



## UvA-DARE (Digital Academic Repository)

### How to design data access for researchers: A legal and software development perspective

van Drunen, M.Z.; Noroozian, A.

**DOI**

[10.1016/j.clsr.2024.105946](https://doi.org/10.1016/j.clsr.2024.105946)

**Publication date**

2024

**Document Version**

Final published version

**Published in**

Computer Law & Security Review

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

van Drunen, M. Z., & Noroozian, A. (2024). How to design data access for researchers: A legal and software development perspective. *Computer Law & Security Review*, 52, Article 105946. <https://doi.org/10.1016/j.clsr.2024.105946>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



## How to design data access for researchers: A legal and software development perspective

M.Z. van Drunen<sup>\*</sup>, A. Noroozian

*Institute for Information Law, University of Amsterdam, Amsterdam, the Netherlands*

### ARTICLE INFO

#### Keywords:

Platforms  
Transparency  
Data access  
Vetted researchers  
Digital services act  
Political advertising regulation

### ABSTRACT

Public scrutiny of platforms has been limited by a lack of transparency. In response, EU law increasingly requires platforms to provide data to researchers. The Digital Services Act and the proposed Regulation on the Transparency and Targeting of Political Advertising in particular require platforms to provide access to data through ad libraries and in response to data access requests. However, these obligations leave platforms considerable discretion to determine how access to data is provided. As the history of platforms' self-regulated data access projects shows, the technical choices involved in designing data access significantly affect how researchers can use the provided data to scrutinise platforms. Ignoring the way data access is designed therefore creates a danger that platforms' ability to limit research into their services simply shifts from controlling what data is available to researchers, to how data access is provided. This article explores how the Digital Services Act and proposed Political Advertising Regulation should be used to control the operationalisation of data access obligations that enable researchers to scrutinise platforms. It argues the operationalisation of data access regimes should not only be seen as a legal problem, but also as a software design problem. To that end it explores how software development principles may inform the operationalisation of data access obligations. The article closes by exploring the legal mechanisms available in the Digital Services Act and proposed Political Advertising Regulation to exercise control over the design of data access regimes, and makes five recommendations for ways in which these mechanisms should be used to enable research into platforms.

### 1. Introduction

Better evidence about the way information is distributed on platforms is key to designing and enforcing platform regulation.<sup>1</sup> Given the scale with which platforms disseminate information, obtaining this evidence often requires computational analysis of large volumes of platform data. Data access has therefore become a central issue in platform governance.<sup>2</sup> Platforms, however, have strong incentives not to provide

access to their walled-off data in a manner that could enable accountability or fuel calls for more stringent regulation. Platforms' voluntary efforts to provide access to data have correspondingly faced heavy criticism for failing to make key data available, hindering the analysis of the data that is available, and lacking the stability necessary for longitudinal and critical research.<sup>3</sup>

In the face of this criticism, both the Digital Services Act (DSA) and the proposed Regulation on the transparency and targeting of political

<sup>\*</sup> Corresponding author. Institute for Information Law, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, the Netherlands

*E-mail address:* [M.Z.vanDrunen@uva.nl](mailto:M.Z.vanDrunen@uva.nl) (M.Z. van Drunen).

<sup>1</sup> Robert Gorwa and T Garton Ash, 'Democratic Transparency in the Platform Society' [2020] *Social Media and Democracy: The State of the Field, Prospects for Reform* 286, 302; Jef Ausloos, Paddy Leerssen and Pim ten Thije, 'Operationalizing Research Access in Platform Governance What to Learn from Other Industries?' (AlgorithmWatch 2020) 15 <<https://hdl.handle.net/11245.1/90e4fa77-d59a-49f1-8ccd-57d0725122bd>>.

<sup>2</sup> Paddy Leerssen, 'The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems' (2020) 11 *European Journal of Law and Technology* <<https://www.ejlt.org/index.php/ejlt/article/view/786/1012>>.

<sup>3</sup> Axel Bruns, 'After the "APIcalypse": Social Media Platforms and Their Fight against Critical Scholarly Research' (2019) 22 *Information, Communication & Society* 1544 <<https://doi.org/10.1080/1369118X.2019.1637447>>; Gorwa and Garton Ash (n 1); Nathaniel Persily and Joshua A Tucker, 'Conclusion: The Challenges and Opportunities for Social Media Research' in Joshua A Tucker and Nathaniel Persily (eds), *Social Media and Democracy: The State of the Field, Prospects for Reform* (Cambridge University Press 2020) <<https://www.cambridge.org/core/books/social-media-and-democracy/conclusion-the-challenges-and-opportunities-for-social-media-research/232F88C00A1694FA25110A318E9CF300>>; 'Impact Assessment Report 1/2 Accompanying the DSA' (European Commission 2020) SWD(2020) 348 final 15, 88.

<https://doi.org/10.1016/j.clsr.2024.105946>

Available online 6 February 2024

0267-3649/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

advertising (RPA) impose a legal obligation on very large online platforms to create data access regimes. Specifically, both regulations require platforms to create publicly accessible databases containing (information about) the advertisements they have displayed (ad libraries), and to make further data available to researchers who are authorised to submit data access requests.<sup>4</sup> The regulations thereby aim to ensure access to (some of) the data necessary for public scrutiny of the way information is distributed on platforms.<sup>5</sup>

However, the DSA and RPA leave open many of the technical details that determine how data access regimes will function. The DSA, for example, does not specify how the APIs it requires platforms to make available should be designed, or in what format data should be made available.<sup>6</sup> These details matter. They determine how researchers can actually observe platforms and which aspects of platforms are observable, and thereby ultimately shape the very meaning of data access.<sup>7</sup> For example, an API that limits researchers' ability to aggregate all ads by a single advertiser can allow researchers to easily call attention to a single illegal ad, but limit oversight of advertisers' long-term behaviour.<sup>8</sup> The lack of detailed rules on the operationalisation of data access regimes in the DSA and RPA is to some extent unavoidable. They are broad legal frameworks intended to regulate a wide variety of online services over a long period of time. Moreover, the needs of the researchers the DSA and RPA empower will also evolve as data analysis techniques improve and researchers gain a better understanding of platforms (and how they can best be scrutinised). At the same time, leaving the operationalisation of data access to platforms' discretion creates a risk that platforms' misaligned incentives simply shift from frustrating whether data is provided, to affecting how data is provided.

This article evaluates how the DSA and RPA should be used to control the operationalisation of data access obligations that enable researchers to scrutinise platforms. It argues that it is necessary to not only approach this question from a legal perspective, but also to explore how software development principles may inform the operationalisation of data access obligations. Providing the data access required under the DSA and RPA requires platforms to make a wide variety of technical design choices, ranging from the development of technical products such as APIs that provide access to ad libraries to the format in which data should be provided. Software development practices provide useful insights into the way the development process in which these technical choices are made should be organised to ensure the end product (data access regimes) meets the needs of its users (researchers). For instance, software development practices emphasise the need for an iterative process in which the developer adapts software following feedback from users.

Section 2 describes why the DSA and RPA now regulate data access, and what data access obligations they impose (including to which

researchers and which services these obligations apply). Section 3 draws on insights from software development to explore how the DSA and RPA require platforms to develop data access tools, and identify overarching principles in software development that are important to ensure such tools are designed to meet the needs of the researchers that the DSA and RPA aim to enable to scrutinise platforms. Section 4 describes the mechanisms in the DSA and RPA with which control can be exercised over the design of data access regimes. Based on this analysis, section 5 makes five recommendations about the operationalisation of data access regimes. This must: ensure consistency between the multiple competent supervisory authorities' approaches to data access; focus on the needs of the researchers that use data access regimes; be an iterative process, adapting to changes in researchers' needs and the platforms they scrutinise; account for the need for timely data access; ensure sufficient transparency about the way data access regimes are designed.

The article focuses on data access obligations that require platforms to enable non-governmental researchers to scrutinise how information is distributed on their service, and more specifically the provisions governing ad libraries and data access requests in the DSA and RPA.<sup>9</sup> The article uses the term 'researchers' to refer to the actors that use the APIs or data provided under the DSA and RPA to scrutinize platforms' impact on society and individuals, whether they are academics, data journalists, or civil society actors. Data access requests, however, can only be filed by a limited group of researchers.<sup>10</sup> The scope of the provisions analysed in this article is largely restricted to platforms with over 45 million monthly active users in the European Union.<sup>11</sup> Section 2.2 will describe both the services regulated under the DSA and RPA as well as the researchers empowered to scrutinise them, in further detail.

## 2. Why and how is data access being regulated

### 2.1. Why (regulated) data access is needed

The RPA argues data access regimes are necessary to "facilitate the work of interested actors including researchers in their specific role to support free and fair elections or referendums and fair electoral campaigns including by scrutinizing the sponsors of political advertisement [s] and analyzing the political advertisement landscape."<sup>12</sup> This quote reflects two goals that characterize the discussion surrounding data access regimes more broadly: the need for accountability for and a better understanding of the impact of the information that is distributed on platforms.

Making online information flows more transparent is a precondition for the enforcement of a wide variety of norms. For example, though ad libraries were introduced following controversies over the way voters were targeted with political advertisements in the Brexit campaign and 2016 US election, they also allow researchers to scrutinise whether job or housing opportunities are shown in a discriminatory manner, or whether companies advertise unhealthy food to minors.<sup>13</sup> Similarly the DSA grants researchers access to data necessary to understand systemic risks and assess "the adequacy, efficiency and impacts of the risk mitigation measures" taken by platforms. Systemic risks defined broadly, and include the dissemination of illegal content, as well as actual or foreseeable negative effects to (among others) the exercise of

<sup>4</sup> Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) 2022 art. 39, 40; Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising 2021 (COM/2021/731 final) art. 7(6), 11, 13. This article is based on the initial Commission proposal for the RPA. Relevant amendments in the Council's and Parliament's positions are noted where appropriate.

<sup>5</sup> DSA recitals 95-96; RPA recitals 42, 46.

<sup>6</sup> Ben Wagner and others, 'Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act' 11; Mathias Vermeulen, 'Researcher Access to Platform Data: European Developments' (2022) 1 *Journal of Online Trust and Safety* <<https://tsjournal.org/index.php/jots/article/view/84>>.

<sup>7</sup> Fernando N van der Vlist and others, 'API Governance: The Case of Facebook's Evolution' (2022) 8 *Social Media + Society* 20563051221086228 <<https://doi.org/10.1177/20563051221086228>>.

<sup>8</sup> Bruns (n 3); Matthew Rosenberg, 'Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised' *New York Times* (25 July 2019) <<https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html>>.

<sup>9</sup> DSA articles 39, 40; RPA articles 7(6), 11, 13.

<sup>10</sup> DSA article 40; RPA article 11, 13.

<sup>11</sup> Apart from RPA article 7(6), none of the provisions analysed in this article apply exclusively to platforms. The provisions on data access requests in the DSA and ad libraries also apply to very large online search engines, and the data access requests in the RPA also apply to political advertising service providers.

<sup>12</sup> RPA recital 42.

<sup>13</sup> Ausloos, Leerssen and ten Thije (n 1); Leerssen (n 2); RPA recitals 32, 39; DSA recital 95. The DSA refers for example to advertising's negative effects on public health, political participation, or equality.

fundamental rights, democratic processes, public security, minors, gender-based violence, and public health.<sup>14</sup>

Providing access to non-governmental researchers can support public authorities' enforcement of legal norms. The DSA creates particularly strong links between data access and enforcement by public authorities. The data to which vetted researchers are granted access concerns systemic risks that articles 34 and 35 DSA require platforms to assess and mitigate. The DSA explicitly references how the work of the researchers it empowers can be used by public authorities when they identify the most prominent and recurrent systemic risks, as well as by the auditors platforms must hire to assess their compliance with the DSA.<sup>15</sup> Of course, the Commission may also draw on such research to start its own investigation into platforms' compliance with the DSA, verifying whether the impact on a systemic risk or adequacy of a platform's mitigation found by researchers indeed violates a specific legal norm. Data included in ad libraries is not as explicitly linked to a legal norm as the DSA's data access right. Yet much of this data is also relevant to uncover evidence of legal violations. For example, researchers can use data on the funding of political ads allows to scrutinize whether political parties violate campaign spending laws; analyse their content to assess whether an ad uses subliminal techniques or promotes medicinal products; or analyse its targeting to assess whether companies unlawfully target alcohol or political ads to minors.<sup>16</sup> In short, by making data available that is relevant to assess whether a legal violation has taken place, the law stimulates research by non-governmental actors that can be used by public authorities to further investigate and enforce platform regulation.

Yet the primary function of the researchers the DSA and RPA empower is typically not to assess platforms' compliance with legal norms. Instead, the journalists, NGOs, and academics the DSA and RPA empower also play an important societal role by providing evidence in the public debate on platforms' impact. This can facilitate accountability for a wide variety of legal and non-legal norms by increasing public pressure on platforms, particularly when data access laws empower actors that already have a prominent role in the public debate, such as journalists and NGOs.<sup>17</sup> Accountability to the public plays a particularly important role in the context of political advertising, given the sensitivity of regulating political speech. In that context, ad libraries have for example been used to assess whether political campaigns use micro-targeting to spread negative messages, or reach out to different voter groups with different promises.<sup>18</sup> Such negative or conflicting messages are not (necessarily) illegal under political advertising law. Yet as the RPA emphasizes, data access laws nevertheless play an important role here by facilitating researchers' overarching role to support fair elections by scrutinizing the way political advertisements are disseminated. That is not to say there is no limit to the norms with which the DSA and RPA aim to enable accountability. Both the DSA and RPA limit the

accountability researchers can promote by limiting the data that these researchers can access. The DSA's right to request data can only be used to obtain data necessary to study systemic risks, and may not be used to request data to hold individual users accountable.<sup>19</sup> And while ad libraries contain information about the way ads are targeted by advertisers, they contain limited information about (and as such are limited in their ability to generate accountability for) the way platforms determine which ads to show to which users.<sup>20</sup>

In addition to accountability, data access is also expected to enable a better understanding of platforms' impact on individuals and society.<sup>21</sup> As platforms assume an increasingly central role in daily life, their societal impact becomes more important to understand. Yet the information necessary to fully understand how the dissemination of content on platforms impacts individuals and society is often not publicly accessible.<sup>22</sup> Both the DSA and RPA echo calls from regulators, civil society, and researchers to "bridg[e] information asymmetries" by providing access to the data necessary to understand platforms' societal impact.<sup>23</sup> This data is not only necessary to better understand the extent and impact of dangers often associated with platforms and identify new or emerging risks;<sup>24</sup> it is also necessary to tackle the normative implications of the way information is disseminated on platforms, and to determine how platforms should be regulated.<sup>25</sup> In that sense, enabling a better understanding of the way information is distributed on platforms is intertwined with accountability. A better understanding of platforms' impact is a precondition for an evidence-based debate about the norms to which platforms should be accountable in the first place.<sup>26</sup>

Accountability and understanding are familiar goals in transparency discussions. What sets data access regimes apart is that they adapt the way platforms are made transparent to the scale at which information is disseminated on platforms. This scale firstly requires platforms to tailor the information each individual sees to their characteristics. As a side-effect of the fact each user sees different information in a different order, researchers can no longer determine what information the audience sees on platforms simply by using the service (as they could on print or TV). Transparency measures such as ad libraries recreate some of the transparency that was present by default in mass media by aggregating

<sup>19</sup> DSA recital 96, article 40; RPA recital 46, article 11(2).

<sup>20</sup> DSA recital 95; RPA Annex II.

<sup>21</sup> Katharine Dommett and Mehmet Emin Bakir, 'A Transparent Digital Election Campaign? The Insights and Significance of Political Advertising Archives for Debates on Electoral Regulation' 73 *Parliamentary Affairs* 208, 221 <<https://doi.org/10.1093/pa/gsaa029>>; Juan Carlos Medina Serrano, Orestis Papyriakopoulos and Simon Hegelich, 'Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections', *International Conference on Social Media and Society* (Association for Computing Machinery 2020) 111 <<https://doi.org/10.1145/3400806.3400820>>; Ausloos, Leerssen and ten Thije (n 1) 15; Laura Edelson and others, 'An Analysis of United States Online Political Advertising Transparency' [2019] arXiv:1902.04385 [cs] <<http://arxiv.org/abs/1902.04385>>.

<sup>22</sup> Persily and Tucker (n 3) 314; Ausloos, Leerssen and ten Thije (n 1) 16.

<sup>23</sup> DSA recital 96; RPA recital 42.

<sup>24</sup> Colin Bennett and Jesse Gordon, 'Understanding the "Micro" in Micro-Targeting: An Analysis of Facebook Digital Advertising in the 2019 Federal Canadian Election' (2021) 46 *Canadian Journal of Communication* 431, 3 <<https://www.utpjournals.press/doi/full/10.22230/cjc.2021v46n3a3815>>; Daphne Keller and Paddy Leerssen, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation' in Joshua A Tucker and Nathaniel Persily (eds), *Social Media and Democracy: The State of the Field, Prospects for Reform* (Cambridge University Press 2020) <<https://www.cambridge.org/core/books/social-media-and-democracy/facts-and-where-to-find-the-m-empirical-research-on-internet-platforms-and-content-moderatio>>

<sup>25</sup> Keller and Leerssen (n 24) 222. Gorwa and Garton Ash (n 1) 302. Ausloos, Leerssen and ten Thije (n 1) 15.

<sup>26</sup> Bennett and Gordon (n 24).

<sup>14</sup> DSA article 34(1).

<sup>15</sup> DSA recital 92, article 35(2)(a).

<sup>16</sup> Ausloos, Leerssen and ten Thije (n 1); Leerssen (n 2); RPA recitals 32, 39; DSA recital 95; Paddy Leerssen and others, 'News from the Ad Archive: How Journalists Use the Facebook Ad Library to Hold Online Advertising Accountable' (2021) 0 *Information, Communication & Society* 1 <<https://doi.org/10.1080/1369118X.2021.2009002>>. Such advertising would potentially violate advertising rules in the EU's Audiovisual Media Services Directive (articles 28b(2) and 9). The ban on targeting political ads to minors is part of both the Council's and Parliament's position on the RPA.

<sup>17</sup> The personal scope of the data access laws at issue in this article is discussed in further detail in Section 2.2.

<sup>18</sup> Fabio Votta and others, 'Going Micro to Go Negative?: Targeting Toxicity Using Facebook and Instagram Ads' (2023) 5 *Computational Communication Research* 1 <<https://www.aup-online.com/content/journals/10.5117/CCR2023.1.001.VOTT>>; T Dobber and CD Vreese, 'Beyond Manifestos: Exploring How Political Campaigns Use Online Advertisements to Communicate Policy Information and Pledges' (2022) 9 *Big Data and Society*.



content displayed to platform users in a single publicly accessible database.<sup>27</sup> Secondly, the scale with which information is distributed on platforms also means that information about what content is available is less useful to understand platforms' impact than it was for media that produce less content, such as newspapers. Instead, data about how content reaches the audience (i.e., information about targeting, amplification, and reach) plays a more important role. Finally, the amount of information available on platforms means their impact can only be effectively scrutinised with computational methods.<sup>28</sup> Transparency measures that do not allow for computational analysis can enable researchers to scrutinise individual pieces of information (such as whether an ad includes discriminatory content), but not platforms' broader impact (such as whether the way information is targeted produces new inequalities in information access).<sup>29</sup> In short, the way information is made available is not an afterthought in data access regimes. Adapting transparency measures to the scale and complexity of the way information is disseminated on platforms is central to the goal of data access regimes, and hinges on the very way such regimes are technically designed.

Another new development in the DSA and RPA is the explicit focus on the need to use legislation to empower researchers to scrutinise platforms.<sup>30</sup> Public authorities have traditionally played a key role in investigating and enforcing compliance with legislation that applies to platforms.<sup>31</sup> A burgeoning field analyses the use of digital enforcement techniques in identifying violations of e.g. antitrust, copyright, consumer protection, and advertising law on platforms.<sup>32</sup> Public authorities continue to have strong investigatory powers – for example, the Commission's power to request data from platforms is not subject to the proportionality requirements, refusal grounds, or focus on systemic risks that apply to vetted researchers' right to request data.<sup>33</sup> Reserving strong data access powers for public authorities can be justified by fact that some of the data necessary to scrutinise platforms can also be used to compromise, for example, the security of platform services, the privacy rights of users, or platforms' trade secrets. These risks are more easily limited when data is only shared with public authorities, as opposed to a potentially large group of vetted researchers (as is the case for data access requests) or the public (as is the case for ad libraries).<sup>34</sup> Yet reserving data access exclusively to regulators also raises risks. Leerssen usefully distinguishes between pragmatic and principled concerns. Pragmatically, scrutinising platforms' impact over 450 million residents and 27 Member States in all the sectors in which platforms

operate is a daunting task that requires considerable expertise in the methodologies necessary to scrutinise platforms' impact. As noted above, providing data access to non-state researchers enables research that may point public authorities to potential violations of advertising law or the DSA's systemic risk provisions. Furthermore, providing such access also allows a much wider group of actors to develop the tools, methodologies, and expertise necessary to scrutinise platforms. This expertise is not only valuable to the research community, but may also be useful to the public authorities that must analyse large amounts of data to assess whether platforms have violated a specific legal obligation.<sup>35</sup> More principled concerns focus on the question of who should be able to hold platforms accountable. Reserving data access to regulators or a limited group of authorised private actors centralises the power to determine with which norms platforms should comply in the hands of a few actors. At the same time, it limits oversight over the way public authorities design and enforce platform regulation, as the information necessary to assess the underlying platform behaviour to which public authorities (fail to) respond is not available to the public.<sup>36</sup>

Platforms have a strong commercial incentive not to create data access regimes that cost money to develop and maintain, and are used to enable researchers to supply evidence to calls for more stringent regulation or to hold them or their users accountable.<sup>37</sup> As such, there is an inherent tension involved in relying on platforms to determine how researchers can scrutinise the impact of their services.<sup>38</sup> In addition to the limits researchers have experienced when using data access regimes platforms have provided voluntarily or under self-regulation (described in further detail in the next section), the lack of data access obligations meant some platforms could choose not to provide, or even remove, access to data.<sup>39</sup> Twitter, for example, has discontinued its ad library and significantly increased to cost of the API researchers used to analyse

<sup>27</sup> Leerssen (n 2).

<sup>28</sup> Bruns (n 3).

<sup>29</sup> Aaron Rieke and Miranda Bogen, 'Leveling the Platform: Real Transparency for Paid Messages on Facebook' (Upturn) 13 <<https://www.upturn.org/stat-ic/reports/2018/facebook-ads/files/Upturn-Facebook-Ads-2018-05-08.pdf>>.

<sup>30</sup> New at least in the context of EU platform legislation; other areas of EU law have historically had a greater focus on empowering researchers, while the self-regulatory (strengthened) code of practice on disinformation also includes commitments regarding platform to researcher data access. Ausloos, Leerssen and ten Thije (n 1).

<sup>31</sup> Leerssen (n 2) 27.

<sup>32</sup> Thibault Schrepel, 'Computational Antitrust: An Introduction and Research Agenda' <<https://papers.ssrn.com/abstract=3766960>>; Catalina Goanta and Jerry Spanakis, 'Discussing The Legitimacy of Digital Market Surveillance' [2022] Stanford Journal of Computational Antitrust 44 <<https://law.stanford.edu/wp-content/uploads/2022/04/goanta-spanakis.pdf>>; Liz Coll and Christine Riefa, 'Exploring the Role of Technology in Consumer Law Enforcement' (2022) 34 Loyola Consumer Law Review 359 <<https://heinonline.org/HOL/P/h=hein.journals/lyclr34&i=361>>.

<sup>33</sup> DSA article 40(1).

<sup>34</sup> Leerssen (n 2) 27.

<sup>35</sup> Goanta and Spanakis (n 32) 53; DSA recital 137. The DSA similarly refers to the need to develop Union expertise to be able to assess platforms' impact, and notes the Commission may invite a wide group of stakeholders including vetted researchers to share their expertise as it builds up this capacity.

<sup>36</sup> Margot E Kaminski, 'Understanding Transparency in Algorithmic Accountability' in Woodrow Barfield (ed), *The Cambridge Handbook of the Law of Algorithms* (Cambridge University Press 2020) 127 <<https://www.cambridge.org/core/books/cambridge-handbook-of-the-law-of-algorithms/understanding-transparency-in-algorithmic-accountability/D355F8D31BF1778431D92D2E79917093>>; Leerssen (n 2) 30.

<sup>37</sup> Ausloos, Leerssen and ten Thije (n 1) 88. See e.g. the studies cited in Keller and Leerssen (n 24) 222.

<sup>38</sup> European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Tackling Online Disinformation: A European Approach' (2018) COM(2018) 236 7–8 <<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236>> accessed 8 February 2022; European Commission, 'EU Code of Practice on Disinformation' (2018) <[https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=54454](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454)> I.I.E.12.Facebook, 'Facebook Report on the Implementation of the Code of Practice for Disinformation' (European Commission 2019) 4, 30 <[https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=62681](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62681)>; Google, 'EU Code of Practice on Disinformation Google Annual Report' (European Commission 2019) 13, 28 <[https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=62680](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62680)>; Twitter, 'Twitter Progress Report: Code of Practice against Disinformation' (European Commission 2019) 11, 31 <[https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=62682](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62682)>. Gorwa and Garton Ash (n 1) 302.

<sup>39</sup> 'Impact Assessment Report 1/2 Accompanying the DSA' (n 3) 50; Laura Edelson and others, 'Universal Digital Ad Transparency' (Knight First Amendment Institute 2021) ID 3898214 9 <<https://knightcolumbia.org/content/a-standard-for-universal-digital-ad-transparency>>; Medina Serrano, Papakyriakopoulos and Hegelich (n 21) 119; Leerssen (n 2) 48; Bruns (n 3) 1548.

the platform.<sup>40</sup> Similarly, TikTok has historically made very little data available to researchers. Though it has recently begun to offer an API to researchers, at time of writing the geographic scope of this API is limited to the US, and research projects using the API must be approved individually by TikTok.<sup>41</sup> This precarious and limited nature of voluntary data access is at odds with investment in longitudinal or critical research projects. It is against that context that the EU has begun to enshrine data access into the DSA and RPA.<sup>42</sup>

## 2.2. How data access is regulated in the DSA and RPA

The DSA requires very large online platforms and search engines (which have over 45 million monthly active users in the EU) to create ad libraries: publicly available repositories that contain all advertisements that have appeared on their service, as well as further information about these advertisements. This information includes the identity of the advertiser, when the advertisement was displayed, how many individuals it reached, and the main targeting or exclusion parameters. The RPA requires very large online platforms to include additional information about political advertisements, including information about the way the ad was funded, the advertising campaign and election with which it is linked, and more detailed information about the way the ad was targeted.<sup>43</sup> The European Parliament and Council have additionally proposed that political advertisements are included in ad libraries in real time.<sup>44</sup> The DSA regulates the functioning of ad libraries by requiring they are accessible both through an API as well as “a searchable and reliable tool that allows multicriteria queries”.<sup>45</sup> The RPA only requires that information about political ads is “where technically possible, machine readable”.<sup>46</sup> Given that the RPA requires information about political ads to be included in the ad libraries established in the DSA, the information about political ads is also subject to the form requirements

the DSA imposes on ad libraries.

The DSA and RPA also provide a limited group of private actors with a right to request further data, in addition to the data that is available to the public through ad libraries. The personal scope of the DSA’s data access right is relatively restrictive. It requires researchers to submit an application for ‘vetted researcher’ status for specific research into a very large online platform or search engine to the competent supervisory authority (called the Digital Services Coordinator, or DSC) in the Member State where a very large online platform or search engine is established. The application must demonstrate that the researchers can put in place appropriate technical and organisational data security measures, that the (timeframe for) the requested data access is necessary and proportionate, and that the requested data is only used to research systemic risks and platforms’ efforts to mitigate them.<sup>47</sup> Additionally, the application must disclose the funding of the research, include a commitment to make the results of the research publicly available free of charge, and demonstrate that the researchers are independent from commercial interests and affiliated with a research institution (including universities but also certain civil society organisations). If researchers’ application meets these criteria, the DSC shall issue a “reasoned request” for data access on researchers’ behalf.<sup>48</sup> Platforms may request that requests are amended when they do not have access to data, or if providing access will create significant vulnerabilities to the security of their service or the protection of confidential information. In those cases, platforms must propose “appropriate and sufficient” alternative means to access the requested data within 15 days of receiving the request, and the DSC must decide on the platform’s requested amendments within another 15 days.<sup>49</sup>

The data access right in the DSA offers important flexibility. In contrast to the other data access provisions at issue in this article, the DSA does not spell out what data must be provided and how. Instead, researchers must demonstrate in their application to the member state that certain data is necessary and proportionate to understand systemic risks on platforms. Additionally, the access request can specify “appropriate interfaces” through which data must be provided.<sup>50</sup> Under the most expansive interpretation, this would mean the request can require platforms to develop an API to meet their exact specifications in each request. Arguably, however, platforms may request that such requests are amended on grounds that they do not have access to the data (or more specifically, that they do not have the precise form of access requested by researchers). In that case, the way in which data access is provided would depend on what “appropriate and sufficient” alternative means to access data platforms propose (and the DSC accepts).<sup>51</sup>

Finally, the RPA’s data access right is limited compared to its

<sup>40</sup> Though the company discontinued its ad library after formally banning political advertising on its platform, the lack of data access makes it difficult for researchers to scrutinise the enforcement of this policy. Twitter has since lifted its ban on political advertising. At time of writing, it has not reinstated the ad library, and instead allows individuals to request data on specific political ads and advertisers. ‘Impact Assessment Report 1/2 Accompanying the DSA’ (n 3) 50; Edelson and others (n 39) 9; Medina Serrano, Papakyriakopoulos and Hegelich (n 21) 119; Leerssen (n 2) 48; Bruns (n 3) 1548; Heidi Ledford, ‘Researchers Scramble as Twitter Plans to End Free Data Access’ (2023) 614 *Nature* 602 <<https://www.nature.com/articles/d41586-023-00460-z>>.

<sup>41</sup> ‘Terms Of Service Research API’ (TikTok, 10 February 2023) <<https://www.tiktok.com/legal/page/global/terms-of-service-research-api/en>> Section II.1; Joe Bak-Coleman, ‘TikTok’s API Guidelines Are a Minefield for Researchers’ (*Tech Policy Press*, 22 February 2023) <<https://techpolicy.press/tiktoks-api-guidelines-are-a-minefield-for-researchers/>>.

<sup>42</sup> ‘Impact Assessment Report 1/2 Accompanying the DSA’ (n 3) 17–18; European Commission, ‘Impact Assessment Report Accompanying the Proposal for a Regulation of the European Parliament and of The Council on the Transparency and Targeting of Political Advertising’ (2021) SWD(2021) 355 47 <[https://ec.europa.eu/info/sites/default/files/4\\_1\\_177594\\_pol-ads-ia\\_en\\_0.pdf](https://ec.europa.eu/info/sites/default/files/4_1_177594_pol-ads-ia_en_0.pdf)>. See in detail on the Commission’s concerns about the limits of voluntary data access Mathias Vermeulen, ‘The Keys to the Kingdom. Overcoming GDPR-Concerns to Unlock Access to Platform Data for Independent Researchers’ 10 <<https://osf.io/vnswz/>>.

<sup>43</sup> DSA article 39; RPA article 7(6), Annex I and II.

<sup>44</sup> Amendments adopted by the European Parliament on 2 February 2023 on the proposal for a regulation of the European Parliament and of the Council on the transparency and targeting of political advertising 2023 [P9\_TA(2023) 0027] article 7b(2); General approach on the Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising 2022 [16013/1/22 REV 1] article 7(6).

<sup>45</sup> DSA article 39(1).

<sup>46</sup> RPA article 7(4), 12(6).

<sup>47</sup> DSA article 40(4), 40(8)(e).

<sup>48</sup> See in detail on the vetting process under article 40 DSA Vermeulen (n 6).

<sup>49</sup> DSA article 40(5-6).

<sup>50</sup> DSA article 40(7). Who exactly specifies how data access is to be provided is somewhat ambiguous under article 40(7) DSA. Arguably, the word “request” implies the DSC (who submits a data access request to platforms following an application for data access by researchers) can specify how data is to be provided. However, the DSA also requires researchers to describe the access to data they need in their application (enabling them to specify the type of access that should be provided) and refers to the vetting by the DSC as an administrative task (recital 124), implying the DSC does not make substantive changes to the way researchers access data. Such an interpretation would also fit with the goal of article 40 – as we argue below, researchers’ ability to determine how access is provided is crucial to their ability to research platforms under article 40.

<sup>51</sup> DSA article 40(6).

counterpart in the DSA, and covers largely the same data as is included in the ad library.<sup>52</sup> However, this data can be requested from a much wider group of actors (including smaller platforms and ad-tech providers) and by a much wider group of actors (including journalists, electoral observers, and political actors).<sup>53</sup> Data providers can reject unclear, excessive, or manifestly unfounded requests, and may also aggregate amounts to protect their commercial interests. The provisions governing the right to request data under the RPA do not spell out how data must be provided. Though the requested data is arguably covered by the RPA's general requirements for the way in which information must be provided, these requirements merely specify that the provided information is machine readable where technically possible and "shall be in writing and may be in electronic form".<sup>54</sup>

### 3. A software development perspective on operationalising data access regimes

The DSA and RPA require platforms to create software products. The creation of an ad library like the one already voluntarily implemented by Meta for instance, involves the implementation of automated mechanisms to categorize ads into political and non-political, implementation of code to gather, process, and store data (and meta-data) related to each ad, the creation of mechanisms to update and refresh databases, as well as mechanisms for serving the data, for instance through the design and implementation of an API or even through visual components like a website for users to interact with via a web browser. As discussed in the previous section, neither the DSA nor the RPA provides concrete guidance on how platforms' data access obligations must be implemented. Rather, both acts set minimum technical requirements leaving platforms considerable room for interpretation to implement APIs or ad libraries as they see fit. Examples of the types of discretionary decisions platforms are left with include, for instance, the need to decide which subset of ads displayed on their service constitute political content about which more information needs to be included in the ad libraries, to what level of aggregation advertiser spending data is made available, whether researchers can search the text in a video or the image use in an ad, or how fast non-political advertisements are included in the ad library.<sup>55</sup>

Within such discretionary design choices lies a danger that platforms could limit research by shaping "how" access is given instead of by deciding whether access to data is provided (as they could prior to the DSA and RPA). Case in point, the development of the EU's self-regulatory policy on disinformation (where much of the regulatory pressure on platforms to provide researchers access to data was exercised prior to the DSA and RPA) shows continued conflict over the operationalization of data access. In particular, though platforms committed themselves to making data available, the Commission

pushed for more specific requirements such as standard data formats, real-time data, multiple search criteria, and the importance of ensuring "functionalities of the APIs are regularly updated to meet researchers' needs."<sup>56</sup> This assessment echoes criticism of platforms' existing, voluntary data access regimes from researchers, who have argued the information provided by platforms is insufficiently precise,<sup>57</sup> inconsistently included,<sup>58</sup> removed,<sup>59</sup> does not include key data necessary to understand how information is targeted and amplified,<sup>60</sup> and that the technical tools used to access information have limited functionalities<sup>61</sup> and present data in a manner that makes it difficult to aggregate.<sup>62</sup>

By giving researchers the legal right to research and scrutinize large online platforms, the DSA and RPA recognize researchers' crucial role in establishing transparency, and the need for legislation to enable researchers to fulfil this role.<sup>63</sup> However, a critical question going forward is to what extent these acts not only recognize researchers' role to scrutinise platforms, but also account for the ways in which data access regimes can be operationalised in a way that ensures they meet the needs of the researcher the DSA and RPA aim to enable to scrutinise platforms.

From a software development perspective, and note that APIs and ad libraries are examples of software, involving end-users by eliciting their needs as part of the design and development process is an essential cornerstone of the engineering practice. This is underpinned by an understanding that systems that do not account for and adapt to their end-users' needs fail, whereas those that involve their end-users in their

<sup>56</sup> European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions European Commission Guidance on Strengthening the Code of Practice on Disinformation' (2021) COM(2021) 262 final 11 <<https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2021:262:FIN>>; European Commission, 'Assessment of the Code of Practice on Disinformation - Achievements and Areas for Further Improvement' (2020) SWD(2020) 180 final. The Code of Practice has been strengthened following the Commission's assessment, and now includes a requirement that "[r]elevant Signatories will engage with researchers and update the functionalities of the APIs and other interfaces to meet researchers' reasonable needs where applicable" European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (2022) 14 <<https://ec.europa.eu/newsroom/dae/redirection/document/87585>>.

<sup>57</sup> Edelson and others (n 21) 1.

<sup>58</sup> Edelson and others (n 21) 9; Niamh Kirk and Lauren Teeling, 'A Review of Political Advertising Online during the 2019 European Elections and Establishing Future Regulatory Requirements in Ireland' (2021) 0 Irish Political Studies 1, 91, 98 <<https://doi.org/10.1080/07907184.2021.1907888>>.

<sup>59</sup> Laura Edelson, Tobias Lauinger and Damon McCoy, 'A Security Analysis of the Facebook Ad Library', 2020 *IEEE Symposium on Security and Privacy (SP)* (2020) 667.

<sup>60</sup> Daniel Kreiss and Bridget Barrett, 'Democratic Tradeoffs: Platforms and Political Advertising' 506 <<https://kb.osu.edu/handle/1811/92273>>; Philippe R Dubois, Camille Arteau-Leclerc and Thierry Giasson, 'Microtargeting, Social Media, and Third Party Advertising: Why the Facebook Ad Library Cannot Prevent Threats to Canadian Democracy' in Holly Ann Garnett and Michael Pal (eds), *Cyber-Threats to Canadian Democracy* (McGill Queen University Press 2021); Rieke and Bogen (n 29). 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>>; Medina Serrano, Papakyriakopoulos and Hegelich (n 21) 118.

<sup>61</sup> Medina Serrano, Papakyriakopoulos and Hegelich (n 21) 18; Leerssen and others (n 16) 12; Edelson and others (n 21) 5.5; Ausloos, Leerssen and ten Thije (n 1) 19.

<sup>62</sup> Rieke and Bogen (n 29) 13; Leerssen and others (n 16).

<sup>63</sup> DSA recital 96; RPA recitals 42, 46.

<sup>52</sup> RPA article 6. In addition to what is included in the ad library, the RPA's right to request data also covers information about the specific service that has been provided to a political advertiser and the amount that has been invoiced.

<sup>53</sup> RPA recital 1, article 2(5), 2(12); Max Zeno van Drunen, Natalie Helberger and Ronan Ó Fathaigh, 'The Beginning of EU Political Advertising Law: Unifying Democratic Visions through the Internal Market' [2022] *International Journal of Law and Information Technology* eaac017, 8 <<https://doi.org/10.1093/ijlit/eaac017>>.

<sup>54</sup> RPA article 6, 7(4), 12(6). In contrast, RPA article 10(1) (which entitles regulators to request data) does require that transmitted information is (inter alia) complete, intelligible, and machine readable where technically possible.

<sup>55</sup> The real-time inclusion of political advertisements in the ad library has been proposed by the Parliament and Council in their positions on the RPA. If these amendments are rejected, the point above applies to all advertisements.



design and implementation have a higher degree of success.<sup>64</sup> Current industry best practices reflect this by the ubiquitous shift to so-called *agile* iterative processes, which embed attention to functional end-users needs as well as other cross-functional requirements into daily engineering operations and processes. Popularized by the so-called “Agile Manifesto” in the early 2000s, the term *agile* as used in industry vernacular (and also how it is being used in the context of this article), is an umbrella term referring to practices designed to iteratively improve software by eliciting requirements and feedback during all stages of the software development lifecycle.<sup>65</sup> The agile manifesto puts this as: “[the] highest priority is to satisfy the customer through early and continuous delivery of valuable software”.<sup>66</sup>

Agile practices have become pervasive as industry practitioners gather in academic-style dedicated conferences<sup>67</sup> to discuss their experiences with various agile development methodologies such as *Continuous Intergration and Delivery (CI/CD)*<sup>68</sup>, and “*DevOps*”<sup>69</sup>. In that context, academic research has explored the efficacy and challenges of various agile development methods<sup>70</sup> as well as their application in various contexts such as in data-driven software projects<sup>71</sup> and even safety critical systems<sup>72</sup> or even how to further develop agile practices by accommodating for even more requirements, for instance security requirements, into their processes.<sup>73</sup> Agile practices being common place is also recognizable by the fact that large cloud platforms have entire suites of tooling and services built and organized around their concepts. “*Azure DevOps*” or “*AWS DevOps*” - respectively Microsoft and

Amazon’s offerings in this space - are but a few examples purporting to “simplify [the] provisioning and managing [of] infrastructure, deploying application code, automating software release processes, and monitoring [of] application and infrastructure performance”.<sup>74</sup> The concept of agility is in fact so commonplace among software developers that it even appears in their lobbying efforts towards and requested of policy makers to make sure regulation “remains *agile* in capturing ongoing developments”.<sup>75</sup> To our knowledge, there is no research that explores how agile development processes are employed by platforms in the context of data access regimes. This may be due to the relative recency of platforms’ data access regimes, or part of platforms’ general tendency to limit external insight into their internal operations. Yet as we have argued above, platforms are also not incentivised to develop data access regimes in a way that maximises their value to the end-user; as we will argue below, enforcement and standard-setting mechanisms can play an important role by ensuring platforms operationalise data access regimes in accordance with effective software design principles.

Agile practices’ close attention to the need to continuously identify and adapt software to end-users’ needs is grounded in the notion that it is not always clear at the start of the development process what end-users’ exact needs are and how software can best be developed to meet them. Particularly in the research context on which this article focuses, end-users’ needs also change over time. Researchers’ needs evolve as the way information is distributed on platforms changes, and as researchers gain a better understanding of the problems facing platforms and develop new tools and methodologies to study these problems. This evolution is inherent to the research process the DSA and RPA aim to enable. With respect to platform transparency, despite researchers’ technical requirements not being clear and even dynamically shifting over time, the widely adopted iterative processes of the software industry are well suited for dealing with such uncertainty. Agile/iterative processes typically capture requirements through repeated cycles of elicitation, followed by a concretisation of requirements via for instance so-called ‘user stories’ which in simple terms summarise what users need and how they expect a system to function. Current industry best practices are also characterised by the routine integration and delivery of small sets of features, over short time windows of typically a handful of days to weeks – thus gradually growing, adapting, and shaping a digital product or service into what its end-users consume.<sup>76</sup> Best practices also dictate routine automated testing of software (e.g., ‘beta’ and ‘acceptance’ tests) to ensure that it functions correctly as well as meets its end-user demands.<sup>77</sup> We indeed see such practices throughout the industry, including with respect to many of the products and services platforms themselves provide to their users, such as their users’ social media news feeds, timelines, and even the backend recommender systems populating their users’ timelines with content.

From a software development perspective, if data access regimes are expected to meet the needs of the researchers that the DSA and RPA aim to empower by imposing data access obligations, the implementation of these obligations should closely and iteratively account for researchers’ feedback on (for example) what data they provide, what features their APIs implement, and how data access regimes function, inter-operate, and perform. However, the envisioned processes by which platforms provide data under the data access obligations in the DSA and RPA are

<sup>64</sup> Muneera Bano and Didar Zowghi, ‘A Systematic Review on the Relationship between User Involvement and System Success’ (2015) 58 *Information and Software Technology* 148 <<https://www.sciencedirect.com/science/article/pii/S0950584914001505>>; Dennis Pagano and Bernd Bruegge, ‘User Involvement in Software Evolution Practice: A Case Study’, *2013 35th International Conference on Software Engineering (ICSE)* (2013); Muneera Bano and Didar Zowghi, ‘User Involvement in Software Development and System Success: A Systematic Literature Review’, *Proceedings of the 17th International Conference on Evaluation and Assessment in Software Engineering* (Association for Computing Machinery 2013) <<https://doi.org/10.1145/2460999.2461017>>; Muneera Bano, Didar Zowghi and Francesca da Rimini, ‘User Involvement in Software Development: The Good, the Bad, and the Ugly’ (2018) 35 *IEEE Software* 8.

<sup>65</sup> Degif Tekka, Yvonne Dittrich and Mesfin Kifle, ‘Adapting Lightweight User-Centered Design with the Scrum-Based Development Process’, *Proceedings of the 2018 International Conference on Software Engineering in Africa* (Association for Computing Machinery 2018) <<https://dl.acm.org/doi/10.1145/319552.8.3195530>>.

<sup>66</sup> ‘12 Principles Behind the Agile Manifesto | Agile Alliance’ (*Agile Alliance* |, 4 November 2015) <<https://www.agilealliance.org/agile101/12-principles-behind-the-agile-manifesto/>>.

<sup>67</sup> ‘Agile Events Calendar’ (*Agile Alliance* |, 18 January 2019) <<https://www.agilealliance.org/events-calendar/>>.

<sup>68</sup> Mojtaba Shahin, Muhammad Ali Babar and Liming Zhu, ‘Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices’ (2017) 5 *IEEE Access* 3909.

<sup>69</sup> Christof Ebert and others, ‘DevOps’ (2016) 33 *IEEE Software* 94; Manish Virmani, ‘Understanding DevOps & Bridging the Gap from Continuous Integration to Continuous Delivery’, *Fifth International Conference on the Innovative Computing Technology (INTECH 2015)* (2015).

<sup>70</sup> Shahin, Ali Babar and Zhu (n 68).

<sup>71</sup> Astri Barbala, Tor Sporse and Viktoria Stray, ‘Data-Driven Development in Public Sector: How Agile Product Teams Maneuver Data Privacy Regulations’ in Christoph J Stettina, Juan Garbajosa and Philippe Kruchten (eds), *Agile Processes in Software Engineering and Extreme Programming* (Springer Nature Switzerland 2023).

<sup>72</sup> Kevin Gary and others, ‘Agile Methods for Open Source Safety-Critical Software’ (2011) 41 *Software: Practice and Experience* 945 <<https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.1075>>.

<sup>73</sup> Håvard Myrbakken and Ricardo Colomo-Palacios, ‘DevSecOps: A Multivocal Literature Review’ in Antonia Mas and others (eds), *Software Process Improvement and Capability Determination* (Springer International Publishing 2017).

<sup>74</sup> ‘DevOps - Amazon Web Services (AWS)’ (*Amazon Web Services, Inc.*) <<https://aws.amazon.com/devops/>> accessed 3 July 2023.

<sup>75</sup> OpenAI, ‘OpenAI White Paper on the European Union’s Artificial Intelligence Act’ (2023) 7 <<https://time.com/6288245/openai-eu-lobbying-ai-act/>>.

<sup>76</sup> Manuel Brhel and others, ‘Exploring Principles of User-Centered Agile Software Development: A Literature Review’ (2015) 61 *Information and Software Technology* 163 <<https://www.sciencedirect.com/science/article/pii/S0950584915000129>>.

<sup>77</sup> Bertrand Meyer, ‘Seven Principles of Software Testing’ (2008) 41 *Computer* 99.



markedly different from the typical software development industry practices described above. For instance, regular end-user requirement elicitation, iterative and continuous development of the APIs or other suitable data access mechanisms, as well as regular end-user acceptance and/or beta testing are not implicitly nor explicitly encoded within these acts. Instead, both acts focus on requiring platforms to make data available, and at least initially leave platforms considerable discretion to determine how data can be made available in a way that enables researchers to scrutinise their service. As such, a clear risk remains that in a context where incentives are misaligned, researchers' efforts to scrutinise platforms are frustrated via how data access is implemented.

#### 4. Controlling the operationalisation of data access regimes through the DSA and RPA

The preceding analysis presents a dilemma. The DSA and RPA require platforms to provide access to data in order to enable researchers to scrutinise how information is distributed, but do not address the technical design choices that determine how useful these data access regimes will be to researchers. However, the DSA and RPA do contain mechanisms with which control can be exercised over the operationalisation of data access regimes. The following section will briefly describe these mechanisms, distinguishing between enforcement and rule-making processes.

##### 4.1. Enforcing access to data

EU law generally relies on national authorities, subject to national procedural law, to apply (and thereby concretize) its norms in specific cases.<sup>78</sup> The RPA follows this traditional approach, and contains few rules on how Member States must enforce its obligations. It requires that data protection authorities oversee compliance with the RPA's chapter on targeted political advertising, including rules on the provision of information about the way advertisements are targeted in response to data access requests or in ad libraries.<sup>79</sup> The RPA does not specify which supervisory authority is competent to ensure compliance with the other provisions, including whether information about political advertisements is included in the ad library, what information about the context of political advertisements is included, and whether data access requests about the context of political advertisements are complied with, though it notes Member States may assign DSCs as the competent authority to supervise online platforms.<sup>80</sup>

The DSA departs from EU law's reliance on national supervisory authorities when it comes to very large online platforms. The DSC of the member state where a very large platform is established retains one important power, namely (as Section 2.2 describes in detail) the ability to submit data access requests to platforms on researchers' behalf. The Commission, however, has the exclusive power to enforce the provisions that apply to very large online platforms, including article 39 DSA (on ad libraries) and article 40 DSA (on data access requests). Ultimately, the Commission can sanction platforms for violating these provisions with a fine of up to 6 % of annual turnover. However, before the Commission formally decides a platform has failed to comply with the DSA, it must inform the platform about the measures it intends to take, and the measures it believes the platform should take to comply with the DSA.<sup>81</sup> Platforms may offer to make binding commitments intended to ensure compliance with the DSA, including participation in the otherwise

voluntary standards and codes of conduct described in the next section.<sup>82</sup> If the Commission does formally decide a platform has violated the DSA, the platform must draw up an action plan outlining the measures with which it will remedy the infringement.<sup>83</sup> In short, the DSA's enforcement mechanism formalize a process through which Commission and platforms determine what concrete actions platforms should take to comply with the data access provisions both before and after an infringement has been found.

The DSA and RPA focus on enforcement by public authorities to enforce researchers' access to data. They provide researchers limited options to contest data access when public authorities opt not to enforce data access obligations. While the DSA creates a right to lodge a complaint with a DSC, this right does not apply to the Commission (which has the exclusive power to enforce the data access obligations) or a DSC's decision whether or not to grant researchers access to data under art. 40 DSA (as the right to lodge a complaint only covers infringements by intermediaries).<sup>84</sup> The DSA does explicitly enable private enforcement by ensuring platform users have the right to sue for any damages they suffered as a result of a violation of their data access obligations.<sup>85</sup> However, even if researchers can claim to have suffered damages because platforms have violated their data access obligations, this does not necessarily provide researchers access to the data they need to fulfil their societal role.

##### 4.2. Rule-making processes

Both the DSA and RPA foresee the creation of more specific rules about the way data access regimes should be operationalised. Mostly, these rules take the form of non-binding policy recommendations and self-regulation which governed data access prior to the DSA and RPA. Specifically, the DSA requires the Commission and the European Board for Digital Services (an institution chaired by the Commission that consists of the different DSCs, hereafter: the Board) to support the development of voluntary standards and codes of conduct regarding APIs through which data in ad libraries is accessed, the interoperability of ad libraries, and the way online advertising intermediaries transmit information to ad libraries.<sup>86</sup> The Commission is also required to support updates to the voluntary standards on APIs and interoperability in light of technological developments.<sup>87</sup> In addition to these industry standards, the Board has a more general task to issue recommendations and promote European standards.<sup>88</sup> It can use this power (as its counterpart in data protection law, the European Data Protection Board, has long done) to issue authoritative guidance on the way the DSA should be applied. Finally, the Commission is specifically empowered to adopt guidelines on the structure, organisation, and functionalities of ad libraries.<sup>89</sup>

As these guidelines and codes of conduct are non-binding, they do not directly address the misaligned incentives that led the DSA and RPA to move away from voluntary data access. However, the DSA and RPA do introduce a salient difference: they have complemented the threat of stricter regulations with the threat of stricter enforcement as a way to pressure platforms to comply with non-binding standards.

<sup>78</sup> DSA article 71(1).

<sup>79</sup> DSA article 75.

<sup>80</sup> That is also not to say researchers may not be able to challenge decisions by the DSC under national procedural law (see also DSA article 51(6)), also taking into account the EU principles of equivalence and effectiveness, as well as the right of access to justice.

<sup>81</sup> DSA article 54.

<sup>82</sup> DSA article 44(1)(d), 44(1)(f), 46(2)(b), 45(1). DSA article 63(1)(e) also empowers the Board to issue opinions about the standards set out under DSA article 44.

<sup>83</sup> DSA article 44(2), 45(4).

<sup>84</sup> DSA article 63(1)(c), recital 132.

<sup>85</sup> DSA article 39(3).

<sup>78</sup> Rob Widdershoven, 'National Procedural Autonomy and General EU Law Limits' (2019) 12 *Uitgeverij Paris 5* <<https://www.uitgeverijparis.nl/nl/reade/r/206691/1001470455>>.

<sup>79</sup> RPA article 15(1), 12(3), 13.

<sup>80</sup> RPA article 7(2), 7(6), 11, 15(2).

<sup>81</sup> DSA article 73(3).

The Commission's guidelines on ad libraries in particular can signal what the Commission believes broad terms such as APIs require platforms to do in practice, and therefore how it will enforce the DSA's data access obligations. However, it should be noted that regulatory pressure to comply with voluntary standards is more likely to be effective when those non-binding standards expand on an explicit obligation in the DSA (such as the requirement to provide APIs) than when those standards cover aspects of data access not regulated in the DSA (such as interoperability requirements).

The DSA and RPA also introduce a tool that is new to platform governance, at least in the EU context: delegated acts. With delegated acts, the EU legislator empowers the Commission to amend and/or supplement non-essential elements of legislation. This is generally done when the legislator foresees a need to make fast, legally binding changes to legislation requiring technical expertise without repeating the full legislative process.<sup>90</sup> Typical examples are rules on pollution levels or food safety. The Parliament and Council retain some control, as they have the power to veto delegated acts (within six months under the DSA and four under the RPA).<sup>91</sup> Additionally, the Commission must consult experts appointed by Member States while drafting delegated acts; the Council's and Parliament's experts may also access these expert group meetings.<sup>92</sup>

Under the DSA, the Commission may use delegated acts "laying down the technical conditions under which providers of very large on-line platforms are to share data" provided in response to data access requests under art. 40 DSA. Specifically, these delegated acts can concern the conditions under which data is shared, indicators of the objectives for which data may be shared, and procedures through which data can be shared with vetted researchers.<sup>93</sup> Additionally, the RPA empowers the Commission to use delegated acts to amend, remove, or (in the case of transparency requirements that do not concern targeting) add information elements platforms must make available under the RPA. Delegated acts concerning targeting may be adopted when necessary due to technological developments in scientific research and developments in supervision, and otherwise when necessary for the advertisement's wider context to be understood in light of technological developments.<sup>94</sup>

## 5. Designing data access to enable research

The DSA and RPA take a two-pronged approach to address platforms'

<sup>90</sup> TFEU article 290. Robert Schütze, "'Delegated' Legislation in the (New) European Union: A Constitutional Analysis" (2011) 74 *The Modern Law Review* 661, 682 <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2230.2011.00866.x>>; Merijn Chamon, 'Limits to Delegation under Article 290 TFEU: The Specificity and Essentiality Requirements Put to the Test' (2018) 25 *Maastricht Journal of European and Comparative Law* 231, 239 <<https://doi.org/10.1177/1023263X18760548>>; Alexander H Türk, 'Legislative, Delegated Acts, Comitology and Interinstitutional Conundrum in EU Law – Configuring EU Normative Spaces' 26 *European Law Journal* 415, 2 <<https://onlinelibrary.wiley.com/doi/abs/10.1111/eulj.12400>>.

<sup>91</sup> Though delegated acts are rarely vetoed, the threat of a veto can push the Commission to adapt drafted delegated acts. Michael Kaeding and Kevin M Stack, 'Legislative Scrutiny? The Political Economy and Practice of Legislative Vetoes in the European Union' (2015) 53 *Journal of Common Market Studies* 1268, 1269 <<https://onlinelibrary.wiley.com/doi/abs/10.1111/jcms.12252>>.

<sup>92</sup> Katrijn Siderius and Gijs Jan Brandsma, 'The Effect of Removing Voting Rules: Consultation Practices in the Commission's Delegated Act Expert Groups and Comitology Committees' (2016) 54 *JCMS: Journal of Common Market Studies* 1265 <<https://onlinelibrary.wiley.com/doi/abs/10.1111/jcms.12380>>; Paul Craig, 'Delegated and Implementing Acts', *Delegated and Implementing Acts* (Oxford University Press 2018) <<https://oxford.universitypressscholarship.com/10.1093/oso/9780199533770.001.0001/isbn-9780199533770-book-part-26>>.

<sup>93</sup> DSA article 40(13).

<sup>94</sup> RPA article 7(8), 12(8).

disincentive to provide access to data: they impose broad data access obligations, and enable public authorities to control the way platforms operationalize data access through standard-setting and ultimately enforcement. The scope of both the data access obligations and the concretization mechanisms is limited. In particular, key limitations to the functioning of voluntary ad libraries are not regulated (such as information on why individuals see an ad or the obligation to include ads in ad libraries in real-time), or left to self-regulation (such as interoperability).<sup>95</sup>

Nevertheless, the DSA and RPA take a significant step in platform transparency regulation by explicitly recognizing researchers' societal role to scrutinize how information is disseminated on platforms, as well as the role that law and public authorities can play to facilitate access to the data researchers need to do so. In the next section we explore, based on the previous analyses of the goals and operationalization of data access regimes, how the mechanisms with which public authorities can exercise control over the operationalization of data access can best be applied to ensure researchers are enabled to scrutinize platforms.

### 5.1. Ensuring consistency in the face of fragmented enforcement and standard-setting

The patchwork of enforcement and standard-setting mechanisms described above is complex. Different aspects of ad libraries are subject to enforcement by the Commission and national authorities, as well as self-regulation, delegated regulations, and Commission guidelines. Similarly, multiple DSCs have a responsibility to vet data access requests for the platforms established within their member state, creating a risk of diverging approaches. This creates the danger that multiple different parties formulate different and potentially conflicting requirements for what the tools with which platforms facilitate researchers' data access should look like. This can frustrate research into platforms when the different requirements imposed on a single software product a platform is required to develop (e.g., Twitter's ad library) is insufficiently aligned; at the same time, it can frustrate research into the platform ecosystem as a whole when data access tools provided by different platforms are subject to different voluntary codes of conduct, or DSCs under article 40 and national authorities under the RPA exercise their power to compel data access in diverging ways.

In software design, the issue of multiple stakeholders influencing the design process is addressed by for instance designating actors that are responsible for collecting the needs of the different stakeholders, and communicating them to the developers in order to ensure they create a coherent end product. This approach cannot simply be transplanted to the legal context. The question of who enforces (for example) political advertising law not only depends on efficiency and consistency, but also on political and normative factors. Nevertheless, insights from software design indicate the importance of coordination between the different actors that can influence how platforms design data access regimes. In that context, the Commission can play an important role. It not only has its own enforcement and standard-setting powers, but must also support the creation of industry standards under the DSA, and chairs the Board in which DSCs collaborate. This combination of powers affords the Commission a unique position with which it can identify the risk of fragmentation in data access design, and push for a consistent approach across the different standard-setting and enforcement mechanisms that require platforms to provide access to data.

<sup>95</sup> See e.g. on the need for cross platform studies in context of disinformation Petra de Place Bak, Jessica Gabriele Walter and Anja Bechmann, 'Digital False Information at Scale in the European Union: Current State of Research in Various Disciplines, and Future Directions' [2022] *New Media & Society* <<https://doi.org/10.1177/14614448221122146>>.

## 5.2. Integrating researchers' needs in data access regimes

The DSA and RPA's goal to provide researchers with data they need to scrutinize platforms raises a logical follow-up question: how can the design process for data access regimes be set up to account for researchers' needs? This is easier said than done. Researchers (and their needs) are not homogenous. Data access regimes are used by a variety of researchers, including academics, NGOs, and (in the case of ad libraries and access requests under the RPA) journalists.<sup>96</sup> Software development principles indicate the importance of a process through which the needs of the different users of data access regimes are identified, prioritised or balanced when they differ from one another, and incorporated in the design of data access regimes.

The different operationalization mechanisms in the DSA and RPA are not equally suitable to take different researchers' needs into account. Enforcement by courts and supervisory authorities can play an important role by establishing minimum standards with which data access regimes should comply to enable researchers to scrutinize how information is disseminated on platforms. However, their individualized focus makes enforcement mechanisms poorly suited to identify the needs of the broader research community that uses data access regimes. There are mechanisms to mitigate this issue. Supervisory authorities and courts can draw on guidelines and delegated regulations to take a broader perspective of the ways in which a legal rule should be operationalised, or consult external parties when adjudicating a case. However, with the exception of delegated regulations there is no legal obligation for supervisory authorities or courts to take these sources of information into account when interpreting what the DSA or RPA require, and any ruling is focused on the specific violation that is alleged to occur.<sup>97</sup> In short, enforcement's individualized focus is at odds with software development practices that indicate the importance of a process that integrates users' multiple and potentially diverging requirements.

In contrast, standard-setting mechanisms are designed to create generally applicable rules and as such, often require or allow for the consultation of the different stakeholders affected by these rules. This enables the more comprehensive identification of researchers' needs that is lacking in individual enforcement procedures. However, the increased involvement of external private parties in the rulemaking process also opens up the risk that rules are shaped to meet the needs of the industry to which they apply, rather than the researchers that use data access regimes. Just as platforms have an incentive to shape how researchers can scrutinise their services by influencing how they make data available, so too do they have an incentive to influence the rules that determine how data access must be provided.

This risk is particularly large in the self- and co-regulatory mechanisms, where platforms take a lead role in the rulemaking process.<sup>98</sup> However, platforms can also influence government-led rulemaking processes such as delegated regulations, particularly by providing the information public authorities rely on to understand what problems

their regulations should address and how they can do so.<sup>99</sup> Public authorities can be particularly reliant on industry expertise where there are few other independent entities who can provide expertise – as is arguably the case in the context of data access. The only ad libraries in existence have been built by a small number of platforms, and platforms have information at the ready about the technical, security, and privacy issues involved in sharing large amounts of data with researchers, as well as how data access regimes interact with the platform systems from which data is aggregated to transmit to researchers.<sup>100</sup> In other words, there is a risk that platforms' power to influence how researchers scrutinise their impact shifts from controlling if and how they voluntarily make data available, to using their expertise as a source of power to influence the delegated regulations and guidelines concretising data access obligations.<sup>101</sup>

The typical remedy for an overreliance on industry expertise is to draw on a diverse set of perspectives.<sup>102</sup> The basic idea behind this approach is simple: information provided by the industry is not inherently problematic, and may even produce better regulatory outcomes. However, where industry expertise is the only information available it narrows regulators' perspective. Taking non-industry perspectives into account resolves this issue. In the context of data access, it can be particularly useful to draw on the expertise that has been accumulated through researchers' experience with platforms' voluntary data access regimes, as well as experience with data access regimes under different legal frameworks.<sup>103</sup> We argue, however, it is necessary to go a step further, and determine what expertise platforms are qualified to offer. Platforms may provide useful input on the way data access regimes interact with other software systems, for example by indicating how access to data spread across multiple separate systems can be provided, or how data access impacts the security of their other systems. Article 40 also requires that the delegated act laying down conditions for data access accounts for platforms' rights and interests. But on core questions in data access (e.g., what information researchers need to scrutinise online information distribution, and how data can be provided in a way that best enables such research), the experts are not platforms, but the researchers themselves. To ensure data access regimes allow researchers to fulfil the role the DSA and RPA set out for them, it is crucial to afford researchers a central role in the standard-setting process.

Only when it comes to the guidelines for ad libraries is there a clear legal obligation to consult researchers and the public. However, though consulting researchers is not required with regard to the other standard-setting processes, doing so would promote the DSA's and RPA's aim of enabling researchers to scrutinise platforms. The Commission again has a large role to play here. It is the authority that the DSA and RPA empower to draft guidelines on ad libraries and delegated regulations, and to support the self-regulatory process. However, Member states and the European Parliament can also ensure researchers' expertise is taken into account when appointing experts the commission must consult

<sup>96</sup> Leerssen and others (n 16); Brian McKernan and others, 'A Human-Centered Design Approach to Creating Tools to Help Journalists Monitor Digital Political Ads: Insights and Challenges' (2022) 0 Digital Journalism 1 <<https://doi.org/10.1080/21670811.2022.2064321>>.

<sup>97</sup> Moreover, the CJEU does not formally allow amicus curiae briefs or third-party interventions.

<sup>98</sup> Sandra Eckert, *Corporate Power and Regulation: Consumers and the Environment in the European Union* (Springer International Publishing 2019) 26 <<http://link.springer.com/10.1007/978-3-030-05463-2>>.

<sup>99</sup> Jan Beyers and Sarah Arras, 'Who Feeds Information to Regulators? Stakeholder Diversity in European Union Regulatory Agency Consultations' (2020) 40 Journal of Public Policy 573 <<https://www.cambridge.org/core/journals/journal-of-public-policy/article/who-feeds-information-to-regulator-s-stakeholder-diversity-in-european-union-regulatory-agency-consultation/s/C2FC172CF9AB3C4AEB606D9F3A81372>>.

<sup>100</sup> Cary Coglianese, Richard Zeckhauser and Edward Parson, 'Seeking Truth for Power: Informational Strategy and Regulatory Policymaking' [2004] Minnesota Law Review <<https://scholarship.law.umn.edu/mlr/662>>; Beyers and Arras (n 99).

<sup>101</sup> Eckert (n 98) 33.

<sup>102</sup> Eckert (n 98) 35; Beyers and Arras (n 99).

<sup>103</sup> Ausloos, Leerssen and ten Thije (n 1).

when making delegated regulations.<sup>104</sup> Finally, researchers themselves can participate in the drafting of standards about data access. They may do so in cooperation with platforms – the key example is the EDMO draft code of conduct on platform-to-researcher data access, which was the result of a collaboration between academics, NGOs, and platforms.<sup>105</sup> Such collaboration is an important way to assess if data access can be designed in a way that allows researchers to scrutinise platforms while preventing platforms from contesting access to data (for example due to threats to the security of their service). It also continues to be important for voluntary codes of conduct, whose success depends on buy in from platforms. Yet the shift toward regulated data access also means that access becomes less dependent on platforms' voluntary cooperation, and more dependent on the ways in which public authorities (and particularly the Commission) enforce and set standards on data access. To enable public authorities to pressure platforms to operationalise data access in a way that supports researchers' needs, it is key that researchers clearly stake out their own position to indicate to public authorities what kind of data access they require to scrutinise platforms.

One way to incorporate expertise on data access issues is to set up an intermediary body that vets individual research projects. The need for such a body was identified in the EDMO draft code of conduct; the strengthened code of practice on disinformation now commits signatories to develop, fund and cooperate with such a body, while the delegated act on article 40 DSA also allows for "independent advisory mechanisms in support of sharing of data" to be established.<sup>106</sup> An independent vetting body could provide expert advice on the nuances of data access that an overburdened DSC may lack. In so doing it could also ensure article 40 can be used by a wider variety of researchers, by decreasing the time needed for vetting, supporting under-resourced projects, and fully assessing the permissibility of more experimental access requests that a DSC might cautiously reject out of a fear it cannot adequately access all their risks. Yet such an intermediary body would fulfil an important gatekeeping role. As such, in line with the analysis above, it is key that the intermediary body represents and can respond to the needs of the wide variety of researchers entitled to use article 40.

### 5.3. Data access can and should be an iterative process

A key challenge to legislation that is used to enable researchers to fulfil their societal role is that researchers' needs are not static. They evolve as researchers develop new tools and methodologies, as the platforms they study change, and as researchers gain a better understanding (ideally in part due to the data provided due to the DSA and RPA) of platforms and the questions they should ask. From a software design perspective, failing to account for this evolution would mean data access regimes are developed based on a single snapshot of the needs of their users, even though the precise nature of these needs is subject to change and unclear at the start of the development process. From a legal perspective, this would in turn jeopardise data access regimes' ability to fulfil their function, namely to enable researchers to scrutinise and detect (emerging) risks concerning the way information is distributed on platforms. Research into platforms would instead be limited to the problems as they are currently understood and with the tools currently

available. In part for these reasons, the EDMO draft code of conduct on platform to researcher data access was designed to be flexible, so as not to foreclose future research projects and methodologies.<sup>107</sup>

The question, however, is to what extent the provisions governing data access in the DSA and RPA are flexible enough to impose different requirements on platforms in light of researchers' evolving needs. The DSA's right to request data stands out in this context. In contrast to the other data access provisions at issue in this article (which list what data must be provided), the DSA's data requests take a goal-oriented approach by providing access to the data that is necessary to understand (platforms' efforts to mitigate) systemic risks. Additionally, the request may specify the way in which data must be provided. This approach makes it possible to require access to different data depending on what is needed to understand information flows on platforms at any given time. Moreover, data access requests can be filed repeatedly, and the delay between filing a data access request and obtaining access to data is relatively short by legal standards (30 days, provided the DSC's decision is not challenged before a court). This further increases researchers' ability to adapt the access to data depending on what is needed to understand information distribution on platforms.

Standard-setting mechanisms offer another opportunity to adapt data access regimes to researchers' dynamic needs. The DSA requires the Commission to support the update of voluntary standards in light of technological developments and user behaviour.<sup>108</sup> Delegated regulations also enable the Commission to update the information that must be added to ad libraries under the RPA, and the technical conditions for sharing data with vetted researchers under the DSA. The need to quickly update rules to account for new developments such as changes in the way research is done is precisely the logic behind delegated regulations, which allow the Commission to respond to detailed issues and technological developments not foreseen in the underlying regulation.<sup>109</sup> Software development practices indicate a number of principles that could ensure updates to the standards on data access better facilitate researchers' ability to scrutinise platforms. It is important for the updates to take account of researchers' experience with data access regimes, the limits they encounter, and how their use of data access regimes changed following the updated standards. This allows for future updates to build on existing experience with the way data access regimes' design impacts their use by researchers. Such information could be gathered by active engagement with the research community, but also by providing tools with which users can submit feature requests or bug reports to platforms and public authorities. To speed up this iterative improvement process and ensure that data access regimes are adapted to address changes in platforms' services, it is important for updates to take place regularly.

By updating delegated regulations and its guidance on ad libraries, and fulfilling its legally required role to support updates to voluntary standards, the Commission can ensure data access regimes are adapted to meet researchers' needs without potentially lengthy enforcement procedures of individual data access provisions in the DSA and RPA. This is particularly important for the standards on ad libraries. Otherwise, the relative flexibility of the data access right available to vetted researchers

<sup>104</sup> RPA recital 66.

<sup>105</sup> EDMO, 'Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access' (EDMO 2022) <<https://edmo.prod.wpengine.com/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf>>.

<sup>106</sup> EDMO (n 105) 12; European Commission, 'The Strengthened Code of Practice on Disinformation 2022' (n 56) 28.

<sup>107</sup> In particular, the EDMO draft code of conduct referred to the fact researchers currently do not know exactly what data platforms collect, or how platforms' data collection and generation practices will change. EDMO (n 105) 11.

<sup>108</sup> DSA article 44(2).

<sup>109</sup> RPA articles 7(8), 12(8).



would place researchers who can only make use of publicly available ad libraries at a greater and greater disadvantage when it comes to scrutinising the impact of platforms.<sup>110</sup>

#### 5.4. Ensuring timely data access

Closely related to the need for an iterative approach is the question how fast platforms comply with their data access obligations. While the DSA and RPA contain mechanisms to control whether and how access to data is provided, they contain only limited rules on how fast data must be made available. The ad library provisions in the DSA do not explicitly address whether ads must be included in the ad library in real-time. The RPA specifies only that best efforts must be made to provide data or a reasoned response to a data access request within a month.<sup>111</sup> Crucially, however, the Parliament and Council have proposed that political ads are included in ad libraries in real time.<sup>112</sup> Finally, the DSA requires researchers who request data to justify the necessity and proportionality of the timeframe in which they request data, and sets out a 30-day window after which a DSC must decide on any amendments to the request proposed by platforms.<sup>113</sup> However, neither the RPA nor the DSA regulates the timeline with which any conflicts over the platforms' compliance with data access requests or their obligation to provide ad libraries must be resolved.

Though the contestation of data access requests provides an important safeguard to ensure data access regimes' impact on privacy, trade secrets, or the security of platforms' services remains proportionate, they also offer the opportunity to delay access to research data. As noted above, software development principles indicate the importance of quick iterations in which software products are changed due to user feedback or changing circumstances, information is collected on the way the changed product is used, and further improvements are made on the basis of this feedback. Delaying researchers' access to data or changes to data access regimes themselves slows down this cycle. It limits the speed with which researchers can determine how data access regimes allow them to scrutinise platforms, how changes to data access regimes further improve or limit this scrutiny, or how existing data access regimes no longer meet researchers' needs because the platform systems about which data is provided have changed.

The speed with which data access is provided is also key to data access regimes' ability to enable accountability and a better understanding of platforms' impact. This is particularly true in the political context. If ad libraries cannot be used to scrutinise political advertising or data access requests cannot be used to scrutinise platforms' impact on electoral processes before the election has taken place, accountability does little good. However, speed is also important in contexts where elections do not set a strict timeline. For example, the DSA's aim of facilitating the detection of emerging risks so they can be addressed before they have a large societal impact is hampered by delays in research access.<sup>114</sup> Delays that result from conflicts over the

<sup>110</sup> See on the normative implications of restricting data access to a limited group of vetted researchers Leerssen (n 2) 38.

<sup>111</sup> RPA article 11(3).

<sup>112</sup> Amendments adopted by the European Parliament on 2 February 2023 on the proposal for a regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (n 44) article 7b(2); General approach on the Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (n 44) article 7(6).

<sup>113</sup> DSA article 40(5-6).

<sup>114</sup> DSA recitals 95, 98.

operationalisation of data access obligations may impact data access requests in particular. Because ad libraries provide the same information in the same way to the general public, any conflict over their scope or design only needs to be resolved once. Conversely, data access requests require researchers to demonstrate (and enable platforms to contest) in each individual case that they should have access to the data. Even if these conflicts are ultimately resolved, addressing them individually costs time and resources, hampering the ability of researchers to scrutinise platforms.<sup>115</sup>

Standard-setting mechanisms can play an important role in this context by reducing the space for conflicts that delay researchers' access to data. They can do so by providing authoritative (and in the case of delegated regulations, legally binding) guidance about what is concretely required to comply with data access obligations. Though some issues (like the way data access can be provided under article 40 in a GDPR-compliant way) still require a case by case assessment, in other cases standard-setting can remove some of the ambiguities that would otherwise have to be resolved through lengthy enforcement processes. This is particularly important for the Commission's power to lay down delegated regulation on the conditions under which platforms are to share access to data with researchers under article 40 DSA. Not only does this reduce the potential for conflict between platforms and researchers over, for example, what security measures researchers must put into place; providing more detailed and legally binding regulatory guidance on the purposes for which researchers are to be provided with access to data also limits the space for conflicts between the DSC and the researchers who wish to obtain vetted researcher status.

Where conflicts over data access do emerge, however, the speed with which researchers are able to access data depends on the speed with which supervisory authorities and courts enforce EU law. Traditionally, and still for the RPA, this is a matter of national procedural autonomy, with EU law only requiring national laws to comply with general principles such as equivalence, effectiveness, and effective judicial protection.<sup>116</sup> Though as we have argued above the effectiveness of the DSA and RPA's data access obligations requires timely access to data (and as such timely enforcement), it is difficult to distil clear and justiciable timelines from these general principles. However, the Commission's exclusive competence to enforce the data access obligations in the DSA opens up the possibility of (and arguably requires) more detailed regulation concerning the way in which the Commission exercises this power.

#### 5.5. Evaluating data access regimes

The data access obligations in the DSA and RPA focus on providing researchers access to data about the way platforms disseminate information. This may sound self-evident, but misses a key aspect of transparency identified in the existing self-regulatory approach to data access: the need for transparency about data access regimes themselves.<sup>117</sup> Such transparency is a precondition for researchers' ability to use data access regimes. For example, without documentation about the data that is provided and how it was collected, it is difficult to gauge the

<sup>115</sup> Ausloos, Leerssen and ten Thije (n 1) 90.

<sup>116</sup> DSA article 50(1), recital 111; Widdershoven (n 78); Steve Peers and others, *The EU Charter of Fundamental Rights: A Commentary* (Nomos Verlag 2022) 1319; Wojciech Piątek and Lukáš Potěšil, 'A Right to Have One's Case Heard within a Reasonable Time before the Czech and the Polish Supreme Administrative Courts – Standards, the Reality and Proposals for the Future' (2021) 17 *Utrecht Law Review* 20 <<http://www.utrechtlawreview.org/articles/10.36633/ulr.586/>>.

<sup>117</sup> Edelson, Lauinger and McCoy (n 59) 664; Ausloos, Leerssen and ten Thije (n 1) 84, 88; Wagner and others (n 6).

data's quality, determine what generalizable conclusions can be drawn from it, and how such conclusions relate to data collected earlier.<sup>118</sup> This in turn limits researchers' ability to determine with any certainty whether any effects they found (or failed to find) are identified correctly. Additionally, transparency about the way researchers can use data access regimes (such as API documentation and information on the available data) is necessary for researchers to be able to use data access regimes to scrutinize platforms.

The DSA and RPA pay little explicit attention to the need to provide researchers with transparency about data access regimes. The list of information to be made available under the data access right in the RPA and the ad libraries provisions in the DSA and RPA does not include meta-information about the way data was collected or provided. However, the DSA's data access right is more open-ended and requires that researchers are given access to the data necessary to understand systemic risks on platforms. This data is not necessarily restricted to information that directly concerns the way content is disseminated on platforms. On the contrary, as argued above, transparency about data access regimes is a precondition for researchers' ability to use these regimes to understand platforms' impact. Article 40's obligation to provide data necessary to understand (platforms' efforts to mitigate) systemic risks therefore arguably also covers information necessary to be able to use data access regimes. This information can for example include API documentation that clarifies what data researchers can request through an API, how they can formulate queries, in what format data is returned, and whether any rate limits apply. Without such documentation, researchers are forced to learn about the API through trial and error and are unable to analyze how the API has evolved and what new capabilities it has (or which it has lost). It is therefore key that article 40 at least requires platforms to provide the documentation on the API's they provide under this article that allow researchers to fully understand how they can use the API, and what data they can request and in what format. Standardization can play an important role in this context. It allows researchers to more easily assess the capabilities of the API's offered by different platforms, allowing them to assess what API is most suitable for their research, and enabling them to critique the different capabilities of the API's offered by platforms.

A lack of transparency can also prevent researchers from knowing what data they can request from platforms. After all, it is often not clear to outsiders how platforms' internal systems are designed, what data they generate and store, and as such what data researchers may be able to request to scrutinize platforms. Without insight into the data that is available, researchers must formulate research questions about a system into which they have no insight. To be able to use article 40 to understand (platforms' efforts to mitigate) systemic risks, researchers may therefore first submit relatively broad data access requests that allow them to obtain an overview of the available data that is relevant to their research into (platforms' efforts to mitigate) systemic risks, and progressively formulate more specific research questions to determine the existence of and gain access to the specific data they require for their research. It should be noted that having multiple researchers formulate broad research requests designed to find out about the existence of certain data creates significant redundancy, especially if each request must be vetted by the DSC where the platform is established. The EDMO code of conduct in that context referred to the possibility that an intermediary body informs researchers about the availability of data. In the context of article 40, an intermediary vetting body could for example

<sup>118</sup> van der Vlist and others (n 7) 4.

provide researchers information on the data access requests that have already been granted and what data was provided in response.<sup>119</sup>

Finally, article 40 could also be used to request information necessary to interpret the data that is provided, where such data is necessary to understand (platforms' efforts to mitigate) systemic risks. Concretely, this could for example cover information necessary to assess the representativeness of a sample of data platforms provide in response to a request for data under article 40 DSA, or information on when and how the data was collected. This logic also applies to ad libraries, which similarly play an important role in identifying systemic risks.<sup>120</sup> Data about their functioning (such as how platforms determine what ads are included, how many and which ads have been found to have been omitted or removed/updated from/in the ad library, and how the functionalities of ad libraries have been updated) is particularly important to understanding systemic risks on platforms, as it allows researchers to better evaluate studies into systemic risks that rely on data from ad libraries. In short, one important function of article 40 DSA is that it can provide access to metadata about the functioning of data access regimes where such data contributes to the understanding of systemic risks on platforms.

### 5.6. Beyond platform operated data access regimes

Finally, it is worth briefly addressing an alternative approach to data access put forward in the context of the RPA. In that context, the Parliament proposed to require the Commission to set up an ad library to which very large online platforms and as well as publishers such as smaller platforms would be required to transmit the political ads that have appeared on their service.<sup>121</sup> The ads in this ad library would be subject to a "common data structure, common standards and a common application programming interface for the exchange of information with publishers", to be concretised by the Commission in a delegated act.<sup>122</sup> The Council did not follow the EP's proposal. Instead, it suggested that other providers of political advertising services should be encouraged to set up political ad libraries<sup>123</sup> and that the need for ad libraries for all political ads should be re-evaluated within two years of each election for the European Parliament.<sup>124</sup>

Having the Commission operate an ad library circumvents many of the software development issues identified in this article. Rather than public authorities relaying the needs of researchers to platforms through standard-setting and enforcement, the Commission could engage with researchers directly to ensure its ad library meets the needs of different researchers and is iteratively updated based on changes in researchers'

<sup>119</sup> EDMO (n 105) 13. The possibility of a centralized data base of data access requests and the data granted in response to them to provide greater insight into the data that is available was also floated during a workshop organized by the Institute for Information Law on article 40 on 15 March 2023.

<sup>120</sup> See for example DSA article 35(2)(a).

<sup>121</sup> The

<sup>122</sup> Amendments adopted by the European Parliament on 2 February 2023 on the proposal for a regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (n 44) article 7b(1), 7b (6).

<sup>123</sup> General approach on the Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (n 44) recital 42d.

<sup>124</sup> General approach on the Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising (n 44) recital 67.

needs or experience with the way the ad library is used in practice. This allows for faster iterations based on users' experiences and behaviour that are much closer to normal software development practices. The Commission's political ad library could also have spill-over effects for the more general commercial ad libraries platforms operate under article 39 DSA. To ensure consistency in an ad library that includes advertisements from all platforms, it is important (and required under the Parliament's proposal) that platforms supply data in a common format. This requires platforms to adapt their advertising systems to ensure their political advertisements follow this common data structure. Given that political ads are typically disseminated through platforms' general advertising systems, it may be more efficient for platforms to use the same common data structure for both political and non-political ads. In this case, the requirements imposed for the format of political advertisements create an incentive for a harmonized format of the advertisements in platforms' general ad libraries.

Finally, it should be noted that having the Commission assume direct control over an ad library comes with its own complications. It puts the Commission into the role of a software developer, and requires it build up the necessary expertise and connections with the research community. It also offers the Commission more direct control over the way political advertisements can be scrutinized. Given that the Commission is not only an executive but also a political body, this power should be subject to a clear governance structure that ensures the design of the ad library is separated from political influence.<sup>125</sup> Neither complication is new, and both the need for increased technical expertise and clear governance structures that insulate executive functions from political influence arguably result from the Commission's more direct role in the enforcement of the DSA as a whole.

## 6. Conclusion

By imposing broad data access obligations, the DSA and RPA shift platforms' misaligned incentives away from the choice of *whether* to make data available, and to the more technical choices involved in *how* to make data available. This article has drawn on insights from software development practices and the goals of the DSA's and RPA's data access provisions to analyse how the enforcement and standard-setting powers available in the DSA and RPA should be used to ensure data access obligations are operationalised in a way that meets researchers' needs.

There is a large gap between principles of effective software development, and the way in which data access design can be controlled through the DSA and RPA. Software development is typically characterised by fast-paced and direct contact from users to developers, enabling a feedback loop that increases the likelihood software products meet users' needs. As a typical example, Facebook users can file bug reports or make direct feature requests for many of Facebook's products, ranging from Newsfeed to Login.<sup>126</sup> However, they cannot do so for Facebook's ad library, nor is such direct contact between platforms and researchers required under the DSA or RPA. Instead, public authorities act as intermediaries that translate researcher's needs to platforms through delegated regulations and enforcement decisions. This makes researchers' ability to fulfil the role the DSA and RPA set out for them dependent on public authorities' continuous effort to translate researchers' data access requirements into more detailed standards and

enforcement decisions.<sup>127</sup> Even if public authorities do so, the speed with which researchers' needs are relayed to platforms is likely measured in months or years rather than the days or weeks that characterize successful software development. Future work could explore alternative approaches, such as regulatory efforts to push platforms to directly collaborate with researchers to ensure effective data access. Another alternative approach already exists elsewhere in the DSA: platforms are required to transfer data about the content they take down to the Commission, which the Commission aggregates in its own database.<sup>128</sup> Though governmental control over access to data concerning the way information is distributed on platforms comes with its own challenges, it also circumvents the inherent tension involved in relying on platforms to create the data access regimes used to scrutinise their services.

For the operationalisation of data access regimes the DSA and RPA require platforms to create, we have drawn five main conclusions. First, the operationalisation of data access regimes should focus on whether platforms grant researchers access in a way that allows them to effectively fulfil the role the DSA and RPA lay out for them: to scrutinise and provide a better understanding of information dissemination on platforms. This requires close attention to the needs of researchers. Second, this is an evolving obligation. The type of data access needed to enable effective scrutiny is not the same now as it will be 5 years after the DSA has been passed. Operationalising data access regimes solely based on what is currently needed to scrutinise platforms is therefore likely to result in data access regimes that fail to adequately enable research into platforms in the long term. Third, the complex patchwork of instruments and authorities with which data access obligations are operationalised potentially limits the extent to which these obligations meaningfully enable research. The Commission has an important role to play to ensure the different instruments and authorities involved in operationalising data access take a consistent approach. Fourth, researchers need (and can request under article 40 DSA) information about the way data is provided to ensure they are able to understand and draw conclusions from this data. And fifth, platforms continue to have an incentive to delay access to data, and there as of yet few hard rules that lay out how fast supervisory authorities must enforce data access obligations.

While enforcement can play an important role in laying out minimum standards, standard-setting mechanisms (and especially delegated regulations) are especially suited to incorporating these conclusions. They allow for the needs of a broad group of researchers to be taken into account and balanced; enable data access regimes to be regularly updated to respond to researchers' changing needs; and pre-empt conflicts between researchers and platforms over data access that would otherwise have to be resolved individually. As such, there is an important need for researchers to take part (and for public authorities to involve researchers) in the standard-setting processes that will determine what data access regimes will look like in practice.

Finally, it is worth emphasizing the linchpin role public authorities now fulfil in researchers' access to data and ability to analyze platforms. The Commission fulfils a particularly important role. It has the exclusive power to enforce the DSA's data access obligations, the power to concretize data access through delegated regulations and guidelines, and the task of promoting additional self-regulation. This combination of powers affords the Commission a central place in determining how platforms can be scrutinized. It is in the best interest of both researchers (who rely on the Commission for data access) as well as the Commission (so as to enhance the effectiveness of data access regimes) to collaborate closely. However, the role of the Commission also raises fundamental questions about its ability to shape the research agenda. Researchers'

<sup>125</sup> This is of course part of a larger discussion on the way the Commission should exercise its increasing power over platforms under the DSA. See e.g., Rachel Griffin and Carl Vander Maelen, 'Codes of Conduct in the Digital Services Act: Exploring the Opportunities and Challenges' <<https://papers.ssrn.com/abstract=4463874>>.

<sup>126</sup> 'Give Us Feedback About a Facebook Feature' (Facebook) <<https://www.facebook.com/help/contact/268228883256323>>.

<sup>127</sup> Ausloos, Leerssen and ten Thije (n 1) 83.

<sup>128</sup> DSA article 24(5).

dependence on public authorities to gain access to data necessary to scrutinize platforms, as well as the rules that should govern public authorities exercise of their power to determine how researchers can scrutinize platforms, must be closely studied in future research.<sup>129</sup>

#### **Declaration of competing interest**

The authors declare that they have no known competing financial

interests or personal relationships that could have appeared to influence the work reported in this paper.

#### **Data availability**

No data was used for the research described in the article.

---

<sup>129</sup> Suzanne Vergnolle, 'Enforcement of the DSA and the DMA - What Did We Learn from the GDPR?', *To Break Up or Regulate Big Tech? Avenues to Constrain Private Power in the DSA/DMA Package* (2021) 107 <<https://hal.archives-ouvertes.fr/hal-03605110>>. See on the need for independent regulators and governance structures Ausloos, Leerssen and ten Thije (n 1) 83.