RESEARCH PAPER

Outside the Black Box

From Algorithmic Transparency to Platform Observability in the Digital Services Act

Paddy Leerssen

University of Amsterdam
p.j.leerssen@uva.nl

ABSTRACT

Algorithmic transparency is high on the agenda for social media regulation. However, recent work in Science and Technology Studies questions whether this endeavor of "opening the black box" is feasible or even meaningful due to the sociotechnical contingency of platform behavior. To address these shortcomings, Bernhard Rieder and Jeanette Hofmann have proposed a move from algorithmic transparency to platform observability: a pragmatic and sociotechnical perspective aimed at securing structural, real-time access to the means of platform knowledge production. This paper applies the concept of observability to recent legislative developments in the EU's new Digital Services Act. Reviewing that legislation's transparency rules demonstrates how familiar algorithmic principles rules are starting to be complemented by innovative new observability policies and how these reflect revised understandings of transparency's possible subjects, functions, and formats. This review also surfaces normative tensions in observability policy. In terms of substance, observability demands access to content but struggles to discern public from private discourses in semi-public social media channels. In terms of function, observability aims to act as a companion to regulation, but tensions arise between a broad concept of knowledge production and a narrow concept of regulatory compliance monitoring. In terms of format, observability's drive for infra-

structural and real-time access entails new API governance tradeoffs between, for example, scope and scalability. Along these lines, observability paves the way for a more constructive debate around platform data access laws and the dead ends of algorithmic transparency.

1   Introduction

Social media governance has taken a regulatory turn. The policies and standards set by dominant platforms have become deeply politicized, increasingly becoming the target of government regulation. Whereas earlier social media regulation focused on the comparatively modest task of combating unlawful content, new measures, such as the EU's Digital Services Act (DSA), reflect far more comprehensive attempts to govern how social media platforms moderate and curate content and align those platforms with public interest principles. In these reforms, the principle of transparency occupies an awkward position: Platform regulation is rife with algorithmic disclosure and explanation duties, while at the same time the very principle of transparency is undergoing a critical reassessment.

The centrality of transparency to platform regulation may seem unsurprising given that it has been a core principle of governance theory for decades. For platforms especially, information asymmetry represents the very foundation of economic and societal power. Their hypercomplex algorithmic decision-making processes are particularly poorly understood. Nonetheless, over this same period, transparency policy has been losing its luster. Since transparency's turn-of-the-millennium heyday, a growing body of critical research and a slew of failed regulatory experiments have highlighted transparency's many limitations and failure modes; rarely the self-executing policy panacea that was expected and often a distraction from more robust behavioral regulation.[1] In parallel, a growing body of work in critical algorithm studies has questioned whether the algorithmic transparency ideal of "opening the black box" is at all meaningful or feasible.[2]

---

[1]   David Pozen, 'Transparency's Ideological Drift' (2018) 126 Yale Law Journal 100. Gregory Michener, 'Gauging the Impact of Transparency Policies' (2019) 79 Public Administration Review 136. Sandrine Baume and Yannis Papadopoulos, 'Transparency: from Bentham's Inventory of Virtuous Effects to Contemporary Evidence-Based Scepticism' (2018) 21 Critical Review of International Social and Political Philosophy 169.

[2]   e.g., Mike Ananny and Kate Crawford, 'Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973. Nick Seaver, 'Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems' (2017) 4(2) Big Data & Society. Mike Ananny, 'Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness' (2017) 41 Science, Technology, & Human Values 93.

Synthesizing these critiques, Bernhard Rieder and Jeanette Hofmann have proposed "observability" as a pragmatic alternative to algorithm-centric models of platform transparency.[3] Their account seeks to recast platforms' "algorithmic" decision making not as a mechanistic product of monolithic algorithms but rather as the contingent outcome of complex and distributed sociotechnical systems. For Rieder and Hofmann, observability aims to de-center explanations of algorithms in favor of real-time, automated access to data on platform behavior and outcomes. Theirs is a pragmatic program that adopts as a starting point the analytical capacities of platforms and data needs of knowledge production institutions instead of aiming to pre-determine which questions are sufficiently meaningful to require answers.

This paper analyzes recent developments in platform regulation, specifically the DSA, from the perspective of observability. As a recent landmark reform, the DSA exemplifies EU policymaking around platform regulation with a strong emphasis on data access and algorithmic accountability. This paper asks how the DSA regulates *for* and *with* observability and how this differs from established approaches to algorithmic transparency.

To answer these questions, Section 2 begins with an overview of algorithmic transparency and its criticisms, followed by a detailed discussion of Rieder and Hofmann's concept of observability as a critical alternative and the paradigm shift it entails in terms of transparency's contents, its aims and audiences, and its formats. Section 3 reviews the DSA's disclosure rules and considers how they regulate *for* observability in terms of three key algorithmic systems: content curation, content moderation, and ad targeting. In each case, the DSA's conventional algorithmic explanation rules are compared to its more innovative observability solutions. Finally, Section 4 revisits observability's principles in light of the DSA's new policies, highlighting the normative tensions and trade-offs faced by the legal project of observability regulation.

2   From Algorithmic Transparency to Platform Observability

This section reviews recent debates on algorithmic transparency, with particular attention to critical algorithmic studies and how these lay the groundwork for Rieder and Hofmann's move toward observability. The observability program is then discussed in detail.

---

3   Bernhard Rieder and Jeanette Hofmann, 'Towards Platform Observability' (2020) 9(4) Internet Policy Review https://doi.org/10.14763/2020.4.1535 accessed 19 September 2022.

2.1 Algorithmic Transparency and its Critics

Transparency has been described as a "quasi-religious principle" of modern governance.[4] By providing external access to internal information, transparency promises to make organizations more accountable and efficient.[5] Although this principle of transparency can be traced as far back as the progressive era or even the enlightenment (then referred to as "publicity"), its popularity reached new heights around the turn of the millennium.[6] Since the turn in the 1980s and 1990s to global governance, legal debates around transparency have expanded from their conventional focus on governments to also address private corporations.[7] Subsequently, however, transparency has undergone a reappraisal as a new field of critical transparency studies has started problematize transparency policy. In practice, transparency's products are often incomplete or unreliable, its audiences inattentive or powerless, its effects negligible or biased, its costs excessive.[8]

For its harshest critics, transparency policy is not merely ineffective at realizing its goals but actively counterproductive in that it distracts from more meaningful substantive reforms.[9] On the often unrealistic assumption that "more information will lead to better behavior," transparency policy and has been invoked as a pretext for avoiding binding duties.[10] In this way, transparency policy has come to be associated with a neoliberal faith in market competition and individual choice as superior ordering mechanisms to top-down legal command.[11] For David Pozen, a more mature engagement with transparency should "desacralize" this article of faith and approach it not as a miracle cure but instead as a support or catalyst for binding, substantive, or behavioral

---

[4] David Heald and Christopher Hood (eds.). *Transparency: The Key to Better Governance?* (Oxford University Press for the British Academy 2006).

[5] Mikkel Flyverbom, *The Digital Prism: Transparency and Managed Visibilities in a datafied world* (Cambridge University Press 2019).

[6] Pozen, 'Transparency's Ideological Drift' (n 1). Robert Gorwa and Timothy Garton Ash, 'Democratic Transparency in the Platform Society,' in Nate Persily and Joshua Tucker (eds.), *Social Media and Democracy: The State of the Field and Prospects for Reform* (Cambridge University Press 2020). However, cf., Emmanuel Alloa, 'Why Transparency Has Little (If Anything) To Do With The Age Of Enlightenment,' in Emmanuel Alloa (ed.), *This Obscure Thing Called Transparency: Politics and Aesthetics of a Contemporary Metaphor* (Leuven University Press 2022).

[7] Heald and Hood, *Transparency* (n 4).

[8] e.g., Pozen, 'Transparency's Ideological Drift' (n 1). Michener, 'Gauging the Impact of Transparency Policies' (n 1). Baume and Panadopoulos, 'Transparency: from Bentham's Inventory of Virtuous Effects to Contemporary Evidence-Based Scepticism' (n 1).

[9] ibid. See also: Catharina Lindstedt and Daniel Naurin, 'Transparency is Not Enough: Making Transparency Effective in Reducing Corruption' (2010) 31 *International Political Science Review* 301.

[10] Flyverbom, *The Digital Prism* (n 5).

[11] Pozen, 'Transparency's Ideological Drift' (n 1). Monika Zalnieriute, '"Transparency-Washing" in the Digital Age: A Corporate Agenda of Procedural Fetishism' (2021) *Critical Analysis of Law* 8:1, 39–53.

regulation.[12] Thus, the overall tenor in transparency discourses has shifted from quasi-religious fervor to "evidence-based scepticism."[13]

During the same period, however, transparency reforms have become central to platform regulation. As such this is not unique, given the ubiquity of transparency as a regulatory tool. However, the domain of platform regulation does raise specific concerns that make transparency especially salient. First, platforms are essentially information services with business models and regulatory powers that hinge on the appropriation and exploitation of large datasets.[14] Profound information asymmetries are a structural feature of platform markets, prompting many regulatory proposals to aim to adjust this imbalance and open up access to outsiders.[15] As such, transparency is often considered an indispensable precondition for the exercise of democratic control over these services.[16] Second, platform governance occupies the forefront of new machine-learning technologies in areas including content moderation and curation, which are uniquely complex and resistant to human explanation and comprehension.[17] This renders opacity a characteristic feature of platforms' algorithmic governance methods, leading to many debates about platform regulation converging on the project of "opening the black box" and subjecting these systems to outside scrutiny.[18] Given that these algorithms generally surpass human cognition in their complexity, exhaustive explanations are not feasible. Instead, the challenge is to produce accounts that highlight the most salient factors; salience, of course, is in the eye of the beholder.[19] All this leads to uneasy ambivalence around transparency in platform governance; an ideal past its prime, but in some sense more relevant than ever.

---

[12] Pozen, 'Transparency's Ideological Drift' (n 1).

[13] Baume and Panadopoulos, 'Transparency: from Bentham's Inventory of Virtuous Effects to Contemporary Evidence-Based Scepticism' (n 1).

[14] Julie Cohen, *Between Truth and Power: The Legal Constructions of Informational Capitalism* (Oxford University Press 2019).

[15] Gorwa and Garton Ash, 'Democratic Transparency in the Platform Society' (n 6).

[16] Gorwa and Garton Ash, 'Democratic Transparency in the Platform Society' (n 6).

[17] Gorwa, Binns and Katzenbach, 'Algorithmic content moderation' (n 10).

[18] Pasquale, *The Black Box Society* (n 32). Yeung, 'Algorithmic regulation' (n 32

[19] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2017) 3(1) Big Data & Society https://doi.org/10.1177/2053951715622512 accessed 15 September 2022. Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 2. Lilian Edwards and Michael Veale, 'Slave to the algorithm? Why a "Right to an Explanation" is probably not the remedy you are looking for' (2017) 16 Duke Law & Technology Review 18.

At the intersection of these trends stands a growing field of critical algorithm studies that, over the past decade, has problematized the precepts of algorithmic governance in general and algorithmic transparency in particular.[20] Scholars in this field have sought to resist the preoccupation with algorithms as sites of power and objects of study, and, with it, the dominant frame of "opening the black box." Building on Science and Technology Studies (STS), platforms' automated decision-making are instead viewed as sociotechnical process, with meanings and outcomes shaped not only by features inherent to the algorithmic artifact but also in large part by the actions of users and other stakeholders who interact with these systems.[21] In practice, automated decision-making systems often comprise many different subsystems, making 'the algorithm' an object of scrutiny without clear boundaries and subject to interpretation. Furthermore, the behavior of these systems is emergent and constantly in flux; user interactions provide new inputs and learning resources to the system, and these outputs, in turn, influence users in their composition, their tastes, and their habits.[22] Platform operators also intervene and steer the system in various ways, although they too typically fail to fully predict or control the complex sociotechnical processes that they unleash.[23] For these reasons, a sociotechnical perspective demands that inquiries into "algorithms" consider the specific social context of their usage.

2.2 Transparency and Observability as Competing Metaphors

Building on critical transparency studies and critical algorithm studies, Rieder and Hofmann have attempted to re-articulate the program of platform disclosure regulation in a way that surpasses the limitations of the algorithmic transparency paradigm. Under the banner of "observability" regulation, they make several recommendations for disclosure regulation that respond to the particular epistemic challenges posed by platforms and their automated decision-making systems. In the following sections, I will first discuss semantic and conceptual differences between transparency and observability before turning to regulatory implications.

---

[20] e.g., Ananny and Crawford, 'Seeing Without Knowing' (n 2). Ananny, 'Towards an Ethics of Algorithms' (n 2). Seaver, 'Algorithms as Culture' (n 2). Gillespie, 'The Relevance of Algorithms' (n 18).

[21] Rieder and Hofmann, 'Toward Platform Observability' (n 3).

[22] e.g., Bernhard Rieder, Ariadna Matamoros-Fernández and Òscar Coromina. "From Ranking Algorithms to 'Ranking Cultures' Investigating the Modulation of Visibility in YouTube Search Results" (2018) Convergence 24(1).

[23] Rieder and Hofmann, 'Toward Platform Observability' (n 3) 8.

Transparency and observability are both metaphors. Both imply a capacity for sight or observation, but they do so in subtly different ways. Rieder and Hofmann do not discuss these metaphorical considerations at length, and below I will elaborate how their *pragmatic* view preferences observability over transparency. Building on their insights I will also discuss how observability implies a different *directionality* than transparency, which aligns it with a socio-technical perspective.

The central distinction for Rieder and Hofmann is that observability is more pragmatic than transparency, in that it highlights the act of viewing. Whereas transparency refers to a physical property ascribed to materials (i.e., diaphaneity, the capacity for letting through light), observability contains the potential for an act or practice of viewing performed by observers. Accordingly, transparency's material metaphor may imply a "view from nowhere" that describes objective and measurable facts about the world, whereas observability draws attention to the viewer(s) and their perspective(s). As such, observability highlights subjectivity and the communicative process of disclosure and interpretation, which, for Rieder and Hofmann, intends to "draw attention to and problematize the process dimension inherent to transparency as a regulatory tool." If transparency is passive, static, and holds pretenses of objectivity, then observability is active, pragmatic, and a departure from subjectivity.

As an aside, and at the risk of overinterpreting the metaphor, I will diverge from Rieder and Hofmann at the claim that observability foregrounds the *mediated* nature of disclosure practices. It strikes me that their chosen metaphor of "observability" does not necessarily work in their favor here. After all, observability draws on the same language of sight, the same coupling of seeing and knowing, as transparency.[24] If anything, transparency suggests the *more* mediated perspective of the two metaphors; after all, it describes a view that is literally mediated by a diaphanous material or medium. Hence, the language of transparency *can* accommodate critical interrogations of mediation, as, for instance, in Flyverbom's "digital prism" of distorted and refracted light or Emmanuel Alloa's biblical invocation of "seeing as through a glass, darkly."[25] Observability, by contrast, suggests no mediation at all – at least in its semantics.

I will also add that the observability metaphor has important spatial implications. Although Rieder and Hofmann do not address this point, it does align powerfully with their sociotechnical perspective. Namely, transparency and observability suggest different objects or directionalities. The transparency metaphor implies an act of seeing *through* materials or boundaries and, hence,

---

[24] Ananny and Crawford (n 3).

[25] Alloa (ed.), *This Obscure Thing Called Transparency* (n 6). Flyverbom, *The Digital Prism* (n 5). Ananny and Crawford, 'Seeing Without Knowing' (n 3). For Ida Koivisto, the transparency metaphor is not merely iconoclastic but iconic-ambivalent; it necessarily implies a mediation, even though this mediation aims to render its target as clear as possible and hence to escape notice. Ida Koivisto, *The Transparency Paradox: Questioning an Ideal* (Oxford University Press 2022).

inside (of an organization or system under scrutiny).[26] It is centripetal. This intrinsically connects the metaphor of "opening the black box" to the metaphor of transparency: Both seek to look inside and make externally visible that which is internal. Observability, by contrast, lacks this centripetal directionality. It does not imply seeing *through* but merely seeing, making it far more capacious. Whereas transparency always peers inside, observability can also look at, on, under, around, or across its object.[27] Applied to algorithmic systems, observability suggests that we expand our view from the algorithm as such to encompass other aspects, such as inputs, outputs, outcomes, and interventions, that is, the social context of usage of or, to extend the metaphor, wwthe people around the black box.[28] If transparency keeps our nose pressed up against the glass, then observability permits us to take in our surroundings.

Table 1: Semantics of transparency and observability

	Transparency	Observability
Denotes:	› Material property (diaphaneity) › Capacity to see *through* and inside	› Practice (observing) › Capacity to see
Connotes:	› Capacity to inspect › Objective, passive, static › Accounting of (internal) reasons › Algorithms (technical perspective)	› Capacity to regard, locate › Subjective, active, pragmatic › Awareness of outcomes, effects, context › Assemblages (sociotechnical perspective)

Ultimately, observability's pragmatism and its decentered directionality respond to two major critiques of algorithmic transparency reforms. Its pragmatism dispels the naïveté of transparency as a source of objective (empirical) truth, calling attention to the selectivity, subjective reception, and contingent effects of disclosure. Its decentered directionality averts the focus on algorithms as objects of study and encourages a more holistic and contextual assessment of algorithmic decision-making as embedded in a social context of usage. I now turn to the practical, policy-oriented implications of this reformulation.

---

26 David Heald wrote a detailed commentary on the directionality of transparency. He uses transparency more capaciously, with an inward but also an outward meaning. This outward variant is rarely invoked in the contexts of algorithmic and platform governance, and I, therefore, restrict myself to the more salient inward variant. See David Heald, 'Varieties of Transparency,' in David Heald, Christopher Hood and David Heald (eds.), *Transparency: The Key to Better Governance?* (Oxford University Press for the British Academy 2006).

27 Ananny, 'Toward an Ethics of Algorithms' (n 2).

28 See also the call for social transparency, "not just about opening the closed box of AI, but also about who is around the box..." Upol Ehsan and others, 'Expanding Explainability: Towards Social Transparency in AI Systems' (2021) Proceedings of the 2021 *CHI Conference on Human Factors in Computing Systems*, 1–19.

2.3 Observability as a Regulatory Program

To achieve observability, Rieder and Hofmann outline three main principles. The first is to "expand the normative horizon" regarding what transparency can achieve.[29] Platformization represents a fundamental reshaping of many societal domains, with its information asymmetries challenging not only regulatory enforcement, fair dealing, and user choice but also entailing a more profound shift in society's capacity for knowledge production. Even platforms themselves do not possess total knowledge regarding their service's functioning, but they do control the infrastructure and possess the data necessary to study them. In this way, platformization "deprives society of a crucial resource for producing knowledge about itself."[30] More concretely, this suggests an approach to transparency that goes beyond regulatory auditing or individual disclosures and also aims to empower researchers and civil society actors to gain access to platform resources. Meanwhile, in terms of substance, it re-imagines transparency regulation not as merely divulging knowledge, but as forcing access to the means of knowledge production.

The second principle is to "observe platform behavior over time."[31] Given the volatility of platform ecosystems – their constant adaptation to changing social contexts – disclosures need to move beyond the conventional "snapshot logic" of periodical auditing or reporting.[32] Instead, society requires structures to study platforms over time, and ideally in real-time. Rieder and Hofmann list four access methods that reflect this approach: (1) Data access agreements, where specific researchers are granted access to select datasets, typically under conditions of confidentiality; (2) Accountability Interfaces, which provide automatic transparency functions to a wider (public) audience; (3) Developer APIs, which are similar to accountability interfaces but are designed primarily with commercial usage in mind (rather than accountability per se); (4) data scraping, where researchers collect platform data via end-user interfaces and independently of the platform.

The third principle is to "strengthen foundations for collaborative knowledge creation." Part of observability's pragmatic approach is an attentiveness to the different information needs of specific actors in making sense of platform data. Transparency rules that prefigure the relevant facts or norms under scrutiny will likely fail to accommodate these different perspectives, needs, and interests. Access to the underlying platform data through interfaces such as those just mentioned, can help "different actors to develop their own observation capacities, adapting their analytical methods to the questions they want to

---

[29] Rieder and Hofmann, 'Toward Platform Observability' (n 3) 10.

[30] ibid 11

[31] ibid 7

[32] ibid 13

ask."[33] Still, Rieder and Hofmann recognize that meeting third party needs is a key challenge for observability policy that "raises the complicated question of how data and analytical capacities should be made available, to whom, and for what purpose."[34] Beyond mere access to data, effective observability also calls for institutional capacity-building and adequate funding of researchers and other societal watchdogs.

Rieder and Hofmann present observability as tightly coupled to regulation. First, regulation is necessary to achieve observability policies, with platforms unlikely to carry them out effectively of their own accord, at least not without the threat of government regulation to motivate them. Second, observability should be conceived of as a companion to regulation. Here, Rieder and Hofmann are responding to criticisms that transparency often functions as a deregulatory wedge against substantive rulemaking. Far from an alternative to regulation, they emphasize that the ultimate goal of observability should be "to assess platform behavior against public interest norms" and to "undergird the regulatory response to the challenges platforms pose."[35] We will see below that the DSA explicitly attaches data access to regulatory enforcement and evidence monitoring.

Summarizing this discussion, Table 2 contrasts key principles of algorithmic transparency and platform observability as regulatory programs, in terms of their distinct disclosure contents, audiences, and formats. The following section then examines how these approaches are realized by the rules of the DSA.

Table 2: Algorithmic transparency and platform observability as regulatory programs

	Transparency	Observability
Contents	› Algorithms	› Sociotechnical systems
	› Parameters, specifications	› Usage, interaction
	› Reasons, explanations, logics ('why?')	› Outcomes, decisions ('what?')
	› Facts / knowledge	› Knowledge production
Audiences	› Users	› Regulators
	› Market actors (consumers, complementors, competitors)	› Academics, journalists
		› Citizens, publics
Formats	› Reports, statements, disclosures	› Infrastructures, databases, APIs
	› Periodical	› Real-time

33 ibid 20

34 ibid 21

35 ibid 23

3   The DSA's Disclosure Rules

This chapter examines the DSA through the lens of observability. After briefly introducing the DSA's overall design and objectives, it reviews its key disclosure rules relating to three important forms of algorithmic decision-making by platforms: content curation, content moderation, and ad targeting. It then discusses how and where the DSA reflects observability principles and how they relate to its more conventional algorithmic explanation rules.

3.1 The DSA's Key Features and Objectives

The DSA is a regulation of the European Union that became fully applicable on February 17, 2024. Alongside the Digital Markets Act (DMA), it is considered a landmark platform regulation reform. Whereas the DMA focuses on online competition and market governance, the DSA is primarily concerned with the governance of user-generated content and associated harms. Its stated objective is to establish "rules for a safe, predictable and trusted online environment that facilitates innovation and in which fundamental rights […] are effectively protected."[36] Text and recitals both refer repeatedly to transparency and accountability as core principles.

The DSA's main ruleset applies to "hosting providers": services that involve the storage of user-generated content (e.g., Dropbox, Gmail) and, more specifically, to "online platforms" that make this content publicly available (e.g., YouTube, TikTok).[37] Its most stringent ruleset applies to "Very Large Online Platforms," which have more than 45 million monthly average active users.[38]

The DSA introduces many new rules for online platforms, but its key elements are as follows. First, it restates the EU's basic rules around intermediary liability for user content (of limited relevance for our purposes). Second, the DSA introduces a set of individual rights and procedural safeguards for content moderation decisions. Third, the DSA contains a set of risk management obligations for large platforms, which are wide-ranging but include obligations to periodically assess and mitigate "systemic risks" related to the use of their service, such as threats to fundamental rights, public health, and civic discourse.[39] As we will see below, both the due process and risk management aspects of the DSA are rife with transparency rules.

---

[36]   DSA, Article 1(1).

[37]   DSA, Article 3.

[38]   DSA, Article 3(i).

[39]   DSA, Article 34 & 35.

In terms of algorithm governance, the DSA targets at least three types of algorithmic decision-making: content moderation, content recommendation, and ad targeting.[40] As used in the DSA, content moderation refers to activities aimed at detecting, identifying, or addressing illegal content and content violating the platform's Terms of Service.[41] The DSA's concept of content moderation is broad and includes not only content removal or account suspension but also measures such as demonetization, fact-checking, and, as mentioned, content demotion or "shadow banning."[42] Content recommendation refers to fully or partially automated systems used to "suggest in its online interface specific information to recipients of the service or prioritize that information."[43] Ad targeting is not defined in the DSA, but for our purposes describes techniques used to "determine the recipient to whom the advertisement is presented."[44]

We know from critical algorithm studies that these processes are not necessarily governed by discrete, identifiable algorithms – each can entail many different sets of "algorithms" or subsystems and these respective processes are also interrelated and integrated in various ways (e.g., recommendation "algorithms" can also incorporate moderation decisions and ad targeting considerations[45]) When one discusses algorithmic systems in these terms, as the DSA does, it is referring primarily to governance functions or decisions rather than discrete technical artifacts.

Other than transparency rules, the key obligations for moderation, recommendation, and targeting are as follows. First, the systemic risk management duties for large platforms impose cross-cutting responsibilities for all these algorithmic systems. These rules are open-ended by design – capturing, for example, illegal content, fundamental rights, civic discourse, public health, public security, physical health and mental well-being – and their precise requirements are difficult to predict in advance.[46] This system is also undergirded by a third-party auditing process that aims to guarantee that the assessments conducted by platforms are complete and accurate.[47]

---

[40] Generally, on these concepts and their role in platform governance, see Thomas Poell, David Nieborg and Brooke Erin Duffy, *Platforms and Cultural Production* (John Wiley & Sons 2021).

[41] DSA, Article 3(t)

[42] Paddy Leerssen, 'An End to Shadow Banning' (2023) *Computer Law and Security Review* 43, 105790.

[43] DSA, Article 3(s).

[44] DSA, Article 26(1)(d).

[45] Section 2.1 *supra*.

[46] DSA, Article 34(1).

[47] DSA, Article 37.

Besides general risk management, specific obligations also apply for recommendation and ad targeting, although they are relatively modest. Recommender systems must disclose their key parameters and, in the case of large platforms, offer "at least one option for each of their recommender systems which is not based on profiling."[48] Ad targeting systems must likewise explain their key parameters, and there are limited prohibitions on the use of certain sensitive data for targeting purposes (building on existing restrictions in the GDPR).[49]

The most detailed suite of rules applies to content moderation. For instance, the DSA requires platform content rules to be clearly defined and enforced consistently and with due regard to fundamental rights; third parties should be able to submit takedown requests for unlawful continent; individual decisions must be notified to the affected parties; and there are mechanisms for internal appeal and external dispute resolution.[50] Overall, this amounts to a broadly procedural approach, that sees the DSA avoids prescribing new categories of content or harm that must be actioned but instead establish the conditions under which platforms are permitted to formulate and enforce their own content rules.[51] Across each of these rulesets, as we shall see, transparency is an integral element.

3.2 The DSA's Disclosure Rules

What kinds of transparency does the DSA demand about the algorithmic systems it governs? In the following, I review key elements, starting with the general data access framework before focusing on more specific rules for moderation, curation, and targeting.

3.2.1 General Data Access Framework (Article 40)

One of the DSA's most ambitious steps is to mandate general data access rights vis-à-vis very large platforms, not online for regulatory authorities but also for researchers. Established in Article 40 DSA, this is not constrained to any specific topic, making it applicable to all algorithmic systems operated by platforms. Importantly, this provision delineates distinct access rights for regulators, vetted researchers, and researchers.

---

[48] DSA, Articles 27, 38.

[49] DSA, Article 26.

[50] DSA, Articles 14, 16, 17, 20, 21.

[51] Martin Husovec and Irene Roche Laguna, 'Digital Services Act: A Short Primer,' in Husovec and Roche Laguna, Principles of the Digital Services Act (Oxford University Press forthcoming 2023) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4153796 accessed 25 September 2022.

First, regulatory access under Article 40(1) of the DSA entitles the DSA's competent authorities to request "access to data that are necessary to monitor and assess compliance with this regulation." This must be the sole purpose of the request, and it must take "due account of the rights and interests" of the service provider.[52] This provision can be seen as an extension of conventional regulatory investigative powers and is not given special attention in this paper.

Second, regarding vetted researchers access (40(4)–40(12)), upon request from researchers, large platforms can also be ordered by DSCs to provide access for independent research purposes. This right is likewise broad in the sense that it applies in principle to any "data held by platforms."[53] However, some important limitations should be noted:

\ Researcher vetting: Researchers must be vetted by the DSC based on numerous requirements, including affiliation with a research organization, independence from commercial interests, and capacity to comply with data protection law.[54]

\ Purpose limitation: The data must be used solely for the purpose of research that contributes to the understanding of "systemic risks" as per the DSA. This is a broad and ambiguous category that represents a substantial departure from the typical starting point of academic freedom and independent inquiry.[55]

\ Limitations to data protection compliance associated with the protection of trade secrecy and service security: To protect these interests, researcher access requests must contain "appropriate safeguards." Legal debates on the relationship between data protection laws and researchers are relatively advanced, but there is comparatively little guidance or precedent for trade secrecy or security or how these interests can be balanced against researcher access.[56]

Regarding format, it is notable that, for both regulators and researchers, Article 40 demands that platforms "facilitate and provide access to data pursuant […] through appropriate interfaces specified in the request, including online databases or application programming interfaces."[57]

---

[52] Article 40(2).

[53] DSA, Article 40(4).

[54] DSA, Article 40(4) and 40(7).

[55] DSA, Article 40(4) and 40(7).

[56] DSA, Article 40(5) and 40(7). On data protection and researcher access outside of the DSA context, see European Digital Media Observatory (2022), Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access. https://edmoprod.wpengine.com/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf

[57] DSA, 40(7).

Third, there is a separate right for researcher access to publicly available data (40(12)) that is broader in the sense that it is not limited to researchers affiliated with a research organization (opening the door to, for example, journalists and NGOs) and that researchers can invoke this right directly against platforms without having to petition any regulatory authority. It is narrower in the sense that it is limited to publicly available data, which (per the DSA) may include data "for example on aggregated interactions with content from public pages, public groups, or public figures, including impression and engagement data such as the number of reactions, shares, comments from recipients of the service."[58]

Many questions remain about the precise process for access under Article 40(12), chiefly whether it entails a duty for platforms to maintain dedicated public access APIs or other access infrastructures ("the Crowdtangle provision" as some have described it) and/or whether it might also be implemented by permitting researchers to independently scrape platform interfaces.[59]

3.2.2 Recommender System Transparency (Article 27)

Article 27 DSA is dedicated to recommender system transparency. Its main principle is that platforms must "set out in their terms and conditions, in plain and intelligible language, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters." At a minimum, this must include "the criteria which are most significant in determining the information suggested" and "the reasons for the relative importance of those parameters."[60]

Another aspect of recommender system transparency is the question of content demotion, also known as "shadow banning." Demotion is not addressed expressly in Article 27 DSA, but it is considered a form of content moderation to which due process safeguards apply. This is discussed further below. Here, we see how nominally distinct "algorithms" such as content recommenders comprise many different subsystems fulfilling different governance functions.

---

[58] DSA Recital 97.

[59] Paddy Leerssen, 'Call for Evidence on the Delegated Regulation on Data Access Provided for in the Digital Services Act Summary & Analysis'. European Commission 2023. Available at: https://digital-strategy.ec.europa.eu/en/library/digital-services-act-summary-report-call-evidence-delegated-regulation-data-access

[60] DSA, Article 27.

3.2.3 Ad Targeting Transparency (Article 27, 39)

The rules for advertising transparency are more far-reaching. Like recommender systems, there must be "meaningful information […] about the main parameters used to determine the recipient to whom the advertisement is presented."[61] However, unlike recommender system transparency, this must be offered not only as a general policy statement in the Terms and Conditions, but "for each specific advertisement presented to each individual recipient […] in real time" and "directly and easily accessible from the advertisement."[62] In addition, advertising transparency also includes more basic requirements related to the advertisement itself – rather than its targeting mechanisms – such as disclosing "that the information is an advertisement" and the "natural or legal person on whose behalf the advertisement is presented."[63]

The most significant difference with advertising, however, is that the DSA also contains a dedicated framework for the creation of a public repository or ad archive. This rule is established in Article 39 of the DSA and applies only to large platforms. It requires them to create public repositories documenting, for each advertisement, (inter alia) the content of the ad, the persons on whose behalf it is presented and who paid for it, demographic statistics on the audience reached, and, again, "the main parameters" used to target this information. This information must be made available "through their online interface, through a searchable and reliable tool that allows multicriteria queries and through application programming interfaces."[64]

The DSA's ad archiving regulations are not entirely new, building as they do upon precedents in self-regulation and national rulemaking.[65] However, the DSA's approach is notably more broad in the sense that it applies to *all* ads sold by the service. Previous versions were typically limited political ads and the publication of targeting criteria for each ad represents a novel requirement.

---

[61]   DSA, Article 26(1).

[62]   DSA, Article 26(1).

[63]   DSA, Article 26(1).

[64]   DSA, Article 39(1).

[65]   Paddy Leerssen and others, 'Platform ad Archives: Promises and Pitfalls' (2019) 8(4) *Internet Policy Review* https://doi.org/10.14763/2019.4.1421 accessed 5 November 2022.

3.2.4 Content Moderation Transparency

The most extensive transparency framework applies to content moderation actions. Key elements of the due process framework are, in fact, disclosure rules. The content policies that platforms enforce must be codified in specific and clear language (Article 14 DSA), and individual decisions must also be notified and explained to the affected users in a "Statement of Reasons" (Art 17 DSA). These rules contain, but are not limited to, explanations of the algorithmic logic involved. Terms and Conditions must contain "information on any restrictions that they impose" in relation to user-generated content, which "shall include information on any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review, as well as the rules of procedure of their internal complaint handling system."[66] The Statement of Reasons for individual decisions must contain information regarding not only "the use made of automated means in taking the decision" but also, for example, the nature of the sanction imposed, the facts and circumstances relied upon to reach the decision, reference to the contractual or legal ground for action, and opportunities for redress.[67]

Besides these due process principles, content moderation transparency also includes periodical reporting.[68] This entails an extensive ruleset that also builds on a long-standing self-regulatory practice. Platforms are expected to publish regular reports with various forms of qualitative and quantitative information, with greater levels of detail required from larger platforms. The focus is on content moderation metrics such as the number of complaints received, the number of actions taken, and appeals heard, broken down for different categories of content norms (e.g., hate speech, copyright, and nudity.). Various operational details are alsorequired, including information on decision times and staffing. As in the case of ad archives, the DSA's rules here are not entirely new, but they do expand the scope and depth of existing self-regulatory practice.

Finally, a remarkable new disclosure principle from the DSA is its so-called Statement of Reasons Database, a type of content moderation archive.[69] This database operates as follows: when platforms issue a Statement of Reasons for a content moderation decision, they must disclose it not only to the affected parties but also to the European Commission. The European Commission then publishes each of these statements in a public, machine-readable database.[70] The data is redacted to avoid including personal data. For instance, it does not include URLs for the items in question. Instead, it is limited to data points

---

[66] DSA, Article 14(1).

[67] DSA, Article 17.

[68] DSA, Articles 15, 24 and 42.

[69] DSA, Article 24(5).

[70] The database can be viewed here: https://transparency.dsa.ec.europa.eu/statement

representing, for example, the nature of the decision, its legal or contractual grounds, and its territorial scope. At the time of writing (in early 2024), it contains over 4 billion statements. A somewhat comparable precedent in self-regulation is the Lumen database, which hosts takedown requests submitted to various participating platforms.[71]

3.2.5 Other Transparency Rules

The DSA contains many other disclosure rules. For instance, platforms are also required to verify and disclose the identity of business users, to share information with law enforcement upon valid requests, and to publicly designate points of contact for regulatory authorities and for users. Besides platforms, other governance entities – such as regulators, trusted flaggers, and dispute resolution bodies – have their own reporting duties. However, further exploration of these rules is beyond the scope of this paper, which specifically concerns platforms and their algorithmic systems.

4 Discussion: Transparency and Observability in the DSA

Having reviewed the DSA's transparency rules for algorithmic systems, we see remnants of the algorithmic transparency paradigm as well as more innovative observability policies. Table 3 sketches how these rules might be classified in terms of algorithmic transparency and observability. The following section discusses the differences between these approaches in further detail. This discussion is structured around the three distinguishing features of algorithmic transparency and platform observability: their contents, their audiences and aims, and their formats. In each domain, the DSA reflects a shift from algorithmic transparency and platform observability while also surfacing new policy for the observability project.

---

[71] On content moderation archiving, see John Bowers, Elaine Sedenberg and Jonathan Zittrain, 'Platform Accountability Through Digital "Poison Cabinets,"' Knight First Amendment Institute (13 April 2021). https://cyber.harvard.edu/story/2021-04/platform-accountability-through-digital-poison-cabinets accessed 16 September 2022. MacKenzie Common, 'Fear the Reaper: How Content Moderation Rules are Enforced on Social Media' (2020) 34 *International Review of Law, Computers & Technology* 126. David Erdos, 'Disclosure, Exposure and the "Right to be Forgotten" after Google Spain: Interrogating Google Search's Webmaster, End User and Lumen Notification Practices' (2020) 38 *Computer Law & Security Review* 105437. Daphne Keller and Paddy Leerssen, 'Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation,' in Nathaniel Persily and Joshua Tucker (eds.), *Social Media and Democracy: The State of the Field and Prospects for Reform* (Cambridge University Press 2020).

Table 3: The DSA's disclosure rules for key algorithmic systems

	Algorithmic transparency	Platform observability	
Content curation	Explanations for the "main parameters" of recsys algorithms (Art 27)	Public content APIS & scraping (Art 40(12))	Access rights for vetted researchers (Art 40(4))
Ad targeting	Explanations for the "main parameters" of ad targeting decisions	Ad archives (Art 39)	
Content moderation	Explanations for content moderation policies and decisions (Art 14, 17)	Statement of Reasons database (Art 24(5))	
	Aggregate reporting on content moderation (Art 15, 24, 42)		

4.1 From Algorithms to Systems (and the Problem of Content Access)

Regarding *the substance* of disclosures, although some rules focus squarely on algorithms, others take a broader perspective. Notably, the term "algorithm" is rarely used in the DSA's disclosure rules, outside of a prominent reference in Article 14 of the Terms and Conditions provision requiring information about "tools used for the purpose of content moderation, including algorithmic decision-making and human review."[72] However, algorithmic transparency does reveal itself as the focus for recommender systems and ad targeting, where the DSA requests information on the "main parameters" involved in automated decision-making.[73]

The criticisms reviewed in Section 2 suggest that such rules are unlikely to produce particularly meaningful disclosures. These disclosures are intended to serve a general audience, including individual users, but it is all but impossible to do so without simplifying the explanation past any meaningful de-

---

[72] DSA, Article 14(1).

[73] DSA, Articles 26, 27, 39.

scription of the algorithm's logic.[74] Even if more detailed documentation were provided, the algorithm's parameters as such would be of limited meaning for grasping the drivers of platform content trends as they occur in practice. This is illustrated powerfully by X's decision to "open source" its content ranking algorithm, a step that already seems to go further than the "main parameters" requirements of the DSA.[75] Although this disclosure does answer some basic questions about the system's architecture, it mostly confirms that X's methods were largely similar to the known state of the art. Arvind Narayanan commented that "releasing the code isn't that revealing" – and the DSA's explanation rules demand far less than this.[76]

These user-facing explanation rules about the "main parameters" of automated decision-making can be contrasted with more contextual and outcome-oriented solutions elsewhere in the DSA. Ad archives represent perhaps the most obvious example here, providing systemic overviews of *what* outputs are being shown through ad-targeting systems and what inputs have been selected by the ad targeted. Also significant here is Article 40(12)'s transparency provision regarding publicly available data, which emulates a Crowdtangle-type output transparency aimed at observing overall content curation trends and outcomes. To the extent that Article 40(12) permits it, data scraping can fulfill the same function in tracking content curation trends (although each approach has its own limitations).[77] Furthermore, Article 40's vetted researcher access framework can be invoked to request similar data, including non-public information, such as user interaction data and removed content.

In terms of content moderation decisions, the Statements of Reasons database fulfills a comparable role, even if it falls short in one important respect. Like ad archives or content APIs, it provides a systemic overview of individual outputs (i.e., content moderation decisions). Unlike these tools, however, the Statements of Reasons database does not specify the context for its decisions, that is, the particular (for example) content, user profile, and page items affected. For this reason, it is open to the same critiques as conventional content moderation reporting: The database merely reproduces the platform's own findings – indicating whether *the platform* found the information to be, for example, disinformation or hate speech – but provides no basis for third parties to assess whether this decision was justified or to ask other broader contextual

---

[74] Works cited in Section 2.1 supra, e.g., Veale and Edwards, 'Slave to the Algorithm?'; Burrell, 'How the Machine Thinks.'

[75] Brandi Geurkink, 'Twitter's Open Source Algorithm Is a Red Herring' (Wired April 7 2023) https://www.wired.com/story/twitters-open-source-algorithm-is-a-red-herring/ accessed April 7 2024

[76] Arvind Narayanan, 'Twitter Showed Us Its Algorithm. What Does It Tell Us?' (2023) Knight First Amendment Institute https://knightcolumbia.org/blog/twitter-showed-us-its-algorithm-what-does-it-tell-us

[77] e.g., Balazs Bodó and others, 'Tackling the Algorithmic Control Crisis: The Technical, Legal, and Ethical Challenges of Research into Algorithmic Agents' (2018) 19 Yale Journal of Law and Technology 133. Eduardo Hargreaves and others, 'Biases in the Facebook News Feed: a Case Study on the Italian Elections' (2018) International Symposium on Foundations of Open Source Intelligence and Security Informatics, In conjunction with IEEE/ACM ASONAM https://hal.inria.fr/hal-01907069 accessed 19 September 2022.

and qualitative questions about the decision.[78] In this sense, the Statement of Reasons database arguably falls short of true observability for content moderation decisions. A more ambitious approach would have included URLs or other unique identifiers, where possible, to link decisions to specific content items. Of course, if the content were taken down, it would no longer be visible, but it would still illuminate a range of non-removal moderation actions (e.g., demotion, delisting, demonetization, and labeling). Alternatively, researchers can still try to access data about individual moderation decisions via Article 40(4)'s general data access framework.

The Statements of Reasons database reflects a problem of user privacy with regard to content moderation. Content moderation decisions can be stigmatizing and reveal private information about targets, victims, and other parties.[79] This might represent an important reason for a database to not link to specific items of content, at least for certain decisions. Nonetheless, for many categories of moderation decisions, such privacy interests are relatively minimal and arguably outweighed by the public interest in observability. For instance, moderation related to political speech, such as disinformation fact-checking or demonetization of controversial topics, faces a relatively strong *prima facie* case for public archiving. The opposite is true for moderation concerning privacy-based harms – such as non-consensual sexual imagery. Furthermore, decisions affecting ordinary end-users will typically have a stronger privacy interest than decisions affecting public figures (e.g., politicians) or commercial actors (such as media organizations, retailers, and content creators). However, the Statements of Reasons database does not even start to engage with such distinctions and instead errs on the side of privacy in each case. If, over time, the database were able to evolve to carve out areas of non-sensitive content moderation and place the affected content on the public record, it could radically expand the observability of content moderation.

These problems with the Statements of Reasons database touch on a general challenge for observability policy: access to content in semi-public social media communications. Given that observability is concerned with the *contexts* of automated decision-making, it is concerned with specific items of content such as user posts and profiles. How can this be reconciled with user privacy? This is a difficult question because privacy interests around social media content differ wildly. A characteristic feature of social media is that it entangles eminently public forms of communication, such as political campaigning and advertising, with deeply intimate and personal communications. Individual users shift between different contexts and roles, with end-users participating

[78] Regarding conventional moderation reporting, see, e.g., Ben Wagner and others, 'Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act' (2020) Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 261. Keller and Leerssen, 'Facts and Where To Find Them' (n 72).

[79] David Erdos, 'Disclosure, Exposure and the "Right to be Forgotten" after Google Spain: Interrogating Google Search's Webmaster, End User and Lumen Notification Practices' (2020) 38 Computer Law & Security Review 105437.

as both consumers and producers of content.[80] Meanwhile, what is "public" or "private" in a social or practical sense does not correspond neatly to what is accessible in a technical sense. There are technically private channels or groups that are so large and accessible as to attain an effectively public or semi-public function.[81] Conversely, there is also a great volume of technically "public" content that serves intimately private relations.[82] Social media *entangles* public and private modes of communication, often without a stable location or clear dividing line, in a dynamic flow of porous and ever-shifting *publicness*.[83] These blurred boundaries pose a serious challenge for observability policy: solutions that might seem eminently suitable for some items might be highly questionable for others, even within the same technical channels.

This problem of semi-publicness in social media recurs in the Statement of Reasons database but also in other parts of the DSA's observability framework. For starters, online platforms under the DSA are *defined* as services that "disseminate[s] information to the public."[84] Furthermore, article 40(12)'s researcher access provision focuses on information that is "publicly accessible in their online interface." Given the above, implementing these concepts in practice may be complex and may require a more nuanced concept of public content than mere technical accessibility. A starting point might be self-regulation. For instance, the CrowdTangle observability tool developed its own own definitions of public pages and "influential figures" for whom documentation is appropriate.[85] Certain research APIs also implement threshold values for (*inter alla*) follower count and item views to select content for researcher access. Given the blurry boundaries in this space, it will likely take considerable time and effort to operationalize and implement these concepts in a balanced way.

---

[80] This blurring of boundaries is already recognized in early Web 2.0 discussions about the 'prosumer' and 'produsage.' Such concepts reflect the idea that social media users straddle the line between consumers and producers, private and amateur – and, in discursive terms, between public and private communication. User-generated content, for José van Dijck, has always been 'a trade market in potential talents and hopeful pre-professionals,' being 'neither exclusively produced by amateurs nor by professionals' but rather a 'blending of work and play.' See Axel Bruns, 'Produsage' (2007) C&C '07: Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition 99. José van Dijck, 'Users like you? Theorizing Agency in User-Generated Content' (2009) 31 Media, Culture & Society 41.

[81] Joelle Swart, Chris Peters and Marcel Broersma, 'Shedding Light on the Dark Social: The Connective Role of News and Journalism in Social Media Communities' (2018) 20 New Media & Society 4329. Rafael Evangelista and Fernanda Bruno, 'WhatsApp and Political Instability in Brazil: Targeted Messages and Political Radicalization' (2019) 8(4) Internet Policy Review https://doi.org/10.14763/2019.4.1434 accessed 27 September 2022.

[82] Danah Boyd, 'Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications,' in Zizi Papacharissi (ed.), A Networked Self (Routledge 2010). Woodrow Hartzog and Frederic Stutzman, 'The Case for Online Obscurity' (2013) 101 *California Law Review* 1.

[83] Thomas Poell, Sudha Rajagopalan and Anastasia Kavada, 'Publicness on Platforms: Tracing the Mutual Articulation of Platform Architectures and User Practices,' in Zizi Papacharissi (ed.), *A Networked Self and Platforms, Stories, Connections* (Routledge 2018).

[84] DSA, Article 3(i). 'Dissemination to the public' is then defined under Article 3(k) as 'making information available, at the request of the recipient of the service who provided the information, to a potentially unlimited number of third parties.' See also DSA, Recital 14.

[85] Chris Miles, 'What data is Crowdtangle Tracking?' (2022) CrowdTangle.

4.2 Audiences and Aims: From Individual Empowerment to Knowledge Production (and Its Complex Relationship to Regulatory Enforcement)

There is also a noteworthy shift in the *audiences* and *aims* of the DSA's transparency rules. Rules are being designed, either explicitly or implicitly, to serve independent knowledge production by researchers and other civil society actors. No longer are individual users or market actors the focal point, as was the case in earlier legislation, such as the General Data Protection Regulation (GDPR) and the Platform-to-Business Regulation.

This shift is not entirely novel: Content moderation reporting, a long-standing practice, has always been directed at an audience of researchers and watchdogs. However, the DSA takes a new step in the scale and depth of such ambitions. The focus on knowledge production is most explicit in Article 40 DSA, being expressly limited to (vetted) researchers. However, a similar attitude is also evident in, for instance, the Ad Archive and Statement of Reasons Database, which are clearly designed for a similar expert audience.

The DSA now faces an important challenge in operationalizing the concept of "researchers." Who qualifies? Article 40(4)'s vetting framework already contains a long list of criteria to be further elaborated in a delegated act. Article 40's text also already reflects some controversial decisions, such as limiting vetted researcher access to "research organizations," potentially excluding NGOs and journalists. (Notably, a last-minute change did re-include those groups in the public content data rules of 40(12), in response to widespread criticism from those same NGOs and journalists). Concerns have also been raised about the risk of abuse by either or both law enforcement and commercial actors, which researcher vetting must rule out.[86] These and several other vetting criteria all reflect the fundamental challenge of institutionalizing the law's relationship to independent knowledge production and expertise.

Perhaps the most controversial principle in this research access framework concerns purpose limitation: Research must be permitted for the sole purpose of conducting research "that contributes to the detection, identification, and understanding of systemic risks in the Union."[87] What to make of this from an observability perspective? Critics of neoliberal and deregulatory transparency approaches, including Rieder and Hofmann, have emphasized that transparency and binding behavioral regulation should go hand in hand, underscoring that transparency alone

---

[86] Caitlin Vogus, 'Defending Data: Privacy Protection, Independent Researchers, and Access to Social Media Data in the US and EU' (2023) *Center for Democracy and Technology*. https://cdt.org/insights/new-cdt-report-documents-how-law-enforcement-intel-agencies-are-evading-the-law-and-buying-your-data-from-brokers/ accessed April 7 2024.

[87] DSA, Article 40(4).

is rarely sufficient as a guarantee of accountability.[88] Still, the DSA's approach may be more than even these critics bargained for, with its insistence that research occur for the *sole purpose* of regulatory-defined aims. For academics and journalists to be confined in their research to legallydefined regulatory topics and purposes, seems to run contrary to principles of independence and free inquiry. In practice, the tensions here are eased somewhat by the fact that "systemic risks" represents a remarkably broad category. Furthermore, the requirement that research should "contribut[e] to the *understanding*" of these risks is also broad, and opens the door for fundamental research going beyond mere compliance monitoring. Still, these principles speak to a deep question about transparency's relationship to knowledge production and to regulation and law enforcement.

The relationship between researcher access and regulation is multifaceted and difficult to measure or observe directly. The archetypal scenario is a "fire alarm" or "smoking gun" incident that sees research uncover direct evidence of unlawful activity, triggering legal sanctions. However, in practice, this is relatively rare, and research tends to serve regulation in more indirect and subtle ways. To begin with, a "fire alarm" scenario might also prompt platforms to reform their actions *before* legal sanctions are ever imposed. Research can also provide the *inspiration or starting point* for regulatory action, such as an investigation or rules reform, serving an agenda-setting function rather than providing concrete evidence of illegality. Furthermore, to repeat, research can also prompt platforms to reform their own rules or policies in anticipation of concrete regulatory interventions. Indeed, by extension, it can be posited that the *mere presence* of observability tools through which relevant research might take place can have a deterrent effect on platform conduct. To complicate things further, research is itself an iterative process that builds on prior art, and projects without any immediate or obvious link to regulatory issues (e.g., descriptive, exploratory, and theoretical research) might later still prove instrumental in giving rise to future research with regulatory relevance. For all these reasons, it is not straightforward to determine – either as a legal condition for access or as an evaluative criterion for the success of the observability project – how specific research projects contribute to regulation.

In practice, observability's role in legal accountability is therefore difficult to pin down and also overlaps with social and reputational forms of accountability. For instance, if a platform changes its policies in response to new research, this can be read as either an anticipation of regulatory risks or a matter of reputation management and corporate social responsibility. In practice, this is barely a meaningful distinction, all the more so now that VLOs must operate in the shadow of the DSA's open-ended risk management framework. Research projects can also mobilize content creators, media actors, and countless other actors in platform governance, each of whom may attempt to influence legal processes in their own way. In this sense, transparency's role as a regula-

---

88   Sections 2.1 and 2.2 *supra*.

tory monitoring tool is always already entangled in complex social, political, and discursive accountability dynamics. These broader social accountability dynamics may be "gradual, diffuse, and indirect," as per Gregory Michener, making them "indirect and challenging to measure," but they are no less significant for understanding how observability affects platform governance.[89] To focus solely on legal enforcement actions is to miss much of the action.

Ultimately, a realistic and pragmatic theory of observability should not go so far as to reduce the value of independent knowledge production to mere enforcement or compliance monitoring alone. The impulse is understandable. Transparency, given its neoliberal legacy, faces pressure to demonstrate its impacts and head off charges of naïveté and ineffectuality.[90] But realizing observability's full potential calls for a looser coupling with regulation, which contextualizes legal effects and outcomes within broader dynamics of social and democratic accountability. Without this nuance, Article 40's purpose limitation risks undermining much of what makes independent research valuable in the first place. Rieder and Hofmann propose observability as a "companion to regulation," the DSA shouldn't treat it like a deputy sheriff.[91]

Observability's non-legal outcomes and effects are, arguably, especially significant in the context of social media and the governance of lawful content. Constitutional limits on legal ordering are especially strong on issues such as disinformation and political extremism.[92] Precisely where hard legal ordering is inappropriate, observability can also aim to contribute to softer, non-legal social accountability.

Another important point regarding the DSA's aims is that it also innovates in ways that do not relate to observability. Specifically, its content moderation rules are interesting in that they aim not only to explain or describe algorithms but also to justify them. DSA Article 14's content policies and Article 17's statement of reasons may not produce fully accurate accounts of their automated decision-making internal logics (for reasons already discussed), but what they *can* do is provide the basis for subsequent appeals, dispute resolution, and litigation. Following Margot Kaminski, one might say that content moderation due process mainly pursues *justificatory* transparency rather than *instrumental* or explanatory transparency.[93]

[89] Michener 'Gauging the Impact of Transparency Policies' (n 1).

[90] Baume and Panadopoulos, 'Transparency: From Bentham's Inventory of Virtuous Effects to Contemporary Evidence-Based Scepticism' (n 1). Igbal Safarov, Albert Meijer and Stephan Grimmelikhuijsen, 'Utilization of Open Government Data: A Systematic Literature Review Of Types, Conditions, Effects and Users' (2017) 22 Information Polity 1. Maria Cucciniello, Gregory Porombescu and Stephan Grimmelikhuijsen, '25 Years of Transparency Research: Evidence and Future Directions,' 77 Public Administration Review 32.

[91] Rieder and Hofmann, 'Toward Platform Observability' (n 3), p. 23.

[92] Ronan Ó Fathaigh, Natali Helberger and Naomi Appelman, 'The Perils of Legally Defining Disinformation' (2022) Internet Policy Review 10(4).

[93] Margot Kaminski, 'Understanding Transparency in Algorithmic Accountability,' in Woodrow Barfield (ed.), *Cambridge Handbook of the Law of Algorithms* (Cambridge University Press 2020).

Whether these disclosures *accurately describe* the functioning of relevant algorithms, then, is less germane than how they *articulate standards*. This means that such justificatory transparency – aimed at establishing and vindicating users' individual rights and legitimate interests – plays a distinct role in the DSA framework. Observability, by contrast, remains primarily instrumental in nature.

Finally, it is also worth noting that the DSA's user-facing algorithmic transparency and research-facing observability can interact and overlap. Although algorithmic transparency is simplified and decontextualized, it can still provide a starting point for more in-depth inquiry (e.g., a recommender system update is announced via Article 27 DSA, researchers turn to Article 40(12) or 40(4) to study its effects on content trends). Here, it is worth recalling Seth Kreimer's claims about the *ecological* nature of transparency: Disclosures can never be evaluated in isolation but always exist in broader ecosystems that can contradict, corroborate, and cascade, thereby producing greater insights than the mere sum of their parts.[94] Related ideas are also expressed in theories of "tiered transparency," with varying levels of detail serving different audiences.[95] Algorithmic transparency and platform observability can be mutually reinforcing.

4.3 Formats: From Reports to Infrastructures (and the Problem of Scale)

Transparency law is also changing in terms of its format. Besides conventional user-facing policy statements and periodical public reporting, the DSA emphasizes automated and real-time disclosure. APIs, web interfaces, and public databases are all mentioned as relevant disclosure formats in various contexts. For vetted researchers, privacy-protective "data vaults" are also referenced as a possible safeguard.[96] The most detailed rules of this sort are for ad archives, which demand publication "in a specific section of their online interface, through a searchable and reliable tool that allows multicriteria queries and through application programming interfaces."[97] These demands respond to criticisms of self-regulatory archives, which have been held back by limited search functionalities and other usability issues.[98]

---

[94] Seth Kreimer, 'The Freedom of Information Act and the Ecology of Transparency' (2007) 10 *University of Pennsylvania Journal of Constitutional Law* 1011. See also René Mahieu and Jef Ausloos, 'Harnessing the Collective Potential of GDPR Access Rights: Towards an Ecology of Transparency', Internet Policy Review (6 July 2020) https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487 accessed 22 September 2022.

[95] Kaminski, 'Understanding Transparency in Algorithmic Accountability' (n 93).

[96] DSA, Recital 96.

[97] DSA, Article 39(1).

[98] Leerssen and others, 'Platform Ad Archives' (n 66).

In other words, we are seeing a shift in algorithmic transparency regulation from the disclosure of knowledge to the facilitation of data access. In this context, the technical interfaces of access take on a newly significant role, and lawmakers find themselves in a complex new terrain of infrastructure building or "API governance."[99] Several delegated acts are now forthcoming to hash out the details of these rules. If they are unsuccessful, observability may be undermined not on substance but on format and useability.

DSA Article 40's researcher access framework already surfaces some difficult questions for the road ahead. With its reference to APIs and databases, it is clear that this provision intends to generate the production of new researcher access infrastructures. These are, by their nature, scalable solutions that feature high investments up-front but low marginal costs for additional users. In this light, there are choices to be made between the scale and diversity of researcher access. This means considering, for instance, whether the main priority is to develop usable, scalable infrastructures that can support a large volume of research on a number of key topics or to serve a diversity of different perspectives with more bespoke, non-automated access grants. A comparable debate has taken place for decades in the field of government transparency around proactive open data policies versus reactive freedom of information laws.[100] Proactive disclosure models have the benefit of accommodating large numbers of users at the cost of constraining the potential scope and subject matter. Scholarship remains divided as to which model is more effective and efficient in practice.[101] Analogically, one might ask whether DSA Article 40 is to be approached more like a (proactive) open data law or a (reactive) freedom of information law. This is only one instance of the complex infrastructural politics inaugurated by the DSA's observability regulation.

---

[99] Fernando van der Vlist and others, 'API Governance: The Case of Facebook's Evolution' (2020) 8(2) *Social Media+ Society* https://doi.org/10.1177/20563051221086228 accessed 20 September 2022. See also Poell, Nieborg and Duffy, Platforms and Cultural Production (John Wiley & Sons 2021).

[100] cf., Albert Meijer, 'Transparency,' in Mark Bovens (ed), *Oxford Handbook on Public Accountability* (2014) Oxford University Press. Pozen, *Transparency's Ideological Drift* (n 1). David Pozen, 'Freedom of Information Beyond the Beyond the Freedom of Information Act' (2017) *University of Pennsylvania Law Review* 165, 1097. Margaret Kwoka, Saving the Freedom of Information Act (Cambridge University Press 2021).

[101] ibid.

5   Conclusion

Via the concept of observability, this paper has attempted to bring legal debates on platform transparency into conversation with insights from critical transparency studies and critical algorithm studies. For legal scholarship, I have aimed to show how observability offers a new language for articulating how and why information access matters for platform governance while reorienting discussion away from a largely fruitless preoccupation with algorithmic transparency. At the same time, this dive into legal practice has also surfaced new challenges for theorists of observability.

For legal scholars, the move from algorithmic transparency to observability entails a shift in the possible substance, audience, and formats of disclosure. Whereas previous legal scholarship and policymaking have focused on rendering explanations for specific *algorithms*, an observability-based approach instead considers specific governance *systems* or functions, such as content curation, content moderation, and ad targeting. Where these disclosures do touch on algorithms, they are situated in specific contexts, that is, in relation to concrete users, interactions, decisions, and items of content. In parallel, aims and audiences have shifted from individual empowerment to knowledge production and regulatory accountability, and formats from periodical reporting to data access infrastructures. I have argued that the most innovative and promising disclosure rules in the DSA all reflect this observability-based approach.

For observability theorists, the DSA's rulemaking also surfaces new challenges and complications. In terms of substance, access to user-generated content emerges as a central tension. Content access is often crucial to providing qualitative, sociotechnical context for automated decision-making on social media platforms, but it can also raise important privacy objections. Recent scholarship in social media studies and public sphere theory reminds us that it is by no means self-evident – in theory or in practice – which aspects of social media content should be observable. Therefore, the legal work of observability regulation entails a complex weighing of access demands against privacy concerns to disentangle – to the extent possible – social media's observable public discourses from more intimate and private contexts of usage.

In terms of audiences and aims, the DSA also reveals that ideas about the relationship between observability and law enforcement are undertheorized. Critical transparency studies' binary discussions of transparency as either *supporting or supplanting* binding behavioral regulation – that is, transparency as an instrument of either legal or social accountability – may establish a false dichotomy. In practice, social and legal accountability dynamics are rarely isolated; instead, they intermingle. Social accountability often aims to threaten legal consequences, and legal accountability is often mobilized through informal social forums. These nuances are lost when – in the DSA as much as in

critical transparency studies – data access is reduced to an instrument of mere evidence gathering or compliance monitoring. Insisting on direct regulatory payoffs or legal effects not only endangers academic freedom but also misunderstands what observability can offer for platform governance by sidelining more descriptive, exploratory, fundamental, and critical lines of research and inquiry. The task for observability theory now is to articulate with clarity and nuance how observability's role in knowledge production can interrelate with its role in legal accountability while maintaining their respective independence and without subordinating the former to the latter. In addition, we have seen that observability is not the *sole* type of transparency pursued by the DSA: It exists alongside a novel framework for due process transparency aimed at establishing and vindicating individual user rights around content moderation. Observability is not the only relevant lens for examining platform transparency, and it can exist alongside other alternative transparency ideals, including Margot Kaminski's "justificatory" transparency.

Of course, none of the DSA's observability promises will be kept without the political will and institutional backing to ensure robust implementation. A lack of enforcement – or an overly strict interpretation of exceptions and limitations – could scuttle the project before it ever gains traction. This paper has not yet addressed, for instance, how limiting clauses on trade secrecy and service security should be interpreted. What it has tried to do is present the case for why data access matters for the DSA framework in the first place and why it deserves to be given full weighting when balanced against these countervailing costs and interests. The strongest case, I have argued, requires us to move past our preoccupations with algorithmic transparency and the opening of black boxes. Look around; there's lots more to see.