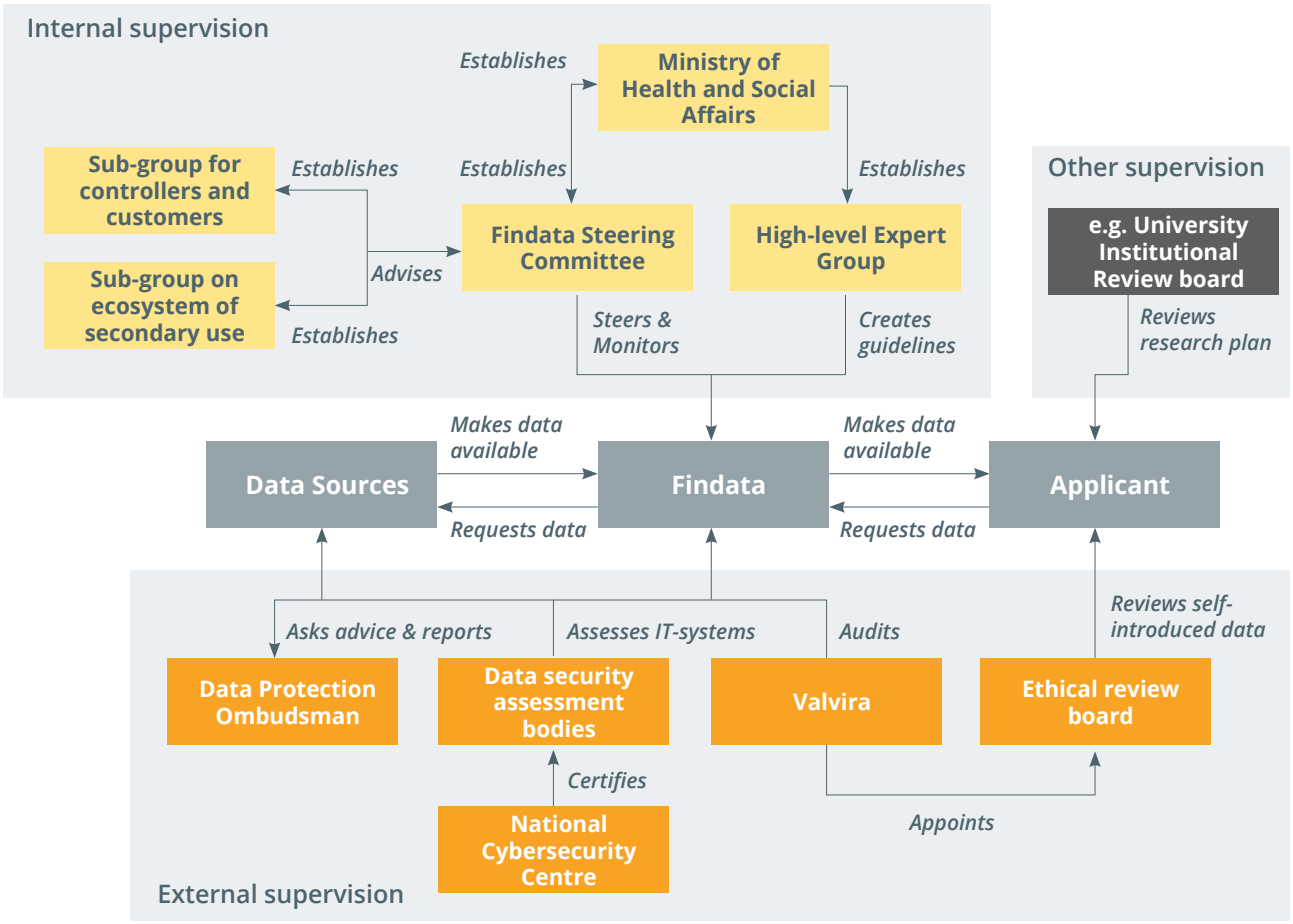


Figure 12 – Governance relations and supervision in the Findata access regime

and Health. The members are appointed for a three-year period and need to comprise: six representatives from the organizations that provide data through Findata; one representative of the Finnish municipalities; one representative of municipalities as organizers of preventive care and one representative of private companies that provide social and health care services.[330]

The Steering Committee has four main tasks and may additionally perform three others. The main responsibilities are to make a proposal to the Finnish Institute for Health and Welfare and the Ministry on: (1) the annual action plan of Findata and the associated budget; (2) the development of resources for its tasks together with controllers; (3) the allocation of resources to all parties cooperating in the development of information systems, and (4) to report on the operations and financial statements of Findata.[331] Additionally, the steering committee has the power to: (5) set goals for Findata's processes and initiate audits of them; (6) make proposals about Findata's operations; and (7) establish new expert groups to support Findata's operations.[332] Currently, the steering committee has established two extra subgroups. A subgroup for controllers of data and applicants to work on the improvement of meta-data descriptions and a subgroup on the ecosystem of secondary

330 ASU, section 8(1).
331 ASU, section 8(2).
332 ASU, section 8(3).

use.[333] Findata itself is obliged to report to the Steering Committee on cases in which it deviates from the standard deadlines for the processing of applications for data or disclosure of data for granted applications.[334]

The Ministry has appointed a high-level expert group for Findata. This group creates guidelines for Findata on how it should handle anonymization, data protection and data security. It must consist of members with expertise in the fields of: artificial intelligence, data analytics, data security, data protection, suitable research and statistical services. It must also contain a representative of Findata.[335] Both the Steering Committee and the high-level expert group are thus formed from a diverse group of representatives from stakeholders in Findata's access regime.

4.3.2 External supervision

The National Supervisory Authority for Welfare and Health (Valvira) supervises compliance of the secure operating environments with data security and data protection requirements laid down in the ASU, Finnish Data Protection Act and GDPR.[336] Valvira must also maintain a public register of compliant operating environments.[337] It has three types of powers: investigative, corrective and punitive. These different powers are further explained below.

For its investigations, Valvira can obtain all information necessary for its supervision free of charge.[338] Valvira may also carry out audits at all premises (except those of permanent residence) where activities are employed that relate to Findata's access regime,[339] or employ experts to do so.[340] A party that is audited must fully cooperate and provide all documents necessary for that audit. The audit's results must remain available for 10 years after completion.[341]

Based on the results of an audit, Valvira can order a service provider to correct defects in an operating environment in use. If such an environment jeopardizes data protection or does not comply with legal requirements, Valvira can prohibit its use until these defects have been corrected.[342] To give force to such a decision, Valvira can: (a) order the service provider to issue a (public) notification on the defects of its services; (b) issue a conditional fine to the service provider; (c) threaten to terminate the operation of the operating environment partially or completely; and (d) order unperformed actions to be performed by others at the service providers' cost.[343] Valvira can use all the same powers against authorities or other parties processing data in accordance with the ASU.[344]

Findata has to submit a detailed report to the Finnish Data Protection Authority ('DPA') at least once a year. In this report Findata must discuss which data has been processed through its systems and discuss

333 ten Thije, 'Interview 3' (n 311).
334 ASU, section 57.
335 ASU, section 8(4) (Further details on this expert group's tasks, its members and their eligibility can be determined by decree of the Ministry of Social Affairs and Health.).
336 ASU, section 30.
337 ASU, section 30(1) (This relates to the IT environment of individual data sources that opt to maintain an access regime for data request and permits that pertain *only* to their own data. If data is requested from multiple sources it is always provided via Findata.).
338 ASU, section 32.
339 ASU, section 30(2).
340 ASU, section 31.
341 ASU, sections 30(3)-(4).
342 ASU, sections 33(1)-(2).
343 ASU, section 33(4).
344 ASU, section 34.

the logs of actions performed on its systems.[345] Findata is also obliged to immediately notify the DPA if it suspects that a party processing data under a data permit does not process personal data in compliance with the law.[346] Furthermore, the DPA has all investigative, corrective and authorization and advisory powers provided to it by Article 58(1)-(3) of the GDPR. In addition to these powers the Finnish DPA can also consult experts,[347] and under certain conditions inspect premises used for permanent residence.[348] As noted before, the DPA can also advise Findata on data requests and permits, if the latter requests so.

The Finnish National Cyber Security Centre only indirectly supervises Findata's IT-systems.[349] One of NSSC-FI's tasks is to certify other parties, which can use their certification to assess the quality of government IT-systems.[350] These parties, called 'data security assessment bodies,' assess the quality of IT systems based on the applicable requirements in Finnish law.[351] Findata's secure hosting service and secure processing environment must be certified by such an assessment body at least every five year to be allowed to operate. The certification process includes a report of the result of such an assessment.

Finally, parties that apply for data through Findata may have own supervising authorities unrelated to Findata. For example, scientific researchers generally have to obtain approval for research involving any type of personal data from their institution's institutional review board and ethics commission. These checks of course differ per applicant, but come on top of the supervisory regime that supervises and audits the operation of Findata's access regime (steering committee, Valvira) and its IT-systems (DPA, Valvira, data security assessment bodies).

4.3.3 Liability & contractual relations

Since Findata is an authority under Finnish administrative law, it does not need to sign contracts or processing agreements with parties providing access to their data through it, or with parties which it grants access to data. Findata's decisions, for example, to grant or deny data requests or data permits, are legally binding administrative decisions in themselves.[352]

The liability for data breaches and other types of GDPR violations will be carried by the relevant data controller. For an explanation of the distribution of responsibilities under data protection law in the Findata access regime, see section 4.4.1. In summary, Findata can generally be considered a (co-)controller throughout the different phases of the access regime. Findata's administrative decisions contain standard provisions describing procedures for if, or when, a data leak happens.[353] With this in mind, Findata can

345 ASU, section 53.

346 ASU, section 56.

347 Data Protection Act, section 19.

348 Data Protection Act, section 18 ('An inspection may be carried out in premises used for permanent residence only if this is necessary ... and if a well-founded and specific reason exists in the case for suspecting that provisions on the processing of personal data have been or are being infringed in a manner that may be sanctioned with an administrative fine or a punishment provided in the Criminal Code (39/1889).').

349 ASU, section 56.

350 TRAFICOM Finnish Transport and Communications Agency National Cyber Security Centre, 'Accredited Information Security Inspection Bodies | NCSC-FI' (*Accredited information security inspection bodies*) <https://www.kyberturvallisuuskeskus.fi/en/our-services/assessment-accreditation-and-guidance/accredited-information-security-inspection> accessed 15 May 2020 (Provides a list of accredited bodies by the NSSC-FI.).

351 ASU, section 3.

352 Pim ten Thije, Interview with Antti Piirainen, Head of Communications, Findata (via Zoom videoconferencing, 17 April, 2020).

353 ibid.

be considered the main entity responsible for GDPR compliance in this data access regime.[354] It must be noted that the original data source always remains a controller with regard to its own processing operations.

Finally, Findata does have a contract and data processing agreement with the company that provides its IT-systems: the data request management system, secure hosting environment and secure operating environments. This is CSC – IT Centre for Science,[355] a Finnish center with expertise in information technology owned by the Finnish state and higher education institutions.[356] This contract and the data processing agreement with CSC are subject to checks by the internal and external supervisory parties mentioned in section 4.3.1 and 4.3.2.[357]

4.3.4 Sanctions & enforcement

The Act on Secondary Use of Health and Social Data does not specifically list sanctions or enforcement measures for non-compliance with Findata's administrative decisions. Findata can, however, at all times revoke access of parties using its services for infringing applicable laws or standards.[358] As mentioned before, Valvira has far reaching competences to audit and fine all parties involved with the Findata access regime and the Finnish Data Protection Authority can impose administrative fines as laid down in the GDPR.[359]

4.3.5 Funding

As Findata is still in its start-up phase,[360] it is fully funded by the Finnish government. Eventually, the goal is that Findata be fully self-funded, covering its operation costs through fees. In 2019 Findata received a budget of €2,5 million,[361] for 2020 that budget is circa €5 million. This budget and the forthcoming budget for the first years in operation are meant to cover the start-up costs of setting up Findata's operations and create its new IT-systems.[362]

Findata charges a fee for four of the services it provides. Its helpdesk service is free.[363] First, Findata charges anyone who applies for a data permit or data request. Findata charges EEA citizens €1.000 for decisions on an application for a data request or data permit. Students from within the EEA, working on their university degree thesis get a €500 discount on either a data request or permit.[364] Applicants for data permits from outside the EEA pay €3.000, since these applications take considerably more effort to fulfil in

354 During this period, Findata is a controller and must comply with all standard GDPR procedures for data leaks as listed in GDPR. arts 33-34 and faces the fines laid down in GDPR, art 83 and Data Protection Act, section 24.

355 CSC, 'About Us' (*About Us*) <https://www.csc.fi/en/about-us> accessed 14 May 2020.

356 ten Thije, 'Interview 3' (n 311).

357 ten Thije, 'Interview 2' (n 352).

358 ASU, section 34(4),

359 Section 4.3.2.

360 ten Thije, 'Interview 2' (n 352) (Findata's helpdesk services became operational in November 2019. It has started accepting applications for data requests per January 2020 and applications for data permits since April 2020. It will start accepting data permit applications for data from Kanta Services when these go live in 2021.).

361 Parikka and others (n 286).

362 ten Thije, 'Interview 1' (n 280) (Findata's budget is established on a yearly basis).

363 ASU, section 50; Findata may charge for: 'the picking, delivery, combination, pre-processing, pseudonymisation and anonymisation … as well as for the use of a secure operating environment.' These fees are laid down in a decree by the Ministry of Social Affairs and Health: Regulation on Findata's fees (nr. 1500/2019) (Sosiaali- ja terveysministeriön asetus: Sosiaali- ja terveysalan tietolupaviranomaisen suoritteiden maksullisuudesta) (Fi).

364 ten Thije, 'Interview 1' (n 280) (Students must be working on their thesis and from within the EEA, as applications from outside the EEA cost a considerable amount of extra work due to compliance checks.).

a GDPR compliant manner.[365] Changing an already submitted and approved data request (which means adding authorized persons or extending the permit's period) costs €350, when a change requires new or updated data, Findata will consider and charge this as a new application.[366]

Findata fees

- A fixed fee for reviewing a data permit (€500–3.000) or data request (€500–1.000);
- An hourly rate for gathering and pre-processing data (€115/hour);
- A fixed fee for the secure processing environment (€2.250–8.500/year);

Figure 13 – Findata's fee structure

Secondly, Findata charges a €115/hour fee for the time its employees use to gather and pre-process data. Any work of Findata's employees on a data permit of which the deadline to deliver the data to the applicant has lapsed costs €75/hour. Thirdly, the use of the secure operating environment where data can be analyzed, costs €2.500–€8.500 per year, based on the computing power of the machine. Installing an applicant's own analysis software in the secure operating environment costs €115/hour.[367]

Fourthly, Findata also collects the fees that original sources of data may charge for their work in collecting the requested data. A data source is obliged to provide an estimate of the costs to fulfil a specific data permit of an applicant to Findata. Findata uses this figure to estimate the total costs for carrying out an application.[368] After receiving this estimation of total costs, an applicant can decide to accept the costs or forgo its application. In Finland the height of these fees is regulated per sector: public health care providers have to follow the cost principle, which means that they can only charge the actual costs of collecting and transmitting the data.[369] Private healthcare providers are, however, not subject to this principle. They can therefore charge higher fees, although they still need to provide some basis for their fee in the costs of collection and transmission of the data.[370] The height of private healthcare providers' fees is not clear yet, since Findata only started accepting applications for data permits as of April 1st 2020.

4.4 Interface with data protection law

Findata's access regime involves the processing of (sensitive) personal data, triggering the application of data protection law. This section describes the interplay between European (and Finnish) data protection law[371] and the *Act on Secondary Use of Health and Social Data*, which establishes both Findata as Data Permit Authority for Health and Social Data and the access regime around it. The interplay between these laws is

365 Data requests from applicants outside the EEA are not more expensive, since the requested data is aggregated and anonymised anyway, wherever the applicant originates from.

366 ten Thije, 'Interview 3' (n 311).

367 Findata, 'Hinnasto' (*Hinnasto*) <https://www.findata.fi/palvelut/hinnasto/> accessed 29 April 2020. (There are four different sizes of the secure operating environment available: small (8GB RAM, 4 cores, €2.250/year), medium (16GB RAM, 6 cores, €2.750/year), large (32GB RAM, 8 cores, €3.500/year) and MaxPower (90GB RAM, 20 cores, €8.500/year).).

368 ASU, section 50(3).

369 ASU, section 50(2); Ministry of Social Affairs and Health (n 363) (The fees for public health care providers and public registers are determined based on regulation (1500/2019) by the Ministry of Social Affairs and Health on Findata's fees.).

370 ASU, section 50(3) ASU; ten Thije, 'Interview 1' (n 280).

371 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) (Text with EEA relevance) [2016] OJ L119/1 (GDPR); Data Protection Act (nr. 1050/2018) (Tietosuojalaki Dataskyddslag) (Fi).

described in in four parts. First, section 4.4.1 focusses on the assignment of the processor and controller roles when data is shared for the purpose of scientific research and examines the relationship between these two roles in the different phases of the access regime. Section 4.4.2 briefly highlights how Findata complies with the GDPR's six principles for processing of personal data (Article 5). Section 4.4.3 summarizes the legal bases for Findata's seven purposes for the processing of health data and describes some of the additional legal requirements laid down in the ASU and Data Protection Act. Lastly, section 4.4.4 briefly highlights Findata's adherence to data protection by design and by default and discusses how Findata handles data subjects' rights and applications that want to derogate from these rights.

4.4.1 Distribution of responsibilities under the GDPR

- **Controller Art.4(7) GDPR:** 'The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.'
- **Processor Art.4(8) GDPR:** 'The natural or legal person, public authority, agency or other body which performs any operation or set of operations on personal data or on sets of personal data, whether or not by automated means, on behalf of the controller.'

Figure 14 – GDPR definitions of 'controller' and 'processor'

The Findata access regime involves different actors that act as controllers and/or processors of personal data at various stages of the data sharing process.[372] From a data protection perspective, it is important to identify which of these actors can be considered controllers and/or processors, with regard to what processing operations, involving what personal data. This is crucial in determining the distribution of responsibilities.[373] This section will clarify these distinctions specifically for the scenario of a scientific researcher applying for a data request or data permit with data from multiple sources through Findata. The three key actors in this scenario for the processing of personal data are: (a) the data sources; (b) Findata itself; and (c) researchers/applicants.

For the purpose of identifying the relevant controller and processor in the context of **data requests**, two phases can be discerned: data collection and pre-processing. With regard to the data collection (i.e. data gathering and transfer to Findata by the data sources), Findata can be considered the data

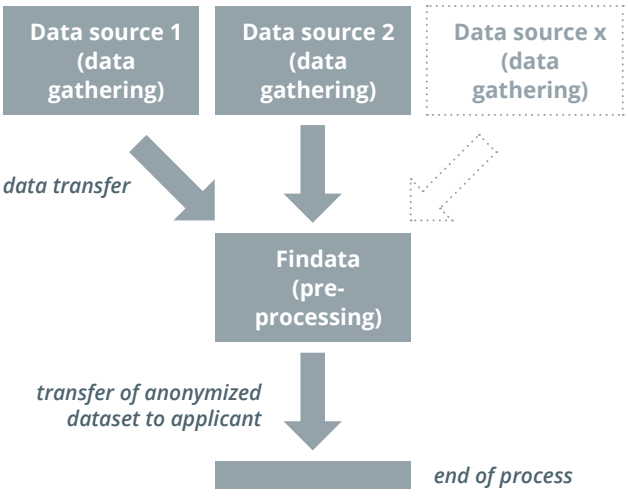


Figure 15 – phases in fulfilling a data request

372 Section 4.2.4 (For the description of the entire process of the access regime).

373 Responsibilities of processors and controllers are described in GDPR, chap IV (arts 24 – 34) and deal with their general obligations and responsibilities for the security of personal data, the data protection impact assessment and prior consultation, data protection officers and codes of conduct and certification; For more information, see: Article 29 Working Party, 'Opinion 1/2010 on the Concepts of "Controller" and "Processor"' (2010) WP 169 <https://ec.europa.eu/justice/article-29/documentation/>; Brendan van Alsenoy, *Data Protection Law in the EU: Roles, Responsibilities and Liability*, vol 6 (1e edn, Intersentia 2019).

controller: it determines the purpose and means of the collection (notably including which data must be gathered to fulfil the application of the researcher). The various data sources can be considered data processors with regard to the data collection: they gather and share the relevant personal data *on behalf of* Findata and transfer it into the secure hosting environment.

In the pre-processing phase, Findata can be considered data controller as well, determining both purpose and means of the respective personal data processing operations (i.e. aggregating, combining, matching, cleaning and anonymizing the data). In this phase, data processors do not fall within the GDPR's scope anymore. Finally, Findata transfers the aggregated statistics created from the personal data to the applicant and deletes the personal data it holds, bringing the process and its obligations as a controller and processor to an end (see Figure 15).

The distribution of responsibilities in the '**data permit** process' is slightly more complicated. In this process, three different phases of data processing can be identified, each with different controller/processor constellations: (a) data collection (data gathering + data transfer), (b) pre-processing, and (c) data analysis (see Figure 16). With regard to the data collection and pre-processing phases (a & b), the allocation of responsibilities is the same as in the context of data requests (i.e. Findata: controller for both collection and pre-processing; data sources are processors only for the data collection). Data permits imply that after personal data has been pre-processed, it is transferred into a secure operating environment where it can be accessed and analyzed by the researcher(s). In this third phase (c), the researcher can be considered a controller, determining the purpose and means of the analyses/ research conducted with the relevant personal data. Findata can be considered the processor, specifically with regard to the processing operations for research purposes in the secure operating environment during this third phase, as Findata's system of virtual machines carries out the processing *on behalf of* the researcher.[374] The process ends when Findata deletes the data from the secure operating environment.

The distribution of responsibilities as described above is not as clear cut as it may seem. One might also argue that in all three phases (and especially the first and last phases), the researcher/applicant is a *co-controller*, jointly with Findata. After all, the entire processing operation (from data collection to analysis) would not have taken place without the researcher/ applicant making a request and determining the overall 'purpose' for conducting research with the respective personal data in the first place.

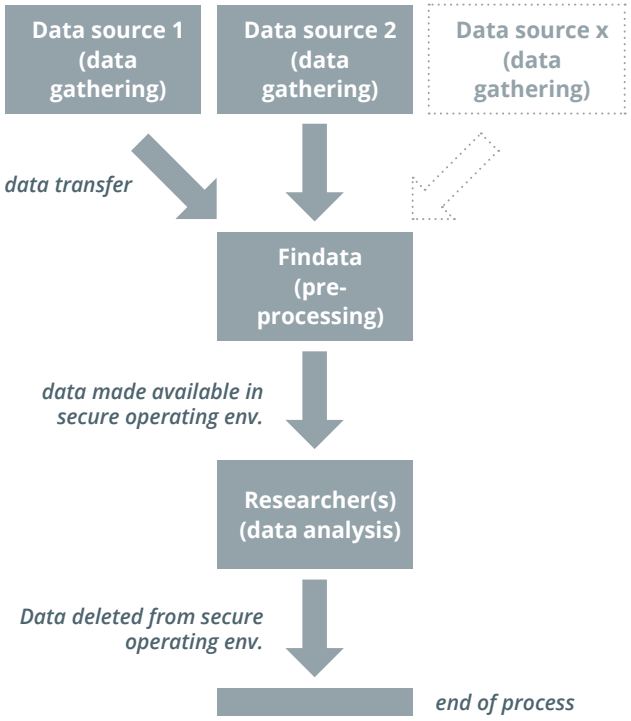


Figure 16 – phases in fulfilling a data permit

374 A further complicating factor is that the secure operating environments may be provided by third parties, contracted by Findata (cf. Section 4.2.4). Without going into detail on this relationship, it can be assumed that as subcontractors, tied to strict contractual requirements, these third parties can be considered processors as well.

CJEU case law states that when a party has a *decisive influence* over the collection by and transmission of personal data to a third party which would not have occurred without its actions, it must be considered a (co-)controller with regard to those processing operations.[375] It can be argued that in Findata's access regime the researcher/applicant has a *decisive influence* over the entire processing operation (from data collection to analysis). The researcher/applicant's data utilization plan could be seen as laying down the purpose for the data collection from the data sources by Findata, as well as the pre-processing by Findata and the actual data analysis in the last phase. With this in mind, why not consider the researcher/applicant the sole controller and Findata a mere processor throughout the entire processing operations? Following the same CJEU case law,[376] Findata can also be considered to have a *decisive influence* over the processing operations in all three phases. Indeed, even with regard to the actual research/analysis (third) phase, Findata provides the secure operating environment, maintaining full control over the researchers' access to the environment as well as logging all of their actions in that environment. Moreover, Findata itself claims that: '[It] becomes a controller of personal data when it receives data from the aforementioned operators'.[377] This wording implies that Findata remains a (co-)controller even during the data analysis phase and until the data is deleted from the secure operating environment. Findata thus explicitly assumes controllership for the whole data permit process.

This argument shows that Findata and the researcher could be considered joint controllers from the data collection until the analysis phase. Clearly, a complex data sharing framework such as Findata, involving (sensitive) personal data and many actors, raises a lot of hard questions as to the distribution of responsibilities under the GDPR. Having a central entity at the heart of the entire ecosystem (*in casu* Findata), stepping up and assuming controllership for the entire process (from data collection to deletion) generates considerable legal certainty (and therefore trust between all stakeholders). It is important to note however, that it is not possible to contractually assign controllership (and all associated responsibilities under data protection law) to an entity that does not fulfill the legal requirements for being considered a controller (i.e. determining the purpose and means of data processing). Put differently, identifying the controller for any given data processing operation requires a functional assessment.[378] As apparent from the previous paragraphs, however, Findata does appear to satisfy these requirements, at least for the three phases identified in Figure 16.[379] As Findata explicitly assumes controllership, the question of whether or not the researcher/applicant can be considered a co-controller becomes less urgent.

4.4.2 Compliance with data protection principles

Article 5(1) GDPR lists six data protection principles: (a) lawfulness, fairness and transparency; (b) purpose limitation; (c) data minimization; (d) accuracy; (e) storage limitation and (f) integrity and confidentiality.[380] Every (joint) controller has to comply with

375 Case 40/17 *Fashion ID GmbH & Co KG v Verbraucherzentrale NRW eV* [2019] EU:C:2019:629 [78]; Case 210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig- Holstein GmbH* [2018] EU:C:2018:388.

376 ibid.

377 Findata, 'Data Protection and the Processing of Personal Data' (*Findata*) <https://www.findata.fi/en/about-us/data-protection-and-the-processing-of-personal-data/> accessed 1 May 2020.

378 Article 29 Working Party (n 373), van Alsenoy (n 373).

379 It goes without saying that the data sources are controllers with regard to the processing of personal data files before they are requested by Findata. Similarly, researchers/applicants are to be considered controllers in case they would manage to extract personal data from the secure operating environment (which would also constitute breach of contract).

380 GDPR, art 5(1).

these principles for every processing purpose they are responsible for. Even if the previous sub-section demonstrated that researchers/applicants as well as data sources can be considered (co-)controllers or processors for certain processing operations, this sub-section will focus on Findata in particular, as it is the central party in this access regime, assuming the main responsibility to assure these principles are met in all phases of processing.

- **Lawfulness** – As further discussed in the next section (4.4.3) all data processing operations facilitated by Findata must be in line with one of the seven purposes determined in the ASU.[381] When it comes to the lawfulness of its processing operations, Findata relies on the fifth ground in Article 6(1) GDPR; i.e. necessity for the performance of a task carried out in the public interest or for the exercises of official authority (Article 6(1)(e), GDPR),[382] and, to the extent it processes 'special categories of personal data' (health data), necessity for reasons of substantial public interest (Article 9(2)(g), GDPR).[383] These grounds ensure the lawfulness Findata's processing activities for the access regime.[384]

- **Fairness and Transparency** – Additionally, processing must be both fair and transparent, the latter of which means that information on the processing must be 'concise, easily accessible and easy to understand' and in clear and plain language.[385] Findata provides this transparency mainly via its website, describing the different steps in its access regime.[386] The fairness principle is to be interpreted in relation to several other data protection obligations,[387] and appears to be complied with by Findata which takes numerous steps to minimize negative externalities stemming from the processing of personal data.

- **Purpose Limitation** – In principle, personal data may only be 'collected for specified, explicit and legitimate purposes and [may] not [be] further processed'.[388] Health data rendered accessible through Findata has been initially collected/processed for the purpose of providing health care to patients. Under the purpose limitation principle, this original purpose for data collection would be the only purpose this data could be lawfully processed for. However, the GDPR allows the Member States to expand this primary purpose with secondary purposes, such as archiving, scientific or historical research or statistical research;[389] for reasons of (substantial) public interest;[390] for preventive or occupational medicine[391] or public health.[392] The ASU provides the lawful grounds for the seven secondary data processing purposes allowed by Findata.[393] The Finnish Data Protection

381 ASU, section 2.
382 Findata, 'Data Protection and the Processing of Personal Data' (n 377) ('Findata as a controller').
383 ibid.
384 GDPR, art 5(1)(a).
385 GDPR, recital 58.
386 Findata, 'Findata – Health and Social Data Permit Authority | Tervetuloa!' (*Findata*) <https://www.findata.fi/en/> accessed 11 June 2020.
387 Damian Clifford and Jef Ausloos, 'Data Protection and the Role of Fairness' (2018) 37 Yearbook of European Law 130.
388 GDPR, art 5(1)(b).
389 GDPR, art 5(1)(b) ('[F]urther processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes'); GDPR, arts 9(2)(j) & 89(1).
390 GDPR, arts 6(e) & 9(2)(g).
391 GDPR, art 9(2)(h).
392 GDPR, art 9(2)(i).
393 ASU, section 1.

Act expands further on the purpose of scientific and statistical research.[394] Findata also appears to put clear safeguards in place so that the further data processing is strictly limited to those seven legally allowed purposes only.

- **Data minimization –** The GDPR prescribes that processing of personal data must be in accordance with the data minimization principle.[395] This principle means that Findata should only provide the minimum amount of data that is necessary to fulfil the purpose of a data request or permit. For example, a researcher plans to investigate the frequency of a specific treatment in Finnish hospitals. This research goal can be achieved using aggregated data of patients who suffer from this illness. While additional personal data of these patients, such as their place of residence, could show whether they chose specific hospitals to receive their treatment, the collection of this extra data is not necessary to carry out the original research plan and is therefore not provided. Findata may only provide a data permit (access to individual-level personal data) if a data request (aggregated statistical data) cannot answer the research question.[396]

- **Accuracy** – Findata must also ensure the accuracy of the data it provides via its access regime.[397] Findata's setup guarantees a high level of accuracy and up-to-dateness. Findata always retrieves data anew from the primary source (hospital or register) for every applicant, just before the data is processed by that applicant.[398] Therefore, it assures that the applicant has the last available version of health data for its research. There is, of course, the possibility that the (validity of this) data changes at the source while research is already being conducted, but this is the case for any access regime.[399] While individuals can exercise their rights as data subject to rectification of data and change incorrect data that Findata holds on them, this would have a short-lived effect (data is deleted after the research is finished).[400]

- **Storage Limitation** – The Findata access regime also complies with the storage limitation principle[401] by ensuring that health data is not made available and stored for longer than necessary in light of the approved data utilization plan of the researcher/applicant. It does so by providing every data permit only for a limited amount of time after which Findata permanently deletes the data it concerns.[402] Additionally, those with access to health data through a data permit are incentivized to limit the time they spend analyzing this data, as a small part of the fee they pay for using the secure processing environment is time-dependent.[403]

- **Integrity and Confidentiality** – Finally, Findata's access regime must ensure the integrity and confidentiality of the health data it provides access to.[404] This entails that Findata provides appropriate security and protection against unlawful

394 Data Protection Act, sections 4(3)-(4). 6(4) & 6(7)-(8),
395 GDPR, art 5(1)(c).
396 ASU, section 43(5).
397 GDPR. art 5(1)(d).
398 Section 4.2.4.
399 For example, if a care provider corrects data while Findata has already gathered the data from its system. This is similar for another access regime or case where the researcher collects data itself at a certain point in time, which can contain flaws.
400 Section 4.4.4.
401 GDPR, art 5(1)(e).
402 Section 4.2.4.
403 Section 4.3.5.
404 GDPR, art 5(1)(f).

processing, accidental loss, destruction and damage of data it holds. Security measures are further discussed below (Section 4.4.4: data protection by design and default). It should be noted here that accidental loss, destruction and damage of data are rendered unlikely to occur, due to the structure of Findata's access regime and role as an intermediate party: providing access to only a (temporary) copy of the health data in a closed and monitored IT-environment.[405]

4.4.3 Findata's seven purposes for data sharing

The ASU defines seven purposes for which it allows secondary use of health and social data as facilitated by Findata. Put briefly, these purposes are: (1) statistics, (2) scientific research, (3) development and innovation activities, (4) education, (5) knowledge management, (6) steering and supervision of social and health care by authorities, and (7) planning and reporting duty of an authority.[406] These purposes will briefly be discussed below, highlighting their lawful ground for processing under the GDPR and any specific requirements the ASU and/or Finnish Data Protection Act set for them to be applied.

- **Data Sharing for Statistics & Scientific Research** – Statistics and scientific research are legitimate purposes for the sharing of data personal held by Finnish national data registers and private health care provides via a data permit.[407] Processing for these purposes appears to be implicitly based on Article 9(2)(j) GDPR ('processing necessary for scientific research or statistical purposes).[408] The ASU states that 'freedom of scientific research must be ensured when a data permit is procured', but does not further specify what this freedom entails exactly.[409] Importantly, from the seven purposes, only data processing for the purpose of statistics and scientific research and for the purpose of education can justify derogations from data subjects' rights.[410]
- **Data Sharing for Development and Innovation** – Researchers/applicants can also request aggregated statistics (not data permits) from Findata for development and innovation purposes other than scientific research.[411] These requests must have one of three specific purposes listed in the accompanying data utilization plans: (1) to promote public health or social security; (2) to develop social and health care services or the service system; or (3) protect the health or wellbeing of individuals or secure their rights and liberties associated with health or wellbeing.[412]Applicants can only be granted a data request innovation purposes for the above-mentioned purposes.
- **Data Sharing For Educational Purposes** – The lawfulness ground for the use of health data in education is Article 9(2)(g) ('reasons of substantial public interest').[413] Data should be used to produce educational material for people working in social and health care, or those studying to work

405 Section 4.2.4.
406 ASU, section 2.
407 ASU, section 38.
408 European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (2020), 17 <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>.
409 ASU, section 38.
410 ASU, section 39; See also: Section 4.4.4.
411 ASU, section 37.
412 ibid.
413 ASU, section 39.

in those areas. To legitimize an application with this purpose, it must meet three conditions: (1) the data must be necessary to meet the goals of education; (2) the education cannot be provided with anonymous data (for example, through the rarity of the case or the nature of teaching) and (3) the teacher must always inform students of their statutory secrecy obligations and the sanctions for breaching these.[414]

- **Data Sharing for Knowledge Management** – Knowledge management refers to the processing of data by a health or social care provider or a (joint) municipality (authority) in its customer, service and production processes to support its operations, management and decision making.[415] For example, for a hospital to compare itself with others who provide similar health or social care, or for a municipality to check its offering of care with others. For the evaluation of its operation through the comparison with others, a care provider can only get statistical data through a data request.[416] However, health service providers do not need a data permit or data request to use the data generated while providing primary care itself,[417] and a (joint) municipality (authority) may process and combine data from its joint registers, without a data permit or request.[418] The lawful ground for this purpose is Article 9(2)(h) GDPR.[419] To evaluate its operations with its own data, a care provider or municipality does not need a data permit. The use of data for this purpose is further facilitated by the Ministry of Social Affairs and Health in the Toivo-program.[420]

- **Data Sharing For Steering and Supervision of Social and Health Care by Authorities** – A steering or supervisory authority for social and health care that needs combined data, based on personal data in social and health care registers or personal data in other registers, can request this data from Findata.[421] The lawful grounds for this processing purpose are Article 9(2)(g) and Article 86 GDPR (processing of and public access to official documents).[422] When Findata considers such a data request, it must follow the normal procedure for deciding on a data request, and therefore also take into account the guidelines of its high-level expert group on anonymization, data protection and data security.[423]

- **Data Sharing for Purpose of The Planning and Reporting Duty of An Authority** – A supervisory authority such as Valvira, The Finnish Supervisory Authority for Welfare and Health, must have access to information on health and social care providers to carry out its tasks.[424] It must, for example, be able to report on the performance of public hospitals and plan its supervisory tasks accordingly. The legal grounds for the sharing of

414 ibid.
415 ASU, section 41.
416 ASU, section 41(2).
417 ASU, section 41(1).
418 ASU, section 41(3).
419 GDPR, art 9(2)(g) ('[P]rocessing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services …').
420 ten Thije, 'Interview 3' (n 311) (This government program is meant to develop knowledge management in the provinces of Finland and to produce knowledge management information by national authorities, data resources and the tools that support them.).
421 ASU, section 42(1).
422 ASU, sections 45(1)-(2),
423 ASU, section 8(4); See also Section 4.3.1.
424 ASU, section 40.

data with supervisory authorities, such as Valvira, is Article 9(2)(g) GDPR.[425]

This subsection has described the legal grounds and additional conditions for the processing of health and social care data for the seven purposes laid down by the ASU. 'Investigative' purposes, such as scientific research; the planning and reporting duty; and the steering and supervision of social and health care authorities are legitimized by the ASU under Article 9(2)(j) (scientific research) and Article 9(2)(g) (public interest or public health) of the GDPR, as expected. While the ASU and Finnish Data Protection Act do lay down some specific provisions for each of these purposes, most of these purposes merely emphasize the general conditions that apply to all processing purposes allowed under Findata's access regime. Significantly, however, the law does constrain one purpose – i.e. development and innovation activities – to data requests only (and not data permits) and limits the purpose of knowledge based management to data requests only when health care providers want to compare their data to that of others.

4.4.4 Appropriate safeguards: technical & organizational measures

Article 89 of the GDPR states that processing for scientific research purposes shall be subject to appropriate safeguards, which must ensure technical and organizational measures are in place to ensure the rights and freedoms of data subjects.[426] This section highlights two examples of data protection by design and default measures in Findata's access regime. Finally, it discusses the possibility for Findata to derogate from data subjects' rights.

Data protection by design & default

The Findata access regime is structured in such a way that it complies with the Data Protection by Design and Default requirement of the GDPR: 'only personal data which are necessary for each specific purpose of the processing are processed.'[427] This is, for example, assured at the start of the application process for a data request or permit: in judging the application, Findata has the obligation to check whether the purpose of a data permit can also be fulfilled through a data request.[428] If this is the case, the personal data in the permit application will never be accessed by the applicant.

Another example of data protection by design and default in Findata's access regime is Findata's role as an intermediary between the health data and the applicant. Findata is present in every step of the process. Either to check plans (application, data utilization plan) or actions (analysis in the operating environment, results from analysis to be published) of the applicant.[429] And, Findata plays a crucial role in the pseudonymization and anonymization of the personal data before making it available for analysis, preventing researchers/applicants from accessing directly identifiable personal data.[430] All these examples show the access regime's structure is built with security and data protection by design and default in mind.

425 GDPR, art 9(2)(g) (furthermore processing must be: 'proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;').
426 GDPR, art 89(1).
427 GDPR, art 25(3).
428 ASU. section 43(5).
429 Section 4.2.4.
430 ibid.

Derogations from data subject rights

The GDPR provides data subjects with several rights, including the right of access, to rectification, erasure, data portability, and to object.[431] Findata has implemented an IT-tool through which Finnish citizens whose data is processed can exercise all of these rights.

The GDPR explicitly provides Member States with the opportunity to lay down rules to derogate from four data subject rights (access, rectification, restriction of further processing and to object) 'where personal data are processed for scientific or historical research purposes or statistical purposes'.[432] For example, if a researcher wants to investigate a rare but deadly illness, but a considerable amount of patients have exercised their right to object to processing of their (personal) medical data. In such a case, Findata could decide not to grant those rights (if the processing relates to scientific research purposes). Findata gets this mandate from Article 31, Data Protection Act providing for derogations in case scientific research would be hindered by data subjects' rights.

In practice, 'Findata does not restrict data subjects' rights by its own initiative. Restrictions are only applied if a decision has been made to restrict the rights of a data subject in connection with the research project for which the data permit is applied at Findata.'[433] In that case, the application must adhere to four conditions laid down by the Finnish Data Protection Act:[434]

- the processing is based on an appropriate research plan;
- a person or group responsible for the research has been designated;
- the personal data are used and disclosed only for scientific or historical research purposes or other compatible purposes, and the procedure followed is also otherwise such that data concerning a given individual are not revealed to outsiders;[435] and
- a data protection impact assessment[436] must be carried out for the intended combination of health data gathered by the researcher and obtained via Findata, or a specific code of conduct, which sufficiently deals with derogations to data subjects' rights, is complied with.[437]

The first three conditions are ensured through Findata's general procedures for the granting of a data permit or access request. These require a data utilization plan (appropriate research plan), which list a limited number of persons getting access to the shared data (responsible person or group); data is pseudonymized and draft results are checked before publications (ensure secrecy of personal data). To comply with the last condition, the researcher has to provide a data protection impact assessment (DPIA) to Findata to inform about the specific processing for scientific research. The DPIA also has to be submitted to the Finnish Data Protection Authority before the data processing has started.[438] Permit applicants can therefore never restrict any rights by themselves.

---

431 GDPR, ch III.

432 GDPR, art 89(2) ('Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.').

433 Findata, 'Data Protection and the Processing of Personal Data' (n 377).

434 Data Protection Act, sections 31(1) & 31(3).

435 Data Protection Act, section 31(1).

436 GDPR, art 35.

437 GDPR, art 40.

438 Data Protection Act, section 31(3).

As stated above, Findata uses an online environment to manage the requests of Finnish citizens who want to exercise their rights under data protection law. Citizens can log in and send requests about data pertaining to them via Suomi.fi. This is a website operated by the Digital and Population Data Services Agency. After this system has verified their identity, they can send in e.g. access requests, which will be handled by Findata. Citizens can choose to opt out from use of their data for all future research. Opting out of the use of one's data via this system only results in the erasure of Findata's copy of the subject's data. The choice to opt-out of future research can be reverted later on if a data subject decides they are willing to participate in new research.[439]

It is important to note that the choices in Findata's rights management system do not affect the copies of citizens' data held at the data sources, e.g. the hospitals, private clinics or health registers. To exercise their rights on that data, citizens have to do so at the individual sources, for example, their public or private health care provider or a public register. These parties might have other systems and procedures to deal with these requests. Findata's rights management system thus only has an effect on the copies of data it stores temporarily to fulfil the data permits and requests it receives.[440]

The GDPR and the Finnish Data Protection Act lay down several conditions before Findata may derogate from data subjects' rights to access, rectification, restriction of further processing and to object. In practice, Findata is reluctant to do this and even provides an online rights management system for citizens to exercise their rights in an easy manner. This system of course only effects citizens' data under Findata's control and does not affect copies of data kept by the data sources. To exercise their rights over that data, individuals have to contact these controllers separately. Overall, Findata's proactive and accommodating position vis-à-vis data subject rights can be seen as a further manifestation of its data protection by design/default obligations, also generating a more trustworthy ecosystem overall.

439 ten Thije, 'Interview 3' (n 311).
440 ten Thije, 'Interview 1' (n 280).

4.5 Lessons Learned

Request-based, adaptive regime	One of the interesting features of the Findata access regime is its request-based, adaptive nature: it puts in place a comprehensive, demand-driven and purpose-based access framework where data access is provided on a case-by-case basis. Such an approach is particularly valuable for regimes that involve (sensitive) personal data and/or where that data might not be static at its source. The comprehensive infrastructure put in place by the Findata regime is designed to maximize research potential of medical data, while at the same time safeguarding privacy/data protection interests. As highlighted in Chapter 2, a meaningful platform research access regime will also include personal data (notably in relation to social media users). The Findata model could serve as a blueprint for enabling access to that data without compromising those data subjects' rights and freedoms.
Mandatory data sharing	The Findata access regime gives the independent institution at the center, i.e. Findata, a legal mandate to demand data from a selected number of data sources. This is important in light of the ad hoc nature of how access requests are accommodated (i.e. each request is evaluated and accommodated on a case-by-case basis). In light of platforms' powerful position, and the experience of non-binding data access initiatives as detailed in Chapter 2, a clear legal obligation to share data upon request is essential in creating a robust and meaningful access regime for platform data. It is also important that this obligation only relates to a well-defined, legally circumscribed, and trusted independent institution (instead of, for example, researchers requesting access directly from the platform of interest).
Iterative regulation	Before drafting the law which eventually called Findata into life, a 'digital health hub' was created. This hub was specifically designed to explore the needs of a comprehensive access framework. It also launched eight pilot projects aimed at exploring various aspects of a potential access regime for health data. The insights gained through this start-up phase were key in the design of the eventual law and Findata's operational specificities. Designing a robust access framework for platforms will be no easy feat. Policy-makers may wish to consider a similar approach to Findata, starting with an exploratory phase where different aspects of a potential access regime are piloted and tested.

Pre-processing	A crucial feature of the Findata access regime is that data is never sent directly from the source to the researcher/applicant. The data is first collected from the relevant sources and combined where relevant, as well as pseudonymized, after which it is made accessible within a secure operating environment. This process also includes a proportionality assessment, where Findata evaluates, in dialogue with the applicants/researchers, whether their request can also be accommodated with less data (e.g. a data request instead of a data permit). Overall, this pre-processing phase offers assurances to data sources (trust), researchers (data quality), and data subjects (privacy/data protection). For these reasons, such a pre-processing phase may also prove quite useful in a platform research access regime. It might look slightly different however, as the combining of data from different sources may not be as straightforward as in the Findata context (where medical files can generally be tied back to national identification numbers). In any case, the pre-processing phase should be made as transparent as possible (i.e. describing in detail the different operations data underwent), so as to ensure accountability and trustworthiness of the access framework overall.
Different forms of data access	Access requests in the Findata regime can be accommodated in two ways: a data permit or a data request. A data request only provides access to a dataset with aggregated, anonymized statistics. A data permit, on the other hand, gives the applicant/researcher access to a pseudonymized dataset (in a secure operating environment) on which they can run their own analyses. The applicant/researcher will need to clearly justify their choice for one or the other, and Findata performs a proportionality assessment on applicants' research plans. This approach is intended to minimize privacy and data protection issues, but could also render the access regime useful to a wider audience; i.e. a data request enables a less expert-audience to ask a question, while Findata actually performs the analysis. Given the complexity of the platform data ecosystem, as well as their wide societal impact, a similar approach would be valuable in a platform access regime as well. It would render that regime more accessible to the public at large and offer additional safeguards for competing rights, freedoms and interests. These forms should not necessarily be limited to two types only, but more approaches could be explored.
GDPR Compliance	The Findata regime demonstrates that a comprehensive, widely accessible research access regime can provide very rich data and be GDPR compliant. This is perhaps the most important lesson learned from this case study. The establishment of an independent institution (section 5.1.2) rigorously guarding GDPR compliance plays a vital role in this. So does the elaborate technical (e.g. secure operating environment) and procedural (e.g. exhaustive list of seven purposes) infrastructure as well as the multiple levels of oversight. The Findata model neutralizes the many GDPR-based objections platforms often raise to block access for researchers (cf. Chapter 2). Policy makers may therefore take inspiration from this regime in order to develop a robust and GDPR-proof framework for platform research access.

5 Lessons learned for research access in platform governance

The deep-dive into two case studies has illustrated how operational data access regimes in other sectors have tackled some of the same challenges also faced in the context of platform governance. The first case study was specifically aimed at learning from another sector with strong disincentives for data disclosure, since data access is intended to enable corporate accountability. The second case study was aimed at exploring the legal and policy requirements for building a robust transparency framework for sharing (highly sensitive) personal data in pursuit of (scientific) research. Figure 17 and Figure 18 reiterate the key takeaways from these case studies.

In addition to these discrete takeaways, specifically targeting the respective case study challenges, **this Chapter will reflect on some cross-cutting best practices** that appear between these case studies. It closes by identifying open questions that remain unanswered, but nonetheless require careful attention in the particular context of platform governance.

Before turning to these lessons, it is important to reiterate that these case studies are not intended to provide integral or comprehensive blueprints for data access in platform governance debate. Rather, they are intended to highlight specific challenges and potential avenues for tackling them. As such, the case studies should instead be seen as helping to design specific cogs within the larger platform governance machinery. They can, for instance, offer answers to questions on suitable governance structures and procedural safeguards when designing scientific access frameworks, but they cannot tell us what the substantive issues within platform governance are that deserve greater transparency. With this disclaimer

Countering the incentive problem (p.49):

- Specific disclosure rules
- Standardised methods for data generation
- Quality assurance
- Size-based regulation
- Transparency by default
- Public transparency by default
- Tiered oversight structure
- Sanctions / penalties
- Explicit call for awareness raising
- Strict timing

Figure 17 – Lessons learned from the E-PRTR case study

Countering data protection concerns (p.79):

- Request-based, adaptive regime
- Mandatory data sharing
- Iterative regulation
- Pre-processing
- Different forms of data access
- GDPR Compliance

Figure 18 – Lessons learned from the Findata case study

in mind, **several best practices can be derived from** both **case studies, central to tackling both the 'incentive problem'** and **'data protection concerns'**:

- Binding rules;
- Independent institutions;
- Tiered regulation;
- Proactive support for researchers
- Public transparency by default;
- Verification and pre-processing by independent institutions

5.1 Cross-cutting best practices

5.1.1 Binding rules

> '[I]n many societies—especially democratic capitalist societies—the major threats to citizens' interests come not from government, but rather from corporations and sometimes secondary associations. In such societies, citizens' main informational interest—democratically speaking—is in the kinds of information that can help them to manage the risks imposed by those organizations and to tame them. Their government is often the only organization with the wherewithal to wrest this information from powerful corporate and social actors.' [441]

Both case studies are characterized by a binding regulatory framework setting out the key obligations and responsibilities relating to the data sharing regime. In the E-PRTR case study, binding regulation is crucial in order to ensure complete, consistent and comparable information is shared by industry. In the Findata case study, binding regulation is important in order to enable access to otherwise sealed-off data by ensuring a high level of privacy and data protection. Both legal frameworks specifically lay down in law a number of key elements;

- the *governance structure* of the independent institution regulating the day-to-day operation of the data sharing framework;
- *who* can directly access data or can apply for access;
- *what* specific data is accessed;
- *how* and *by whom* that data is to be gathered and checked before disclosure;

A clear and consistent legal framework is vital in establishing a robust research access framework in the platform governance context. Indeed, as discussed in Chapter 2, self-regulatory initiatives – e.g. Facebook oversight board,[442] ad archives,[443] transparency reports[444] – are far from sufficient to meet the needs of a meaningful research access regime.[445] Yet, the more these platforms integrate themselves into our economy and society, the more critical a robust and sustainable access framework becomes in order

441 Archon Fung, 'Infotopia: Unleashing the Democratic Power of Transparency' (2013) 41 Politics & Society 183, 190.

442 Catalina Botero-Marino and others, 'We Are a New Board Overseeing Facebook. Here's What We'll Decide.' *The New York Times* (6 May 2020) <https://www.nytimes.com/2020/05/06/opinion/facebook-oversight-board.html> accessed 13 May 2020.

443 Paddy Leerssen and others, 'Platform Ad Archives: Promises and Pitfalls' (2019) 8 Internet Policy Review <https://policyreview.info/articles/analysis/platform-ad-archives-promises-and-pitfalls> accessed 7 February 2020.

444 Daphne Keller and Paddy 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* (Cambridge University Press 2019) <https://papers.ssrn.com/abstract=3504930> accessed 8 January 2020.

445 Also see: Council of Europe Committee of Ministers to Member States, Recommendation CM/Rec (2018)1 on Media Pluralism and Transparency of Media Ownership (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies), paras 4.8-4.13; European Regulators Group for Audiovisual Media Services, 'ERGA Position Paper on the Digital Services Act ERGA 2020 Subgroup 1 – Enforcement' (2020) <http://erga-online.eu/wp-content/uploads/2020/06/ERGA_SG1_DSA_Position-Paper_adopted.pdf>, para 26.

to study their impact and hold them accountable. This is especially pressing in light of the incredibly powerful position these platforms find themselves in, essentially constituting (significant portions of) society's informational infrastructure. In this context, the law can play a crucial role in mitigating power asymmetries, levelling the playing field between platforms and (in this context) researchers. Moreover, platforms themselves also benefit considerably from the legal certainty offered by a sound regulatory framework.

It also appears evident that regulation ought to be undertaken at the EU level. Similar to the environmental protection context, where externalities ignore national borders, a meaningful research access framework in the platform governance context will equally require a wider, European approach. As recognized in the E-PRTR Regulation, the 'need for comparability of data throughout the Member States argues for a high level of harmonization'.[446]

Despite hard law being essential in developing a robust and effective research access regime, it is also important to recognize the limitations of relying on hard law alone to make such a regime operational and effective. What can be learned from the case studies is that while the law might set down the baseline structure (of principles, obligations and governance), a considerable part of the eventual access regime will be given shape through a range of self- and co-regulatory tools, such as standard setting, best practices and through stakeholder dialogues. As demonstrated by the case studies – and will be discussed further below in Section 5.1.2 – strong independent institutions can play a central role in making this possible. Finally it is also important to flag challenges related to scope and legislative competencies, complex legal issues that are tackled in the Cornils Report.[447] An open question here, in light of platforms' ever-expanding reach – e.g. into media, transportation, health, etc. – is the intersection of horizontal and sectoral approaches to data access regulation (see section 5.3.3 for further discussion).[448]

5.1.2 Independent institutions

A clear common denominator in both case studies is the presence of a strong, legally ordained, independent institution at the center of the data access framework. The data access framework in the E-PRTR case study relies on proactive independent institutions (i.e. environmental authorities) both at the Member State and EU level. These authorities have a legal duty to enable, facilitate and promote the access framework among the public at large. Moreover, they have a clear legal mandate to take (enforcement/investigatory) action in case of non-compliance. Similarly, in the Findata case study, the independent institution is well-established at the heart of the access regime (i.e. Findata) and plays a vital role in making it operational. Indeed, Findata has clear legal duties to facilitate the data access between a multitude of different entities, while at the same time safeguarding privacy and data protection.

This constitutes an important 'lesson learned' for developing an access regime in the platform governance context. The case studies demonstrate that **a strong independent institution can act as an important bridge builder between those holding the data and those wishing to get access to that**

446 E-PRTR Regulation, recital 18.

447 Matthias Cornils, 'Designing Platform Governance: A Normative Perspective on Regulatory Needs, Strategies, and Tools to Enhance the Information Function of Intermediaries' (AlgorithmWatch 2020) <https://algorithmwatch.org/en/governingplatforms/legal-study-cornils-may-2020>.

448 Cf. In light of platforms' ever-expanding reach, policymakers may wish to explicitly allow for sector-specific rules to build on top of a generic data access framework. A relevant example here is the GDPR, which constitutes a very wide-reaching generic framework laying down key principles, but also explicitly inviting for sector-specific rules to be built on top of it.

data. Not only can they enforce access to data under the threat of sanctions; they can also act as a neutral arbiter in deciding on requests for confidentiality from the disclosing party (based on e.g. intellectual property or data protection law), and in periodically auditing disclosing parties to verify the accuracy of disclosures. Indeed, such a role is particularly relevant in the platform context, characterized by strong power asymmetries. As such, **independent institutions are needed to level the (data research) playing field with the force of law behind them**. Moreover, and as alluded to in the previous subsection (5.1.1), such an institution is also essential in making sure that abstract rules are translated into an operational infrastructure for research access. Achieving this will of course require a very clear governance structure in law; distributing responsibilities, liabilities and resources. Both case studies offer valuable illustrations of how this can be achieved in practice.

Crucially in the platform context**, the legal mandate of these independent institutions should be constrained to enabling the (research) data access framework as such**. As mentioned before, platform activities reach into many different regulatory frameworks, some of which are not harmonized and differ strongly across Member States. This is notably the case with regard to content moderation issues. In order to prevent competency-issues and minimize the politicization of a potential platform data access framework, it is advisable that the role of institutions is limited to being 'transparency facilitators'. The resulting transparency can then feed into enforcement actions by the competent (e.g. consumer protection, data protection, competition, media, transport) authorities.

Even if constrained to merely enabling transparency, legitimate concerns may still arise over the EU and/or Member States establishing such an institution. Indeed, trust in government bodies differs widely across Member States and it is therefore important to install sufficient safeguards and guarantees for independence. The case studies offer valuable lessons in this regard, from different layers of accountability-mechanisms (cf. Section 5.1.3), to constitutional hooks (notably to the obligation to ensure freedom of scientific research in the Findata case study[449]), awareness-raising obligations and proactive support for researchers (cf. Section 5.1.4). Furthermore, it is also advisable to install adequate transparency of the institution's operations itself, including information such as the number and type of requests, enforcement actions (c.f. Section 5.2.1).

An important element to consider when learning from the E-PRTR case study, is that the **jurisdiction and enforcement challenges** in the platform context will differ significantly. The high volume of corporations falling within the E-PRTR's access regime, which numbers in the tens of thousands, necessitates delegation to Member States in order to render it operational. In the platform context however, the number of corporations falling within a potential access framework will presumably be much lower, shifting enforcement challenges. With this in mind, a more centralized EU-level institution might be more advisable, also considering the political-economic power and multinational dimensions of platform operators. That said, Member State level authorities might still have an ancillary role to play in raising awareness and providing support to researchers (cf. Section 5.1.4).

A final element to consider relates to the funding of these independent institutions. **In light of the complexity and scale of many platforms, a comprehensive data access framework will require significant resources**. The case studies offer two potential models – publicly funded (E-PRTR) and fee-based (Findata) – but several other models are imaginable (e.g.

449 ASU, section 38.

a tax on the targeted operators). It is hard to make conclusive recommendations on what would be the most appropriate model in the platform context. Further research on this aspect is needed, considering in particular the resources needed in function of the eventual scope of the access framework (i.e. *what* data, *which* actors, *how* will access be facilitated) and rendering it effective (i.e. capacity and resources needed by those using the data). In any case, it should be recognized that significant (public) resources *are* required in order to confront platform power and produce a more accountable (and trustworthy) platform environment. A robust data access framework constitutes a vital component in this effort.

5.1.3 Tiered regulation

The two case studies show that access frameworks themselves require checks and balances to ensure accountability of (and trust in) the oversight system itself. In the case of the E-PRTR regulation, for instance, the regulatory structure involves multiple layers of oversight, with the activities of national authorities overseen by the EEA and ultimately the European Commission. This tiered structure centers local regulators who are more attuned to local sensibilities, and are likely more approachable for regulated parties, researchers, and other affected stakeholders, whilst ensuring a degree of uniformity based on the overarching EU framework.

In the regulation of dominant global platforms, the enforcement challenges are different than with the thousands of domestic industrial facilities at issue in the E-PRTR. Given the relatively smaller number of regulated entities, a singular EU-level access framework may be more feasible and more efficient. This being said, it is worth exploring how this framework can continue to leverage the authority and expertise of national authorities. For instance, in the area of media regulation, national authorities play a central role and EU competence is relatively limited. If data access is then regulated at the EU level, it is **important to consider how this framework can best interface with national regulators to assist in their respective policy and enforcement agendas.**

In the case of Findata, one also sees multiple layers of oversight, with many different government entities responsible for overseeing different aspects of the operation. This also demonstrates how effective governance of complex, ambitious access frameworks may necessitate multiple governmental perspectives (e.g. data protection authorities, cyber-security agency, media regulators). **Collaboration between these different oversight bodies can be made a structural feature of data access regimes**, in order to ensure that all relevant concerns are addressed and the risk of failure or negligence by any given authority (such as through capture or negligence) is minimized. With this in mind, it is advisable for a future platform (research) data access framework to explicitly lay down these different levels of oversight and how they interact, so as to ensure adequate accountability.

Closely related to the principle of accountable public institutions, is **the principle of research independence**. As part of the checks and balances imposed on government oversight, they should not attempt to influence research agenda's and outcomes from third party researchers. This is also reflected in both case studies. For instance, the Findata regime includes a clause requiring 'freedom of scientific research' to be ensured.[450] Such features allow for relatively broad usage of the regime, and reduce the risk of bias or capture affecting the overall research agenda. Another approach, seen in the E-PRTR case study, is to focus on public datasets. Since public

450 ibid.

datasets are accessible unconditionally, they protect research independence by their very design. That said, researchers are of course still required to comply with applicable laws and ethical research standards. A platform (research) data access framework can also promote the development of such standards for responsible data use.

5.1.4 Proactive support for researchers

State policy has a role not only in making data *accessible*, but also in actively stimulating its use by researchers. In both of the above case studies, the government assumed a proactive stance to ensure that their data access frameworks found uptake. The E-PRTR Regulation requires both the Commission and Member States to raise awareness of the framework at national level, and also to provide assistance in accessing, understanding and using the framework.[451] The Irish Environmental Protection Agency, for example, has created an 'Environmental Queries Unit' answering the public's questions about the environment. Findata provides a free helpdesk via email and phone, which can advise researchers on the data sources available via the access regime that might fit their research ideas. It can also assist applicants in the process of creating a solid data utilization plan and applications for a data request or permit.

These proactive measures underscore that governments can do more to enable research than merely create accessible datasets. They also have an important role to play in **ensuring that researchers have the resources necessary to actually conduct research.** At a minimum, this proactive government role can include awareness-raising amongst, and advice for, relevant stakeholders. Given that these are expert tools, in contrast to many other forms of transparency, this awareness-raising need not necessarily focus on the average citizen or end-user. Instead, it can be tailored to relevant stakeholders such as academics and investigative journalists, and other stakeholders who, as Archon Fung writes, 'have material or mission-driven interest in obtaining information (such as information about their competitors) and who possess the analytic capacity to make sense of it'.[452]

Beyond awareness raising, funding relevant research organizations can also be key to ensuring the success of data access frameworks. Subsidies, grants and other forms of material support are central in, for instance, the EU's new AI White Paper,[453] and also deserve further exploration in the context of platform governance efforts such as the Digital Services Act. Worth exploring, for instance, is a tax on systemic platforms which serves to fund independent public interest research into these services. Whilst the issue of research funding is not central in this Report, it should be clear that **data access frameworks go hand-in-hand with the broader cultivation of a robust and democratic civil society, which is adequately funded and guaranteed of its independence.** More fundamentally, this also requires that researchers are given due recognition in subsequent policymaking, so that their findings are not merely published but reflected in public policy dialogues.[454] Simply put: platforms are unlikely to take criticism from civil society seriously, as long as regulators fail to do so.

---

451 E-PRTR Regulation, art 15.

452 Fung (n 441) 187–88.

453 European Commission, 'White Paper on Artificial Intelligence – A European approach to excellence and trust 2020' (White Paper) COM (2020) 65 final.

454 On the interaction between civil society and governments in platform governance, see e.g.: Natali Helberger, Jo Pierson and Thomas Poell, 'Governing Online Platforms: From Contested to Cooperative Responsibility' (2017) 34 The Information Society 1 ('[I]t is by enabling and shaping substantive public deliberations by crucial stakeholders on how to balance different public values in the management of contentious content that governments can and have to play a crucial democratic role'.).

5.1.5 Public transparency by default

A crucial lesson learned from both case studies is that the respective access regimes position public transparency as the default. Put differently, the starting assumption is that **the predefined data should be accessible to anyone who should request it**. This appears vital in order to ensure the validity of the framework for use as an accountability mechanism as well as a tool for research more broadly. Exceptionally, certain data can be excluded, to the extent necessary in order to safeguard competing rights, freedoms or interests. The nature of the default rule implies that such exclusion needs to be evaluated on a case-by-case basis (cf. the vital role of independent institutions, Section 5.2.3).

Depending on the challenges raised in the respective access regime, the requestors may have to fulfil certain minimum criteria. Ideally such criteria should not be over-burdensome. Indeed, when looking at the Findata case study, there are no *a priori* constraints as to who can apply for access, as long as they have a valid data utilization plan and pay the relevant fees. This being said, Findata does provide a limitative list of purposes for which access can be requested,[455] which do constrain, if not the persons involved in research, then at least the potential topics of research. While these access criteria might still be too onerous for some, these appear to be relatively subsidiary and proportionate restrictions in light of the urgent privacy and data protection interests involved, and do little to detract from the general principle of public access.

In short, the accessibility of these data access regimes should be seen as a spectrum from no restrictions whatsoever, to entirely opaque. While the E-PRTR case study can be situated on the left extremity of this spectrum,[456] the Findata case study can be located slightly left from center, and more privileged access regimes such as the proposed EU Digital Media Observatory right from center. Where exactly to position a given access regime will depend on the trade-off between the challenges and risks raised in a given case, including the benefits of broader research access, and the potential risk of abuse of this data. This trade-off should also consider the potential uses for commercial purposes. In light of the above, a key recommendation for an access regime in the platform context is to consider those trade-offs in light of all of the rights, freedoms and interests involved: concern for privacy and data protection should not lead us to disregard entirely, or dismiss out of hand, the ideal of open and inclusive data access.[457] Independent institutions at the center of the data access regime (cf. Section 5.1.2) can play a vital role in enabling these trade-offs, as will appear below.

5.1.6 Verification and pre-processing by independent institutions

Both case studies reveal how **the independent institutions or public bodies at the heart of the access regimes can play a vital role in the disclosure process**, acting as an intermediary between the disclosing corporations and ultimate recipients. Not only can they maintain relevant access infrastructures, such as virtual machines (in the case of Findata) and

455 ASU, section 2 (The purposes for which Findata provides access to personal health data are: (1) conducting scientific research; (2) the creation of statistics; (3) development and innovation that furthers public health or the health sector; (4) health education of students and professionals; (5) knowledge-based management, e.g. the benchmarking of different healthcare providers and municipalities that provide healthcare services; (6) steering and supervision of healthcare services by authorities and (7) to assist in the planning and reporting duties of those authorities.).

456 Cf. E-PRTR Regulation, recital 14 (Access to information provided by the European PRTR should be unrestricted and exceptions from this rule should only be possible where explicitly granted by existing Community legislation.).

457 Indeed, the Charter of Fundamental Rights of the EU protects information freedoms (article 11), scientific research and academic freedom (article 13) as well as privacy (article 7) and data protection (article 8).

public databases, websites and forums (in the case of E-PRTR), they also play an important role in **verifying and pre-processing corporate data in order to ensure it is suitable for disclosure**.

Verification entails that corporate disclosures are periodically audited, in order to discourage false or incomplete disclosures. This responds to the 'incentive problem', outlined in Section 2.3.1, which indicates that private corporations such as industrial facilities or online platforms may lack the incentives to voluntarily disclose the (entire) truth. This is further illustrated in the E-PRTR case study by authorities laying down a standardized (albeit non-binding) methodology for industrial facilities to follow when generating the required data. These issues are relatively less urgent in the context of Findata, which is directed at institutions with fewer incentives to unduly modify their disclosures about personal health data.

Closely related is the concept of **pre-processing**. In both the Findata and E-PRTR case studies, public bodies play an important role in preparing datasets for disclosure and preventing the unwarranted disclosure, e.g. of sensitive/personal data. Findata notably helps researchers in combining data from different sources, while at the same time ensuring data disclosures are GDPR compliant. With the E-PRTR, companies may request confidentiality for certain data, but this is subject to approval from the national regulatory authority. Crucially, the regulatory authority therefore has full access to the data which the corporation refuses to disclose, and can thereby make a full assessment on the merits of their confidentiality claim. This independent check on confidentiality claims is essential, since the incentive problem predicts that these claims are not always made in good faith. In order to safeguard integrity and trustworthiness of the system**, it is vital for the respective institutions to be transparent about how exactly data has been pre-processed before being made accessible**.

Finally, it is also important to emphasize how the institutions at the center of both case studies play an active role in rendering data meaningful. It is not unlikely that the data that researchers may wish to get access to (for their research and/or to hold relevant companies accountable), may not exist within those companies, or may not be readily interpretable from the data they do have.[458] This may be the case because of the different objectives for valorizing data that companies have when compared to civil society actors such as academic researchers. **The access regimes – and independent institutions that manage them – play a vital role in ensuring otherwise unavailable or uninterpretable data to be made accessible**. A clear example is the standardized methods for generating pollutant data, enforced by authorities in the E-PRTR case study, data that might otherwise not have been produced in the first place. Within the platform context, it will be important to clearly identify *what* data is needed to ensure the desired levels of accountability in the respective policy areas. In doing so, **the fact that data is not readily available and/or produced by the respective platforms, should not be a reason to discard including that data into the access regime**. It may, however, necessitate additional efforts from the independent institutions to make sure that data is rendered useful for those who access it. In the second case study, for example, Findata plays a vital role in helping researchers to combine otherwise disperse datasets. As explained by Ananny and Crawford, '[w]hat systems are or mean depend upon the tools and perspectives people employ while looking'.[459] Any robust research access framework will therefore require an institution offering such tools – and an enabling environment more broadly – for rendering

458 Cf. Fung (n 441) 187–89.

459 Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973. 982.

the relevant data interpretable in a variety of ways. Simply put, **transparency regulation is not strictly a question of creating access to data, but also in ensuring that useful data is produced in the first place**.

5.2 Open questions

Even if the case studies offer a lot of relevant best practices for research access in the platform governance context, a number of open questions remain. These relate to the following important issues: (1) Ensuring transparency *within* the access framework; (2) Deciding on an appropriate balance of proactive and reactive disclosures; (3) Devising a clear and workable allocation of liability; (4) Determining the substantive scope of data access frameworks.

5.2.1 Being transparent about being transparent

In order to assess the performance of data access frameworks, they should be transparent in their own operations. Such transparency has at least two dimensions. Firstly, **the framework should require openness about *how* data was generated**. Documenting how the disclosed datasets came to be is important for researchers (and civil society more broadly) to judge its precise quality for their research aims. In the e-PRTR Regulation, for instance, companies are required to state the *methodology* they employed to record their data. Similarly, Findata is also transparent about what operations data underwent in the pre-processing phase, so as to give researchers the necessary methodological assurances.

Secondly – and not immediately apparent from the case studies per se – **a platform research access framework should also ensure transparency of its internal operations**. Our case studies showed that it is not straightforward to study in a systematic fashion what types of research have been conducted on the basis of these access frameworks, which creates obstacles in determining the potential downsides or pitfalls of these systems.

Data access frameworks should therefore be open about their operations, e.g. by publicly reporting the information requests they receive and whether or not they were granted. At present, the impact of programs such as Findata and the E-PRTR remains cumbersome to assess, since there is no clear overview of what types of uses it has enabled -- or indeed failed to enable.[460] To offer one best practice in this space, it is worth noting that the self-regulatory Social Science One program maintains a public registry of the research projects that have received their data.[461] Of course, these types of documentation are relatively straightforward in the context of on-request frameworks such as Findata, where a record exist of each data request, rather than in public data frameworks such as E-PRTR where access is unconditional and anonymous. Further research may be needed to explore how the impact of public transparency frameworks can be tracked and documented.

5.2.2 Proactive vs. reactive disclosure

A noteworthy distinction between the two case studies relates to the timing of disclosures: Findata's health data is only shared *reactively*, in response to requests by researchers with specific demands, whereas E-PRTR discloses data *proactively*, in an online register where it is available to the general

460 It should be said though, that Findata appears to report at least some of this information (albeit entirely voluntary): Findata, 'Data Requests' (*Findata*, 9 June 2020) <https://www.findata.fi/en/services/data-requests/> accessed 11 June 2020.

461 Social Science One, 'Researchers' (*Researchers*) <https://socialscience.one/researchers> accessed 11 June 2020.

public. This choice between reactive and proactive frameworks involves important trade-offs for the overall framework design.

The challenge with proactive frameworks is identifying data which is in high demand amongst researchers. Since proactive disclosures do not respond to specific requests, there is a risk that time and money will be invested into developing datasets that ultimately find little uptake and create little impact. Indeed, as noted above, it may be difficult to track the overall usage of public datasets such as those studied in the PRTR. This makes it all the more **important that regulators remain in close dialogue with public interest researchers in order to connect their rulemaking to existing data demands**.

For data that is indeed in high demand, proactive frameworks have the potential to be more efficient and accessible. Reactive frameworks, conversely, have the downside that every individual must enforce their own specific demands, introducing possible delays and denials from the disclosing party. In economic terms, one might say that reactive frameworks typically have lower up-front costs, but higher marginal costs.

In light of the 'incentive problem' discussed earlier, such delays and denials are especially likely to occur in the context of platforms: we may assume that platforms will do whatever they can (within the laws or perhaps even exceeding this) to delay, deny or even refuse access.[462] Indeed, the GDPR's reactive framework of data access rights faces a similar problem, with many requestors reporting lackluster compliance which undermines the right's effectiveness.[463] One way to improve compliance could be to entrust the competent independent institution or supervisory body to enforce third party access requests; their greater expertise, experience and wherewithal would likely help to push back against unfounded deflections by the platform. Nonetheless, this approach would introduce additional costs which might be better spent elsewhere.

Experiences with government transparency may be instructive here, since it includes both proactive and reactive approaches. The reactive approach is manifested in freedom of information laws, which have contributed to transparency and accountability in countless ways but are also frequently undermined by inadequate compliance. The proactive approach can be seen, for instance, in open data policies, which have proliferated over the past decades but have been accused of inefficiency: many open government datasets remain underused and fail to respond to existing demand. This being said, proactive approaches to government transparency can be found not only in 'open government' initiatives but also in many mainstays of modern public governance such as, for instance, public land ownership registrars and public court documents.

The case studies in this Report do not allow us to draw conclusive lessons on what approach would be most appropriate in the platform context, and more research on this important issue is required. What can be said however, is that **an access framework for platform governance must try to find an appropriate balance between proactive and reactive disclosure**. Reactive disclosure can serve a more general purpose or catch-all function, whereas the proactive approach must be targeted strategically towards key datasets with known public interest values and existing research demand.

462 Cf. ch 2.

463 Jef Ausloos and Pierre Dewitte, 'Shattering One-Way Mirrors – Data Subject Access Rights in Practice' (2018) 8 International Data Privacy Law 4; Jef Ausloos, Réne Mahieu and Michael Veale, 'Getting Data Subject Rights Right A Submission to the European Data Protection Board from International Data Rights Academics, to Inform Regulatory Guidance' (2020) 10 JIPITEC 294.

5.2.3 Liability for disclosed data

Data access frameworks raise complex liability-related questions, resulting from the multitude of actors, interests and data involved. In any data access framework, there is a risk of liability emerging from e.g. deliberate or accidental breaches of data protection or intellectual property laws. **The division of these liabilities can have powerful effects on the incentives of its participants, and thus the overall success of the framework**. On the one hand, excessive liability imposed on the disclosing party may discourage them from full disclosure, and lead them to underreport relevant information for fear of incurring liability under other applicable laws. On the other hand, far-reaching immunities for disclosed data could create a moral hazard and push platforms towards carelessness, increasing the risk of harmful content being disclosed. Indeed, a worst-case scenario, bearing in mind the incentive problem identified in Section 2.3, could be that platforms would undermine the framework and its legitimacy by disclosing harmful data for which they bear no legal responsibility.

An additional complexity in the context of platforms is that these services tend to operate on a transnational or even global scale, whilst liability rules are largely grounded in national law. For reasons of legal certainty, this may offer an additional argument to prioritize transparency regulation at the EU level. Indeed, attempts to impose transparency duties at national level could potentially even conflict with EU-level attempts to harmonize the liability of information society services, and limit their liability for user-generated content.[464] **An EU-level approach would help to clarify the interaction between transparency-related liabilities and such generic liability frameworks.**

While the Findata case study offers some insights on a potential model for the distribution of responsibilities under data protection law, the analysis in this Report does not offer firm conclusions regarding the different facets of liability more broadly. Ultimately, the appropriate division of liability will depend on the nature of the governance framework and the parties involved. Nonetheless, these are crucial questions that we strongly recommend further consideration for in any policy-making effort.

5.2.4 Subject matter and scope

A fundamental question which remains outside of the scope of this Report, is determining the *subject matter* covered by data access frameworks. **Platforms are active in a constantly growing range of economic and social fields, and the types of data one demands depend on the type of phenomenon one wants to study.** For instance, those concerned about 'filter bubbles' and the diversity of online media will focus on audience viewing patterns and the role of platform recommendation algorithms. Those concerned about hate speech will want to study hateful content and communities, whilst those interested in freedom of expression will want to about platform decisions to remove allegedly illegal content. In other words, each of these different research interests will generate different informational demands, and this makes substantive scope a key question in designing data access frameworks.

A **clear formulation of both the purpose and scope of the respective data access framework is crucial to its effectiveness.** If the relevant disclosure obligations are too ambiguous, vague and/or high-level, this can undermine the research utility of the dataset.[465]

---

464 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 Internal Market ('Directive on electronic commerce') [2000] OJ L178/1; See also, by analogy, attempts by Facebook and Google in the US to invalidate state-level regulation of ad archives based on CDA 230.

465 See *a contrario*: Jennifer Cobbe, Chris Norval and Jatinder Singh, 'What Lies beneath: Transparency in Online Service Supply Chains' (2020) 5 Journal of Cyber Policy 65, 65.

For instance, in the context of political advertising, the demand to disclose all 'political advertising' has led to much uncertainty and criticism, simply because 'political advertising' is an ambiguous concept which is difficult for platforms to enforce at scale, leading to the disclosure of datasets which hare of questionable quality for scientific research.[466] A more objective approach would be for platforms to disclose *all* advertising, thereby allowing independent researchers to determine for themselves what qualifies as 'political' or not, based on their own scientific standards.[467]

As this example shows, it may be more effective to base disclosure obligations on the technical functionalities of the platform service, rather than more ambiguous and politically-charged conceptions of harm such 'disinformation', 'political advertising', 'hate speech' and so forth. Along these lines, key technical features that may deserve closer scrutiny might include: high-level aggregate audience metrics; advertising and microtargeting; search features; feeds, ranking and recommendation; and content moderation (including removal but also other measures such as demonetization or fact-checking).

Overall, however**, it is essential that disclosure rules remain flexible and subject to updates and revisions. This can be achieved by delegating to oversight bodies the task of identifying new areas of interest**; rather than fixing them exhaustively in legislation. In deciding on these issues, a degree of flexibility can be created by delegating to independent institutions or oversight bodies (discussed in section 5.1) the authority to identify new areas of interest. Such flexibility is essential in order to make the framework future-proof, and fit to tackle the policy challenges of tomorrow. After all, data access also plays a crucial role in diagnosing harms. If a research access framework is targeted exclusively on *known* harms, then it will necessarily fail to assist in the detection of new and unknown harms. Furthermore, platform services are highly dynamic and adjustable, and changes to the service architecture can serve to evade and undermine earlier definitions and rules.[468] Precisely in the context of platforms, with their rapid rate of change, the possible topics of transparency should not be set in stone but instead be embedded in a flexible and iterative structure.

Relatedly, it is worth exploring whether and how transparency policies should operate horizontally. After all, social media governance may demand other forms of research access than, for instance, ridesharing or e-commerce governance. One way forward would be to develop a baseline of generally applicable rules that apply to all major (or 'systemic') platforms, whilst leaving room for tailored sectoral rules targeted specifically at certain subcategories of platforms.

***

Platforms constitute an increasingly central infrastructure for modern society, collecting vast amounts of data about individuals, and shaping how they interact with each other and their environment. So far,

466 Section 2.2.4.

467 ibid; Panoptykon Foundation, 'Who (Really) Targets You?' (2020) <https://panoptykon.org/political-ads-report> accessed 20 April 2020; Leerssen and others, 'Platform Ad Archives' (n 443).

468 Bridget Barrett and Daniel Kreiss, 'Platform Transience: Changes in Facebook's Policies, Procedures, and Affordances in Global Electoral Politics' (2019) 8 Internet Policy Review <https://policyreview.info/articles/analysis/platform-transience-changes-facebooks-policies-procedures-and-affordances-global> accessed 11 June 2020.

strategic secrecy has prevented robust accountability mechanisms from being established. While there might be various (legal, economic, technical) reasons for restricting transparency and data access, it is clear that platforms' interests in maintaining exclusivity may not always align with public interests in transparency – and that their arguments for maintaining secrecy may not always be valid or in good faith. Against this backdrop, there is a clear need for a more robust data access framework that is legally enforceable. Taking inspiration from other sectors with operational data access frameworks in place already, this Report has formulated concrete recommendations to push the debate forward, and to move from generic calls for 'transparency' to concrete policies for public interest research access. Indeed, the unprecedented reach and complexity of online platforms should not distract from the fact that the challenges they pose are not all new. Rather than reinventing the wheel, we should build on Europe's rich experience with transparency frameworks that handle sensitive data, and hold powerful actors to account.

References

Table of Authorities (Primary Sources)

EU Cases

Fashion ID GmbH & CoKG v Verbraucherzentrale NRW eV (C-40/17) [2019] Second Chamber ECLI:EU:C:2019:629

Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig- Holstein GmbH (C-210/16) [2018] Grand Chamber ECLI:EU:C:2018:388

EU Legislation

Charter of Fundamental Rights of the European Union [2012] OJ C326/391

Commission Decision (EC) 2000/479 of 17 July 2000 on the implementation of a European pollutant emission register (EPER) according to Article 15 of Council Directive 96/61/EC concerning integrated pollution prevention and control (IPPC) (notified under document number C(2000) 2004) (Text with EEA relevance) [2000] OJ L192/36

Commission Implementing Decision (EU) 2019/1741 of 23 September 2019 establishing the format and frequency of data to be made available by the Member States for the purposes of reporting under Regulation (EC) No 166/2006 of the European Parliament and of the Council concerning the establishment of a European Pollutant Release and Transfer Register and amending Council Directives 91/689/EEC and 96/61/EC (Text with EEA Relevance) [2016] OJ L267/3

Council Directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control [1996] OJ L257/26

Council Regulation (EC) No 1224/2009 of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy, amending Regulations (EC) No 847/96, (EC) No 2371/2002, (EC) No 811/2004, (EC) No 768/2005, (EC) No 2115/2005, (EC) No 2166/2005, (EC) No 388/2006, (EC) No 509/2007, (EC) No 676/2007, (EC) No 1098/2007, (EC) No 1300/2008, (EC) No 1342/2008 and repealing Regulations (EEC) No 2847/93, (EC) No 1627/94 and (EC) No 1966/2006 [2009] OJ L343/1

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 Internal Market ('Directive on electronic commerce') [2000] OJ L178/1

Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use [2001] OJ L121/34

Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC [2003] OJ L41/26

Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC (Text with EEA relevance) [2003] OJ L275/32

Directive 2004/35/CE of the European Parliament and of the Council of 21 April 2004 on environmental liability with regard to the prevention and remedying of environmental damage. [2004] OJ L143/56

Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU (Text with EEA relevance) [2014] OJ L173/349

Regulation (EC) No 166/2006 of the European Parliament and of the Council of 18 January 2006 concerning the establishment of a European Pollutant Release and Transfer Register and amending Council Directives 91/689/EEC and 96/61/EC (Text with EEA relevance) [2006] OJ L33/1

Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety [2002] OJ L31/1

Regulation (EC) No 223/2009 Of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities (Text with relevance for the EEA and for Switzerland) [2009] OJ L87/164

Regulation (EU) 2015/847 of the European Parliament and of the Council of 20 May 2015 on information accompanying transfers of funds and repealing Regulation (EC) No 1781/2006 (Text with EEA relevance) [2015] OJ L141/1

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) (Text with EEA relevance) [2016] OJ L119/1

Regulation (EU) 2019/1010 of the European Parliament and of the Council of 5 June 2019 on the alignment of reporting obligations in the field of legislation related to the environment, and amending Regulations (EC) No 166/2006 and (EU) No 995/2010 of the European Parliament and of the Council, Directives 2002/49/EC, 2004/35/EC, 2007/2/EC, 2009/147/EC and 2010/63/EU of the European Parliament and of the Council, Council Regulations (EC) No 338/97 and (EC) No 2173/2005, and Council Directive 86/278/EEC (Text with EEA relevance) [2019] OJ L170/115

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 (Text with EEA relevance) [2019] OJ L186/57

Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC (Text with EEA relevance) [2014] OJ L158/1

Regulation (EU) No 600/2014 of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Regulation (EU) No 648/2012 (Text with EEA relevance) [2014] OJ L173/84

Official EU Publications/Parliamentary Papers

Article 29 Working Party, 'Opinion 1/2010 on the Concepts of 'Controller' and 'Processor'' (2010) WP 169 <https://ec.europa.eu/justice/article-29/documentation/>

Commission (EC), 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee of the Regions and the Committee of the Regions: A European strategy for data' (Communication) COM (2020) 66 final, 23 September 2016

Commission (EC), 'Proposal for a regulation of the European Parliament and the Council on the alignment of reporting obligations in the field of environment policy and thereby amending Directives 86/278/EEC, 2002/49/EC, 2004/35/EC, 2007/2/EC, 2009/147/EC and 2010/63/EU, Regulations (EC) No 166/2006 and (EU) No 995/2010, and Council Regulations (EC) No 338/97 and (EC) No 2173/2005' (Proposal) COM (2018) 381 final, 31 May 2018

Commission (EC), 'Report from the Commission to the European Parliament and the Council on progress in implementing Regulation (EC) 166/2006 concerning the establishment of a European Pollutant Release and Transfer Register (E-PRTR)' (Report) COM (2013) 111 final, 5 March 2013

Commission (EC), 'Report from the Commission to the European Parliament and the Council on progress in implementing Regulation (EC) 166/2006 concerning the establishment of a European Pollutant Release and Transfer Register (E-PRTR)' (Report) COM (2017) 810 final, 13 December 2017

Commission (EC), 'White Paper on Artificial Intelligence – A European approach to excellence and trust' (White Paper) COM (2020) 65 final, 19 February 2020

Commission (EC) and High Representative of the Union for Foreign Affairs and Security Policy, 'Joint Communication to the European Parliament, The European Council, The Council, the European Economic and Social Committee and the Committee of the Regions: Tackling COVID-19 disinformation – Getting the facts right' (Joint Communication) JOIN (2020) 8 final, 10 June 2020.

European Data Protection Board, 'Guidelines 04/2020 on the Use of Location Data and Contact Tracing Tools in the Context of the COVID-19 Outbreak' (2020) <https://edpb.europa.eu/our-work-tools/our-documents/linee-guida/guidelines-042020-use-location-data-and-contact-tracing_en> accessed 23 April 2020

European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (2020) <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>

European Environment Agency, 'Mercury in Europe's Environment' (2018) Publication 11/2018 <https://www.eea.europa.eu/publications/mercury-in-europe-s-environment> accessed 1 May 2020

— A Decade of Industrial Pollution Data' (European Environment Agency 2019) Briefing 4/2019 <https://www.eea.europa.eu/ds_resolveuid/b8208000593e49d3aabaa8500b31b087> accessed 1 May 2020

— Quality Assurance Logic EU Registry on Industrial Sites – Document for Users – Version 5.0' (European Topic Centre for Air pollution, Transport, Noise and Industrial Pollution (ETC/ATNI) 2020) <https://cdr.eionet.europa.eu/help/euregistry/Documents/QAQC%20Master%20Document_CID_V5_January2020.pdf>

European Parliament Committee on Civil Liberties, Justice and Home Affairs, 'Draft report on the Digital Services Act and fundamental rights issues posed', PE650.509v01-00, 24 April 2020

European Parliament Committee on Legal Affairs, 'Draft report with recommendations to the Commission on a Digital Services Act: adapting commercial and civil law rules for commercial entities operating online', PE650.529v01-00, 22 April 2020

European Parliament Committee on the Internal Market and Consumer Protection, 'Draft report with recommendations to the Commission on Digital Services Act: Improving the functioning of the Single Market', PE648.474v02-00, 24 April 2020

European Regulators Group for Audiovisual Media Services, 'ERGA Position Paper on the Digital Services Act ERGA 2020 Subgroup 1 – Enforcement' (2020) <http://erga-online.eu/wp-content/uploads/2020/06/ERGA_SG1_DSA_Position-Paper_adopted.pdf>

— ERGA Report on Disinformation: Assessment of the Implementation of the Code of Practice' (2020) <http://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf>

Finnish Legislation & Statutory Instruments

Act on Secondary Use of Health and Social Data (nr. 552/2019) (Laki sosiaali- ja terveystietojen toissijaisesta käytöstä) (Finland)

Data Protection Act (nr. 1050/2018) (Tietosuojalaki Dataskyddslag) (Finland)

Regulation on Findata's fees (nr. 1500/2019) (Sosiaali- ja terveysministeriön asetus: Sosiaali- ja terveysalan tietolupaviranomaisen suoritteiden maksullisuudesta) (Finland)

Statistics Act (nr 280/2004) (Tilastolaki) (Finland)

Other Jurisdictions

Council of Europe Committee of Ministers, Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes (Adopted by the Committee of Ministers on 13 February 2019 at the 1337th meeting of the Ministers' Deputies)

Council of Europe Committee of Ministers to Member States, Recommendation CM/Rec (2018)1 on Media Pluralism and Transparency of Media Ownership (Adopted by the Committee of Ministers on 7 March 2018 at the 1309th meeting of the Ministers' Deputies)

Protocol on Pollutant Release and Transfer Registers to the Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters (adopted 21 May 2003, entered into force 8 October 2009) 2629 (UNTS) 119

'Reports of the United Nations Conference on Environment and Development' (Rio de Janeiro, 3 June-14 June 1992) (12 August 1992) UN Doc A/CONF.151/26 (Vol. I)

UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters (adopted 25 June 1998, entered into force 30 October 2001) 2161 (UNTS) 447 (Aarhus Convention)

Secondary Sources

Books

Benkler Y, Faris R and Roberts H, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford University Press 2018)

Bünger D, *Deficits in EU and US Mandatory Environmental Information Disclosure: Legal, Comparative Legal and Economic Facets of Pollutant Release Inventories* (Springer-Verlag 2012)

Cohen JE, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019)

Dijck J van, Poell T and Waal M de, *The Platform Society: Public Values in a Connective World* (Oxford University Press 2018)

Erdos D, *European Data Protection Regulation, Journalism, and Traditional Publishers: Balancing on a Tightrope?* (Oxford University Press 2019)

Gorwa R and Ash TG, 'Democratic Transparency in the Platform Society (Draft Chapter)' in Nate Persily and Josh Tucker (eds), *Social Media and Democracy: The State of the Field (Forthcoming)* (Cambridge University Press 2019)

Gürses S and van Hoboken J, 'Privacy after the Agile Turn' in Jules Polonetsky, Omer Tene and Evan Selinger (eds), *Cambridge Handbook of Consumer Privacy* (CUP 2018)

Keller D and Leerssen P, '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)

Mashaw J, 'Accountability and Institutional Design: Some Thoughts on the Grammar of Governance' in Michael Dowdle (ed), *Public Accountability: Designs, Dilemmas and Experiences* (Cambridge University Press 2006)

Moore M and Tambini D (eds), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (Oxford University Press 2018)

van Alsenoy B, *Data Protection Law in the EU: Roles, Responsibilities and Liability*, vol 6 (1e edn, Intersentia 2019)

Zuboff S, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019)

Journal Articles

Ananny M and Crawford K, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973

Ausloos J, 'GDPR Transparency as a Research Method' (Institute for Information Law (IViR), University of Amsterdam 2019) Draft Paper <https://papers.ssrn.com/abstract=3465680> accessed 17 October 2019

Ausloos J and Dewitte P, 'Shattering One-Way Mirrors – Data Subject Access Rights in Practice' (2018) 8 International Data Privacy Law 4

Ausloos J, Mahieu R and Veale M, 'Getting Data Subject Rights Right A Submission to the European Data Protection Board from International Data Rights Academics, to Inform Regulatory Guidance' (2020) 10 JIPITEC 294

Ausloos J and Veale M, 'Researching Through Data Rights' [2020] Forthcoming

Barrett B and Kreiss D, 'Platform Transience: Changes in Facebook's Policies, Procedures, and Affordances in Global Electoral Politics' (2019) 8(4) Internet Policy Review <https://policyreview.info/articles/analysis/platform-transience-changes-facebooks-policies-procedures-and-affordances-global> accessed 11 June 2020

Bodo B and others, 'Tackling the Algorithmic Control Crisis -the Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents' (2018) 19 Yale Journal of Law and Technology 4

Borgesius FJZ and others, 'Should We Worry about Filter Bubbles?' (2016) 5(1) Internet Policy Review <https://policyreview.info/articles/analysis/should-we-worry-about-filter-bubbles> accessed 9 June 2020

Bovens M, 'Analysing and Assessing Accountability: A Conceptual Framework' (2007) 13 European Law Journal 447

Bruns A, 'After the 'APIcalypse': Social Media Platforms and Their Fight against Critical Scholarly Research' (2019) 22 Information, Communication & Society 1544

— 'Filter Bubble' (2019) 8(4) Internet Policy Review <https://policyreview.info/concepts/filter-bubble> accessed 9 June 2020

Clifford D and Ausloos J, 'Data Protection and the Role of Fairness' (2018) 37 Yearbook of European Law 130

Cobbe J, Norval C and Singh J, 'What Lies beneath: Transparency in Online Service Supply Chains' (2020) 5 Journal of Cyber Policy 65

Cohen P and others, 'Using Big Data to Estimate Consumer Surplus: The Case of Uber' (National Bureau of Economic Research 2016) Working Paper 22627 <http://www.nber.org/papers/w22627> accessed 9 June 2020

Diakopoulos N, 'Accountability in Algorithmic Decision Making' (2016) 2 Communication of the ACM 56

Fikru MG, 'Does the European Pollutant Release and Transfer Register Enable Us to Understand the Environmental Performance of Firms?' (2011) 21 Environmental Policy and Governance 199

Freelon D, 'Computational Research in the Post-API Age' (2018) 35 Political Communication 665

Fung A, 'Infotopia: Unleashing the Democratic Power of Transparency' (2013) 41 Politics & Society 183

Haukka J and Gissler M, 'Finnish Health and Social Welfare Registers in Epidemiological Research' (2004) 14 Norsk epidemiologi 113

Helberger N, Pierson J and Poell T, 'Governing Online Platforms: From Contested to Cooperative Responsibility' (2017) 34 The Information Society 1

Hirsch DD, 'Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law' (2006) 41 Georgia Law Review 1

Hirsch DD and King JH, 'Big Data Sustainability: An Environmental Management Systems Analogy' (2015) 72 Washington and Lee Law Review Online 409

Leerssen P, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems' (Institute for Information Law (IViR), University of Amsterdam 2020) Preprint Paper <https://papers.ssrn.com/abstract=3544009> accessed 9 June 2020

Leerssen P and others, 'Platform Ad Archives: Promises and Pitfalls' (2019) 8(4) Internet Policy Review <https://policyreview.info/articles/analysis/platform-ad-archives-promises-and-pitfalls> accessed 7 February 2019

Munger K and Phillips J, 'A Supply and Demand Framework for Youtube Politics' (Department of Political Science, Pennsylvania State University 2019) Preprint Paper <https://osf.io/73jys/download>

Oberski DL and Kreuter F, 'Differential Privacy and Social Science: An Urgent Puzzle' 2(1) Harvard Data Science Review <https://hdsr.mitpress.mit.edu/pub/g9o4z8au/release/2> accessed 29 April 2020

Powles J, 'The Case That Won't Be Forgotten' (2015) 47 Loyola University Chicago Law Journal 583

Russo MA and others, 'Shipping Emissions over Europe: A State-of-the-Art and Comparative Analysis' (2018) 177 Atmospheric Environment 187

Shaddick G and others, 'Towards an Assessment of the Health Impact of Industrially Contaminated Sites: Waste Landfills in Europe' (2019) 3 Environmental Epidemiology 324

Staunton C, Slokenberga S and Mascalzoni D, 'The GDPR and the Research Exemption: Considerations on the Necessary Safeguards for Research Biobanks' (2019) 27 European Journal of Human Genetics 1159

van Wezel AP and others, 'Impact of Industrial Waste Water Treatment Plants on Dutch Surface Waters and Drinking Water Sources' (2018) 640–641 Science of The Total Environment 1489

Zuiderveen Borgesius FJ and others, 'Online Political Microtargeting: Promises and Threats for Democracy' (2018) 14 Utrecht Law Review 82

Official Reports

AMEC Environment & Infrastructure UK Limited, 'Contribution of Industry to Pollutant Emissions to Air and Water' (2014) 32790–01 FR 13298i5 <https://circabc.europa.eu/ui/group/06f33a94-9829-4eee-b187-21bb783a0fbf/library/c4bb7fee-46df-4f96-b015-977f1cca2093/details> accessed 1 May 2020

Amec Foster Wheeler Environment & Infrastructure UK and IEEP, 'Supporting the Evaluation of Regulation (EC) No 166/2006 Concerning the Establishment of a European Pollutant Release and Transfer Register and Its Triennial Review: Final Report.' (Publications Office of the European Union 2016) <http://op.europa.eu/en/publication-detail/-/publication/5b347a4a-9ae6-11e6-868c-01aa75ed71a1> accessed 2 May 2020

Lazarus A, 'Explaining the Death Ticker' (European Environmental Bureau 2016) <https://eeb.org/publications/61/industrial-production/1070/explaining-the-death-ticker.pdf>

Arnika and Eko forum Zenica, 'Top Ten of Biggest Environmental Polluters According to Data of Integrated Pollutant Release and Transfer Register (PRTR) of Bosnia and Herzegovina – Report for the Year 2016' (2016) <https://issuu.com/arnika.org/docs/grafy-bosna-en1>

Bertram T and others, 'Three Years of the Right to Be Forgotten' (Google Inc 2018) <https://pdfs.semanticscholar.org/13f5/e3cd0e8e522238f5df2ce279e6188664165e.pdf>

Cornils M, 'Designing Platform Governance: A Normative Perspective on Regulatory Needs, Strategies, and Tools to Enhance the Information Function of Intermediaries' (AlgorithmWatch 2020) <https://algorithmwatch.org/en/governingplatforms/legal-study-cornils-may-2020>

European Environmental Bureau, 'Mercury Emissions from Coal Power Plants in Germany' (2017) <https://eeb.org/library/mercury-emissions-from-coal-power-plants-in-germany-de/> accessed 30 April 2020

Krtková E and others, 'E-PRTR Data Review Methodology – Update 2019' (European Environment Agency – European Topic Centre on Air pollution, transport, noise and industrial pollution) Eionet Report ETC/ATNI 2019/5 <https://www.eionet.europa.eu/etcs/etc-atni/products/etc-atni-reports/etc-atni-report-5-2019-e-prtr-data-review-methodology-update-2019>

Lewis R, 'Alternative Influence: Broadcasting the Reactionary Right on YouTube' (Data & Society Research Institute 2020) <https://datasociety.net/library/alternative-influence/> accessed 9 June 2020

Panoptykon Foundation, 'Who (Really) Targets You?' (2020) <https://panoptykon.org/political-ads-report> accessed 20 April 2020

Parikka H and others, 'A Finnish Model for the Secure and Effective Use of Data – Innovating and Promoting the Secondary Use of Social and Health Data' (2019) 153 <https://www.sitra.fi/en/publications/a-finnish-model-for-the-secure-and-effective-use-of-data/> accessed 23 April 2020

Schaible C and others, 'Lifting Europe's Dark Clouds – How Cutting Coal Saves Lives' (European Environmental Bureau (EEB), Sandbag, Climate Action Network (CAN) Europe, Health and Environment Alliance (HEAL), WWF European Policy Office 2016) <https://eeb.org/lifting-europes-dark-cloud-how-cutting-coal-saves-lives/>

Schaible C, Ogando P and Lazarus A, 'Burning the Evidence: A Case Study on Large Combustion Plants' (European Environmental Bureau 2017) <https://eeb.org/library/burning-the-evidence-a-case-study-on-large-combustion-plants/> accessed 30 April 2020

Stark B and others, 'Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse' (AlgorithmWatch 2020) <https://algorithmwatch.org/en/governingplatforms/communications-study-stark-may-2020>

Tworek H and Leerssen P, 'An Analysis of Germany's NetzDG Law' (Institute for Information Law (IViR), University of Amsterdam 2019) <https://hdl.handle.net/11245.1/3dc07e3e-a988-4f61-bb8c-388d903504a7> accessed 9 June 2020

Blogs

CNN Business and Gold H, 'Facebook Promised Transparency on Political Ads. Its System Crashed Days before the UK Election' (*CNN Business*) <https://www.cnn.com/2019/12/11/tech/facebook-political-ads-uk-election-ge19/index.html> accessed 16 June 2020

Dencik L and others, 'Funding Matters – a Statement about the Corporate Funding of Academic Conferences' (Funding Matters, 2018) <https://fundingmatters.tech/> accessed 9 June 2020

European Advisory Committee Social Science One, 'Public Statement from the Co-Chairs and European Advisory Committee of Social Science One' (Social Science One, 11 December 2019) <https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one> accessed 9 June 2020

European Commission, 'Commission Launches Call to Create the European Digital Media Observatory' (Shaping Europe's digital future – European Commission, 7 October 2019) <https://ec.europa.eu/digital-single-market/en/news/commission-launches-call-create-european-digital-media-observatory> accessed 9 June 2020

Hunter ML and others, 'Special Investigation: How the Common Agricultural Policy Promotes Pollution' (The Ecologist, 23 March 2018) <https://theecologist.org/2018/may/23/special-investigation-how-common-agricultural-policy-promotes-pollution> accessed 1 May 2020

Kappe M, 'CBS: Inquiry into Risks of Data Access' (Centraal Bureau voor de Statistiek, 20 December 2019) <https://www.cbs.nl/en-gb/corporate/2019/49/cbs-inquiry-into-risks-of-data-access> accessed 15 May 2020

King G and Persily N, 'Unprecedented Facebook URLs Dataset Now Available for Academic Research through Social Science One' (Social Science One, 13 February 2020) <https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one> accessed 4 March 2020

Krafft TD and others, 'Filterblase geplatzt? Kaum Raum für Personalisierung bei Google-Suchen zur Bundestagswahl 2017' (AlgorithmWatch, 8 September 2017) <https://algorithmwatch.org/filterblase-geplatzt-kaum-raum-fuer-personalisierung-bei-google-suchen-zur-bundestagswahl-2017/> accessed 9 June 2020

Kuklis L and Wagner B, 'Disinformation, Data Verification and Social Media' (Media@LSE, 7 January 2020) <https://blogs.lse.ac.uk/medialse/2020/01/07/disinformation-data-verification-and-social-media/> accessed 10 June 2020

Merrill JB and Tobin A, 'Facebook Moves to Block Ad Transparency Tools — Including Ours' (ProPublica, 28 January 2019) <https://www.propublica.org/article/facebook-blocks-ad-transparency-tools> accessed 9 June 2020

Parikka H, 'One-Stop Shop for Well-Being Data – Isaacus Laid the Foundations for the Future' (Sitra, 9 November 2018) <https://www.sitra.fi/en/articles/one-stop-shop-well-data-isaacus-laid-foundations-future/> accessed 17 April 2020

'SCHUFA, a Black Box: OpenSCHUFA Results Published' (AlgorithmWatch, 29 November 2018) <https://algorithmwatch.org/en/schufa-a-black-box-openschufa-results-published/> accessed 9 June 2020

Silverman C, 'Funders Are Ready To Pull Out Of Facebook's Academic Data Sharing Project' (BuzzFeed News, 27 August 2019) <https://www.buzzfeednews.com/article/craigsilverman/funders-are-ready-to-pull-out-of-facebooks-academic-data> accessed 9 June 2020

Williams J and Gilens N, 'Federal Judge Rules It Is Not a Crime to Violate a Website's Terms of Service' (Electronic Frontier Foundation, 6 April 2020) <https://www.eff.org/deeplinks/2020/04/federal-judge-rules-it-not-crime-violate-websites-terms-service> accessed 9 June 2020

Yokoyama J, 'Closing the Data Divide: The Need for Open Data' (Microsoft on the Issues, 21 April 2020) <https://blogs.microsoft.com/on-the-issues/2020/04/21/open-data-campaign-divide/> accessed 24 April 2020

'YouTube Punishes Star over Suicide Video' (BBC News, 11 January 2018) <https://www.bbc.com/news/world-asia-42644321> accessed 9 June 2020

Newspaper articles

Botero-Marino C and others, 'We Are a New Board Overseeing Facebook. Here's What We'll Decide.' *The New York Times* (6 May 2020) <https://www.nytimes.com/2020/05/06/opinion/facebook-oversight-board.html> accessed 13 May 2020

Cadwalladr C and Graham-Harrison E, 'Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach' *The Guardian* (17 March 2018) <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election> accessed 9 June 2020

Hern A, 'Facebook, Apple, YouTube and Spotify Ban Infowars' Alex Jones' *The Guardian* (6 August 2018) <https://www.theguardian.com/technology/2018/aug/06/apple-removes-podcasts-infowars-alex-jones> accessed 9 June 2020

Hill K, 'The Secretive Company That Might End Privacy as We Know It' *The New York Times* (18 January 2020) <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> accessed 9 June 2020

Paul K and Waterson J, 'Facebook Bans Alex Jones, Milo Yiannopoulos and Other Far-Right Figures' *The Guardian* (2 May 2019) <https://www.theguardian.com/technology/2019/may/02/facebook-ban-alex-jones-milo-yiannopoulos> accessed 9 June 2020

Smith C, 'Diageo Defiant despite Distillery Listed as One of Europe's Worst Polluters' *The Courier* (12 July 2017) <https://www.thecourier.co.uk/fp/news/local/fife/466983/diageo-defiant-despite-distillery-listed-as-one-of-europes-worst-polluters/> accessed 1 May 2020

Web Resources

Berkman Klein Center for Internet & Society at Harvard University, 'Lumen' (Lumen Database) <https://lumendatabase.org/> accessed 9 June 2020

— 'Lumen – About' (About Us) <https://www.lumendatabase.org/pages/about> accessed 9 June 2020

— 'Lumen – Research' (Research) <https://www.lumendatabase.org/pages/research> accessed 9 June 2020

Centraal Bureau voor de Statistiek, 'Gezondheid En Welzijn' (Gezondheid en welzijn – Microdatabestanden) <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata/gezondheid-en-welzijn#id=zorgzvtab-personen-die-zorg-zonder-verblijf-hebben-ontvangen--voorheen-cakzzv---vervangen-vanaf-2009-door-gebzzvtab--0> accessed 15 May 2020

CSC, 'About Us' (About Us) <https://www.csc.fi/en/about-us> accessed 14 May 2020

Digital and Population Data Services Agency, 'Personal Identity Code' (*The personal identity code*) <https://dvv.fi/en/personal-identity-code> accessed 14 May 2020

European Commission, 'Legal Notice' (*European Union*, 16 June 2016) <https://europa.eu/european-union/abouteuropa/legal_notices_en> accessed 7 May 2020

— 'The EU's Fisheries Control System' (*Fisheries – European Commission*, 16 September 2016) <https://ec.europa.eu/fisheries/cfp/control_en> accessed 10 June 2020

— 'Infringement Procedure' *(European Commission)* <https://ec.europa.eu/info/law/law-making-process/applying-eu-law/infringement-procedure_en> accessed 15 May 2020

European Commission DG Environment, 'The European Pollutant Release and Transfer Register (E-PRTR) – Environment – European Commission' (*The European Pollutant Release and Transfer Register (E-PRTR)*, 30 January 2020) <https://ec.europa.eu/environment/industry/stationary/e-prtr/legislation.htm> accessed 10 June 2020

European Environment Agency, '2019 Industrial Pollution Country Profiles' (*2019 Industrial pollution country profiles*, 2 December 2019) <https://www.eea.europa.eu/themes/industry/industrial-pollution/2019-industrial-pollution-country-profiles> accessed 1 May 2020

— 'EEA-33 – Industrial Pollution Profile 2019' (*EEA-33 – Industrial pollution profile 2019*, 2 December 2019) <https://www.eea.europa.eu/themes/industry/industrial-pollution/industrial-pollution-country-profiles-2019/eea33> accessed 1 May 2020

— 'Publications' (*European Environment Agency*) <https://www.eea.europa.eu/themes/industry/publications/publications_topic> accessed 16 June 2020

— 'The European Pollutant Release and Transfer Register (E-PRTR), Member States Reporting under Article 7 of Regulation (EC) No 166/2006' (*European Environment Agency*, 6 February 2020) <https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-23> accessed 15 May 2020

— 'EEA FORUM' (*EEA FORUM*, 15 May 2020) <https://community.eea.europa.eu/search?SearchableText=e-prtr&x=0&y=0> accessed 15 May 2020

— 'EEA Budgets' (*European Environment Agency*) <https://www.eea.europa.eu/ds_resolveuid/6675272a4c594cc0912114008f35dd17> accessed 11 June 2020

— 'E-PRTR' (*E-PRTR*) <https://prtr.eea.europa.eu/#/home> accessed 9 June 2020

— 'E-PRTR FAQ' (*Frequently Asked Questions*) <https://prtr.eea.europa.eu/#/faq> accessed 27 March 2020

European Environment information and Observation Network (Eionet), 'European Topic Centres' (*European Topic Centers*) <https://www.eionet.europa.eu/etcs> accessed 12 June 2020

European IPPC Bureau, 'Reference Documents | Eippcb' (*Reference Documents*) <https://eippcb.jrc.ec.europa.eu/reference> accessed 1 May 2020

Eurostat, 'Overview – Eurostat' (*Overview*) <https://ec.europa.eu/eurostat/web/microdata> accessed 15 May 2020

Findata, 'Data Requests' (*Findata*, 15 June 2020) <https://www.findata.fi/en/services/data-requests/> accessed 16 June 2020

— 'About Us' (*About Us*) <https://www.findata.fi/en/about-us/> accessed 29 April 2020

— 'Data Permits' (*Data Permits*) <https://www.findata.fi/en/services/data-permits/> accessed 29 April 2020

— 'Data Protection and the Processing of Personal Data' (*Findata*) <https://www.findata.fi/en/about-us/data-protection-and-the-processing-of-personal-data/> accessed 1 May 2020

— 'Findata – Health and Social Data Permit Authority | Tervetuloa!' (*Findata*) <https://www.findata.fi/en/> accessed 11 June 2020

— 'Hinnasto' (*Hinnasto*) <https://www.findata.fi/palvelut/hinnasto/> accessed 29 April 2020

'FINLEX ® – Translations of Finnish Acts and Decrees: 1050/2018 English' <https://www.finlex.fi/en/laki/kaannokset/2018/en20181050?search%5Btype%5D=pika&search%5Bkieli%5D%5B0%5D=en&search%5Bpika%5D=Data%20Protection> accessed 16 June 2020

Finnish Information Centre for Register Research, 'Register i alfabetisk ordning' (*Informationscentret för registerforskning – ReTki*, 28 April 2012) <https://rekisteritutkimussv.wordpress.com/register/register-i-alfabetisk-ordning/> accessed 14 May 2020

— 'ReTki Info' (*Finnish Information Centre for Register Research*, 29 April 2012) <https://rekisteritutkimusen.wordpress.com/retki-info/> accessed 14 May 2020

Google Inc., 'Requests for User Information – Google Transparency Report' (*Global requests for user information*) <https://transparencyreport.google.com/user-data/overview> accessed 9 June 2020

Handl G, 'Declaration of the United Nations Conference on the Human Environment – Main Page' (*UN International Library of International Law, 2012*) <https://legal.un.org/avl/ha/dunche/dunche.html> accessed 30 April 2020

Health Data Hub, 'Health Data Hub | Plateforme Des Données De Santé | France' (*Healthdatahub*) <https://www.health-data-hub.fr?lang=en> accessed 20 March 2020

Ministry of Social Affairs and Health (Sosiaali- ja terveysministeriön), 'Secondary Use of Health and Social Data' (*Sosiaali- ja terveysministeriö*) <https://stm.fi/en/secondary-use-of-health-and-social-data> accessed 23 April 2020

Office of the Data Protection Ombudsman, 'Office of the Data Protection Ombudsman' (*Tietosuojavaltuutetun toimisto*) <https://tietosuoja.fi/en/office-of-the-data-protection-ombudsman> accessed 11 June 2020

Project AWeSome, 'For Researchers' <https://www.project-awesome.nl/for-researchers> accessed 9 June 2020

Social Science One, 'Researchers' (Researchers) <https://socialscience.one/researchers> accessed 11 June 2020

South Korean Ministry of Health and Welfare and Health Insurance Review & Assessment Service, '#opendata4covid19' *(#opendata4covid19*) <https://hira-covid19.net/> accessed 11 June 2020

Statistische Ämter des Bundes und des Landes Forschungsdatenzentren, 'Research Data Centre' (*Data supply by topic*) <https://www.forschungsdatenzentrum.de/en#understand-rdc> accessed 15 May 2020

The European Commission (*European Commission E-PRTR Validation tool*) <https://www.eionet.europa.eu/schemas/eprtr/validationtool> accessed 7 May 2020

The Guardian, 'The Cambridge Analytical Files' (*The Cambridge Analytical Files*) <https://www.theguardian.com/news/series/cambridge-analytica-files> accessed 16 June 2020

TRAFICOM Finnish Transport and Communications Agency National Cyber Security Centre, 'Accredited Information Security Inspection Bodies | NCSC-FI' (*Accredited information security inspection bodies*) <https://www.kyberturvallisuuskeskus.fi/en/our-services/assessment-accreditation-and-guidance/accredited-information-security-inspection> accessed 15 May 2020

Other secondary sources

Conference papers

Sandvig C and others, 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms' (Preconference to the 64th annual meeting of the International Communication Association, Seattle, Washington, US, 22 May 2014) <https://www.semanticscholar.org/paper/Auditing-Algorithms-%3A-Research-Methods-for-on-Sandvig-Hamilton/b7227cbd34766655dea10d0437ab10df3a127396> accessed 15 June 2020.

Tufekci Z, 'Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls' (ICWSM '14: 8th International AAAI Conference on Weblogs and Social Media, Ann Arbor, Michigan, USA, June 1-4, 2014) <http://arxiv.org/abs/1403.7400> accessed 9 June 2020

Documents

Atkins Danmark, GIS & IT and Tripledev, 'E-PRTR Validation Tool – User Manual Version 3.0' <https://www.eionet.europa.eu/schemas/eprtr/EPRTRUserManual.pdf>

'E-PRTR Data Completeness and Errors' <https://prtr.eea.europa.eu/docs/Errors%20and%20emissions%20disclaimer%20Oct2011.pdf>

European Commission, 'Guidance Document for the Implementation of the European PRTR' <https://ec.europa.eu/environment/industry/stationary/e-prtr/implementation.htm> accessed 10 June 2020

— 'EU Code of Practice on Disinformation' <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation> accessed 9 June 2020

— 'SERVICE REQUEST – ANNEX 'Specific Terms of Reference': Review of E-PRTR Implementation and Related Guidance' <https://ec.europa.eu/environment/industry/stationary/e-prtr/pdf/terms_of_reference_external_use.pdf>

European Environment Agency, 'EEA Forum Quick User Guide' <https://community.eea.europa.eu/home/Forum_manual_by_EEA.pdf>

— 'Reported Information under Regulation (EC) No 166/2006 on the Establishment of a European Pollutant Release and Transfer Register Information on the Database Structure and Use' <https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-23/database-structure-and-use-information/eprtr_database_metadata_v11.pdf-1/at_download/file> accessed 2 May 2020

Forum & Instant Messages

Sang Woo Park, 'Sang Woo Park on Twitter: 'For More Information: Https://T.Co/4gtJce6RIK" <https://twitter.com/sang_woo_park/status/1247313805752885248> accessed 15 May 2020

Sogalla R, 'Dear EEA Team,

I Would like to Analyze the Emissions of Air — EEA FORUM' <https://community.eea.europa.eu/home/environmental-topics/air-emissions/dear-eea-team-br-br-i-would-like-to-analyze-the-emissions-of-air/view?searchterm=E-PRTR#1571320782> accessed 2 May 2020

vvilla, 'Hi All. Animal by-Products and Derived Products Not Intended for Human — EEA FORUM' <https://community.eea.europa.eu/home/environmental-topics/air-emissions/hi-all.-animal-by-products-and-derived-products-not-intended-for-human/?searchterm=E-PRTR> accessed 2 May 2020

Interviews

Pim ten Thije, Interview with Antti Piirainen, Head of Communications, Findata (via Zoom videoconferencing, 26 March 2020)

——, Interview with Antti Piirainen, Head of Communications, Findata (via Zoom videoconferencing, 17 April 2020)

——, Interview with Antti Piirainen, Head of Communications, Findata (via Zoom videoconferencing, 13 May 2020)

Personal Communications

Letter from Jameel Jaffer and others to Mark Zuckerberg (6 August 2018) <https://knightcolumbia.org/sites/default/files/content/Facebook_Letter.pdf> accessed 9 June 2020

Operationalizing Research Access in Platform Governance

What to learn from other industries?

Jef Ausloos, Paddy Leerssen, Pim ten Thije

25 June 2020

Publisher:
AW AlgorithmWatch gGmbH
Linienstraße 13
10178 Berlin
Germany
Contact: info@algorithmwatch.org

Coordination:
Mackenzie Nelson

Layout:
Beate Autering, Beate Stangl, beworx.de

Published as part of the research project Governing Platforms
Website: algorithmwatch.org/en/governingplatforms

A project by

ALGORITHM WATCH

in partnership with

EPC EUROPEAN POLICY CENTRE

UNIVERSITY OF AMSTERDAM Institute for Information Law

MAINZER MEDIENINSTITUT

with support by

civitates

The sole responsibility for the content lies with the author(s) and the content may not necessarily reflect the positions of Network of European Foundations NEF, Civitates, or the Partner Foundations.