cogitatio

MEDIA AND COMMUNICATION

Volume 11

Issue 2

2023

A Datafied Society: Data Power, Infrastructures, and Regulations

Open Access Journal

ISSN: 2183-2439

Edited by Raul Ferrer-Conill, Helle Sjøvaag, and Ragnhild Kr. Olsen

Media and Communication, 2023, Volume 11, Issue 2
A Datafied Society: Data Power, Infrastructures, and Regulations

Published by Cogitatio Press
Rua Fialho de Almeida 14, 2º Esq.,
1070-129 Lisbon
Portugal

Design by Typografia®
http://www.typografia.pt/en/

Cover image: © NicoElNino from iStock

Available online at: www.cogitatiopress.com/mediaandcommunication

Table of Contents

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 291–295
https://doi.org/10.17645/mac.v11i2.7317

Editorial

Datafied Societies: Digital Infrastructures, Data Power, and Regulations

Raul Ferrer-Conill [1,2,*], Helle Sjøvaag [1], and Ragnhild Kr. Olsen [3]

[1] University of Stavanger, Norway
[2] Karlstad University, Sweden
[3] Oslo Metropolitan University, Norway

* Corresponding author (raul.ferrerconill@uis.no)

Submitted: 20 June 2023 | Published: 28 June 2023

Abstract
The datafication and platformization of social processes further the overall shift from an open, public, and decentralized internet towards a private and siloed realm that establishes power asymmetries between those who provide data and those who own, trade, and control data. The ongoing process of datafying societies embraces the logics of aggregation and automation that increasingly negotiate transactions between markets and social entities, informing governance systems, institutions, and public discourse. This thematic issue presents a collection of articles that tackle the political economy of datafication from three main perspectives: (a) digital media infrastructures and its actors, data structures, and markets; (b) the articulation of data power, public access to information, data privacy, and the risks of citizens in a datafied society; and (c) the policies and regulations for effective, independent media institutions and data sovereignty. It concludes with a reflection on the role of media and communication scholarship when studying sociotechnical processes controlled by giant technological companies.

Keywords
datafication; datafied society; data power; digital infrastructure; media policy; media political economy; media regulation; platforms

Issue
This editorial is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

Societies become datafied by virtue of extensively turning more and more aspects of everyday life into machine-readable data (van Es & Schäfer, 2017). In other words, "to datafy a phenomenon is to put it in quantified form so that it can be tabulated and analyzed" (Mayer-Schönberger & Cukier, 2013, p. 78). Behind this seemingly unstoppable development, there are multifaceted explanations that cater to different philosophical understandings of sociotechnical constructions (van Dijck, 2014). Despite their stance toward datafication, proponents and critics alike tend to agree that collecting, analyzing, and utilizing data in various aspects of life provides various actors with the necessary resources for data-driven decision-making (Kennedy, 2016; Kitchin, 2014; Redden, 2018; Ruppert, 2016) under the (often misplaced) banner of efficiency, accuracy, and effectiveness in various processes.

Much of the media and communication scholarship on datafication has focused on the reasons, objects, and outcomes of the datafication project. Balancing the benefits of data utilization with the potential dangers of data practices becomes a significant challenge. On the one hand, research has shown datafication can be a source of empowerment by democratizing knowledge and decision-making processes, enabling citizen participation and engagement in various domains, such as open data initiatives and participatory governance (Baack, 2015) and structuring patterns of engagement

(Ferrer-Conill et al., 2023). On the other hand, ethical concerns over citizens' privacy and surveillance are at the forefront of datafication scholarship, as governments, organizations, and other entities may have access to vast amounts of personal data, potentially leading to surveillance practices that impact civil liberties and individual autonomy (Cheung & Chen, 2022; Hintz et al., 2018).

In this thematic issue, we are less interested in the who and why, and instead, we focus on the structures that support the datafication of society. We agree with Pellegrino's assessment that datafication amplifies and enriches contradictions already present in modern societies that "are not born with datafication, but rather emphasized and consolidated by it" (Pellegrino, 2019, p. 92). We argue that regardless of intent, the datafication project is supported and enabled by digital infrastructures, power asymmetries that reside in data, and regulatory frameworks. These structures are the symbolic and material constraints that shape how social processes become data points. We believe that our field is not paying enough attention to these three aspects of datafication, and this thematic issue contributes to exploring and highlighting the importance of infrastructures, power, and regulation in the datafied society. Technologies like machine learning, artificial intelligence, and data mining have shown that sometimes intent and outcome are not necessarily correlated during datafication processes. What may start as the marketization of personalized experiences and services tailored to individual preferences and needs may be the source of security vulnerabilities, algorithmic biases, data monopolies, and exacerbating existing inequalities. We thus believe that increased attention to the political-economic aspects of datafication—the material conditions, the power relations, and the regulatory frameworks under which datafication processes take place—will move the field forward as we consider the consequences of increasingly datafied societies.

2. Digital Infrastructures

Digital infrastructures are technological structures with multiple owners, actors, and stakeholders that serve as the backbone for data flows and datafication processes (Parks & Starosielski, 2015) as well as the social processes and practices that organize mundane communication (Hesmondhalgh, 2021). The datafication and platformization of the digital infrastructure, however, shifts the open, public internet towards the private realm, creating power asymmetries between those who provide data and those who own, trade, and control data. This is done through a complex interaction between the political economy of data and the logic of aggregation and automation that increasingly negotiate transactions between markets and social entities, informing governance systems, institutions, and public discourse (Sjøvaag & Ferrer-Conill, 2023).

In their article in this thematic issue, Hesmondhalgh et al. (2023) draw from legal studies research to inform media studies and science and technology studies on the intricate relationship between digital platforms and infrastructures through the lens of political economy and internet governance. Through the case of music, they demonstrate that long-term analysis of infrastructural politics provides a macro-historical account of change and continuity in the shaping of culture. More concretely, they show how platforms have become the main agent of eroding and diminishing the democratizing and emancipatory affordances of an open internet infrastructure.

The role of platforms in capturing the infrastructure that supports datafied societies is also the center of Kristensen and Hartley's (2023) contribution to this thematic issue. As they map the elements that form the digital infrastructure of news media organizations, they offer a compelling overview of how data flows beyond the reach of these organizations and into the infrastructure of platforms and tech companies. This approach reveals how the infrastructures that connect media organizations with the rest of the internet create a set of interdependencies in which the logics of standardization, classification, and datafication articulate manifestations of power between internal and external actors.

To drive the connection between digital infrastructure and power, Flensburg and Lai (2023) elaborate on the concept of "infrastructural power" by following how data flows through the internet infrastructures in Northern Europe and showing how the actors who have control over data can mobilize it into economic profit and societal power. As the flows of data cut across various geopolitical contexts, sectors, and institutional arrangements, they visualize the macro structures that control how data is generated, distributed, and utilized in datafied societies.

3. Data Power

Studying infrastructures of datafication keeps revealing the many ways in which the power and functions exerted by platforms (i.e., Google, Facebook, Apple, and Microsoft) continue to pervade most aspects of everyday life. Governments and advocacy groups have raised concerns about privacy and surveillance fears, threats to freedom of expression, and technological and infrastructure capture (Gillespie, 2018). These issues manifest both at the macro and micro levels, and even on how different actors articulate their understandings of data.

A clear example of how a tech giant such as Google creates a digital innovation that exerts power and creates a dependency on media innovation is at the core of de-Lima-Santos et al. (2023) article in this thematic issue. Through an innovation challenge, the Google News Initiative supports projects in Africa and the Middle East as a form of "philanthrocapitalism," in which Google sets the terms and conditions of the financial and technological grant by expecting projects that replicate the

entrepreneurial approaches of tech companies. As news outlets build technological solutions based on the platforms' technologies, they further depend on platforms' structures to operate, widening the power asymmetries between the tech giants and the news organizations.

These different values and understanding among actors managing datafication processes suggest diverging patterns of decision-making in organizations. In the context of the smart city, Okafor (2023) suggests that the value of data does not reside in the volume of data, but rather in the quality of the data, which is directly connected to its capacity to deliver impactful decision-making and hence its societal power. Importantly, the differences in how technical and governance actors understand datafication, provide a more complex dynamics of power negotiation within organizations.

The power asymmetries are not only felt at the organizational level but also at the micro level. Dutt (2023) shows how Norwegian students negotiate their own experiences with datafication as the entire digital ecosystem pushes social interaction to be mediated through data-driven processes. As the structures of datafied societies shape communication, citizens must contend with digital risks that challenge their perception of wellbeing. Despite internal strategies to manage these risks, concerns at the slow-paced response by governments and digital policies regarding risks over which users have neither control nor power.

4. Regulations

Relying increasingly on platforms and proprietary resources of tech giants places too much responsibility on private actors, threatening to curtail government power (Flynn, 2004). As citizens grow more dependent on corporate platforms for communication, they become bound by the benevolence of private actors, to which states have little recourse for action to regulate the abuse of market power (Hintz et al., 2017).

This becomes particularly apparent in Salonen et al. (2023) research, in which they demonstrate that news workers' editorial decision-making processes are iteratively shaped by the constraints of audience data, platform affordances, working practices, and regulations. More concretely, the authors suggest that broader regulatory frameworks, such as General Data Protection Regulation exert a post-publication gatekeeping power on Finnish media organizations. Enforcement of legislation is seen as a key lever of change and an explicit articulation of data ethics upon which media self-regulation is not equipped to act upon.

But despite the notion that regulation has an important role to play, the reactionary stance of policy initiatives is often perceived as late and fragmented. In their aim to translate normative dimensions of media diversity into a framework for operationalizing exposure diversity into tangible policy goals, Ranaivoson and Domazetovikj (2023) expose the challenges in which EU regulation finds itself in a time of digitalization and datafication. Through a review of policy initiatives and interviews with policy experts in various countries, the authors acknowledge the many potential benefits of regulatory frameworks, but they caution policymakers to include measures, metrics, methods, and data requirements to achieve more diversity.

And while current regulatory practices seem to disadvantage small media organizations in favor of tech giants, Seipp (2023) argues that media concentration law is the relevant legal tool to curb the scale and power gained by platforms due to datafication. The research emphasizes exposing the gaps and promises for a digital media concentration law from the macro to the micro levels. This contribution proposes a united piece of legislation that draws from multiple policy fields with shared policy goals such as normative public values (media pluralism, equality, power dispersal, and transparency) and fair competition.

5. Conclusions

This thematic issue contributes to the debate on datafication by: (a) making the infrastructures that support datafication visible, enabling insight into the power dynamics, data control, and regulatory frameworks that shape citizens' access to information on which inclusive decision-making relies; (b) expanding the empirical basis on which to critically interrogate what the privatization of communication infrastructures and what the data structures mean for citizens' inclusion and communication rights within datafied societies; and (c) providing policymakers insight into the complex dynamics in which datafication rests so that they can incorporate the impact of foreign players on the diversity of the media landscape, and maintain universal communications provisions in policy formations. Together these contributions shed new light on the depth of infrastructural dependencies (cf. Plantin & Punathambekar, 2019) that media organizations face as they datafy their practices.

While datafication has an undoubtedly technological background, the articles in this thematic issue have approached the underlying social and economic dynamics of a process that is rapidly questioning the current social order (Couldry, 2020). In the final commentary of this issue, Gillespie (2023) uses the case of content moderation as a call to social scientists to caution against "solving the platforms' problems for them." We agree with his assessment that media and communication studies is a discipline institutionally caught between its critical commitment to social issues and the actors that set in motion those issues. We hope we have not overplayed the sense of urgency and that instead of solving problems, this thematic issue has deconstructed some of the often unseen issues, failings, and risks associated with datafication, and made them visible for regulators and policymakers, who are tasked with addressing them.

Acknowledgments

This thematic issue is supported by various funding bodies. The guest editors would like to acknowledge the Norwegian Research Council (NFR) through the research project The Datafication of Communicative Power: Towards an Independent Media Policy for Norway's Digital Infrastructures (project No. 314257); The Swedish Research Council (VR) through the research project Digital Infrastructure Sovereignty: Towards a Public Value-Based Media Policy for the Datafied Swedish Welfare State (project No. 2022–05392); and the Anne-Marie och Gustav Anders Stiftelse för Mediaforskning through the research project Digital News Agendas in Scandinavia. We would also like to thank all the reviewers who kindly contributed their time and insight.

Conflict of Interests

The authors declare no conflict of interest.

References

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 1–11. https://doi.org/10.1177/2053951715594634

Cheung, A. S., & Chen, Y. (2022). From datafication to data state: Making sense of China's social credit system and its implications. *Law & Social Inquiry*, *47*(4), 1137–1171. https://doi.org/10.1017/lsi.2021.56

Couldry, N. (2020). Recovering critique in an age of datafication. *New Media & Society*, *22*(7), 1135–1151. https://doi.org/10.1177/1461444820912536

de-Lima-Santos, M.-F., Munoriyarwa, A., Elega, A. A., & Papaevangelou. C. (2023). Google News Initiative's influence on technological media innovation in Africa and the Middle East. *Media and Communication*, *11*(2), 330–343. https://doi.org/10.17645/mac.v11i2.6400

Dutt, B. (2023). Wellbeing amid digital risks: Implications of digital risks, threats, and scams on users' wellbeing. *Media and Communication*, *11*(2), 355–366. https://doi.org/10.17645/mac.v11i2.6480

Ferrer-Conill, R., Karlsson, M., Haim, M., Kammer, A., Elgesem, D., & Sjøvaag, H. (2023). Toward "cultures of engagement"? An exploratory comparison of engagement patterns on Facebook news posts. *New Media & Society*, *25*(1), 95–118. https://doi.org/10.1177/14614448211009246

Flensburg, S., & Lai, S. S. (2023). Follow the data! A strategy for tracing infrastructural power. *Media and Communication*, *11*(2), 319–329. https://doi.org/10.17645/mac.v11i2.6464

Flynn, J. (2004). Communicative power in Habermas's theory of democracy. *European Journal of Political Theory*, *3*(4),433–454. https://doi.org/10.1177/1474885104045914

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Gillespie, T. (2023). The fact of content moderation; Or, let's not solve the platforms' problems for them. *Media and Communication*, *11*(2), 406–409. https://doi.org/10.17645/mac.v11i2.6610

Hesmondhalgh, D. (2021). The infrastructural turn in media and internet research. In P. McDonald (Ed.), *The Routledge companion to media industries* (pp. 132–142). Taylor & Francis. https://doi.org/10.4324/9780429275340-13

Hesmondhalgh, D., Valverde, R. C., Kaye, D. B. V., & Li, Z. (2023). Digital platforms and infrastructure in the realm of culture. *Media and Communication*, *11*(2), 296–306. https://doi.org/10.17645/mac.v11i2.6422

Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance society: Introduction. *International Journal of Communication*, *11*(2017), 731–739.

Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2018). *Digital citizenship in a datafied society*. Wiley.

Kennedy, H. (2016). *Post, mine, repeat: Social media data mining becomes ordinary*. Springer.

Kitchin, R. (2014). *The data revolution*. SAGE.

Kristensen, L. M., & Hartley, J. M. (2023). The infrastructure of news: Negotiating infrastructural capture and autonomy in data-driven news distribution. *Media and Communication*, *11*(2), 307–318. https://doi.org/10.17645/mac.v11i2.6388

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Okafor, C. C. (2023). "I think quality is more important than a lot of data" in cities datafication. *Media and Communication*, *11*(2), 344–354. https://doi.org/10.17645/mac.v11i2.6510

Parks, L., & Starosielski, N. (2015). *Signal traffic: critical studies of media infrastructures*. Oxford University Press.

Pellegrino, G. (2019). Inside "the below": Ambivalences of datafication and infrastructuring of everyday life. *Tecnoscienza*, *10*(1), 89–96.

Plantin, J.-C., & Punathambekar, A. (2019). Digital media infrastructures: Pipes, platforms, and politics. *Media, Culture & Society*, *41*(2), 163–174. https://doi.org/10.1177/0163443718818376

Ranaivoson, H., & Domazetovikj, N. (2023). Platforms and exposure diversity: Towards a framework to assess policies to promote exposure diversity. *Media and Communication*, *11*(2), 379–391. https://doi.org/10.17645/mac.v11i2.6401

Redden, J. (2018). Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. *Big Data & Society*, *5*(2), 1–13.

https://doi.org/10.1177/2053951718809145
Ruppert, E. (2016). Big data economies and ecologies. In L. McKie & L. Ryan (Eds.), *An end to the crisis of empirical sociology? Trends and challenges in social science research* (pp. 13–29). Routledge.
Salonen, M., Ehrlén, V., Koivula, M., & Talvitie-Lamberg, K. (2023). Post-publication gatekeeping factors and practices: Data, platforms, and regulations in news work. *Media and Communication*, *11*(2), 367–378. https://doi.org/10.17645/mac.v11i2.6486
Seipp, T. J. (2023). Media concentration law: Gaps and promises in the digital age. *Media and Communication*, *11*(2), 392–405. https://doi.org/10.17645/mac.v11i2.6393
Sjøvaag, H., & Ferrer-Conill, R. (2023). Digital communications infrastructures and the principle of universality: Challenges for Nordic media welfare state jurisdictions. In P. Jacobsson, J. Lindell, & F. Stiernstedt (Eds.), *The future of the Nordic media model: A digital media welfare state?* Nordicom.
van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776
van Es, K., & Schäfer, M. T. (2017). *The datafied society. Studying culture through data*. Amsterdam University Press.

About the Authors

Raul Ferrer-Conill is an associate professor in the Department of Media and Social Sciences at the University of Stavanger, Norway. His research focuses on digital journalism, media engagement, and sociotechnical processes of datafication. He currently leads the Digital Society Research Group and is chair of the Journalism Studies Division of Nordmedia. Ferrer-Conill received the Bob Franklin Journal Article Award in 2021.

Helle Sjøvaag is professor of journalism at the University of Stavanger, Norway. Her research focuses on media economics, digital infrastructures, datafication, and regulation of the communication industries. She currently leads the Research Council of Norway-funded project The Datafication of Communicative Power: Towards an Independent Media Policy for Norway's Digital Infrastructures. Her most recent book is *The Markets for News: Enduring Structures in the Age of Business Model Disruptions* (Routledge, 2023).

Ragnhild Kr. Olsen is an associate professor at the Department of Journalism and Media Studies at Oslo Metropolitan University. Her research interests are digitisation and value creation in journalism, particularly in local journalism, as well as media innovation, audience perspectives in editorial priorities, and platformisation of news media.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 296–306
https://doi.org/10.17645/mac.v11i2.6422

Article

Digital Platforms and Infrastructure in the Realm of Culture

David Hesmondhalgh *, Raquel Campos Valverde, D. Bondy Valdovinos Kaye, and Zhongwei Li

School of Media and Communication, University of Leeds, UK

* Corresponding author (d.j.hesmondhalgh@leeds.ac.uk)

Submitted: 31 October 2022 | Accepted: 2 March 2023 | Published: 28 June 2023

Abstract
The concepts of (digital) platform and (digital) infrastructure have been widely used and discussed in recent media research, and in neighbouring fields such as science and technology studies (STS). Yet there is considerable confusion about these concepts and the relations between them. This article seeks to bring these concepts together more coherently by showing how "platformisation" might be understood in terms of its impacts on information infrastructure, including on the principles of openness and generativity underlying early internet architecture, and potential further effects on media and culture deriving from those impacts. To develop this perspective, we draw on research from legal studies which: (a) articulates these principles more fully than in recent media studies and STS; (b) understands infrastructures as resources subject to political contestation; and (c) in the work of Julie Cohen, interprets digital platforms as strategies for disciplining infrastructures. We discuss how such a perspective might complement approaches to digital platforms and infrastructures to be found in political economy of media and internet governance research. We then apply the perspective to a case study: the transition of online music from chaotic experiments with alternative models of distribution in the early century to a thoroughly platformised environment in the 2020s.

Keywords
digital information infrastructure; digital platforms; internet architecture; internet regulation; music streaming platforms; platformisation

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction: Infrastructural Turns and Platformisation

Infrastructure appears to be a concept for our times. It is increasingly central to national politics and global geopolitical power struggles ("Is an infrastructure boom," 2021). It is now common to hear terms such as "infrastructure studies" or "infrastructural turn" in academic research, not only in media and communication studies, but across the social sciences and humanities (Edwards et al., 2009). A rich body of research has emerged around the concept in media and communication studies, often fuelled by interest in digital information infrastructures. Valuable ethnographic media research has helped to illuminate the ways in which people in postcolonial contexts have adapted imported infrastructural information and communication technologies (ICTs), decentring western understandings of them (Parks, 2015b; see also Larkin, 2008). Some of the new infrastructural research (e.g., Winseck, 2017) builds on a long history of engagement with media and information infrastructures, for example, political economy research on privatisation and marketisation of telecommunications (e.g., Schiller, 2000). Digital information infrastructures have also been of interest to scholars working in other fields, perhaps most notably anthropologists in science and technology studies (STS; e.g., Bowker & Star, 1999; Burrell, 2018).

Yet there appears to be great confusion among researchers across these fields about how to conceptualise and use the term. Whereas many non-academics

would understand it as referring to something like "the basic systems and services that are necessary for a country or an organisation to run smoothly, for example buildings, transport and water and power supplies'' (Infrastructure, n.d.), social science and humanities researchers increasingly seem to use it in a bewildering variety of ways, including quite often as a synonym, or perhaps a metaphor, for "system" or even "importance," implicitly understanding infrastructure merely as something which has important implications for something else (see Hesmondhalgh, 2022, which catalogues some of the confusions).

Perhaps as a result of this definitional and conceptual vagueness, curiously little of this rich body of research has engaged explicitly and in detail with some fundamental infrastructural aspects of the development of ICTs in the 21st century. One key aspect, relatively neglected in recent research, is that the internet infrastructure that underpins so much contemporary media and communication was initially framed by its developers as a common resource available to all (regardless of private or public ownership of any elements of it), enabling the creation of an international network of networks, based on values of *open-ness* and more broadly what Zittrain (2008, p. 70) called "*generativity*": a "system's capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences."

Such principles of open-ness in internet architecture were discussed widely in writing about the internet and web in the oughties (e.g., Gillespie, 2007) but have largely been neglected since in media studies and STS. During that decade, the term "platform" was hardly used, but from around 2010 onwards, internet and digital scholars began to employ it very widely and it is now pervasive in a range of disciplines and fields (see Gillespie, 2018, pp. 18–21, for a helpful discussion of definitions). Researchers from media studies, urban studies, and geography have highlighted the importance of evolving relations between digital platforms and infrastructures, and between "platform studies" and "infrastructure studies" (Plantin et al., 2018). A large number of media scholars have implicitly or explicitly followed this move (Eriksson et al., 2019; Lobato, 2019; Poell et al., 2022; see also a number of contributions discussed in Section 2). Yet in spite of Plantin et al.'s (2018) suggestive efforts to put these two previously separate domains into dialogue, much of the research citing their important article has not sought to conceptualise the relationships between digital infrastructures and digital platforms. Indeed, confusion reigns here too. It is common to read or hear, in the wake of Plantin et al.'s (2018) influential article, phrases to the effect that "platforms have become…important infrastructures" (Sadowski, 2020, p. 567). But this appears only to mean that platforms are used as the basis of some other set of activities: infrastructure as a vague metaphor rather than a conceptualised tool for analysis. This article seeks to bring these concepts together more coherently by pointing to ways in which "platformisation"—a currently very popular and yet poorly conceptualised notion in media studies, internet studies, STS, and beyond—might be understood as having impacts on information infrastructure. This includes effects that, drawing on debates about open-ness and generativity in internet and web architecture, might be characterised as "closure" or "enclosure" (Cohen, 2019). Such a view is arguably implicit in Plantin et al.'s (2018) discussion of Google and Facebook's impact on the open web, but it does not seem to have been much taken up in discussions of that very widely-cited article.

To develop this perspective further, we make two important moves. First, we draw on research (e.g., Cohen, 2019; Frischmann, 2012; van Schewick, 2010) that helps to conceptualise these values and principles much more fully than has been the case in recent media studies and STS. This research mainly comes from legal studies, but its remit is by no means exclusively or even primarily concerned with law. Second, we apply it to an actual case study of the impacts of platformisation on the possibilities made available by open internet infrastructure, by examining what happened in the realm of online music. Normatively, our concern is that the internet (with access to content provided by the World Wide Web) was designed as an "open" network that for all its problems brought about "an explosion in innovation and content, which in turn is why the openness is considered to be worth protecting as something that has an intrinsic public-interest value" (Horten, 2016, p. 9).

The case of music is particularly illuminating in terms of the above issues, because for a period in the late 1990s and 2000s open internet infrastructure seemed to provide a major challenge to the problematic institutional framework that had sustained the recorded music industry for decades. Yet the same lack of attention to the politics of infrastructural systems and their place within the changing political economy of capitalism is also apparent in research on changes in music production and consumption.

In Section 2, we expand and nuance the above claims by demonstrating the limitations in even the current leading research on digital infrastructure and digital platforms in media and communication studies, and in digital music studies (we note in passing that studies of music and digitalisation have tended to pay much greater attention to media studies than media studies has to digital music research). In Section 3, we turn to the (mainly) legal studies research just mentioned to develop a framework for understanding the (en)closure of internet architecture's open-ness via platformisation and other means. In Section 4, we show how music serves as an early and revealing case study of the closing down of infrastructural potential by platformisation.

2. Recent Treatments of Infrastructure and Platforms in Research on Media, Culture, and Music

In spite of the widespread use of infrastructure as a rather vague metaphor or synonym for "system," as

noted in the previous section, surprisingly little research on media infrastructures has actually analysed media infrastructures *as* systems. The recent "infrastructural turn" in media, communication, and internet studies has provided fine studies of particular infrastructural items, such as the server farms that are part of general IT infrastructure (Holt & Vonderau, 2015; Mayer, 2019), or the pipelines that serve inter-continental digital traffic (Starosielski, 2015). Such analyses have produced significant insights regarding matters such as the environmental consequences of IT infrastructure and the implications for the labour of where particular infrastructural sites are located. But as we show below, very few authors have sought to delineate the systems of internet infrastructure or architecture in a way that illuminates recent and current developments in media content and distribution. This includes a lack of attention to the infrastructures underpinning the streaming platforms that increasingly shape media consumption today.

The necessarily brief review that follows focuses on those rare in-depth treatments of media infrastructural systems that we have been able to locate. A chapter by Ramon Lobato (2019) on the infrastructure underpinning Netflix helpfully delineates how that platform relies on two different kinds of infrastructure: public and private telecommunications networks, and its own internal IT networks and systems. He shows how this leads Netflix to lobby for state investment in internet infrastructure (cf. Elkins, 2018) and he relates this to the debates about net neutrality in the US, especially whether internet service providers (who essentially control telecommunications infrastructure) should be able to charge services that are dependent on them, such as video streaming services. Lobato shows that while many major web services use a commercial content delivery network, Netflix has established their own, bringing this vital infrastructural element under its control. Such issues concerning speed of delivery have undoubtedly been the main way in which the infrastructures underpinning video distribution have featured in public debate (cf. Johnson, 2019, pp. 70–72). But Lobato's conclusion suggests that the main value of an examination of infrastructure is to remind us that phenomena such as Netflix are dependent on many "longer-term, larger-scale social and technical transformations" (2019, p. 103), from the history of electrification and lighting to modern architectural forms and changing family practices. Lobato's point is that recognition of this dependence in turn might throw light on the uneven availability and uptake of Netflix across the globe, helping us see that Netflix is a very culturally specific phenomenon, dependent on television's longer history as a domestic technology.

This is valuable, but Lobato does not explicitly address the infrastructural principles underlying the internet, as indicated above and discussed below, nor how platformisation, along with associated changes such as datafication, have reshaped them. Poell et al. (2022) come closer to addressing this topic in a chapter examining the implications of platform infrastructure for cultural producers, and they do so by ranging across a number of sectors, including games and social media. Their main argument is that the relationship of platform companies to neighbouring industries, including the cultural industries, are characterised by an oligopoly of major tech firms that base their goals on "infrastructural integration" and "interoperability." For many decades, cultural producers have been dependent on infrastructures owned by companies as well as by states, but Poell et al. (2022) claim the balance has become far more tilted towards privately-owned networks in the era of platforms. And because platforms curate, organize, archive, and moderate content, this means that, according to Poell et al. (2022), tech businesses now have a potentially profound effect on cultural producers.

Poell et al. (2022) deserve great credit for delineating a number of elements that are relevant to understanding changing relations between platformisation and infrastructure, including the need to differentiate particular platforms (e.g., Facebook Messenger) from the platform ecosystem that they operate within (i.e., Facebook's systems); and the way in which "boundary resources" such as application programming interfaces and software development kits operate not only as "support" for cultural producers, but also as ways of securing or controlling them. However, their main normative focus is on the potentially pernicious effects of *ownership* rather than on infrastructural principles per se, and on the principle of open-ness mentioned in the previous section.

The most significant writing we have found that applies understandings of infrastructural principles to the distribution of media content is a chapter by Christian Sandvig (2015), which explains how internet architecture, based on a "point-to-point" system oriented towards communication between two nodes, more akin to postal services than to television, had to be radically amended to accommodate a new sector of business organised around the distribution of online videos. The thinkers behind internet architecture, and the commentators that followed them, in Sandvig's words, "expected that providing television via the Internet would transform television, but instead it caused the Internet's distribution architecture to become like television in significant ways" (2015, p. 237). Sandvig outlines multiple efforts to solve the problem of video distribution, via compression, streaming, buffering, server farms, and so on, but how in the end "changes to standards, protocol, and system architectures" had to be made in order to shift to "a more familiar model of mass communication," exemplified by the rise of content delivery networks, methods of caching files to ensure the prioritisation of those that required vast bandwidth (2015, p. 238). The result was a hybrid of open-ness and closure, but increasingly centred on a "closed" mass audience model. Sandvig's is a vital precedent for our take here but it makes strangely little mention of audio content. While video content required much greater change to

accommodate its much greater bandwidth, music served as a key testing ground for the transformative potential of internet architecture's open-ness in the realm of information, entertainment and culture (in addition, radio has arguably been as important historically as television in terms of broadcasting infrastructures).

So how *has* music featured in debates and research about internet infrastructure? There was a huge amount of media coverage of music in the period following the widespread take up of the internet in wealthy countries in the late 1990s and early 2000s, when various websites made it possible for ordinary internet users to share and download digital files of music, threatening the system of copyright that had sustained the recorded music industry. Polarised positions arose between those who emphasised the need to protect copyright, often expressed in terms of the interests of musicians, and those who considered copyright to be a problem. In response to these heated public debates, copyright emerged from the specialist shadows to become a fashionable and widely discussed topic across a range of academic disciplines including media and communication studies, law, and music studies (among many contributions, see Lessig, 2004; Vaidhyanathan, 2001). There were also numerous studies of "piracy," peer-to-peer file sharing, and other related phenomena, some with a strong emphasis on music (Andersson Schwarz, 2013; David, 2010). The dependence of such developments on the open-ness of the internet featured fairly prominently in these accounts.

However, by the time music streaming platforms had emerged as a new lasting basis for recorded music commerce, from about 2015 onwards, the open infrastructure of the internet seemed to disappear from the academic agenda. Morris (2015)'s important book on the digitalisation of music refers to infrastructure many times but barely mentions the infrastructural ideals behind the "open web" or its closure. In more recent accounts, the first decade of the 21st century is often treated as a brief period of chaos before order was re-established, first in the form of Apple's iTunes and then in the form of music streaming platforms such as Spotify (Sun, 2019), with little reference to internet infrastructure. Morris (2021) has a later article on "infrastructures" of discovery, mainly in relation to podcasting on audio platforms, but he seems to use the term as a metaphor for "system." The agenda of recent research has been overwhelmingly to understand the new order based on streaming, and little reference is made to the period preceding it. But even in those accounts of digitalisation published before 2015, which attempted to tell the story of disruption, infrastructural politics were almost entirely missing from the picture. There was considerable attention to particular sites and applications such as Pirate Bay and Napster, but little consideration of the information infrastructure that made them possible (though Andersson Schwarz's, 2013, excellent account of how file-sharers understand and justify their practices recognises the importance of internet architecture). This neglect seems all the more remarkable given how widely the term infrastructure is used in much recent writing on media, music and culture.

On first sight, writings by Paolo Magaudda seem to blend consideration of digital media infrastructures with attention to the concept of platformisation, while making specific reference to music—in line with the aims of this article. For example, Magaudda (2020) explores how the concept of infrastructure might help develop further the notion of music scenes, very widely used in popular music studies to refer to the musical activity built around particular genres in particular geographical areas, mainly cities. Magaudda recounts instances of the use of the term infrastructure in earlier research on music scenes, where it was used to refer rather vaguely to music venues and other institutions such as record shops, and he asserts the need to understand how "new kinds of digital infrastructures and platforms did much more than offer a new space for fandom or new opportunities to link together artists and listeners from different countries and regions" (2020, p. 33). However, it is not clear what Magaudda means by "digital infrastructures and platforms," nor how he understands their different roles and their relationship. Instead, his discussion of platforms is confined to a brief mention of the increasing use of recommendation algorithms based on user data, and some speculative comments about the use of blockchain technology. A later contribution (Magaudda, 2021) includes the thoughts of users on algorithms and musical taste, but the role of infrastructures in such developments and the precise conceptualisation of infrastructure in operation is not made clear there either. As with so many treatments of implications of "digital infrastructure" for media, as indicated in this section, there is no discussion of how the principles and values underlying internet infrastructure might have helped to reshape music, and how later dynamics of platformisation, including the rise of recommendation algorithms, etc., relate to those principles.

One detailed account of the infrastructure underpinning music streaming that we have been able to identify, from a critical social science and humanities perspective (rather than a purely technical one) is a chapter on infrastructure in Eriksson et al.'s *Spotify Teardown* (2019). The approach of that chapter is based on the authors' view that "to understand the logic and rationale of streaming services such as Spotify, we need to ask what exactly happens when data are turned into music and vice versa" (Eriksson et al., 2019, p. 80). Their analysis addresses specific *elements* of infrastructure in some detail. For example, they outline Spotify's "event delivery system," "one of the foundational pieces" of Spotify's data infrastructure—i.e., how data gets transferred between different elements and places within the company. They also discuss (pp. 82–88) data exchange and interaction with other companies, such as the music information retrieval company Echo Nest, which allowed

for automated music recommendation to be integrated into the service (Spotify eventually bought this company); and Spotify's collaboration with Facebook and its opening of its application programming interfaces to external developers, as Spotify attempted to mutate for a while into more of a social media service. They analyse Spotify's systems for storing and retrieving data, including their shift from using their own servers to using Google's cloud services—though the significance of this shift is not elaborated. The role of aggregators (Eriksson et al., 2019, pp. 91–96) or digital distributors is also discussed as part of their consideration of infrastructure though it is not clear why the authors consider such digital distributors or aggregators to be part of infrastructure as such—possibly they too are using infrastructure primarily as a metaphor for *importance*, here applied to organisations rather than infrastructures per se. Yet the role of Spotify and other platforms in closing down cultural possibilities that were fundamentally dependent on key features and principles of internet and web infrastructure are not addressed.

3. Infrastructural Systems as Resources and Digital Platforms as Disciplinary Mechanisms

Our approach in this article is premised on the view that details of particular infrastructural technologies are rarely illuminating with respect to public culture and media outcomes. Instead, our view is that if infrastructure is going to be mobilised as part of truly critical media studies, and to contribute to the understanding of the potential for a more democratic, vibrant media system that serves human flourishing for all, it is necessary to focus on the *politics* of infrastructure and its place within evolving modes of capitalism (whether understood as "informational capitalism," "platform capitalism," or some other formulation; here we borrow Cohen's use of the former term but remain agnostic on its merits).

In line with this approach, we highlight two aspects of infrastructure that have only infrequently been recognised as central issues in the recent infrastructural turn in media and communication studies: the way that infrastructures potentially function as *resources* for many individuals, organisations, communities, and groups; and the infrastructural politics involved in disputes over their *provision*. We derive this dual focus partly from legal scholar Brett Frischmann's (2012) book *Infrastructure*, a source that has been largely ignored in media, communication, and internet studies. While Frischmann's work addresses a range of infrastructures, including transport and environmental ones as well as "intellectual infrastructures," it is his work on internet infrastructure that offers the most potential for understanding the implications of information infrastructures for public culture, because of its emphasis on infrastructure as a resource and on the politics surrounding the provision of that resource.

Frischmann's (2012) approach also happens to offer routes for moving beyond the definitional and conceptual chaos surrounding academic understandings of infrastructure already discussed, and he clarifies the specificity of information infrastructure. Like many recent analysts of infrastructures, Frischmann's approach recognises that they involve more than just pipes and tubes, the "stuff you can kick" (Parks, 2015a). In the context of the internet, Frischmann (2012) borrows a distinction between the physical infrastructure ("a wide variety of physical networks interconnected with each other," p. 319) and the "logical infrastructure" ("the standards and protocols that facilitate seamless transition of data across different types of physical networks," p. 319). This is consistent with the emphasis on standards and protocols in internet governance research that considers infrastructural questions (e.g., Musiani et al., 2016); media studies and STS literature tends to be rather less precise (Hesmondhalgh, 2022).

Compared with the internet governance literature, the approach of another legal scholar, Julie Cohen, is much more macro-historical in focus, and shares our concern (and that of political economy) with changing relations between capitalism, technology, and culture. Drawing on Frischmann's understanding of infrastructure (Cohen, 2019, pp. 40–41), she analyses the role of (US) legal systems in laying the ground for three large-scale shifts characterising what she calls "informational capitalism." The first shift is capitalism's drive to produce property out of intangible resources, including the expansion of intellectual property across copyright, trademark, and branding, the way for which has been paved by a massive expansion of legal entitlements of rights owners. The second is that labour, land, and money (the basic factors of production in a capitalist economy in the classic work of economic historian Karl Polanyi) have been reconfigured into "datafied inputs to new algorithmic modes of profit extraction" (Cohen, 2019, p. 25), centred on digital platforms. Cohen (2019) sees this as a process of de facto appropriation and enclosure with implications as profound for human well-being as the earlier enclosures of land and labour that marked the beginning of the industrial phase of capitalism. The third shift identified by Cohen as characterising the rise of informational capitalism is the way that the emergence of platforms from the new emphasis on data has created a new layer of infrastructure, with huge ramifications for economic exchange. For Cohen, digital platforms do not just enter markets, but replace and rematerialize them (2019, p. 42). Digital platforms have thus come, in Cohen's view, to serve as "strategies for bounding networks and privatizing and disciplining infrastructures" (Cohen, 2019, p. 41).

What might Cohen mean by this latter formulation? A key element (though not the only one) concerns the lost potential of the original principles underlying internet infrastructure or architecture, as discussed by another legal scholar, Barbara van Schewick

(2010). The term "architecture" is potentially as vulnerable to misunderstanding and conceptual confusion as "infrastructure," but van Schewick uses it to mean "the fundamental structures of a complex system as defined during the early stages of product development" (van Schewick, 2010, p. 20); we follow Frischmann in translating van Schewick's "architecture" into "infrastructure." Van Schewick (2010) shows that internet architecture depends on three fundamental principles: modularity, layering, and the end-to-end principle. Modularity is the design principle of making elements of a system independent of each other. Layering represents a particular version of modularity, whereby the modules or elements are organised hierarchically. Scholars nearly always use layered models of the internet to understand the relations of dependence and complementarity underlying its functioning, and, drawing on van Schewick and others, Frischmann (2012) adopts a five-layer model (see Table 1) whereby the physical and logical infrastructure layers provide the foundations for "higher" layers of applications and content (third and fourth layers) but also a "social layer" of networks, affiliations and groups (see Zittrain, 2008, pp. 67–69, for a similar model).

The important point is that the end-to-end principle (or strictly speaking, a particular, "broad" version of it; see van Schewick, 2010) organises this layering so that the lower layers of the network are as general as possible, while all application-specific functionality is concentrated in higher layers at "end" hosts. This means that the lower layers are in effect "blind" as to the way in which the internet is used, and to the identity of the user.

Crucially, in terms of the politics of the internet, end-to-end design "sustains an infrastructure commons by insulating end-users from market-driven restrictions on access and use of the infrastructure" (Frischmann, 2012, p. 322). It does so because if infrastructure providers follow end-to-end principles strictly, they "cannot distinguish between end uses, base access decisions or pricing on how packets may be used, or optimize the infrastructure for a particular class of end-uses" (Frischmann, 2012, p. 322). It is this and other related aspects of internet infrastructure that Cohen is referring to in seeing platforms as "strategies for bounding networks and privatizing and disciplining infrastructures" (Frischmann, 2012, p. 322). Ten Oever (2021) also captures the tangle of issues here by distinguishing three key elements of the "internet architecture imaginary": the end-to-end principle, but also permissionless innovation (that there are no barriers to the development of new protocols), and open-ness (for example, that new computers can be added, and that information can travel freely from node to node).

It perhaps goes without saying that this set of values, which are both technical and ethical, were rapidly compromised, even in the early internet, by forces such as privatisation and the erosion of such principles in governance organisations. And they can be applied in dubious ways by powerful actors, for example by justifying problematic applications of ideas of liberty and autonomy, as shown by Cath's (2021) ethnographic study of the Internet Engineering Task Force's attitudes to the relationship between technological development and human rights. Our goal in this article is not to celebrate these infrastructural principles but to analyse the fate of their potential in terms of media, communication, and culture, focusing on the domain of music.

As indicated in Section 1, one way of summarising this set of values or principles is "open-ness" (though ten Oever, 2021, lists this as merely one of the key elements of the "architecture imaginary"); another is Zittrain's "generativity." Early utopian accounts of the emancipatory potential of the internet made frequent reference to such open-ness (Russell, 2014, traces the emergence of open standards and associated ways of thinking). Perhaps because the internet has so clearly fallen short of the aspirations of its idealistic early proponents, discussion of these principles and their lost potential is rarely found in recent media research on the impact of the internet on culture and communication. Yet it is surprising that, in a media studies context where infrastructure is such a fashionable term, this vital and consequential aspect of internet infrastructure has been so ignored.

The role played by digital platforms in "closing down" this open-ness or generativity also seems poorly understood. Of course, many observers, from inside and outside academia, have a sense that the internet has

Table 1. Five-layer model of the internet.

Layer	Description	Examples (music streaming)
Social	Relations and social ties among users	Integration with social networking platforms (i.e., Facebook)
Content	Information/data conveyed to end users	Music files, playlists
Applications	Programs and functions used by end-users	Desktop and mobile media players
Logical Infrastructure	Standards and protocols that facilitate transmission of data across physical networks	Content delivery networks, event delivery systems
Physical Infrastructure	Physical hardware that comprises interconnected networks	Cable and satellite networks, data centres, routers, and servers

Source: Adapted from Frischmann (2012, p. 320).

"gone wrong" in some way and a huge number of critical accounts have appeared, in academic and trade publishing, in recent years, dealing with both infrastructure and platforms. In some political economy versions analysis of digital platforms, this is sometimes reduced to questions of ownership, to control by massive tech corporations, often with excessive focus on the famous GAFAM (Google, Apple, Facebook, Amazon, and Microsoft) oligopoly, or on the dependence of platforms on the collection of data (e.g., Smyrnaios, 2018; Srnicek, 2016). Ownership and datafication are extremely important, involving vital problems of democratic control and power, and such accounts do attempt to analyse how digital platforms are embedded in a broader context of the evolution of capitalism in the digital age. But they do not in themselves address how the infrastructural open-ness or generativity of the internet, just outlined, has been affected by platforms, nor do they make any sustained attempt to theorise the relationship of digital platforms to media or internet infrastructures. Indeed, many of these accounts, including some of those characterisable as political economy (such as Srnicek and Smyrnaios) often resort to using the term "infrastructure" in the vague metaphorical way that we critiqued earlier. Much more helpful in understanding the dynamics of media infrastructure is research from political economy of media on telecommunications, including that already mentioned above (Schiller, 2000; and Winseck, 2017, also demonstrates the problematic focus on the GAFAM oligopoly in some of the other political economy accounts just mentioned). Also worth noting is work from critical geography media infrastructures on how telecommunications infrastructures such as broadband and mobile telephony powerfully delineate space (Easterling, 2016). But none of these resources conceptualises the role of digital platforms in closing down the architectural principles of open-ness and generativity that underlay the early internet (though Srnicek, 2016, pp. 110–112, makes brief reference to Amazon and Google as "closed platforms"). One way of explaining this move is that platforms "often just work better or fit better" into the lives of consumers ("the screen comes to them, they don't have to go to the screen," in the words of a World Economic Forum publication [Drake et al., 2016, p. 53]). But we still need an understanding of how platforms were able to offer this convenience to consumers, and what was lost in the process. Cohen's (2019, p. 41) vision of platforms as "strategies for bounding networks and privatizing and disciplining infrastructures" opens the door to a greater engagement by media research with such questions, which we now pursue in the section that follows.

4. The Platformisation of Music as a Closing Down of Infrastructural Possibilities

Cohen's macro-historical and densely theorised account does not examine the implications of the enclosure and commodification she identifies for particular domains of culture and communication, at least not in any detail. How then might a version of such an account of infrastructure, embedded within a political economy of informational capitalism, and centred on the three shifts identified by Cohen (which we might cautiously summarise as propertisation, datafication, and platformisation) be developed and applied to media and culture, here using the specific case study of music?

What makes the case of recorded music a particularly revealing one is the stark historical contrast it presents between the current situation and a not-so-distant past when huge amounts of musical activity were afforded by digital architecture organised according to a very different logic, and where music acted as a test case for radical changes that might take place in other culture and information sectors. Music had this role foisted on it because it does not take up much bandwidth, at least compared with video, and also because CDs already contained huge amounts of "unprotected" music which could be shared online, once (easily) compressed into the highly portable MP3 software format. This made music technology a site for experiments in how to use the internet as the basis for new ways of exchanging information and entertainment, based on the infrastructural features and principles outlined in the previous section, including permitting commons-based use of computing resources.

One notable such experiment was the use of peer-to-peer computing, based on the principle that each node in the network is an equal peer which simultaneously functions as both a client and a server. The famous Napster website was not actually based on peer-to-peer but it offered easy search via a display of the files currently available from the computers of logged-on users. This made it popular but its centralised nature meant that courts held it responsible for not preventing infringement. This did not apply to peer-to-peer sharing protocols such as BitTorrent, and these were a much greater threat than Napster. The recorded music industry took action to protect its copyrights against these various technologies by criminalising their use, including the prosecution of ordinary file-sharing users (David, 2010). But the bad publicity generated by such developments meant that rights-holders shifted to different solutions.

Let us now apply a version of Cohen's triad of propertisation, datafication, and platformisation to the case of online music. Propertisation involved a mix of legal and technological means. Following intense lobbying by rights-holders, courts, especially in the US, issued judgements that predictably affirmed these owners' views about intellectual "property" (a term that had by the late twentieth century become naturalised as a way of thinking about cultural products, itself the culmination of a long ideological battle by cultural-industry businesses, going back decades). Meanwhile, record companies introduced means of preventing the kinds of circulation that internet infrastructure permitted, through the introduction of encryption software that sought to control

reproduction and sharing, at least for ordinary users, in the form of digital rights management controls. Publicity about prosecution and "illegality," along with growing problems of spam in the chaos of the open web, helped to create a climate of anxiety among more nervous, more dutiful, and often older groups of citizens, who sought out "safer," more secure, and more seamless online musical experiences, though digital rights management proved clunky and unpopular with consumers.

An even more significant technological development, in terms of the role of propertisation in closing down the generative potential of internet architecture, was the introduction of what at the time were called "trusted systems," which offered copyright owners greater and more precise control over their products (Gillespie, 2007). By contrast with the relatively open protocols associated with connected PCs, these often involve what Zittrain (2008, p. 101) called "tethered appliances," more centrally controlled devices such as mobile phones and games consoles, which no one can tinker with. As Tarleton Gillespie (2007) showed in an excellent account seemingly overlooked by nearly all music scholars and more recently by media and internet researchers, the reason behind the word "trusted" is that software rules protecting files from copying and sharing are built into devices and software systems built around particular devices. Such "black box" devices do more than respond to infringement, they allow for "an incredibly subtle and parsing of the use of information so as to be sold" (Gillespie, 2007, p. 55).

"Trusted systems" in the realm of music paved the way for what soon came to be known as "datafication," in this context meaning the ability to extract profit from the collection and analysis of data. The enormous value of data had been made clear via a whole set of separate developments, including the discovery by Google of the vast profits to be made from what Zuboff (2019, p. 81) has called the "behavioural surplus," the accumulation of vast behavioural data beyond that needed to improve services, instead oriented towards predictions for accurate digital targeting, especially via search and recommendation algorithms. In music, Apple's iTunes system began to provide personalised experiences of recommendation in 2003 and operated as a trusted system, far removed from peer-to-peer. Even more significant was the iPhone (Apple do not sell data to third parties but experimented with personalisation technologies).

Crucially, these tethered devices were increasingly linked to websites operating on similar principles (Zittrain, 2008, pp. 101–106), which eventually came to be known as digital platforms. In music, the innovation of firms such as Spotify and Deezer was to build trusted system architectures on top of internet infrastructure, making it impossible for all but the most sophisticated users to share or tinker, rendering them hardly platforms at all on the user side (cf. Sandvig, 2015). At the same time, they developed business models based on either advertising (already proven to be successful and lucrative in the form of YouTube) or subscription (Rhapsody and the "legal" reincarnation of Napster both showed the potential of this model). Once a tipping point of security for content providers and seamlessness for users was achieved, most notably by Spotify, rights-holders (mainly the major multinational record companies, and their "publishing" arms, i.e., those dealing with song rights) began to license their content to streaming platforms—though at a premium.

The platform model, based on "technical protocols and centralized control to define networked spaces in which users can conduct a heterogeneous array of activities and to structure that space for ease of use," (Cohen, 2019, p. 41) eventually produced substantial revenues (though rarely big profits) for companies and products able to create and exploit "first mover" (or first winner) advantages, across a range of sectors. Vast amounts of investment were necessary to build platform infrastructures on top of the internet, but as investor confidence gradually started to build, financing poured in. Music technology companies became able to attract very high levels of financial investment, nearly a billion dollars between 2011 and 2013 (Mulligan, 2015). Just three music technology companies (Spotify, Deezer, and Beats) accounted for 70% of this sum. All operated on the platform model.

Concurrently, the massive data needs of the platform model required enormous computing power, and the solution that emerged was outsourced "cloud" storage and retrieval. While listening to a track on Spotify always involves musical files being transferred from remote servers to a personal computer, back in the early 2000s, music consumption through BitTorrent only involved file transfer between individual personal computers, without the need for the centralised infrastructure system that underlies the Amazon and Google cloud empires that serve streaming platforms.

The result is a musical ecosystem that now essentially consists of two parallel oligopolies: music platforms owned and controlled by technology companies (with Spotify, Apple, Google, and Amazon dominant across much of the world, and Tencent in China) and a recording sector with corporate rights owners scarcely less profitable and dominant than before the internet. While musicians can now try to make their music available to global audiences without passing through record companies, they are unlikely to be heard or paid much at all without them. Certainly, successful musicians stood to have their share of earnings from rights eroded by the chaos unleashed by internet generativity. But new payment systems would surely have emerged and perhaps under different terms than those in which the major-dominated recording and publishing industries prevail. The threat temporarily posed by the "open" and "generative" architecture of the internet was well and truly contained, and while this "enclosure" preceded platforms, it was only with platformisation, underpinned by propertisation and datafication, that it was more or less fully realised.

5. Conclusions

The perspective developed here is intended to illuminate the politics of information infrastructures in terms of public culture and media, here demonstrated by the case of music. This, we would argue, is a particularly illuminating case study, firstly because it was in the realm of online music that the infrastructural potential of "open" or "generative" architecture was most fully apparent, and secondly, because music represents a remarkably successful and rapid instance of platformisation, across much of the world. But beyond that, our perspective points us to the normative implications of the incorporation of music into information capitalism that the legal studies perspective helps to illuminate. This has involved networks becoming mainly platforms, supported by infrastructure designed to ensure security and seamlessness, rather than the generativity, interactivity, and open-ness envisaged by an earlier generation of internet enthusiasts. While some of the predictions of democratisation were naïve and even silly, they recognised something extraordinary about internet infrastructure, which has in many respects been lost. Instead, platformisation has allowed something more than just oligopolisation of ownership. It has enabled the incorporation of music into what Durham and Born (2022) have called a "rentier" model of musical exchange. Rent here means something different from the rent paid to a landlord by a tenant; it refers to a specialist economic sense that has been defined in a mass of sometimes contradictory ways. At the heart of this use of the term "rent," as Brett Christophers (2020, p. xvi) shows in his book *Rentier Capitalism*, is economic actors receiving rewards "purely by virtue of controlling something valuable." In this case, it involves not only intellectual property assets, long central to music, but also platforms and infrastructures, the latter a resource organised not on a generative and open basis, but as assets to be milked by providing a service that only a very small number of corporations can afford to offer.

How does infrastructure influence or shape culture, if at all? In this article, on the basis of a case study of music, we have sought to demonstrate that one important way to consider infrastructure as part of an account of how culture is shaped and influenced is to examine developments in infrastructural politics over a relatively long duration, as part of a macro-historical account of change and continuity. Specifically, we use the case of music to show how platforms have operated as the main means by which the democratising and emancipatory possibilities afforded by the (always partial) commons-based open-ness of internet infrastructure were eroded or "closed down." What seems strange, in terms of academic research, is that, in spite of the popularity of terms such as platforms and infrastructure, this erosion has hardly been recognised, let alone analysed, in recent media and communications research on those topics.

Acknowledgments

This article was produced as part of a research project—Music Culture in the Age of Streaming—which has received funding from the European Research Council, under the European Union's Horizon 2020 Research and Innovation Programme, in the form of an Advanced Research Grant awarded to Professor David Hesmondhalgh, at the University of Leeds (Grant agreement no. 1010020615). The authors would like to thank Jonas Andersson Schwarz for his comments and suggestions on an earlier version of the article.

Conflict of Interests

The authors declare no conflict of interests.

References

Andersson Schwarz, J. (2013). *Online file sharing*. Routledge.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out*. MIT Press.

Burrell, J. (2018). Thinking relationally about digital inequality in rural regions of the U.S. *First Monday*, *23*(6). https://doi.org/10.5210/fm.v23i6.8376

Cath, C. (2021). The technology we choose to create: Human rights advocacy in the Internet Engineering Task Force. *Telecommunications Policy*, *45*(6), Article 102144. https://doi.org/10.1016/j.telpol.2021.102144

Christophers, B. (2020). *Rentier capitalism*. Verso.

Cohen, J. E. (2019). *Between truth and power.* Oxford University Press.

David, M. (2010). *Peer to peer and the music industry*. SAGE.

Drake, W., Cerf, V., & Kleinwachter, W. (2016). *Internet fragmentation*. World Economic Forum.

Durham, B., & Born, G. (2022). Online music consumption and the formalisation of informality. In G. Born (Ed.), *Music and digital media* (pp. 171–219). UCL Press.

Easterling, K. (2016). *Extrastatecraft: The power of infrastructure space*. Verso.

Edwards, P., Bowker, G., Jackson, S., & Williams, R. (2009). Introduction: An agenda for infrastructure studies. *Journal of the Association for Information Systems*, *10*(5), 363–374. https://doi.org/10.17705/1jais.00200

Elkins, E. (2018). Powered by Netflix: Speed test services and video-on-demand's global development projects. *Media, Culture & Society*, *40*(6), 838–855. https://doi.org/10.1177/0163443718754649

Eriksson, M., Fleischer, R., Johansson, A., Snickars, P., & Vonderau, P. (2019). *Spotify teardown*. MIT Press.

Frischmann, B. M. (2012). *Infrastructure*. Oxford University Press.

Gillespie, T. (2007). *Wired shut*. MIT Press.

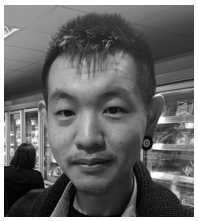
Gillespie, T. (2018). *Custodians of the internet*. Yale University Press.
Hesmondhalgh, D. (2022). The infrastructural turn in media and internet research. In P. McDonald (Ed.), *The Routledge companion to media industries* (pp. 132–142). Routledge.
Holt, J., & Vonderau, P. (2015). "Where the internet lives": Data centers as cloud infrastructure. In L. Parks & N. Starosielski (Eds.), *Signal traffic* (pp. 71–93). University of Illinois Press.
Horten, M. (2016). *The closing of the net*. Polity Press.
Infrastructure. (n.d.). In *Oxford learners' dictionaries*. https://www.oxfordlearnersdictionaries.com/definition/english/infrastructure
Is an infrastructure boom in the works? (2021, January 2). *The Economist*. https://www.economist.com/finance-and-economics/2021/01/02/is-an-infrastructure-boom-in-the-works
Johnson, C. (2019). *Online television*. Routledge.
Larkin, B. (2008). *Signal and noise*. Duke University Press.
Lessig, L. (2004). *Free culture.* Penguin.
Lobato, R. (2019). *Netflix nations*. NYU Press.
Magaudda, P. (2020). Music scenes as infrastructures: From live venues to algorithmic data. In T. Tofalvy & E. Barna (Eds.), *Popular music, technology, and the changing media ecosystem* (pp. 23–41). Springer.
Magaudda, P. (2021). Smartphones, streaming platforms, and the infrastructuring of digital music practices. In A. Hennion & C. Levaux (Eds.), *Rethinking music through science and technology studies* (pp. 241–255). Routledge.
Mayer, V. (2019). The second coming: Google and internet infrastructure. *Culture Machine*, *18*, 1–12. https://culturemachine.net/vol-18-the-nature-of-data-centers/the-second-coming
Morris, J. W. (2015). *Selling digital music, formatting culture*. University of California Press.
Morris, J. W. (2021). Infrastructures of discovery: Examining podcast ratings and rankings. *Cultural Studies*, *35*(4/5), 728–749.
Mulligan, M. (2015). *Awakening: The music industry in the digital age*. MIDiA Research.
Musiani, F., Cogburn, D. L., DeNardis, L., & Levinson, N. S. (Eds.). (2016). *The turn to infrastructure in internet governance*. Palgrave Macmillan.
Parks, L. (2015a). "Stuff you can kick": Toward a theory of media infrastructures. In P. Svensson & D. T. Goldberg (Eds.), *Between humanities and the digital*. MIT Press.
Parks, L. (2015b). Water, energy, access: Materializing the internet in rural Zambia. In L. Parks & N. Starosielski (Eds.), *Signal traffic* (pp. 115–136). University of Illinois Press.
Plantin, J. C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, *20*(1), 293–310. https://doi.org/10.1177/1461444816661553
Poell, T., Nieborg, D. B., & Duffy, B. E. (2022). *Platforms and cultural production*. Polity.
Russell, A. (2014). *Open standards and the digital age*. Cambridge University Press.
Sadowski, J. (2020). The internet of landlords: Digital platforms and new mechanisms of rentier capitalism. *Antipode*, *52*(2), 562–580.
Sandvig, C. (2015). The internet as the anti-television: Distribution infrastructure as culture and power. In L. Parks & N. Starosielski (Eds.), *Signal traffic* (pp. 225–245). University of Illinois Press.
Schiller, D. (2000). *Digital capitalism.* MIT Press.
Smyrnaios, N. (2018). *Internet oligopoly*. Emerald.
Srnicek, N. (2016). *Platform capitalism*. Polity.
Starosielski, N. (2015). *The undersea network*. Duke University Press.
Sun, H. (2019). *Digital revolution tamed*. Palgrave Macmillan.
ten Oever, N. (2021). "This is not how we imagined it": Technological affordances, economic drivers, and the internet architecture imaginary. *New Media & Society*, *23*(2), 344–362.
Vaidhyanathan, S. (2001). *Copyrights and copywrongs*. NYU Press.
van Schewick, B. (2010). *Internet architecture and innovation*. MIT Press.
Winseck, D. (2017). The geopolitical economy of the global internet infrastructure. *Journal of Information Policy*, *7*(2017), 228–267.
Zittrain, J. (2008). *The future of the internet and how to stop it*. Yale University Press.
Zuboff, S. (2019). *The age of surveillance capitalism*. Penguin.

About the Authors

David Hesmondhalgh (he/him) is professor of media, music, and culture in the School of Media and Communication at the University of Leeds. He is the author of *The Cultural Industries* (4th ed., 2019) and *Why Music Matters* (2013) and co-author of *Culture, Economy and Politics: The Case of New Labour* (2015), *Creative Labour: Media Work in Three Cultural Industries* (2011), and of the book-length report *Music Creators' Earnings in the Digital Age* (2021). He is the principal investigator of the ERC-funded research project *Music Culture in the Age of Streaming* (MUSICSTREAM, 2021–2026).

Raquel Campos Valverde (she/her) is a postdoctoral research fellow at the University of Leeds, working on the project MUSICSTREAM (2022–2026) where her research focuses on users, experiences, emotions, and pleasures. Before joining the project, she was a lecturer in digital culture and society at King's College London. Her first article "Online musicking for humanity: The role of imagined listening and the moral economies of music sharing on social media" was published in the journal *Popular Music*.

D. Bondy Valdovinos Kaye (he/they) is a postdoctoral research fellow at the University of Leeds, working on MUSICSTREAM (2022–2026). Kaye is the author of *TikTok: Creativity and Culture in Short Video* (2022, co-written with Jing Zeng and Patrik Wikström). Kaye has published eleven peer-reviewed journal articles on topics including short video platforms and patriotism in China, media censorship in India, and copyright enforcement on YouTube.



Zhongwei Li (he/him) is a postdoctoral research fellow at the University of Leeds. His research explores the ruptures and continuities between the contemporary moment of music streaming and music cultures in the past thirty years. His work with the MUSICSTREAM project (2022–2026) contextualises these changes within broader social, political, and economic transformations on China's path to global capitalist modernity. Before joining MUSICSTREAM, he worked as an LSE fellow in media and communications.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 307–318
https://doi.org/10.17645/mac.v11i2.6388

Article

The Infrastructure of News: Negotiating Infrastructural Capture and Autonomy in Data-Driven News Distribution

Lisa Merete Kristensen * and Jannie Møller Hartley

Department of Communication and Arts, Roskilde University, Denmark

* Corresponding author (lisak@ruc.dk)

Submitted: 30 October 2022 | Accepted: 23 March 2023 | Published: 28 June 2023

Abstract
The platformisation of news has triggered public and scholarly concern regarding the impact of platforms on the news industry and, more importantly, platforms' potential threat to ideals of autonomy and economic independence. Despite ongoing debate and the increasing investment in technologies for automated distribution and artificial intelligence, the material infrastructures of the news media sustaining this artificial intelligence-driven news distribution remain understudied. Approaching the infrastructural relationship as spaces of negotiation this article investigates how the news media is negotiating their own autonomy vis-à-vis infrastructure capture by platforms. The analysis is grounded in a mapping of technologies sustaining the production, distribution, and commercial viability of the media. This is further combined with ethnographic observations from two large Danish news organisations and 19 in-depth interviews with news organisations and digital intermediaries from Scandinavia, the US, and the UK. The research shows how infrastructure capture is manifested and negotiated through three overall logics in the infrastructure of news: logics of classification, standardisation, and datafication.

Keywords
datafication; digital media infrastructures; infrastructure capture; media logics; platformisation

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

At a meeting in the spring of 2022, directors of several departments and two of the key developers at the Danish tabloid *Ekstra Bladet* discussed various taxonomies for describing and categorising news content, for example by topic. They discussed the pros and cons of the industry-standard taxonomy developed for digital marketing by the Interactive Advertising Bureau (IAB) and the taxonomy of the International Press Telecommunications Council (IPTC). Among the more than 700 members of the IAB are Microsoft, Amazon, Nielsen, Spotify, Yahoo, and Twitter, while the IPTC standard is developed for media companies. The editors agreed that some 80% of the categories and subcategories were usable but that new categories had to be added, as the taxonomy seemed "overly commercial." Hence, they embarked on the task of adjusting this taxonomy to their own context. They coded thousands of articles, removed categories, added their own, and eventually negotiated an adjusted taxonomy, which was a combination of categories from the IAB, IPTC, and their own categories adjusted to the needs of *Ekstra Bladet* and the specific context of news.

This scene from our fieldwork took place almost 50 years since Tuchman (1973) convincingly showed how journalists categorise the news in order to "routinise the unexpected" in everyday news production. In her seminal sociological study of newsrooms, she showed how news stories were categorised and hierarchised and how

a given news flow (and thus reality, she argued) is socially constructed in journalistic practice. Today, news are still categorised according to internal journalistic criteria and economic news values, but they are also categorised to allow the application of performance metrics and to ensure distribution to increasingly personalised digital news sites, search engines, and social media platforms. Thus, the opening example from our fieldwork illustrates how the news media link themselves to larger infrastructures and thereby adapt to certain logics of platformisation, thus negotiating their own autonomy, norms, and values in the process.

It is no understatement that in the past 15 years, we have seen an intensified datafication of news industries. Most significantly, the distribution of news has experienced a radical change, as the communicative system and the infrastructural conditions of distribution have moved from being operated by media companies themselves, as was the case with the printing press, or by states, such as with much of telecom and postal infrastructures in the Western context (Flensburg, 2020). Due to its complexity and ability to transfer data on a global scale, much of the material infrastructure is owned by large technology companies, resulting in what van Dijck et al. (2018) termed "platform societies." Research has begun to examine how this development affects other spheres of society, for example, by showing how the news media adapt to these logics of datafication by increasingly basing decision-making practices on the algorithmic processing of audience and user data (Christin, 2020; Kristensen, 2021; Petre, 2021).

In this article, our focus is on how these technologies become deeply ingrained into the organisational structure of the news organisation. We argue that it is important to examine the material basis of news production, scrutinising the interdependencies between news media and infrastructures to understand how new logics are entering the processes of media production and distribution (Simon, 2022, p. 1833). As such, they are not simply value-free plug-and-play packets of software but actants with purposes and values built-in (Friedman & Nissenbaum, 1996; Thurman, 2011). They exist in what Poell et al. (2022) call "spaces of negotiation," which means that news media to a varying degree adapt to the inherent logics and audience constructions and make them fit their own values and norms, for example how they perceive the audiences. These "fittings" are important because they also make the values and norms durable, as they become part of the technical systems.

The article first positions our research question in the existing literature, arguing that we need to look closer at how the materialities and technologies of news distribution are implemented, but also negotiated along the way. Next, we present the conceptual-theoretical framework of infrastructure capture (Nechushtai, 2018) and media logics (Altheide & Snow, 1979), which leads to the formulation of our research question. This is followed by a methods section. The first part of the analysis maps the infrastructural elements of news production, news distribution, and commercial viability of news. The second explores how media organisations negotiate power over dominant logics by designing their tech stacks. Building on this, we argue that infrastructure capture is negotiated and manifested through three overall logics: logic of datafication, standardisation, and classification.

2. Literature on the Infrastructures of News Distribution

In recent years, scholars have theorized and examined the increasing dependency between news organisations and the infrastructures supplied by commercial platforms, a trend which has resulted in the "platformisation of the news" (van Dijck et al., 2018, p. 49). Drawing on software studies, political economy, and business studies, Poell et al. (2022, p. 5) argued that platforms can be understood as "data infrastructures that facilitate, aggregate, monetize, and govern interactions between end-users and content and service providers." This definition illustrates that platforms simultaneously operate as multi-sided markets, data infrastructures, and governance frameworks.

Several scholars have addressed how various systems sustaining news production and distribution influence the work of journalists. For example, studies of how audience measurement data impact editorial choices (Anderson, 2011). Furthermore, an increasing number of studies are examining the use of recommender systems by news organisations. This research emphasises that news organisations vary regarding the use of recommenders (Møller, 2022) and contradict public service values, such as universalism (Sørensen, 2022). In addition, how they impact diversity (Neyland & Möllers, 2017) and can be designed to support democratic values (Helberger, 2019). We argue that the literature has to some degree overlooked the material aspects of both metrification and personalisation, in that it involves building complect tech stacks and tech systems inside media organisations. By taking an infrastructure approach, we contribute with new knowledge on how decisions concerning the implementation of these systems are negotiated in a news organisation domain, questioning more broadly how the autonomy of news organisations is negotiated in the implementation of systems for production, distribution, and monetisation. Poell et al. (2022) theorised the relationship between news media and the providers of tech solutions as a space of negotiation and argued that the relationship between platforms and news media is not one-sided, as news organisations also adopt the platformisation that they encounter. We find this valuable for examining how news organisations approach the development of tech systems and AI-driven distribution differently, depending on their size and type of news organisation.

We take a material rather than a relational approach to the study of infrastructure, in line with what has been

called for by Flensburg (2020) and Flyverbom and Murray (2018). This entails focusing on the interplay between the technologies and organisational cultures to understand how the technologies shape the institutions and practices they sustain. This approach also helps us understand that the infrastructures are "stacked" via a large number of smaller systems or tech stacks, as they are referred to in the industry. This means including everything from the deepest levels of hardware (e.g., data storage) to the more dynamic layers of software development. Hence, the development of AI within news organisations requires us to analytically go beyond observing relationships between publishers and "traditional" social media platforms to include emerging data, code, and model-sharing platforms such as Github, PyTorch, and HuggingFace. Of interest to this study is work that examines the emergence and implications of cloud infrastructures as preconditions for platformisation. Narayan (2022), for example, provided an analysis of the platformisation of computing assets in which she examined how platform infrastructures expand through cloud infrastructures. She referred to this tendency as "radical outsourcing" and pointed out that very little is still known about cloud providers and their practices of expansion through these outsourcing processes (Narayan, 2022, p. 916). From a social perspective, these infrastructures also give rise to new practices. Such studies show how the development of AI analytics is financed through creative practices of reusing data, codes, and models from one context and fitting them into a different context, as well as how these creative practices involve new conditions of infrastructural dependency and vulnerability because of the risk of infrastructural lock-in, infrastructural decay and new licence models (Thylstrup et al., 2022). The present article contributes to these studies by expanding knowledge about how infrastructural development and platformisation processes unfold in the field of news and the social practices they engender.

3. Theoretical Framework: Media Logics and Infrastructure Capture

Infrastructure, in crude terms, refers to an "underlying foundation or basic framework" (Infrastructure, n.d.). In our research, we zoom in on the media backend as an infrastructure of the individual media organisation and its relation to the larger infrastructure of the internet and platforms (Plantin et al., 2018). Although we fully recognise the importance of tangible large-scale infrastructures, such as undersea cables, this study limits its empirical scope to focus on the media backend, an infrastructural micro-perspective one might say. Thus, we position ourselves in previous research that exemplifies infrastructures as "software, data, and technologies from outside newsrooms" (Ananny & Finn, 2020, p. 1600), "search engines and related systems" (Feuz et al., 2011, para. 13), or "protocols (human and computer), standards, and memory" (Bowker et al., 2009, p. 97). Infrastructures are often defined in terms of their affordances and characteristics (Flanagan et al., 2008; Star & Bowker, 2002). They are built on top of previously installed infrastructures; thus, it can seem that we are dealing with a patched system with infinite versions (Star & Ruhleder, 1996). A functioning infrastructure requires standardisation across systems and former versions of systems. This also means that elements of infrastructure are embedded in—and therefore cannot be viewed as separated from—the values of former and current structures. Lastly, following Star and Ruhleder (1996), we view infrastructures as shaped by conventions of a community of practice, but simultaneously, they shape practice.

To connect the infrastructural focus to our interest in media, a conceptual lens is provided by the concept of "media capture" and, more specifically, Nechushtai's (2018) concept of "infrastructure capture." This notion refers to "circumstances in which a scrutinising body is incapable of operating sustainably without the physical or digital resources and services provided by the businesses it oversees and is therefore dependent on them" (Nechushtai, 2018, p. 1043). The capture can be both material and non-material, with the first referring to instances in which a regulator is benefitting financially from the industry it is overseeing (Nechushtai, 2018, p. 1046). Non-material forms are cultural and cognitive capture, which refer to capture through formal channels, for example, public relations efforts, and capture through informal relations, for example, personal relationships (Nechushtai, 2018, pp. 1046–1047). Simon (2022) demonstrated the usefulness of Nechushtai's concept of infrastructure capture in his analysis of how AI technologies are increasingly permeating the phases of journalistic gatekeeping. Simon argued that the capture and potential loss of control and media autonomy occur at different paces in the news industry and in news production (Simon, 2022, p. 1843). Eventually, news organisations risk adopting the logics of the external platforms and actors that are sustaining news production and distribution while simultaneously being competitors in seeking the attention of users. Autonomy is one of the key dimensions upholding the journalistic profession and refers in this article to the ability of the media and journalists to carry out professional routines without being influenced or having obligations to external actors, here the providers of infrastructure (Singer, 2007). In line with Simon (2022, p. 1833), we consider infrastructure capture and autonomy to be on opposite sides of a theoretical continuum which delineates the level of dependence between media and platforms. We acknowledge that our mapping does not give us an exact answer as to whether media are "captured" by infrastructure. Instead, our mapping and subsequent analysis make use of the concepts to illustrate the negotiations happening within news organisations regarding infrastructure.

To operationalise infrastructure capture, we apply an adaptation of the theory of media logic first introduced by Altheide and Snow (1979). Media logic

theory concerns itself with the "assumptions and processes for constructing messages within a particular medium" (Altheide, 2016, p. 1). Altheide and Snow (1979) employed the term "logic" in the singular, but as Thimm et al. (2018, p. 3) noted, today's networked media landscape is far more complex than in the mass media tradition from which Altheide and Snow departed. As such, several logics have been proposed in later years to account for the changes in media technologies and conditions (Couldry, 2008; Klinger & Svensson, 2018; van Dijck et al., 2018). In this study, we align with Klinger and Svensson (2018, p. 1244), who argued that "media logics as specific norms, rules and processes both influence and are influenced by the involved actor." Journalistic logics thus influence and are influenced by multiple competing logics. Extending this thinking, our goal is to investigate which logics are at play in infrastructuring as news media implement and develop systems in their tech stacks. Following this, the study aims to answer the following research question: How are the autonomy and infrastructure captured vis-à-vis external tech providers negotiated in the process of implementing and developing tech systems as infrastructure for the production and distribution of news?

4. Methodology

This research rests on a combination of interviews, fieldwork, and desk research. The first analytical part presents a mapping of the backend systems of several media organisations. This is based on a policy and document analysis, combined with our interviews, fieldwork, the StackShare website (https://stackshare.io), and searching the web for software solutions marketing themselves for the media industry. We also attended industry conferences, WebSummit in 2021 and TechSummit in 2021, and participated in industry networks such as the Nordic AI Network, where news media collaborate and exchange ideas on how to implement various tech systems and put together their "tech stacks," meaning the composition of systems on which news sites, news work, and news distribution are built. Methodologically, the aim of this study is not to provide a full picture of the extent to which these systems are used by different organisations, but the methods allow for an overview of the vast types of systems in all infrastructural corners of the news organisations, including production, distribution, and the commercial part of the news industries.

The mapping and the subsequent analysis are also based on in-depth interviews with 13 European and US-based publishers and intermediaries, an analysis of press releases and software documentation from system providers, and ethnographic observations in the development departments of two large Danish news organisations. We selected these two news organisations based on their publicly announced aim of developing independent data infrastructure platforms and personalised recommender systems. For the interviews, we included US- and UK-based media to assess potential similarities and links in the deployed backend infrastructures in our mapping, and they provided us with a backend understanding helpful for choosing cases for the focus points in the ethnographic observations. The initial fieldwork took place at Jysk Fynske Medier (JFM) from May 2019 to May 2021. JFM has around 1,850 employees and covers parts of North Zealand, all of Fynen, and most of Jutland. It has 15 regional subscription newspapers and 63 local free weeklies. The second fieldwork phase took place at *Ekstra Bladet* in JP/Politikens Hus from February to November 2022. *Ekstra Bladet* is a national newspaper in tabloid format and is one of the most read in its online version. It has around 300 employees, but a total of 2,100 people are employed at JP/Politikens Hus. The observations focused on the development departments and analytics departments and the managerial level to understand which and how systems were chosen, developed, and implemented. For this reason, we focused less on the newsrooms of the two organisations, although the journalists were implicitly present in both the observations and interviews as the "users" of many of the systems implemented during this period. We participated in meetings on project management as well as on everyday work two to three times per week during the observation period.

The interviews were conducted from May 2020 to November 2021 (see Table 1). The news organisations were selected following desk research on which news organisations were and are experimenting with AI in various forms and designs of, for example, recommender systems or in-house metrics and data analytics tools. The system providers were chosen because of their services being aimed at and employed by media companies. The interviews lasted from 40 to 60 minutes and were transcribed shortly after and analysed thematically using the NVivo software package (Braun & Clarke, 2006).

5. Analysis

5.1. Mapping the Backend of News Organisations

In Table 2, we present our mapping of the infrastructural systems sustaining the media. We categorised the systems and services into three levels: (a) production and publishing technologies, (b) distribution technologies, and (c) technologies that sustain the commercial viability of media (monetisation). Publishing technologies refer to the systems that form the basis of the workings of the news site. Production technologies refer to the systems used by journalists in their production processes. A characteristic of these is that they have a user interface, for example, typing in article text in a content management system, choosing photos from a photo library, and using audience measurement systems. There are multiple ways of reaching the audience and we know from previous research that users access content on social

Table 1. Informants.

Identifier	Organisation	Country	Position/Field
1	TV2	Denmark	Journalist and developer
2	Chartbeat	US	Account executive
3	Amedia	Norway	Head of digital development
4	Amedia	Norway	Head of engineering
5	*Ekstra Bladet*	Denmark	Head of research and innovation
6	JFM	Denmark	Editor of data and analysis
7	Mitt Media	Sweden	Data scientist and project manager
8	*The Guardian*	UK	Director of engineering
9	Midtjyske Media	Denmark	Head of digital development
10	*The New York Times*	US	Data science lead
11	*Information*	Denmark	Head of digital development
12	*Altinget*	Denmark	Editor of IT and development
13	*Ekstra Bladet*	Denmark	Data science developer
14	*Ekstra Bladet*	Denmark	Data science developer
15	JFM	Denmark	Digital editor
16	JFM	Denmark	Developer
17	JFM	Denmark	Developer
18	Google	Spain	Partner manager
19	Infomedia	Denmark	Head of data science

media, via search engines, newsletters, and, of course, via the media website itself. These are categorised as distribution technologies. In our mapping, we included technologies that enable monetisation. They may not be directly involved in journalistic practice, but they are key points of exchange of data and standards between systems used in production, distribution, and advertising.

As is the case when dealing with infrastructures, individual technologies sometimes cross categories. For example, content management systems often "solve" several tasks, and other systems are embedded into them. In addition, some technologies, such as cloud services, are foundational for all other systems to run. These deeply rooted infrastructural interdependencies are categorised here as "production and publishing technologies" for simplification purposes.

First, we find it striking that there are so many systems involved on various levels of the news media, which indicates a high level of infrastructuring via tech systems overall. While the printed newspaper also had to be printed and delivered, the infrastructural systems today are increasingly complex and involve many more actors and providers of such services and systems. Interestingly, the mapping further highlights that platform companies are present in all three categories of infrastructural technologies sustaining the media. The representation of Google products is especially striking, suggesting that infrastructure capture can take place on multiple levels. Zooming in on a case from each of the levels of tech systems in the following second part of our analysis allows us to show how different infrastructural logics are at play in the process of implementing, highlighting logics of classification, standardisation, and datafication.

5.1.1. Infrastructure Capture Through Classification Logics

In Section 1, we presented the Danish tabloid *Ekstra Bladet*. Here, the team of developers experimented with large language models, which are machine-learning algorithms that can recognise, predict, and generate human languages on the basis of very large text-based data sets. Automated textual analysis is particularly useful for the implementation of recommender systems, as well as for pairing certain forms of content with advertisers or for coupling articles with supplementary relevant information from the internet. The development usually involves several steps, the end goal being to automatically analyse articles, content, or pictures and to categorise them so that they can be paired with the interests of the specific user and this user's history and profile. By automatically creating ways for content to flow through the systems, the news media link themselves to larger infrastructures of data. Hence, the purpose of the standardised categories is that they allow for integration with, for example, search engines, whose web crawlers require standardised data categories to "understand," store, and subsequently make news content visible:

> Our ranking systems for news content across Google and YouTube News use the same web crawling and indexing technology as Google Search to continually

Table 2. Technologies sustaining news media.

Categories of technologies	Facilitation in practice	Technologies	Commonly used service providers
Monetisation technologies	Advertising; handling of subscriptions, log-ins, customer profiles, and customer engagement; audience insights and business intelligence; strategy/resource allocation	Ad exchange server, customer data platform, audience measurement (for marketing purposes), customer relationship management, data warehouse	Google ads, Oracle responses, Twilio, Tealium, Adform, Google Analytics, BigQuery (Google), ElasticSearch
Production and publishing technologies	Content storage and content delivery	Server/database and content delivery network	Amazon DynamoDB, Fastly, AWS, CloudFront (Amazon), Cloudflare, Akamai, Microsoft Azure, Firebase
	Producing and publishing news to the news site and/or news app	Content management system/publishing platform	Stibo CUE, Wordpress VIP, Sitecore, Drupal
	Adding storytelling elements and data to news articles	Storytelling and visualisation tools	Infogram.com, Google Fusion Tables, Tableau Public
	Testing different versions of headlines on news articles	A/B testing	Chartbeat, Optimizely, Google Analytics
	News selection and prioritisation of day-to-day editorial resources	Audience measurement system real time and aggregated over time (for editorial purposes)	Google Analytics 360 (Realtime Content Insights), Chartbeat, Parse.ly, ComScore, Gemius, Moat, Facebook Insights
	Content organising, analysis for automation, and tagging	Transformer models (NLP) and topic/language modeling	Think Analytics, Cxense, Contentwise, GDP 1, 2, 3 (OpenAi/Microsoft), Google (BERT), Facebook (XLM Roberta), Huggingface (Huggingface), Google Tag Manager
	Automatic curation of news on the website	Recommender systems	Think Analytics, Cxense, Contentwise
Distribution technologies	News links shared to external platforms (by users and journalists)	Social media/debate fora and message apps	Facebook, Twitter, SnapChat, WeChat, Instagram, TikTok, Reddit
	Newsletters	Email newsletter services	MailChimp, HubSpot, SubStack
	Making news articles available on external platforms	Aggregation services (including podcast apps), voice assistants, pre-load article solutions	Apple News, Google News, Nachtrichten.de, Apple podcasts, Alexa, Google Assistant, Google AMP, Facebook Instant Articles, News APIs
	Search	Search engines	Google Search (Alphabet), Bing (Microsoft), Yahoo

Source: Authors' work based on fieldwork and interviews, provider websites, StackShare, and Newman et al. (2021, 2022).

> identify and organize news articles from across the web, taking note of key factors—from keywords to website freshness—and keeping track of it all in the Search index. (Google News Initiative, n.d.)

Our interview with the Danish broadcaster TV2 showed how the organisation was conscious about being able to adapt to outside standards to ensure compatibility across time, platforms and devices: "When we build our

model, we try to look at the open models on the internet, like Google's, work. We try to apply those standards instead of our own to match models and connect content more easily" (journalist and developer at TV2).

In both media organisations in our fieldwork, *Ekstra Bladet* and JFM, the development and implementation of automated text analysis followed a similar pattern. This involved finding a suitable categorisation vocabulary, a taxonomy of content, and suitable tags for content. On a simple level, such tags could be "sports," "finance," or "entertainment," but on a much more finely grained level, text recognition (automated or manual) also involves finding places, names of specific sources, or categories in stories that are linked to a previously covered story. As we indicated in Section 1, the final taxonomy created at *Ekstra Bladet* (originally for different purposes than building transformer models), was a combination of content categories from the IAB, IPTC, and the paper's own categories, a negotiation between outside and inside values.

These negotiations of media autonomy in relation to the taxonomies offered by global marketing organisations, often developed for social media platforms in particular, mainly surfaced as a clash in topics that the audiences were interested in and the interests of the specific audiences of *Ekstra Bladet*, which were somehow not part of the more commercially built taxonomy. As the taxonomy is put into production and used as a cornerstone to train the large language models, it is included and embedded into a larger infrastructure, for example, linked up to databases and existing models provided by other actors available via a site like Huggingface or Github. If a category of content is left out in the first phase, no users will receive the content on this topic, neither as recommended nor as part of a personalised front page, and the developers are acutely aware of this. At both media organisations, the editors often discussed how they would solve the problem of new emerging content, which would then not be recognised by the large language models. For example, it would take a new tag to categorise content on Covid-19, which they argued was of as much democratic importance to the users as other news content, though it did not fit well with the advertising categories in the commercial models. "The IAB taxonomy tends to focus on cars and washing machines, which is far from the content we publish here at *Ekstra Bladet*," a developer said one afternoon, as we discussed how much work had gone into building the adapted taxonomy.

The trained models of text recognition and automated classification are interwoven into complex structures of data, both data on content and audiences, moving users in certain directions through the available content. As touched upon above, industry organs such as the IAB and IPTC, along with platforms such as Google, Facebook, and Yandex, are involved in streamlining categories of content on news websites. This partly pertains to the need to deliver accurate reports of audience data to advertisers. For media and their potential advertisers to make comparisons on the market, the method of measuring and reporting audience data cannot be entirely up to each media organisation. The interdependency here is driven by the industry level by these classification logics, as seen above, but it is also at the level of both commercial and non-profit actors that agree on standardisations for metadata and structured data markup. As previous research has shown, this, however, means that a news organisation might miss out on being distributed via search engines, for example, if it does not follow the mark-up standards provided, for example, by Google and Schema.org (Kristensen & Sørensen, in press). In the following section, we look more into how these standards work and manifest themselves in negotiations around infrastructure capture.

5.1.2. Infrastructure Capture Through Standardisation Logics

As discussed in the previous section, media organisations are faced with outside technical standards required to produce, distribute and monetise news. An interesting case to examine is the audience measurement systems used in the newsroom and for making editorial decisions. These are not, per se, required to fit outside the standards of measuring methodology and taxonomy. However, in our empirical data, we observed that the systems used often originate or migrate from the marketing departments, which adhere to formal standards, such as the IAB, to allow for comparing measurements across media outlets. Furthermore, most systems used for real-time editorial insights, for example, Chartbeat and Parse.ly, are developed with newsrooms and marketers in mind as end users. Whereas the former (e.g., IAB) involves formal standards, here we see a case of informal standards applied through the use of the same system in multiple settings—for example, online marketers focusing on making "content" rather than "news."

The media organisations in our data were trying to different degrees to negotiate and deal with this. The second-largest media organisation in Norway, Amedia, was working to both eliminate external systems that provide access to their own data (e.g., Google Analytics) and to better tailor the measurements to a media organisation with public service ideals:

> What we saw was that the questions we wanted to ask about our data, we couldn't answer in those kinds of systems and, also, we wanted the ability to customise both the data collection and the observations and how data flowed through our systems. And we really didn't want to be sort of sitting there, just as customers of a third-party product and be limited by the solutions that they offered. But, of course, it's still like every cost, or I'm not sure how many man hours or employees are dedicated to working on this, that wouldn't have been working or that we wouldn't

> need it if we have another system. (Head of digital development, Amedia)

As such, Amedia was trying to reclaim its autonomy in setting the standards for operationalising what news is and what could be considered empirical evidence of the "success" of a news story and the media organisation at large. At JFM, the audience measurement system was designed in-house in terms of the user interface, but the data came from Google Analytics and Facebook. Although this meant that the organisation was relying on outside standards of measurement methodology, and content categories, it allowed them to present the data in ways that helped them "qualify how to evaluate the journalism" (head of analytics, JFM). This entailed developing a point system that pooled together relevant metrics and developing custom dimensions in Google Analytics, for example, whether users had spent at least one minute and 30 seconds on the article page.

Standardisation logics come from both universally applied standards, such as taxonomies of Schema.org and Dublin Core, as well as from the systems applied in newsroom analytics. These are grounded in measurement methodologies and fixed metrics, along with visual representations of data with an interface designed by external system providers. Our empirical data reveal that a degree of infrastructure capture is in place. However, media organisations are aware of the potentially contradictory logics between them and the system providers.

5.1.3. Infrastructure Capture Through Datafication Logics

As *Ekstra Bladet* embarked on the project of personalisation and NLP (PIN project) in 2021, they realised that this also meant building their own data platform, eventually named Longboat. The purpose was also to share the data between the different publishers in the same media organisation, in this case, *Ekstra Bladet*, *Politiken*, and *JP*, who are all part of JP/Politikens Hus. Further, it was an attempt to gain autonomy vis-à-vis Google, as the platform was built with its own data analytics system. As the Head of Strategy at *Ekstra Bladet* Kasper Worm-Petersen explained in a press release:

> It's no secret that data is a very central element in the realization of our strategy for the coming years. It is therefore important to us that we have control over and ownership of our data throughout the value chain from collection to processing to activation. Relevance ensures us ownership of the activation. With the PIN project, we are investing heavily in the processing, and with Longboat, we are now also taking ownership of the collection itself. This gives us some completely unique opportunities in the media reality that Ekstra Bladet is moving into.

Across media organisations, we observed an awareness of how developing and maintaining one's own infrastructure is expensive and even risky in the case of software and hardware breakdown. As was the case with Amedia, at *The Guardian*, and JFM, developing proprietary systems was at the forefront to avoid technology "giants" profiting from the media's own data and to be able to define measurement categories themselves to a higher degree. However, the audience data infrastructure "pipeline" remained the same, but the visual and statistical presentation of data and the organisational discourse changed following a push to incorporate audience behaviours and preferences through web measurement reports on email and newsroom dashboards. As another example, the Danish daily *Information* experienced a shutdown of its email automation platform, which also interfered with its ability to send purchase receipts to subscribers. *Information* abandoned their major US-based provider following the incident:

> We cannot have a system that is so critical to our business where the provider does not have a phone number…so, first of all, I want a provider I can call, preferably in Denmark, but Germany, Norway or Sweden would also be okay. (Head of digital development at *Information*)

Eventually, they chose a Danish/Swedish email platform that had more functions but was also more expensive. This suggests that media organisations are aware of interdependencies, constantly negotiating their autonomy vis-à-vis these infrastructural systems and providers. Outsourcing infrastructural tasks to external software providers was a way to minimise spending, as the three people employed in the technical department did not have the resources to develop and maintain systems for subscriber login and payment, but functionality and control were still a priority.

Building transformer models and personalisation algorithms at *Ekstra Bladet* also meant negotiating infrastructure capture by not using an external dataset for machine learning and the training of models. Thus, a great deal of work goes into developing various datasets, including those of content, articles, and different kinds of users. This was not only due to the fact that datasets on open source platforms were often not in Danish but also because they did not feel adapted enough to the specific media organisation. For example, *Ekstra Bladet* has more users who are men and of a certain age and thus the baseline dataset used to recommend content needed to reflect this.

When building the transformer models at *Ekstra Bladet*, it was often discussed how much they could rely on open-source models and data provided via the free platform Hugginface, as it created an infrastructure dependence, which it was hard to foresee the consequences of. When a manager questioned this in a meeting, asking whether Huggingface would, at some point, capitalise on the models and code available, the answer from the developer was: "If Hugginface dies, we

die." Interestingly, although the primary aim of building transformer models is to retain autonomy, they make themselves dependent on other providers for the code, datasets, cloud services, etc.

6. Discussion

In the following, we discuss consequences, based on our findings, for journalistic production, monetisation, and the distribution of news.

Autonomy is a key ideal in journalism and through our analysis, we illustrated the extent to which external systems are sustaining parts of the journalistic production process, creating a form of infrastructural interdependence of these systems. The systems also bring with them certain logics. In the pre-digital age, categorisations of news (e.g., in terms of genre and subject matter) and audience members were to a certain degree standardised across the media industry itself. Today, these standards also occur across fields that operate with different logics than is traditionally the case in journalism. The consequence of this could be that news categorised through the same or similar taxonomies as "content" in general might be assimilated. In the case of audience measurement systems, we observed how metrics and visual representations could be standardised, but the large media organisations in our study acknowledged and negotiated the degree of infrastructure capture by implementing their own systems and tweaking the ones they bought externally.

We observe that audience measurement systems and other tech systems used in media organisations are inseparable from other infrastructures. These interdependencies are expressed, for the most part, through logics of standardisation, that is, an alignment of and path dependencies pertaining to practices around the use of systems, methods and data flow, data reporting, and discourses around the practice in and around the media. As illustrated by our cases, infrastructure capture of the news through the tagging of news content using (often open-source) algorithms, audience measurement metrics, and statistical representations becomes a negotiation between media organisations and the providers, that is, the composition of technologies behind media production and distribution.

In the literature on platformisation, we often see the loss of autonomy over distribution on external platforms. This includes Facebook's changes to algorithms in 2016 to focus more on friend relationships and the 2023 revealing of a function within TikTok that allows employees to override the factors that normally determine the position of posts in the feed. Distributing content on these external platforms thus means conforming to external logics of what content is popular and what the platform owner or employees prefer, resulting in a potentially high degree of infrastructure capture.

We found that the picture is somewhat more complex and that media organisations are aware of infrastructure capture through potentially competing logics. It is worth noting that the organisations in our sample are relatively well-resourced and that smaller or digital native media likely do not have the same opportunities to negotiate the degree of capture. Contrarily, this means that they potentially lag behind legacy media in their visibility on external platforms for not adhering to standards, leading to a lesser degree of infrastructure capture, perhaps at the expense of monetisation of news.

Dependence on advertising platforms is a key indicator of infrastructure capture (Nechushtai, 2018). In our empirical mapping, we observed how platforms are deeply enthralled in sustaining monetisation through advertising on media websites and externally through ad exchanges. Email services, login, and customer platforms are also deeply intertwined with the media and across systems. This would be considered material infrastructure capture in the sense that the media has the role of a scrutinising body (Nechushtai, 2018) and, at the same time, depends financially on platforms for both advertising and distribution. A future avenue for research is thus the potentially impaired ability to scrutinise the very companies that sustain news distribution and operation.

Although infrastructure capture is indicated to a certain degree in our empirical data, our analysis similarly points to the potential benefits of media organisations' backends being related to and embedded into existing systems. For instance, "outsourcing" technology allows media organisations to abandon maintaining servers for hosting and to use programming resources for tasks other than keeping a user database, for example. In the sense that journalism is important for democracy, opportunities to save money and ensure system stability by buying cloud services and embedding externally maintained and developed software from outside system providers can be a positive shift. Based on our empirical data, large media corporations often have the resources to decide whether to "outsource" their infrastructure, which might not be as accessible for digital native outlets and other smaller publishers. In addition, if media organisations did not structure data on news content to be recognisable to search engine web crawlers, they would not be visible in the search results, which would affect their access to users. The representation of Google products in our mapping is especially striking, suggesting that the company is as involved in sustaining the news media industry as it is in sustaining other parts of society (van Dijck et al., 2018). As such, we might not only see a case of infrastructure capture, but a case of what Plantin et al. (2018, p. 4) described as the "platformization of infrastructure" and "infrastructuralization of platforms." Thus, if we consider journalism in the form of legacy media news as a pillar of democracy, Google and the tech providers that are sustaining the backend of media could be approaching the status of the infrastructure of democracy. We find media logics to be a valuable framework to understand these developments while, for the same reasons, a concept that needs to be expanded

on. We believe that we have contributed to the theory by suggesting classification, standardisation, and datafication as entry points for this.

7. Conclusion

By mapping the elements of the digital infrastructure of media organisations from systems that handle the sales and distribution of advertising to systems that classify news, we have illustrated (Table 2) how data flows through systems originating from both within and outside media organisations. Our findings suggest that the latter is most often the case, illustrating that news and news production are increasingly and inevitably part of the larger infrastructure of the internet, provided by big tech companies, which have been theorised elsewhere as infrastructure capture (Simon, 2022; Nechushtai, 2018). Through case studies centering on the development of different parts of the backend tech stack of news distribution, we have shown that these tasks that were previously performed within the news organisation are now "outsourced" to external systems and providers. The analysis of the interviews and fieldwork illustrated how news organisations deal with this reality and how they are negotiated.

Finally, we discussed the consequences of these interdependencies on the autonomy of news media. To summarise, the dominant logics of media organisations' interdependencies with larger infrastructures are (a) standardisation and (b) embeddedness in former and parallel systems and infrastructures—materially and in terms of values and practices. The interdependencies are evident in our mapping and are of concern to media organisations across our fieldwork and interviews. This, in turn, highlights that infrastructure capture should not be seen as a one-way information highway, but as spaces of negotiation in which the infrastructural power manifests itself through logics of standardisation, classification, and datafication.

With the increasing use of AI in news organisations, the network of backend infrastructures is likely to be even bigger, and, from a research perspective, we need to analyse how this infrastructuring unfolds in different media settings. After all, if news media and journalism were ever the backbone of democracy, the infrastructure supporting them should not be overlooked.

Acknowledgments

The authors wish to thank the editors and reviewers of this article for excellent feedback and assistance. This research was funded by the Velux Foundation.

Conflict of Interests

The authors declare no conflict of interests.

References

Altheide, D. L. (2016). Media logic. In G. Mazzoleni (Ed.), *The international encyclopedia of political communication*. https://doi.org/10.1002/9781118541555.wbiepc088

Altheide, D. L., & Snow, R. (1979). *Media logic*. SAGE.

Ananny, M., & Finn, M. (2020). Anticipatory news infrastructures: Seeing journalism's expectations of future publics in its sociotechnical systems. *New Media & Society*, *22*(9), 1600–1618. https://doi.org/10.1177/1461444820914873

Anderson, C. (2011). Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms. *Journalism*, *12*(5), 550–566. https://doi.org/10.1177/1464884911402451

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 97–117). Springer.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Christin, A. (2020). *Metrics at work: Journalism and the contested meaning of algorithms*. Princeton University Press.

Couldry, N. (2008). Mediatization or mediation? Alternative understandings of the emergent space of digital storytelling. *New Media & Society*, *10*(3), 373–391. https://doi.org/10.1177/1461444808089414

Feuz, M., Fuller, M., & Stalder, F. (2011). Personal web searching in the age of semantic capitalism: Diagnosing the mechanisms of personalisation. *First Monday*, *16*(2). https://doi.org/10.5210/fm.v16i2.3344

Flanagan, M., Howe, D. C., & Nissenbaum, H. (2008). Embodying values in technology: Theory and practice. In J. van den Hoven & J. Weckert (Eds.), *Information technology and moral philosophy* (pp. 322–353). Cambridge University Press.

Flensburg, S. (2020). *Det Digitale Systemskifte: En historisk analyse af digitaliseringen af det danske Kommunikationssystem* [The digital system change: A historical analysis of the digitalization of the Danish communication system] [Doctoral dissertation, University of Copenhagen]. Københavns Universitet Forskning. https://forskning.ku.dk/soeg/result/?pure=da/publications/det-digitale-systemskifte(39af4a6b-6dfc-46bd-8ce8-2365dde25191).html

Flyverbom, M., & Murray, J. (2018). Datastructuring—Organizing and curating digital traces into action. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718799114

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information*

Systems, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Google News Initiative. (n.d.). *Organizing news from around the web*. https://newsinitiative.withgoogle.com/hownewsworks/approach/organizing-news-from-around-the-web

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012. https://doi.org/10.1080/21670811.2019.1623700

Infrastructure. (n.d.). In *Merriam-Webster's online dictionary*. https://www.merriamwebster.com/dictionary/infrastructure

Klinger, U., & Svensson, J. (2018). The end of media logics? On algorithms and agency. *New Media & Society*, *20*(12), 4653–4670. https://doi.org/10.1177/1461444818779750

Kristensen, L. M. (2021). Audience metrics: Operationalizing news value for the digital newsroom. *Journalism Practice*. Advance online publication. https://doi.org/10.1080/17512786.2021.1954058

Kristensen, L. M., & Sørensen, J. K. (in press). Classifying the news: Metadata as structures of visibility and compliance with tech standards. In J. M. Hartley, J. K. Sørensen, & D. Mathieu (Eds.), *DATAPUBLICS—Constructing publics in datafied democracies*. Bristol University Press.

Møller, L. A. (2022). Between personal and public interest: How algorithmic news recommendation reconciles with journalism as an ideology. *Digital Journalism*, *10*(10), 1794–1812. https://doi.org/10.1080/21670811.2022.2032782

Narayan, D. (2022). Platform capitalism and cloud infrastructure: Theorizing a hyper-scalable computing regime. *Environment and Planning A: Economy and Space*, *54*(5), 911–929. https://doi.org/10.1177/0308518X221094028

Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, *19*(8), 1043–1058. https://doi.org/10.1177/1464884917725163

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K. & Nielsen, R. K. (2022). *Reuters Institute digital news report 2022*. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022

Newman, N., Fletcher, R., Schulz, A., Andı, S., Robertson, C. T., & Nielsen, R. K. (2021). *Reuters Institute digital news report 2021, 10th Edition*. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2021

Neyland, D., & Möllers, N. (2017). Algorithmic IF…THEN rules and the conditions and consequences of power. *Information, Communication & Society*, *20*(1), 45–62. https://doi.org/10.1080/1369118X.2016.1156141

Petre, C. (2021). *All the news that's fit to click: How metrics are transforming the work of journalists*. Princeton University Press.

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, *20*(1), 293–310. https://doi.org/10.1177/1461444816661553

Poell, T., Nieborg, D. B., & Duffy, B. E. (2022). Spaces of negotiation: Analyzing platform power in the news industry. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2103011

Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, *10*(10), 1832–1854. https://doi.org/10.1080/21670811.2022.2063150

Singer, J. B. (2007). Contested autonomy: Professional and popular claims on journalistic norms. *Journalism Studies*, *8*(1), 79–95. https://doi.org/10.1080/14616700601056866

Sørensen, J. K. (2022). The shortcomings of the diversity diet: Public service media, algorithms and the multiple dimensions of diversity. In J. Meese & S. Bannerman (Eds.), *The algorithmic distribution of the news: Policy responses* (pp. 289–307). Palgrave Macmillan.

Star, S. L., & Bowker, G. C. (2002). How to infrastructure. In L. A. Lievrouw & S. Livingstone (Eds.), *Handbook of new media: Social shaping and consequences of ICTs* (pp. 151–162). SAGE. https://doi.org/10.4135/9781446206904.n12

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Thimm, C., Anastasiadis, M., & Einspänner-Pflock, J. (2018). Media logic or media logics? An introduction to the field. In C. Thimm, M. Anastasiadis, & J. Einspänner-Pflock (Eds.), *Media logic(s) revisited: Modelling the interplay between media institutions, media technology and societal change* (pp. 1–8). Palgrave Macmillan.

Thurman, N. (2011). Making "The Daily Me": Technology, economics and habit in the mainstream assimilation of personalized news. *Journalism*, *12*(4), 395--415. https://doi.org/10.1177/1464884910388228

Thylstrup, N. B., Hansen, K. B., Flyverbom, M., & Amoore, L. (2022). Politics of data reuse in machine learning systems: Theorizing reuse entanglements. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221139785

Tuchman, G. (1973). Making news by doing work: Routinizing the unexpected. *American Journal of Sociology*, *79*(1), 110–131.

van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

About the Authors

Lisa Merete Kristensen is a postdoctoral researcher at the Department of Communication and Arts at Roskilde University, Denmark. Her work is centred around metrics, tech stacks, and technical infrastructures in journalism and the impact of these on media organisations and journalism. She specialises in theories around the journalistic profession and is working with both digital methods and more qualitative research methodologies.

Jannie Møller Hartley is an associate professor at the Department of Communication and Arts at Roskilde University, Denmark. She is currently co-leading the Centre for Big Data at Roskilde University and is the PI of the Velux Funded research project DataPublics (2020–2023). Her research is situated in between the fields of journalism research, audience studies, and data science, and ranges across subjects such as media ethics, #MeToo, and AI in media.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 319–329
https://doi.org/10.17645/mac.v11i2.6464

Article

Follow the Data! A Strategy for Tracing Infrastructural Power

Sofie Flensburg * and Signe Sophus Lai

Center for Tracking and Society, University of Copenhagen, Denmark

* Corresponding author (sofie.flensburg@hum.ku.dk)

Submitted: 7 November 2022 | Accepted: 6 February 2023 | Published: 28 June 2023

Abstract
Recalling the well-known strategy of "following the money" when investigating the underlying power structures and business models of legacy media, this article argues that studies of digital political economies can benefit instead from following the data. Combining perspectives from critical data studies and infrastructure research, we first discuss how direct money flows can be difficult to trace in digital ecosystems, creating a need for alternative analytical approaches for studying and scrutinising contemporary power configurations in digital societies. As a theoretical backdrop, we elaborate on the concept of infrastructural power and apply it in a walkthrough of critical data infrastructures. To illustrate the efficacy of this strategy, we provide perspectives and examples from the political economies of internet infrastructures in Northern Europe and discuss how control over data is translated into economic profit and societal power. In doing so, we argue that increased attention to data infrastructures is needed to advance both critical data and infrastructure studies, improve digital market monitoring, and ground future regulation and policy.

Keywords
critical data studies; data economy; infrastructural power; internet infrastructures; political economy

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

The ubiquitous presence of Big Tech companies and recent political backlashes against their global market dominance have sparked a call for critical studies of the economic transactions and power structures that shape digital communication environments. Over the last decade, we have, for instance, seen a growth in "critical data studies" (Iliadis & Russo, 2016; Kitchin & Lauriault, 2014) aiming to denaturalise processes of datafication (van Dijck, 2014) and surveillance capitalism (Zuboff, 2019), along with an increasing interest in the infrastructures that underlie and support digital societies (Hesmondhalgh, 2021; Plantin & Punathambekar, 2019; Sandvig, 2013). Calling attention to the shortcomings of established methods and analytical frameworks developed in and for analogue media systems, this research has revealed a range of epistemic problems related to obtaining reliable knowledge on the often opaque and black-boxed market structures and modes of governance surrounding digital communication (DeNardis, 2020; Mansell, 2017). As a response to these urgent research challenges, the article suggests that the first step for "following the money" is to "follow the data."

In developing this argument, we first discuss the theoretical implications of studying data infrastructures as political economies by combining classic media and audience studies with perspectives from infrastructures research, thereby advancing the concept of infrastructural power (Mann, 1984; Munn, 2020). Applying this key concept, we then move on to present a walkthrough of the critical components of data infrastructures, arguing that access networks, backbone systems, (first-party) applications, and (third-party) data services provide

valuable insights into how digital power is obtained, exercised, maintained, and amplified in contemporary communication environments. To illustrate the strategy, we provide examples of how the various layers of internet infrastructures are organised and controlled within the context of Northern Europe and discuss how control over data is translated into economic profit and societal power. This leads us to conclude that increased attention to data infrastructures is needed to advance both critical data and infrastructure studies, improve digital market monitoring, and ground future regulation and policy.

2. Data Infrastructures as Political Economies

The "follow the data" strategy builds on and contributes to the legacies of political economist media studies by seeking to understand how capital "changes hands" between various stakeholders, and how these economic exchanges influence the structural conditions in (digital) communication environments (Garnham, 1979; Mosco, 2009). Broadly inspired by DeFleur's (1971) efforts to draw up a "schematic representation of the mass media as a social system," we seek to understand the institutional arrangements that shape the internet as a key communication infrastructure in contemporary societies. Like DeFleur, our goal is ultimately to shed light on the crucial but often overlooked "money arrows" that enable the production and flow of information in modern digitalised societies, influencing the capabilities of public and private institutions as well as individual citizens. That is, we look beyond the creation of (symbolic) content to study the production, distribution, and consumption systems as they are shaped by a wide range of (conflicting) interests and business models. In doing so, we answer recent calls for a strengthening of the bond between the classic political economy of media and communication and state-of-the-art infrastructure studies (Hesmondhalgh, 2021).

2.1. Follow the Money

As the main innovation of DeFleur's model, the "money arrow" broke with decades of media research by challenging the emphasis on one-way flows of mass communication content from senders to receivers and pointing to the flows of information going in the opposite direction—from audiences and back to media institutions in the form of audience measurements including target group analyses and rankings (DeFleur, 1971; Ettema & Whitney, 1994). The mapping of financial transactions connecting viewers, listeners and readers, media institutions, measurement companies, and advertising agencies, among others, served as an important foundation for analysing the structural conditions for (analogue) media and for explaining how some media companies gained dominant market positions and continuously expanded their communicative power (Wasko, 2011). As such, DeFleur's efforts to map out the "hidden'' transactions and dependencies underlying the production and publishing of legacy media content serve as a guiding inspiration for critically assessing and uncovering the underlying value chains and dependencies that shape digital markets.

The gradual shift from analogue to digital distribution of mediated communication has significantly altered the conditions for studying these money arrows since digital outlets are embedded in ecosystems of external service providers, intermediaries, and global distribution networks (Nielsen & Ganter, 2018). Digital business models, value chains, and market structures are notoriously difficult to map, with multitudes of money arrows crisscrossing in complex and often obscure ways. While, for instance, ads are a vital source of income for many online service providers, the economic crossfires supporting contemporary audience measurement and ad sales are not easily drawn up. One reason, among many, is that these services are often supplied in a "freemium" manner, making it difficult to trace their business models and money flows. The task of identifying and following key assets in contemporary digital markets is, in other words, both urgent and critical for political economists in the fields of media, communication, and internet research.

While identifying the economic circuits between user measurement, content provision, and advertising is as crucial as ever for understanding contemporary media and communication structures, such studies often neglect to consider the multitude of underlying production and distribution systems. In effect, the "data market" is often seen as detached from the broader digital economy, while it is, in fact—as we will discuss in the examples below—increasingly entangled in the broader infrastructures supporting digital communication. As such, the remainder of this article discusses and explores how an enhanced understanding of the infrastructures that control digital data flows can help researchers and regulators make sense of and ultimately monitor economic power structures in digital environments.

2.2. Critical Data Infrastructures

Following DeFleur's argument above, data exchanges have always been incremental to the double-sided marketplace of attention (Webster, 2014)—in analogue media systems, data came in the form of representative panels that have evolved into today's census counts, or what is often simply referred to as big data. While former approaches collected under the headings of, for instance, "digital tracing" or "digital footprints" have focused on what the comprehensive collections of big data can—or cannot—be used for (Golder & Macy, 2014; Lambiotte & Kosinski, 2014; Lewis, 2015), infrastructural approaches draw attention to the ways data flows are handled and controlled. The "turn to infrastructure" (Hesmondhalgh, 2021; Parks et al., 2015; Plantin & Punathambekar, 2019; Sandvig, 2013) in media and communication studies thus broadly entails a renewed

attention to the *material* structures rather than the *symbolic* content of digital communication.

When following the data, we thereby refer to a broader, much more fundamental, and infrastructurally embedded resource in digital ecosystems. Whether people make an online appointment with their doctor, search for shoes via Google, or turn on their smart TV to watch the latest episode of their favourite show on Netflix, their activities are materialised in the form of data flowing back and forth in a distributed network of servers. "Data" is thus defined as the packages that all types of digital content are broken down to when transported over the internet—regardless of whether they contain media content in a conventional sense and flow from content providers *to* individual users, or if they flow *from* users to content providers carrying meta-data about the users' online behaviour, browser history, location, and other valuable information. When sketching out what we refer to as data arrows in digital ecosystems, we focus on the infrastructures that allow these data packages to be distributed between dispersed devices, applications, and computer networks rather than on the specific information they contain.

We thereby contribute to a growing research field aiming to uncover and critically assess the ways architectural and technological arrangements shape human capabilities and societal development and vice versa—be it through, for instance, the everyday governance of monumental data centres (Velkova, 2019), the planning, laying, and maintenance of global submarine cables (Starosielski, 2015), or the control over operating systems, web and mobile applications (Dieter et al., 2019; Gerlitz et al., 2019; Weltevrede & Jansen, 2019), and third-party data services (Binns, Lyngs, et al., 2018; Helles et al., 2020). Building on the broad and diverse literature within the field of infrastructure studies, the strategy of following the data opens up for cutting across otherwise separate infrastructural (and academic) domains—for instance, the backbone industry and the platform economy (Plantin et al., 2018)—and engaging in empirical investigations of the spaces in-between them and the critical dependencies they share.

2.3. Infrastructural Power

While refraining from going into a lengthy, albeit interesting, debate about the theoretical definition of "infrastructure" (see, e.g., Lee & Schmidt, 2018), we will employ a more narrow and conventional use of the concept than what is often referred to as a "deeply relational" understanding that leads researchers to ask not "what" but "when" is an infrastructure (Star & Ruhleder, 1996). When using the concept, we instead refer to the physical resources that enable key societal functions—such as communication. Communication infrastructures, in this sense, constitute the components that senders and receivers continuously rely on, regardless of how or when they are used in practice. These components can be controlled by a variety of stakeholders and be more or less transparent to individuals, nation-states, and researchers. Following perspectives from classic media ecology (Innis, 2007), our ultimate goal in uncovering and studying data infrastructures is to understand how the evolution and institutionalisation of these new infrastructures reshape fundamental societal power structures.

In approaching the relationship between infrastructures and societal power, we build on Mann's (1984) work on infrastructural power. Mann distinguishes between despotic power, understood as the ability of states to exert direct power over individuals (e.g., by imprisoning them), and infrastructural power, understood as the ability to "penetrate and centrally coordinate the activities of civil society through its own infrastructure" (p. 190). Similar to the related concept of institutional power, infrastructural power is exerted through the organisation of societal structures that inevitably influence what people can and cannot do—and not least what choices they have as well as their abilities to imagine alternatives (Mansell, 2002). But while institutional power is discursively constructed and serves as a legitimising and self-regulatory mechanism that naturalise behavioural control through physical means (Foucault, 2008), infrastructural power is materially manifested in the organisation of the physical world (the design of buildings, networks, code, and so forth). The efficacy of institutional power is thereby interchangeably dependent on the infrastructural arrangements that prevail at any given time and in any given context—not least in periods where infrastructures undergo significant changes (Beniger, 1986).

While Mann originally used the concept in relation to nation-states and political systems, we argue that it is highly relevant for understanding political and economic power in a broader sense and, in particular, for making sense of the ways digital infrastructures are organised and controlled by both private and public stakeholders. As we will argue further below, Mann's understanding of states' infrastructural power resembles the current role of so-called Big Tech companies in that it cuts across sectors and, through the design of the physical world, influences the capabilities and activities of individuals and institutions that rely on their systems and services. Prominent examples of this include Meta's decision to close down the Facebook Application Programming Interface, which caused ruptures across the digital industry as so many companies (and academics) had come to rely on it for their business (Bruns, 2019); Alphabet's requirement that phone manufacturers include their services in return for licensing the Android operating system; and their insistence that app developers abide by the rules of the world's largest app store in order to publish their products (Lai & Flensburg, 2021).

The following sections provide a systematic walkthrough of how the concept of infrastructural power can be explored by following the data through the different

layers of the internet and investigating how they are owned and controlled.

3. Sites of Infrastructural Data Power: Perspectives From Northern Europe

In mapping the infrastructural arrangements that underly our increasingly datafied societies, we build on former research (Flensburg & Lai, 2019) identifying four main infrastructural layers of digital communication, namely: *access networks* that allow users to connect to the internet and thereby send and receive data; *backbone systems* that enable these networks to exchange data with other operators and networks; *applications* in the form of, for instance, websites and mobile apps that present the data for the user through an interface; and finally, technologies for storing, processing, analysing, and distributing (meta) *data*, often provided by external third-party services. The four layers of internet infrastructure are illustrated in Figure 1, which also depicts

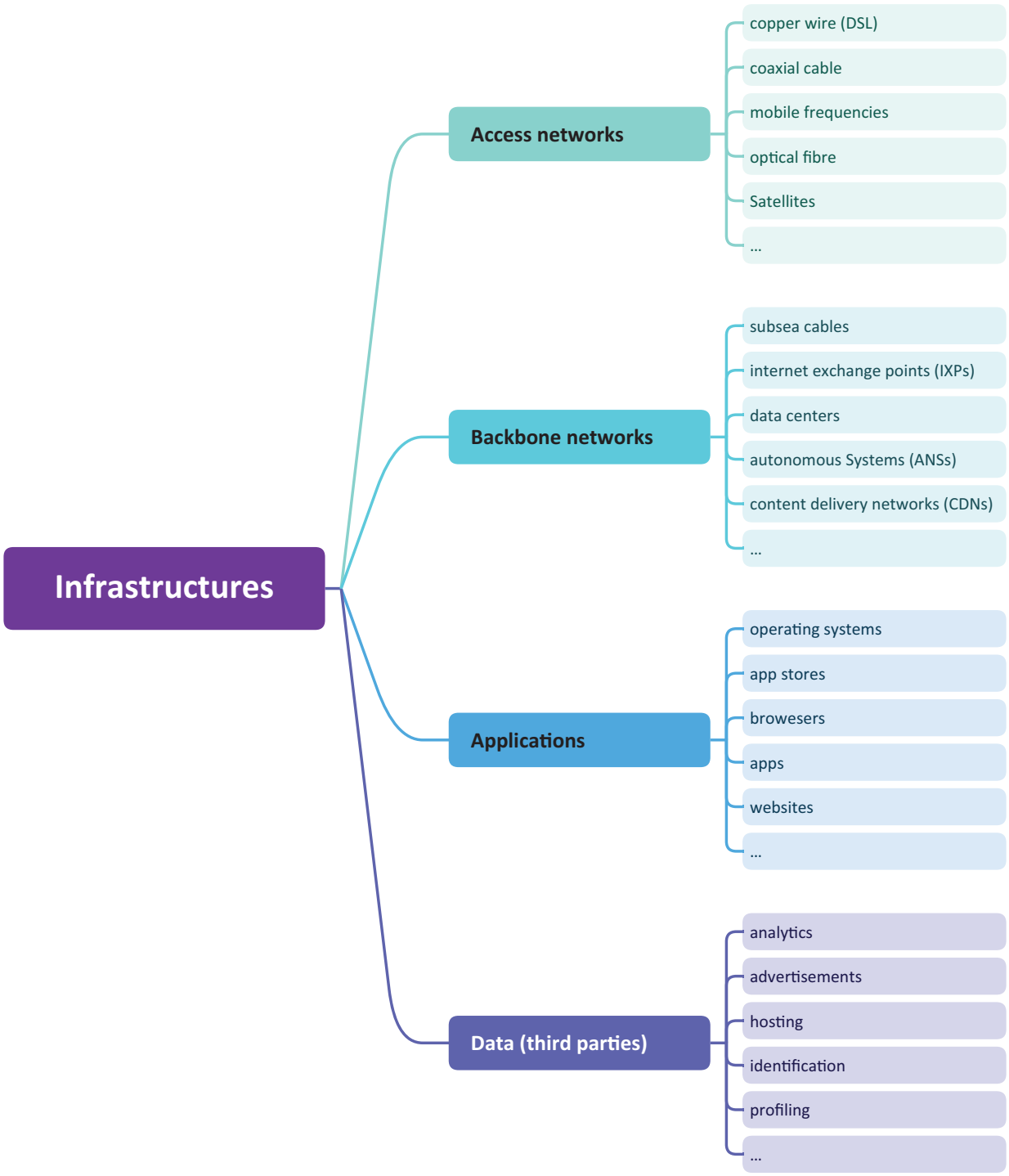


Figure 1. Four layers of internet infrastructure and their infrastructural components.

examples of their specific components (e.g., the different types of access networks, backbone systems, applications, and third-party data services).

This perspective enables us to map out infrastructural dependencies in and across the digital ecosystem. We can, for instance, uncover how access networks depend on backbone systems when exchanging data with other network operators or how providers of online services rely on externally provided tools for data processing, storage, and distribution. We can also begin to shed light on and question the various economic transactions involved in exchanging data between different layers. For example, when access networks charge users but pay other operators for routing and peering; when applications put up paywalls and sell ads but also pay for, for instance, content delivery networks (CDNs), cloud solutions, and analytics; and when third-party services offer tools and services "free of charge" while monetising them in other ways.

Illustrating how the strategy can be applied, the following sections provide examples from the context of Northern Europe to discuss how the different components of the internet serve as economic assets and sites of infrastructural power within specific societal contexts. More specifically, we provide examples of (a) how the economic conditions for running and using a digital service are framed by the data infrastructures in which they are embedded, and (b) how dominant market actors in the digital realm use these data infrastructures to obtain, maintain, and expand their infrastructural power.

3.1. Access Networks

Local access networks constitute the first stop—or last mile—of the internet by allowing individual internet users or services to send and receive data through, for instance, a fixed (copper wire, coaxial, fibre optic) or mobile (e.g., 5G or satellite) connection as sketched out in the upper right corner of Figure 1. Having the power to connect or cut off individual users, specific communication services, or entire communities (Benjamin, 2022; Krapiva et al., 2022), internet service providers (ISPs) are crucial gatekeepers in the digital ecosystem. As a result, the access network layer is rigorously monitored in statistics of internet penetration, connectivity, and coverage (Access Now, n.d.; European Commission, n.d.; ITU, n.d.; OECD, n.d.), making this layer of internet infrastructure relatively transparent to both researchers and regulators.

Since the structural conditions for supplying, and thus using, broadband services differ significantly depending on the underlying technology, access networks are prime cases for investigating the relationships between physical infrastructures and political economies. The earliest forms of broadband connections in the Nordic region were, for instance, based on extensive landline (copper wire) telephone networks and later (coaxial) cable TV systems, making it fairly easy for legacy telecommunications companies to position themselves in the emerging ISP market. As the market has matured and the demand for high-capacity connections has grown, optic fiber networks have been rolled out, many of which utilise existing electricity grids, while mobile networks have gradually improved to the point that they can now offer an alternative to fixed broadband subscriptions. The increasing competition over network traffic has evoked intense power struggles between network operators, who are investing billions of euros in updating and developing their services.

The competition between fixed and mobile networks constitutes one of the clearest examples of infrastructural power at the access network layer and, unlike what we will see in the next infrastructural layers, it is largely politically determined. Since the electromagnetic spectrum used for mobile networks is a scarce resource, it is allocated and assigned by state authorities, meaning that questions about whether or not to release spectrum for mobile communication and how to price and assign frequencies are highly political matters (Ala-Fossi & Bonet, 2018; Martínez-Santos et al., 2021). By holding back the allocation of frequencies or selling them off at high prices, governments can—deliberately or not—delay the spread of mobile broadband and thereby create an advantage for fixed broadband providers. In contrast, the current investments in and rollout of 5G are making mobile data subscriptions a strong alternative to fixed internet, thereby possibly weakening the incentives for users to pay for an additional (fixed) internet subscription. In the Nordic region, these tensions are apparent: Finland tops the global charts in terms of mobile data consumption and subscriptions as a direct result of the country's early allocations of spectrum, which made mobile networks widely available. In contrast, terrestrial (fibre optic) networks are scarce compared to neighbouring countries such as Norway and Sweden, which largely rely on fixed connections (International Telecommunication Union, n.d.). In Norway, in particular, spectrum has been auctioned off at a high cost and mobile data subscriptions are expensive, making fixed broadband more attractive for users as well as operators.

As another example of how infrastructural arrangements play into economic power structures, the internet's success and the Nordic countries' comprehensive digitalisation have altered the basic conditions for developing and supplying access networks. While legacy telcos such as Danish TDC, Finnish Elisa, Norwegian Telenor, and Swedish Telia continue to dominate the ISP market in the Nordic region, their original sources of income and, for some, their ability to make new infrastructural investments have been severely debilitated by the rise of so-called over-the-top services that provide web and app-based alternatives to classic telecommunication services such as telephony and traditional TV distribution. While the access network market, for now, continues to be controlled by traditional telecommunications operators, Big Tech companies originating in

the applications and data layer show increasing interest in building and running access networks outside the Nordic region. Facebook's mother company Meta is, for instance, collaborating with established network operators on mobile connectivity projects where basic versions of Facebook's products can be accessed free of charge (Eisenach, 2015), while Google (owned by Alphabet) supplies high-speed fibre networks in selected American cities (Lam, 2017).

These examples illustrate a central tension in the digital ecosystem where former network operators lose their exclusive positions as service providers, while digital service providers increasingly invest in underlying network infrastructures (Plantin et al., 2018). These investments can be seen as proprietary efforts to control the entire value chain underlying data traffic: from the collection and initial transport of data packages to service operations and third-party service provision. By supplying and controlling their own access networks, Big Tech companies can obtain independence from other service providers while also potentially channelling users towards their own services (e.g., the Google search engine or Facebook's website). Interestingly, these commercial strategies largely resemble the former business models of legacy telcos that would also control entire value chains: from the supply of vital communication services (such as telephony) to the underlying network infrastructures connecting dispersed terminals. As telephone companies increasingly become ISPs, they inevitably give up this market advantage and grow dependent on other infrastructure operators enabling them to enter the global network of networks. Returning to Figure 1, this creates a direct data (and money) arrow between the infrastructural layers of access and backbone networks and the companies that control them.

3.2. Backbone

Often described in vague and obfuscating terms such as the "cloud," the extensive cable networks, exchange hubs, and data centres located beyond the last mile of access networks constitute a materialisation of what the internet essentially is—a (global) network of networks. For this article, we stress three backbone components serving as key sites of infrastructural power in contemporary digital societies, namely: internet exchange points (IXPs), submarine fibre optic cables, and CDNs, each of which is represented in the second branch of Figure 1. While these examples do not provide a comprehensive account of the highly complex global backbone infrastructure (e.g., terrestrial dark fibre networks and data centres are difficult to map), they make up key physical resources supporting the exchange and transport of data and can be studied through various types of publicly available information sources and databases.

Providing the (physical) facilities that enable dispersed access networks to exchange data with each other without having to establish individual peering points, IXPs make up a critical component of the contemporary internet infrastructure. In the Nordic region, the first generation of IXPs was established by national institutions (often universities) in the 1990s, following the growing demand for internet peering and routing. In recent years, the number of exchange points has increased significantly, with some of the largest being placed in Amsterdam, Virginia, and Hongkong, and a number of multinational companies such as the US-based Equinix and Russia-based DATA-IX running facilities in, for instance, Denmark, Finland, and Sweden (see Internet Exchange Map, n.d.). The IXPs hold tremendous infrastructural power by determining the conditions for peering and thereby influencing the global economy of the internet (e.g., what ISPs and other network operators need to pay to interconnect).

As another critical backbone resource, submarine fibre cables create links within and between countries, regions, and continents separated by sea, thereby allowing for the transfer of global internet traffic (Starosielski, 2015). Constituting an important infrastructural foundation for the spread of the internet—and not least the exponential use of US-based services in the 2000s and 2010s—the first generation of submarine internet cables were laid between the late 1980s and early 2000s by large consortia of national telcos. With an anticipated lifespan of 25 years, many of them are currently being retired or superseded by higher-capacity cables (Routley, 2019), often funded by other types of corporations and institutions—including American Big Tech companies such as Alphabet and Meta that in recent years have invested significantly in global cable routes (Clark, 2016). The submarine cable market, however, continues to be inhabited by a diverse group of stakeholders reflecting geopolitical contexts and national power structures (Winseck, 2017, 2019). In Northern Europe, Norway is, for instance, characterised by a relatively high degree of national infrastructure ownership, unlike Denmark, where foreign investments are more common. Further research is needed to conclude whether these differences result from regulation and policy decisions, geographic features, or economic interests.

The final example of backbone infrastructures creating important data—and money—arrows in the digital economy are the CDNs used by, for instance, streaming services to prevent network congestion when distributing high-capacity content. As a crucial innovation in the evolution of the internet, CDNs solved an inherent challenge of the internet's point-to-point architecture by moving content away from the producer and placing it (temporarily) at the opposite edges of the network, close to the end-user (Sandvig, 2015). This requires significant collection and analysis of user data since CDNs need to know—and predict—what content to store where (Helles & Flyverbom, 2019), linking this part of the backbone directly to the data layer, described later in the article. The first and globally leading CDN company is US-based Akamai which supplies services

to a multitude of content providers (BasuMallick, 2022), followed by Amazon's CloudFront, which offers integration with its immensely popular cloud solution (AWS), while Microsoft Azure, CasheFly, and Cloudflare provide alternative CDN solutions. In the highly digitalised Nordic region, the users' growing preference for streaming pushes up the expenses for media outlets and public service institutions, who have to both maintain their original distribution system (e.g., broadcasting and print) and invest in digital network services (including, but by no means limited to, CDN).

All in all, the control of backbone infrastructures constitutes one of the most critical and black-boxed forms of infrastructural power. Without access to central peering points and data highways, access network operators would not be able to exchange data, and users would be unable to use services located beyond the confines of their local networks. Service operators wishing to make use of the global market potentials created by the internet are, in other words, highly reliant on efficient and accessible backbone infrastructures and their service conditions. This, along with the capacity-demanding ambitions such as developing the internet of things and the metaverse, provides explanations for so-called platform companies' recent investments in the backbone market while also creating a link to the next layer of the internet infrastructure, namely that of applications.

3.3. Applications

Moving to the third infrastructural layer of Figure 1, digital applications provide the interfaces and services that make the internet useful for individuals and thereby essentially trigger the sending and receiving of data. Whenever a user activates a digital service, data is sent to the servers hosting the application, requesting that the applications return data to the individual. Application infrastructures thereby refer to the physical servers of websites and apps but also to operating and domain name systems, web browsers, and app stores. Since the content of these services is often hosted on external servers and in CDN caches, cloud services and CDNs are closely linked to the application layer. For this article, however, we emphasise the infrastructural and economic relationships between websites and browsers and between mobile apps and app stores since these constitute clear examples of how infrastructural power is obtained and exercised in contemporary digital economies.

As the "killer application" of the early internet (Naughton, 2016), the World Wide Web (www) holds an important key to understanding the public breakthrough of digital communication and the rise of the digital political economy. Providing a common coding language and hyperlink protocols, it allowed for website programming, web searches, browsing between data stored on different servers, and much more. To access a website, users need a web browser that presents the requested content in a comprehensible format and allows the user to enter URLs and navigate between different websites. As a key characteristic of web-based communication, websites are not browser-specific, meaning that any website can be accessed from any browser and that the supply of browsers does not constitute a business model as such. Most browsers are therefore owned by corporations that are based on other related revenue streams such as advertisement (in the case of Alphabet's Chrome browser), device manufacturing (in the case of Apple's Safari browser), or software development (in the case of Microsoft's Edge).

The introduction of smartphones, mobile networks, and apps constitutes another key moment in application history as it released digital services from their previous reliance on (stationary) computers, fixed network connections, and web browsers. The success of mobile apps has evoked an infrastructural rearrangement of the basic conditions for supplying—and using—digital services since mobile apps are installed on the users' devices and, therefore, must be custom-made for the different operating systems. This means that app stores take over from browsers as the main gatekeepers in the application ecology—but with the important difference that individual apps must develop specific versions for different operating systems and make them available in different app stores (e.g., Google Play for Android devices and AppStore for Apple devices). Contrary to web browsers, app stores require apps to pay a percentage of their profit (typically 30%) and can remove apps as they see fit. As such, mobile apps are developed and published in more closed-off environments than initially imagined with the development of the open web (Berners-Lee et al., 1992).

The implications of these infrastructural differences stand out clearly when we turn to the specific context of Northern Europe, where the web and app ecologies are characterised by similarities but also significant differences. While the national web ecologies, to a wide extent, reflect historically anchored market structures in the different Nordic contexts with a strong presence of, for instance, legacy media institutions (e.g., Norwegian Schibsted, Swedish Aftonbladet, Finnish Alma Media, and Danish JP/Politiken) and national public service institutions, the app ecologies are more similar and globalised across the region. Looking at, for instance, the most used apps, Google (Alphabet), Facebook (Meta), Samsung, and Microsoft dominate, while apps developed for and by state authorities (e.g., health services, identification, public communication platforms) make up most of the (minority) of nationally specific apps amongst the most used.

The application layer, in other words, constitutes a clear arena for studying contemporary infrastructural power exertion, where the architectonical principles and design choices are intrinsically linked to the political economies that evolve around them. Since the prevalence of free of monetary charge services makes digital market dominance and revenue streams difficult to

trace, the amount of data traffic travelling to and from the servers and domains of immensely popular websites and apps constitutes an important object of study that is in urgent need of methodological innovation and systematic analysis. In mapping out these market structures, it is essential to consider the infrastructural environments that shape them (e.g., operating systems and app stores) and to consider how dominating suppliers of devices, operating systems, key applications, and network services directly or indirectly shape usage patterns and competition structures. However, the infrastructural entanglement of applications in the underlying data economy is even more crucial when seeking to understand the data and money arrows of digital ecosystems.

3.4. Data

By now, we hope to have made a convincing argument for seeing all the resources described above as essentially being data infrastructures in so far as they enable (or constrain) the transfer of data packages between senders and receivers. As such, it might seem incongruous to label this last part of the digital infrastructure (the bottom right corner of Figure 1) as the "data layer." Yet, we do so to emphasise that these infrastructural arrangements support what we commonly refer to as the "data economy"—understood as the economic structures supporting the *processing* and *handling* of user (meta) data—be it by controlling network capacity and speed, troubleshooting, tracking users' web history, registering location information, serving ads, or any other purpose (Libert, 2015). Following the data arrows into this often hidden and implicit part of the digital ecosystem reveals important dependencies between third-party operators and application providers, and it enables us to enquire into how and why user data has become one of the most valuable resources in the digital economy.

While data transport is essential to any internet-based communication, the early phases of digitalisation were surprisingly free from registration and tracking. In fact, the anonymity and one-way flow of information characterising the early web was a major obstacle for especially those emergent e-commerce initiatives where registration of purchases and payment information were critical. The later infamous "web cookie" became the solution by making websites capable of storing and tracking user data to, for instance, remember user preferences, profiles, and search history (Naughton, 2016). In time, web cookies became the spine of online advertisement, replacing more or less representative panels with big data collected through users' browsers and providing more granular and wide-ranging information on their preferences and behaviours to encourage them to make future purchases (Zuboff, 2019). The cookie market has gradually been taken over by companies such as Alphabet, Meta, and Amazon that have bought up a range of third-party services while simultaneously using—and nurturing—the user information collected from their own immensely popular applications (Falahrastegar et al., 2014).

Similar to and as a direct consequence of the development in the application layer, the introduction of smartphones and mobile apps has extended and disrupted the data market. First and foremost, the penetration of digital communication into almost all spheres of everyday life has enabled more comprehensive data collection, including location tracking. Furthermore, unlike websites, mobile apps are built in a modular fashion where third-party services are integrated into the very architecture as building blocks rather than as later add-ons (Dieter et al., 2019), making it more difficult to opt out. And finally, the more closed-off environments of mobile apps also mean that the large operating system and app store suppliers (Alphabet and Apple) have even stronger positions in the app-based third-party environment than in the web sphere (Binns, Zhao, et al., 2018). The ongoing concentration of power across the application and data layer is, in other words, infrastructurally rooted as the dominant market actors serve as important gatekeepers controlling operating systems, browsers, and app stores, while also providing the tools and services on which their competitors rely.

Directly reflecting the different power configurations in the Nordic application marked outlined above, the control over third-party services also differs significantly when comparing web and mobile third-party infrastructures. Studies of third-party services in websites (Helles et al., 2020) and apps (Binns, Lyngs, et al., 2018; Kollnig et al., 2022) show that (in)famous third-party services such as Google Analytics and other highly successful products provided by Alphabet, appear on more than half of the top websites and apps. In the Nordic region, the Norwegian legacy media company, Schibsted, is a (not so close) second runner-up in the Nordic web cookie market due to its provision of services to news sites, especially Nordic ones—while Nordic third-party services are next to non-existent in the mobile app market (Flensburg & Lai, in press). This clearly illustrates how the gradual shift from "the open web" (Berners-Lee et al., 1992) to the more walled-off environments of mobile apps entails a significant altering of the infrastructural power structures where the (data) rich (Andrejevic, 2014) become richer while the (data) poor continuously contribute to the success of their largest competitors by relying on their (data-driven) services.

By following the data beyond its "known" destinations (the requested website or app) and identifying the wide range of companies collecting, storing, analysing, and feeding data back to applications and users, we can get a glimpse into a largely hidden but equally crucial part of the digital infrastructure—and market. This allows us to study and scrutinise how design choices and continuous system updates are linked to corporate strategies and can help explain increasing market concentration. The ever-growing data economy, in turn, serves as a foundation for the constant expansion of Big Tech's

infrastructural power as it fuels investment in other parts of the internet infrastructure, thereby ensuring sufficient capacity and efficiency of their increasingly advanced services and drawing up new data arrows from this part of the infrastructures and back to the latter three.

4. Conclusion

Concluding this walkthrough of critical data infrastructures, we hope to have demonstrated what researchers can gain from following and drawing up data arrows as means of identifying economic transactions and money flows in digital ecosystems. Returning to the introductory ambitions of combining the questions asked by critical data studies with the empirical and analytical approaches of infrastructure studies, the strategy and examples discussed above provide a foundation for further investigations of the market structures and economic arrangements surrounding ongoing processes of datafication. Infrastructure research can, in turn, benefit greatly from developing theoretical frameworks and critical research questions to substantiate its strong empirical contributions. Or to sum up, following the flows of data as they travel through and across geopolitical contexts, sectors, and institutional arrangements fosters a broader understanding of the data economy and how it can be studied—and ultimately, regulated.

When cutting across the different internet layers and sites of infrastructural power, we begin to see the contours of a multitude of data and money arrows that ground commercial power structures in datafied societies. The "follow the data" strategy allows us to scale up from specific case studies and particular flows of data to investigate, map, and monitor the macro structures that currently are subject to little democratic scrutiny. By applying the strategy, we gain deeper insight into the conditions for running a digital business and extend our understanding of how and why a handful of companies obtain increasingly powerful positions in the digital ecosystem. Such efforts are pivotal since companies such as Alphabet continuously extend their infrastructural power across the value chain: from being a global leader in the applications and data market to increasingly investing in backbone and even access network infrastructure. Through these investments, Big Tech companies become increasingly independent of other actors while simultaneously making other market actors increasingly dependent on the company's infrastructures.

Conflict of Interests

The authors declare no conflict of interest.

References

Access Now. (n.d.). *Homepage*. https://www.accessnow.org

Ala-Fossi, M., & Bonet, M. (2018). Who's afraid of a pan-European spectrum policy? The EU and the battles over the UHF broadcast band. *International Journal of Communication*, *12*(2018), 337–358. https://ijoc.org/index.php/ijoc/article/view/6713/2238

Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, *8*(2014), 1673–1689. https://ijoc.org/index.php/ijoc/article/view/2161/1163

BasuMallick, C. (2022). *Top 10 content delivery network (CDN) providers in 2021*. Spiceworks. https://www.spiceworks.com/tech/networking/articles/content-delivery-network-providers

Beniger, J. R. (1986). *The control revolution: Technological and economic origins of the information society*. Harvard University Press.

Benjamin, G. (2022, June 23). *Internet shutdowns hide atrocities: People in Myanmar need global action*. Access Now. https://www.accessnow.org/myanmar-shutdowns-hide-atrocities

Berners-Lee, T., Cailliau, R., Groff, J., & Pollermann, B. (1992). World-wide web: The information universe. *Internet Research*, *2*(1), 52–58. https://doi.org/10.1108/eb047254

Binns, R., Lyngs, U., Van Kleek, M., Zhao, J., Libert, T., & Shadbolt, N. (2018). Third party tracking in the mobile ecosystem. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 23–31). Association for Computing Machinery. https://doi.org/10.1145/3201064.3201089

Binns, R., Zhao, J., Kleek, M. V., & Shadbolt, N. (2018). Measuring third-party tracker power across web and mobile. *ACM Transactions on Internet Technology*, *18*(4), 1–22. https://doi.org/10.1145/3176246

Bruns, A. (2019). After the "APIcalypse": Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, *22*(11), 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447

Clark, B. (2016). Undersea cables and the future of submarine competition. *Bulletin of the Atomic Scientists*, *72*(4), 234–237. https://doi.org/10.1080/00963402.2016.1195636

DeFleur, M. L. (1971). Mass media as social systems. In W. Schramm (Ed.), *The process and effects of mass communication* (pp. 63–83). University of Illinois Press.

DeNardis, L. (2020). *The internet in everything: Freedom and security in a world with no off switch*. Yale University Press.

Dieter, M., Gerlitz, C., Helmond, A., Tkacz, N., van der Vlist, F. N., & Weltevrede, E. (2019). Multi-situated app studies: Methods and propositions. *Social Media + Society*, *5*(2), 1–15. https://doi.org/10.1177/2056305119846486

Eisenach, J. A. (2015). *The economics of zero rating*. NERA Economic Consulting. https://www.nera.com/content/dam/nera/publications/2015/EconomicsofZeroRating.pdf

Ettema, J. S., & Whitney, D. C. (1994). The money arrow:

An introduction to audiencemaking. In J. S. Ettema & D. C. Whitney (Eds.), *Audiencemaking: How the media create the audience* (pp. 1–18). SAGE.

European Commission. (n.d.). *Broadband in EU countries*. https://digital-strategy.ec.europa.eu/en/policies/broadband-eu-countries

Falahrastegar, M., Haddadi, H., Uhlig, S., & Mortier, R. (2014). The rise of panopticons: Examining region-specific third-party web tracking. In A. Dainotti, A. Mahanti, & S. Uhlig (Eds.), *Traffic monitoring and analysis* (pp. 104–114). Springer.

Flensburg, S., & Lai, S. S. (2019). Mapping digital communication systems: Infrastructures, markets, and policies as regulatory forces. *Media, Culture & Society*, *42*(5), 692–710. https://doi.org/10.1177/0163443719876533

Flensburg, S., & Lai, S. S. (in press). *Gateways. Comparing digital communication systems in Nordic welfare states*. Nordicom.

Foucault, M. (2008). "Panopticism" from discipline & punish: The birth of the prison. *Race/Ethnicity: Multidisciplinary Global Contexts*, *2*(1), 1–12.

Garnham, N. (1979). Contribution to a political economy of mass-communication. *Media, Culture & Society*, *1*(2), 123–146. https://doi.org/10.1177/016344377900100202

Gerlitz, C., Helmond, A., Nieborg, D. B., & van der Vlist, F. N. (2019). Apps and infrastructures—A research agenda. *Computational Culture*, *7*. http://computationalculture.net/apps-and-infrastructures-a-research-agenda

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*(1), 129–152. https://doi.org/10.1146/annurev-soc-071913-043145

Helles, R., & Flyverbom, M. (2019). Meshes of surveillance, prediction, and infrastructure: On the cultural and commercial consequences of digital platforms. *Surveillance & Society*, *17*(1/2), 34–39. https://doi.org/10.24908/ss.v17i1/2.13120

Helles, R., Lomborg, S., & Lai, S. S. (2020). Infrastructures of tracking: Mapping the ecology of third-party services across top sites in the EU. *New Media & Society*, *22*(11), 1957–1975. https://doi.org/10.1177/1461444820932868

Hesmondhalgh, D. (2021). The infrastructural turn in media and internet research. In *The Routledge companion to media industries* (pp. 132–142). Routledge.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716674238

Innis, H. A. (2007). *Empire and communications*. Rowman & Littlefield.

International Telecommunication Union. (n.d.). *Universal broadband connectivity*. https://www.itu.int:443/en/action/broadband/Pages/default.aspx

Internet Exchange Map. (n.d.). *Homepage*. https://www.internetexchangemap.com

Kitchin, R., & Lauriault, T. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work* (Working Paper No. 2). The Programmable City Project. https://papers.ssrn.com/abstract=2474112

Kollnig, K., Shuba, A., Van Kleek, M., Binns, R., & Shadbolt, N. (2022). *Goodbye tracking? Impact of iOS app tracking transparency and privacy labels*. ArXiv. http://arxiv.org/abs/2204.03556

Krapiva, N., Tackett, C., Zhyrmont, A., Garfield, L., & Skok, A. (2022, August 18). *Updates: Digital rights in the Russia–Ukraine conflict*. Access Now. https://www.accessnow.org/digital-rights-ukraine-russia-conflict

Lai, S. S., & Flensburg, S. (2021). Invasive species of the app ecosystem: Exploring the political economy of mobile communication. *International Journal of Communication*, *15*(2021), 2301–2318. https://ijoc.org/index.php/ijoc/article/view/16906

Lam, C. (2017). FTTH deployment—Google fiber's perspective. In *Optical Fiber Communication Conference 2017* (Tu2K.1). Optica Publishing Group. https://doi.org/10.1364/OFC.2017.Tu2K.1

Lambiotte, R., & Kosinski, M. (2014). Tracking the digital footprints of personality. *Proceedings of the IEEE*, *102*(12), 1934–1939. https://doi.org/10.1109/JPROC.2014.2359054

Lee, C. P., & Schmidt, K. (2018). A bridge too far? Critical remarks on the concept of "infrastructure" in computer-supported cooperative work and information systems. In V. Wulf, V. Pipek, D. Randall, M. Rohde, K. Schmidt, & G. Stevens (Eds.), *Socio-informatics*. Oxford University Press. https://doi.org/10.1093/oso/9780198733249.003.0006

Lewis, K. (2015). Three fallacies of digital footprints. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715602496

Libert, T. (2015). Exposing the invisible web: An analysis of third-party HTTP requests on 1 million websites. *International Journal of Communication*, *9*(2015), 3544–3561. https://ijoc.org/index.php/ijoc/article/view/3646

Mann, M. (1984). The autonomous power of the state: Its origins, mechanisms and results. *European Journal of Sociology/Archives Européennes de Sociologie/Europäisches Archiv Für Soziologie*, *25*(2), 185–213.

Mansell, R. (2002). From digital divides to digital entitlements in knowledge societies. *Current Sociology*, *50*(3), 407–426. https://doi.org/10.1177/0011392102050003007

Mansell, R. (2017). Bits of power: Struggling for control of information and communication networks. *The Political Economy of Communication*, *5*(1), Article 1. https://www.polecom.org/index.php/polecom/article/view/75

Martínez-Santos, F., Frias, Z., & Escribano, Á. (2021).

What drives spectrum prices in multi-band spectrum markets? An empirical analysis of 4G and 5G auctions in Europe. *Applied Economics*, *54*(5), 536–553. https://doi.org/10.1080/00036846.2021.1967277

Mosco, V. (2009). *The political economy of communication* (2nd ed.). SAGE.

Munn, L. (2020). Red territory: Forging infrastructural power. *Territory, Politics, Governance*, *11*(1), 80–99. https://doi.org/10.1080/21622671.2020.1805353

Naughton, J. (2016). The evolution of the internet: From military experiment to general purpose technology. *Journal of Cyber Policy*, *1*(1), 5–28. https://doi.org/10.1080/23738871.2016.1157619

Nielsen, R. K., & Ganter, S. A. (2018). Dealing with digital intermediaries: A case study of the relations between publishers and platforms. *New Media & Society*, *20*(4), 1600–1617. https://doi.org/10.1177/1461444817701318

OECD. (n.d.). *Broadband portal*. https://www.oecd.org/sti/broadband/broadband-statistics

Parks, L., Sterne, J., Starosielski, N., Acland, C. R., Dourish, P., Harris, S., Holt, J., Mattern, S., Miller, T., & Sandvig, C. (2015). *Signal traffic: Critical studies of media infrastructures*. University of Illinois Press.

Plantin, J. C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, *20*(1), 293–310. https://doi.org/10.1177/1461444816661553

Plantin, J. C., & Punathambekar, A. (2019). Digital media infrastructures: Pipes, platforms, and politics. *Media, Culture & Society*, *41*(2), 163–174. https://doi.org/10.1177/0163443718818376

Routley, N. (2019, December 2). Wired world: 35 years of submarine cables in one map. *Visual Capitalist*. https://www.visualcapitalist.com/wired-world-35-years-of-submarine-cables-in-one-map

Sandvig, C. (2013). The internet as infrastructure. In W. H. Dutton (Ed.), *The Oxford handbook of internet studies* (Vol. 1, pp. 86–106). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199589074.013.0005

Sandvig, C. (2015). The internet as the anti-television: Distribution infrastructure as culture and power. In L. Parks & N. Starosielski (Eds.), *Signal traffic* (pp. 225–245). University of Illinois Press. https://www.jstor.org/stable/10.5406/j.ctt155jmd9.14

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134.

Starosielski, N. (2015). *The undersea network*. Duke University Press.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Velkova, J. (2019). Data centres as impermanent infrastructures. *Culture Machine*, *18*. http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-169006

Wasko, J. (Ed.). (2011). *The handbook of political economy of communications*. Wiley-Blackwell.

Webster, J. G. (2014). *The marketplace of attention, how audiences take shape in a digital age*. MIT Press.

Weltevrede, E., & Jansen, F. (2019). Infrastructures of intimate data: Mapping the inbound and outbound data flows of dating apps. *Computational Culture*, *7*. http://computationalculture.net/infrastructures-of-intimate-data-mapping-the-inbound-and-outbound-data-flows-of-dating-apps

Winseck, D. (2017). The geopolitical economy of the global internet infrastructure. *Journal of Information Policy*, *7*, 228–267. https://doi.org/10.5325/jinfopoli.7.2017.0228

Winseck, D. (2019). Internet infrastructure and the persistent myth of U.S. hegemony. In B. Haggart, K. Henne, & N. Tusikov (Eds.), *Information, technology and control in a changing world: Understanding power structures in the 21st century* (pp. 93–120). Springer. https://doi.org/10.1007/978-3-030-14540-8_5

Zuboff, S. (2019). *The age of surveillance capitalism. The fight for a human future at the new frontier of power* (1st ed.). PublicAffairs.

About the Authors

Sofie Flensburg is a tenure track assistant professor in the Center for Tracking and Society at the University of Copenhagen. Her research focuses on the evolution of the internet and digital political economies with a particular emphasis on the relationship between commercial infrastructure providers and state interests across geopolitical contexts. Contributing with empirical, comparative, and historical perspectives on the societal implications of digitalisation, she has published on digital communication systems, over-the-top distribution, and mobile datafication.

Signe Sophus Lai is a tenure track assistant professor at the Center for Tracking and Society at the University of Copenhagen. Her research is placed at the intersection between infrastructure studies, political economy of communication, and critical data studies. Cutting across established sectoral divides and research silos, her published work advances new methods for empirical research on the role of the internet in everyday life as well as the societal implications of big data, digital infrastructures, and emergent business models.

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 330–343
https://doi.org/10.17645/mac.v11i2.6400

Article

Google News Initiative's Influence on Technological Media Innovation in Africa and the Middle East

Mathias-Felipe de-Lima-Santos [1,2,*], Allen Munoriyarwa [3], Adeola Abdulateef Elega [4], and Charis Papaevangelou [5]

[1] Faculty of Humanities, University of Amsterdam, The Netherlands
[2] Digital Media and Society Observatory, Federal University of São Paulo (Unifesp), Brazil
[3] Department of Media Studies, University of Botswana, Botswana
[4] Department of Mass Communication, Nile University of Nigeria, Nigeria
[5] Laboratory of Applied Studies and Research in Social Sciences, University of Toulouse, France

* Corresponding author (m.f.delimasantos@uva.nl)

Submitted: 30 October 2022 | Accepted: 24 February 2023 | Published: 28 June 2023

Abstract
The Google News Initiative (GNI) aims to collaborate closely with the news industry and financially support the creation of quality journalism in the digital age. It also aims to bring technological advancements and innovation into newsrooms' operations. Drawing on journalism innovation and responsible innovation theories, this study examines GNI beneficiaries in Africa and the Middle East. To address this, we analysed GNI projects' descriptions combined with thirteen (n = 13) in-depth interviews with leading actors and beneficiary news organisations to answer two main questions: (a) What are the main characteristics of the technological innovations proposed by GNI Innovation Challenge grantees in Africa and the Middle East? and (b) How are these news media organisations becoming increasingly dependent on these platforms' technological and financial aspects? Anchored in journalism innovation, responsible innovation, and platformisation theories, our findings show that funded organisations heavily depend on Google's technological and financial infrastructure to innovate. Furthermore, we note that some projects do not offer a clear path for sustainability in the future. We further argue that this initiative builds an infrastructure of power and dependency that poses risks to responsible innovation in journalism. Our study contributes to extant scholarship on digital platforms and their role in the infrastructure of news organisations, creating power asymmetries between those who serve as the backbone for data flows and technological processes and those dependent on these institutions.

Keywords
artificial intelligence; business models; data; dependence; Google News Initiative; innovation; news; philanthrocapitalism; platform

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

In 2018, Google officially launched the Google News Initiative (GNI) programme, which was developed to foster collaboration with news institutions to achieve, as Google described, "a stronger future for news" (Schindler, 2018, n.p.). Broadly, the GNI aimed to collaborate closely with the news industry by offering financial and training support for creating quality journalism in the digital age. The programme revolves around three main objectives. First, GNI sought to elevate and strengthen quality journalism. Second, it sought to evolve journalism business models to drive sustainable growth. Last, the initiative sought to empower news organisations through

technological innovation. This programme section was structured at its inception through Digital Innovation Challenge, which focused only on European newsrooms. Since 2018, the GNI has globally expanded through its "Innovation Challenge" scheme to include hundreds of media organisations with a fund of over US $300 million, aiming to develop sustainable business models by diversifying revenue streams, generating audience engagement, and bringing emerging technologies to media organisations (Google, 2021). Thus, Google's long-term aim was to spur technological innovation and advancement in newsrooms worldwide.

The GNI comes at a time when journalism is facing an "institutional crisis" (Reese, 2021, p. iv). This critical point has assumed a variety of manifestations in the news industry, including the hard transition wrought by digitalisation (Paulussen, 2016) and the flight of advertisers from the institutional journalism spaces. As a result, it triggered an existential crisis that has seen newsrooms folding, cutting down on staff (Skowronski, 2009), being "juniorised" (Rodny-Gumede, 2014) and, in some instances, disappearing altogether. Specifically, journalism in the Global South has paved a more complex reality, often complicit in colonial regimes, such as in Africa and the Middle East (see Barratt & Berger, 2007). Faced with such existential threats, journalism institutions have embraced different funding models for survival, including philanthropic (Lugo-Ocando, 2020) and platform funding (Papaevangelou, 2023). These new sources of income attempted to close the financial gap created by the diminished advertisement revenue and cut-throat competition for existing advertisers with other institutions. When Google launched its GNI, it was a welcome source of funding for many journalism institutions (de-Lima-Santos & Mesquita, 2021a) that were already crippling under complex revenue models in ways that stifled innovation (Schindler, 2018). In the Middle East and Africa, about 43 organisations were beneficiaries of the GNI Innovation Challenge fund until 2021.

Through the conceptual lens of journalism innovation, responsible innovation (RI), and platformisation, we explore the GNI Innovation Challenge as a catalyst for journalism innovation in African and Middle East newsrooms. In doing so, we seek to understand how the GNI sought to elevate and strengthen quality journalism through technological innovations. Thus, the aims of this study are threefold. First, we discuss how the GNI Innovative Challenge programme pushes technological innovation in journalism to create "sustainable" business models. Second, to understand if projects funded by GNI have a clear path for sustainability in the future. Last, to examine if these projects have key dimensions of RI. Therefore, this article poses two specific research questions:

> RQ1: What are the main characteristics of the technological innovations proposed by GNI Innovation Challenge grantees in Africa and the Middle East?
>
> RQ2: Are these news media organisations becoming increasingly dependent on these platforms' technological and financial aspects? If so, how?

Findings show that the GNI Innovation Challenge builds an infrastructure of power and dependency that poses risks to the continuity of the developed projects in the region and, thus, of technological development. Furthermore, adopting emerging technologies does not bring key dimensions of RI, as it is challenging for most organisations to deploy them. Additionally, this program limits the deployment of these technologies to a certain extent in these countries. Our study contributes to extant scholarship on digital platforms and their role in the infrastructure of news organisations, creating power asymmetries between those who serve as the backbone for data flows and technological processes and those dependent on these institutions. Furthermore, as most scholarship studying platforms' funding for journalism has primarily focused on the Global North, specifically the European Union and the US, our study broadens this scope by homing in on an understudied geographic area with unique nuances and challenges.

2. Theoretical Grounding

Our research builds on three pillars: (media) innovation theory, the concept of RI, and literature on the dependency of news organisations on tech companies such as Google. We detail each topic in the following subsections.

2.1. Journalism Innovation: Disrupting Innovative Processes and Leading-Edge Technologies in Newsrooms

The increasing need for innovation in rejigging old media platforms, activating creative alterations in content production, and spurring the rise of distribution and commercialisation initiatives within the newsroom has become increasingly apparent in the face of institutional challenges that journalism has suffered since the digital disruption. Be that as it may, journalism innovation of some sort has been notably present in different aspects of both legacy and digital media landscape, which offers symbolic and practical comfort because "innovation is essential to the survival of the news industry" (Posetti, 2018, p. 8).

In a bid to establish the core tenets of innovation, Francis and Bessant (2005) outlined the "four Ps of innovation," which are novel *products*, new *processes* and modus operandi, new *positions*, and *paradigmatic* innovation in the guiding principles for the business model of organisations; all of which are broad groupings with unclear boundaries. Scholars have applied these four Ps to understand the innovation challenges in the news industry (de-Lima-Santos et al., 2022). As far as the journalistic world goes, these four Ps exist within or beyond media products (e.g., media platforms), media processes

(e.g., production and dissemination of media products, such as books, computer games, software, sound, and video recordings), media positions (e.g., brand identity, strategic positioning), and in media paradigms (e.g., models of revenue generation; Morlandstø, 2017).

In the classical Schumpeterian philosophy of innovation, there is a consensus that innovations can be either "incremental" or "radical" based on the extent of innovation and change or value addition (Storsul & Krumsvik, 2013). For journalism, most innovations, especially the early ones, were incremental in that they revolved around creative alterations in content production. However, technological developments over the past decades have spurred the industry to make radical changes within newsrooms and their various markets. For example, the internet and mobile devices have transformed the extant status quo of news media, like their communication model (Küng, 2013). Unlike legacy media, where the model is one-to-many (information gatekeepers), radical innovation has shifted towards a many-to-many model, where information is produced from and received by multiple channels or individuals/collectives (Belair-Gagnon et al., 2019; de-Lima-Santos & Mesquita, 2021b).

Recently, journalism witnessed radical innovation by embracing data practices and artificial intelligence (AI) systems (de-Lima-Santos et al., 2022). The adoption of emerging technologies in newsrooms is part of this journalism innovation, whereby organisations are "doing things (from the incremental to the transformative) that support the digital era development of journalism" (Posetti, 2018, p. 9). Automated news generation and algorithmic dissemination of news content fundamentally disrupt the journalistic culture and tradition (Lokot & Diakopoulos, 2016). Similarly, AI-oriented news tools, such as aggregators or fake news detectors, have emerged worldwide. For example, Tencent, a Chinese tech giant, introduced Dreamwriter in 2015, a news writing bot that many believed could lead to a new disruption in journalism (Kuai et al., 2022). More recently, OpenAI's cutting-edge tools, such as ChatGPT and DALL·E, indicate the potential of AI systems to automatically generate content based on text prompts.

While there is evidence of change, many scholars believe that the news media do not typically embrace radical (transformative) innovations, as most tend to hesitate to change newsroom rituals, procedures, strategies, and norms (Paulussen, 2016). Products and services with such a level of disruption (capable of replacing new ones) have been described as "creative destruction" in the traditional Schumpeterian literature (Hendrickx & Picone, 2020). Radical innovations are typically disruptive in nature; thus, they challenge the status quo, questioning old processes and impeding long-standing discoveries (de-Lima-Santos & Mesquita, 2021b). Other scholars have excused reluctance for various reasons, such as house cultures, lack of necessary resources, organisational tradition, legal requisites (Hodgkinson & Healey, 2011), and job loss (Munoriyarwa et al., 2021). Albeit to the reluctance and challenges, evidence from incremental and cumulative transformations over time shows that innovations in journalism continue to be essential in determining the field's current and future direction of the industry, as they might be a key to finding sustainable business models.

2.2. Thinking About Responsible Innovation in the News Industry

As the power of technology has become more evident, debates concerning responsibility have broadened (Stilgoe et al., 2013) to include benefits and harms, the dilemma of control (Collingridge, 1980), the development of pathologies of path dependency (David, 2007), and technological lock-in mechanisms (Arthur, 1989). Due to the limits of fully recognising the implications of innovations, the adverse effects often become evident with a considerable time delay. Thereby, some managers tend to be more contentious in embracing new technologies in their organisations, "leading to an incremental, not transformative change" (Voegtlin et al., 2022, p. 8). This unpredictability of innovation is inherently linked to its collective nature, where several stakeholders collaborate to develop it (Blaskó et al., 2014). Furthermore, adopters of innovations must deal with potential trade-offs between deploying emerging technologies in their organisations or lacking behind their competitors.

Approaches to RI aim to encompass this discussion by posing questions of uncertainty in multiple forms: purposes, motivations, social and political perspectives, sustainability, trajectories, and directions of innovation—particularly technological ones—as their designs can shape humans' lives by promoting or undermining specific values (van de Poel, 2009). In other words, research on RI promotes reflection on how to develop innovative processes in a transparent, interactive format so that societal actors and innovators become mutually responsive to each other with a view to the acceptability and sustainability of innovations by society and considering ethical values in their development (Von Schomberg, 2011).

In this view, four principles can be adopted to promote RI in organisations: Anticipation, Reflexivity, Inclusiveness, and Responsiveness. Anticipation is a process that faces tensions between prediction—which tends to draw particular futures—and participation, seeking to open them up to foresee potential risks, dangers, and public concerns. Reflexivity means reflecting on underlying purposes and motivations to explore innovations' impacts on society within territorial contexts. Inclusiveness is the interactive process of engaging the public and diverse stakeholders to open discussions, raise dilemmas and provide an open space to create solutions to the underlying problems of innovation. Responsiveness considers innovation's subsequent trajectory and pace to ensure its proper continuity rather

than "just another form of window-dressing" (de Hoop et al., 2016, p. 111).

In journalism, when emerging technologies are deployed, organisations must be aware of the potential risks of developing projects that are not sustainable in the long run (Voegtlin et al., 2022). Furthermore, the responsibility to do no harm should be part of the innovation process. Adhering to this line of reasoning, we contend that RI can best be conceptualised as an endorsement of relevant public values during innovation (Taebi et al., 2014), which are aligned with journalism practices.

Innovation involves creating value from ideas, which subsequently includes establishing relationships with stakeholders to facilitate its incorporation. Tech companies are essential and necessary social change agents (Aguilera et al., 2007), particularly in journalism innovation. However, their role in society comes with relevant responsibilities, like mitigating harmful practices and having reliable governance comprising institutions, structures, and procedures on multiple levels. In this respect, tech companies have long been criticised for leading many news media's business models to fail (Rashidian et al., 2018). In the RI's view, tech companies might find two paths for developing technological innovations in news media: changing the design to accommodate conflicting values or making a value trade-off deciding what should take priority in the design. The key for RI is to find and maintain the right balance between the benefits of development and social disadvantages (Voegtlin et al., 2022).

RI also has its limitations. For example, ethical elements are anchored to circumscribed territorial spaces, as different objects and social situations are ruled by other normative systems (Blaskó et al., 2014). However, material barriers can limit innovations in certain conditions. Innovations can also require abandoning or reducing engagement with various existing practices, which might have cultural roots that are not acknowledged. Additionally, even responsibly, innovation can exacerbate power imbalances as some advancements may depend on specific individuals or groups. As a result, these various stakeholders may have conflicting and opposing goals, making it difficult to develop an effective innovation strategy and therefore hindering the implementation of RI (de Hoop et al., 2016; Voegtlin et al., 2022).

2.3. The Complicated Relationship Between Digital Platforms and News Media Organisations

Digital platforms have contributed to the transformation of news content's online distribution. Many news publishers have largely become dependent on platforms as crucial traffic sources, raising concerns regarding, among others, the monetisation of news content. Additionally, digital platforms' recommendation engines use advanced machine learning algorithms to analyse individual and aggregate user data to deliver the "most relevant" news content, changing audiences' behaviours through filtering and bundling content (Capobianco, 2021).

This phenomenon has broadly resulted from platformisation, that is, "the penetration of online infrastructures, economic processes, and governmental frameworks of online platforms in multiple socioeconomic sectors and spheres of existence" (Poell et al., 2019, pp. 5–6). In other words, the infrastructural status that these platforms have acquired has permitted them to extend their reach in a myriad of domains, making them omnipresent in our online activities (Plantin & Punathambekar, 2019). Journalism has not been able to avoid the impact of platformisation, influencing many facets of editorial processes. Conversely, platforms rely on publishers to exercise "platform power," which is "contingent on [platforms'] ability to maintain relations and sustain them over time" (Nielsen & Ganter, 2022, p. 22).

Through these configurations, platforms have shown how they can enhance publishers' reliance on their services to innovate. However, studies have demonstrated that news outlets are potentially exposed to dangers caused by unanticipated changes in platforms' algorithms or business interests (Nielsen & Ganter, 2022). Therefore, the relationship between platforms and publishers has become particularly complicated. Despite that, both parties desire to collaborate, yet with significant reservations, particularly from the publishers' side, as they are becoming overly dependent on these tech companies. To this end, publishers have been attempting to counterbalance this by reconfiguring their resources' investment in platform services (Meese & Hurcombe, 2021) to "wrangle back control of their audiences, data, and revenues" (Chua & Westlund, 2022, p. 82).

Researchers have also approached this issue from the standpoint of editorial autonomy in the face of online platforms' algorithmic and automated content curation (Simon, 2022). This evokes issues of media capture, a frame used to describe situations where the dependency of news media organisations on other influential stakeholders, such as platforms, might dilute their role of holding power to account (Schiffrin, 2021). However, media capture should not only be treated as a threat to editorial autonomy, which can largely remain intact (Poell et al., 2022). It can also be considered a risk to news media organisations' infrastructural autonomy and innovation capacity (Nechushtai, 2018).

Therefore, crucial to understanding the power asymmetry that underpins the examined relationship is the concept of "infrastructural capture," which describes "situations in which an organisation tasked with scrutinizing another organization, institution, business, or industry is incapable of operating sustainably without the resources or services they provide" (Nechushtai, 2018, p. 1046). Looking at the technological innovation capacity of news media, they became constrained by reinforcing structural advantages of platforms (de-Lima-Santos & Salaverría, 2021). For example, these companies have

not only "become dominant in AI research and provision" but have also made it "difficult for many news organizations to develop AI without having to rely on tools and infrastructures provided and maintained by these companies" (Simon, 2022, p. 7). These concepts provide us with the conceptual tools to understand the implications of funding concerning innovation processes behind the digital platforms' aid to news organisations.

3. Methods

The GNI Innovation Challenge has conducted five regional innovation challenges, funding over 200 projects in 47 countries. According to Google, this scheme aims to "empower news organisations from around the world that pioneer new thinking in online journalism, develop new paths to sustainability, and better understand their communities" (GNI, 2022). To understand the realities and challenges faced by GNI beneficiaries in Africa and the Middle East, specifically regarding the adoption of innovative processes in their newsrooms, this study followed a multi-method qualitative research approach based on the analysis of the projects descriptions available on the GNI portal triangulated with semi-structured and in-depth interviews with 13 leading actors in selected organisations. The interviews were made between July and October 2022 and were conducted and recorded via Zoom. On average, they lasted 42 minutes. Table 1 lists the GNI beneficiaries interviewed and the place they are located. Broadly in these interviews, we sought to understand issues around their dependency on Google as an organisation, the kind of innovation they were supported to undertake, and the project's sustainability post-GNI funding. Answering these questions helped us to understand, at a broader level, the power dynamics that link the GNI to its beneficiaries.

The data-gathering process was fraught with challenges. These organisations were selected based on the list available on the GNI website. The final list of 13 beneficiaries is based on a snowball sampling to identify others, as it was not possible to interview representatives of all GNI beneficiaries due to the unavailability of potential interviewees, as many of our repeated requests went unanswered. Some contacts declined our request, unaware of the GNI project that their organisation had partaken in, while others retracted their participation at the last minute. Thus, we did not follow a purposive sampling strategy. Considering these limitations, we aimed for geographical representation. We secured interviews from at least one representative from North and Central Africa, and the Middle East. In addition, we sought to cultivate a representative sample that reflects the ideological diversity and variety of media types (e.g., digital native and traditional print-first outlets).

For our data analysis, we conducted an inductive thematic analysis after all authors transcribed the interviews. This is a widely used method to draw themes from qualitative data, particularly in datasets composed of interviews (Braun & Clark, 2006). With an inductive approach, the identified themes emerged from the data themselves without trying to fit them into a pre-existing coding frame or the researcher's analytic preconceptions. The inductive thematic analysis used for this study was performed using NVivo, a common software to assist in qualitative research. After that, we reported our findings combining them with our theoretical framework to create a thematic narrative.

Table 1. A breakdown of the organisations and their geographical location whose representatives were interviewed.

Code	Organisation	Country	Type of Organisation
R1	Egab (Official incorporated name: Egab for Digital Content)	Egypt	Digital native media
R2	Africa Uncensored	Kenya	Digital native media
R3	TelQuel Digital	Morocco	Digital native media
R4	Richmond Hill Media Limited (Ripples Nigeria)	Nigeria	Legacy media
R5	Stears News Limited (Operating Company), Stears Information Services (Holding Company)	Nigeria	Legacy media
R6	Food For Mzansi, a digital news platform of Farmers For Change (Pty) Ltd	South Africa	Digital native media
R7	263 Chat	Zimbabwe	Digital native media
R8	WhiteBeard	Lebanon	Start-up
R9	Daraj Media	Lebanon	Digital native media
R10	Community Media Network	Jordan	Digital native media
R11	Sowt Podcasting and Training L.L.C.	Jordan	Digital native media
R12	L'Orient-Le Jour/Société Générale de Presse et d'Édition SAL	Lebanon	Legacy media
R13	Raseef22 dba Levant Laboratories SAL	Lebanon	Digital native media

Note: Codes were used in the presentation of findings.

4. Findings

In Africa and the Middle East, GNI Innovation Challenge had granted funds for 43 projects until 2021, 22 (51,16%) in 2019 and 21 (48.84%) in 2021. These projects were concentrated in 18 countries in this geographic region. As shown in Figure 1, many countries were not contemplated with these grants, expanding the power asymmetry in this region.

The datafication and platformisation of the digital infrastructure create power asymmetries between those who embrace technological knowledge and those who are surviving in the digital age (Nielsen & Ganter, 2022; Poell et al., 2022). Using the typology proposed by de-Lima-Santos and Mesquita (2023), we classified these projects according to their aims to identify these discrepancies. According to the project description, our categorisation focused on aspects of the data we were most interested in exploring. To avoid projects overlapping multiple categories, we followed the priority based on what was mentioned. First, if the description states the use of emerging technologies, such as AI or immersive tools, it would be classified as Technological Innovation. Second, if the project focused on reaching new or niche audiences, Audience Building was used. Lastly, the Business Model applies to projects that mention new revenue streams or strategic goals.

According to this typology, most of these projects focused on developing a business model for these organisations (about 42%). To a lesser extent, these projects aimed to bring emerging technological innovations (30%) or new audiences (28%). As shown in Table 2, the media organisations that used GNI grants to introduce novel technologies in their newsrooms were predominantly found in Israel and Jordan, which might suggest a more enabling environment for digital media infrastructures. As stated in their descriptions, most projects aimed to bring AI solutions to newsrooms.

4.1. The Reflexivity and Motivations Behind Technological Innovations

Having a broader view of these projects and looking at smaller geographic regions, we can see that the Middle East led the emerging technologies development. For example, the Turkish company Demirören Media proposed an AI system that categorises news content and offers personalised options to readers on the topics they are interested in, aiming to increase readers' engagement on its platforms through microsegment level. Similarly, the Jordan news outlet Al Bawaba proposed a solution to leverage its digital archive using Google Cloud and a trained AI system capable of semantically understanding and tagging Arabic content.

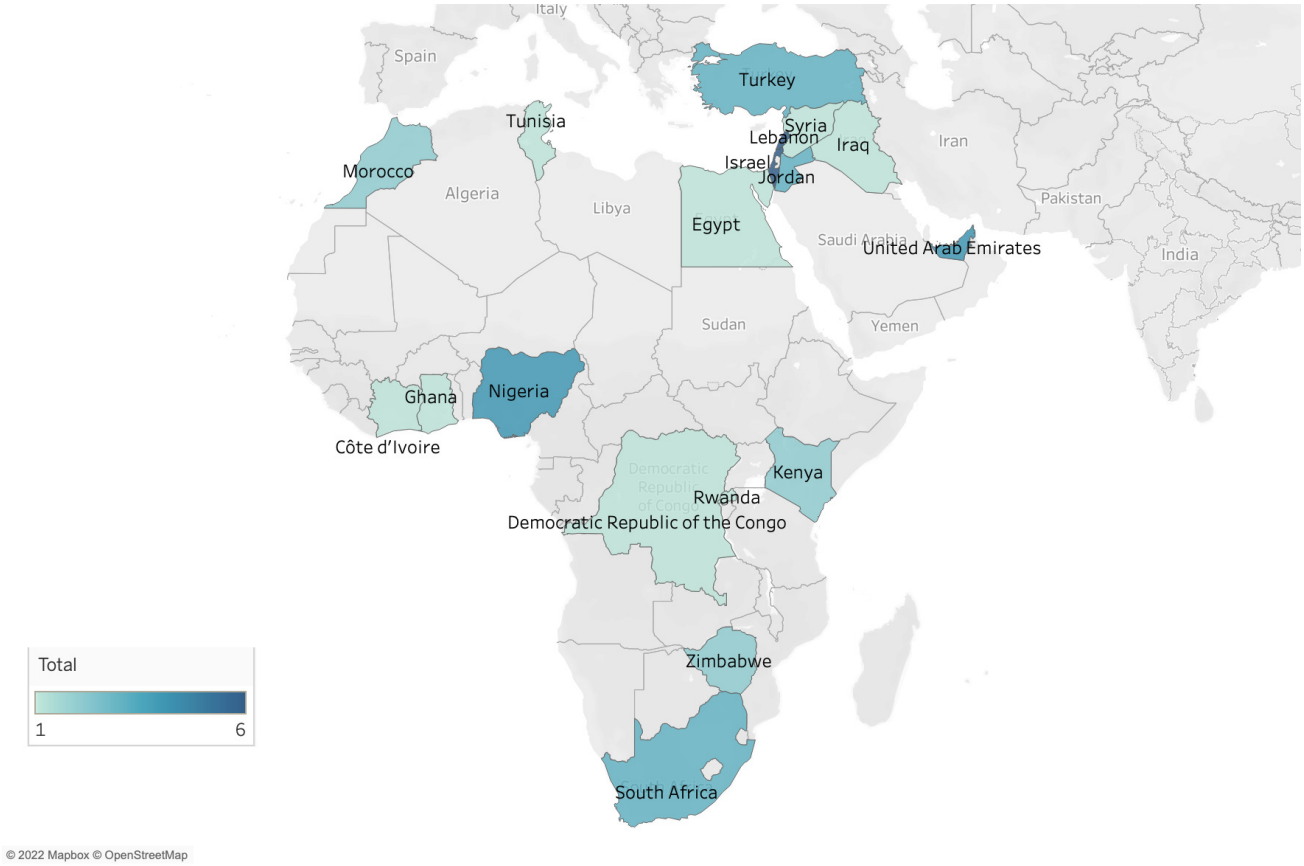


Figure 1. Countries awarded by GNI Innovation Challenge grant in Africa and the Middle East.

Table 2. Typology of projects in different countries in Africa and the Middle East.

Typology	Country	Count
Technological Innovation	Israel	5
	Jordan	2
	United Arab Emirates	1
	Syria	1
	Nigeria	1
	Morocco	1
	Kenya	1
	Iraq	1
Audience Building	United Arab Emirates	3
	Zimbabwe	2
	Turkey	2
	South Africa	2
	Rwanda	1
	Nigeria	1
	Lebanon	1
Business Model	Lebanon	5
	Nigeria	2
	Turkey	1
	Tunisia	1
	South Africa	1
	Morocco	1
	Kenya	1
	Jordan	1
	Israel	1
	Ghana	1
	Egypt	1
	Côte d'Ivoire	1
	Congo	1

Other organisations embraced the logic of aggregation and automation to negotiate transactions between markets and social entities, informing the public about governance systems and institutions. For example, Community Media Network (Jordan) used Google's grant to build Rabet, which means "link" in Arabic, a platform that aggregated various data of the top 500 Jordanian officials and members of parliament, former ministers, and influential figures (such as writers, party people, artists, and sports people) to show the connection between them, based on their family, work, and business. According to R10, this tool "allows us to have more transparency when appointments are made at any level." For example, one law in Jordan states "that a member of parliament or the government is not allowed to own any company that deals with the government" nor can "be a minister or a member of Parliament and then apply for a tender or work tender" (R10). This allows them to investigate power abuses in the government.

In North and East Africa, organisations focused on developing solutions that could help them to create a sustainable business model, such as an online platform to structure, validate, and enable their projects. According to R1, GNI Innovation Challenge helped her to scale her business from a minimum viable product to a valuable, scalable process that drives growth. Egab is a platform that connects local journalists across Africa and the Middle East and editors of regional and international media outlets to pitch stories in any format or language. Focusing on solutions journalism, the platform's big motto is "learning by doing" (R1), which includes feedback for journalists to learn and implement in the next pitch:

> We reject many pitches, but the difference is that we say why. We want the journalist to learn. I always say we're a business. Yes, we are for profit, but we have solid social goals. Our ultimate goal is for these journalists to be able to pitch and produce stories without our help. (R1)

For two years, "instead of putting a lot of money into building an online platform that no one would use, my idea was to start with emails and a Google workspace. This has been our operation until we got the grant" (R1).

In Kenya, Africa Uncensored brought new ways to produce news content by giving voice to communities not always heard by the news media. Focusing on

crowdsourcing information from people living in informal settlements in Nairobi, Africa Uncensored created a direct channel for the public to air their issues while providing a channel for the organisation to map these issues and better cover them: "We decided this channel will be via SMS because it is very cheap in Kenya. It's accessible to many people, and many mobile plans have it for free" (R2). The solution created by Africa Uncensored would aggregate these messages received from the public in a portal that journalists could check and build stories from these texts and images.

In Southern Africa, examples of building new audiences are commonly found. Focused on citizen journalism, Food For Mzansi (South Africa) targeted three agricultural communities in its pitch for Google, aiming to increase the literacy of South African youth and young small-scale farmers in these regions by involving them in news reporting processes. The organisation was in a "start-up" phase when its team saw the GNI call:

> We are three and a half years old at the moment. When we applied, we were in our first or early second year. As a start-up, we were always looking for cash, which is very limited [in the news industry]. So, we're always looking around for grant opportunities. (R6)

Similarly, the news outlet 263Chat (Zimbabwe) proposed to build new audiences by presenting an alternative to radio by establishing a podcast network, as it was eager to find another way to create audiences beyond its website limited to those people who have internet access. 263Chat has complemented its offer by creating an e-paper—which is sent out to its "54,000 subscribers daily (Monday to Friday)" (R7)—, an SMS platform, and a podcast, allowing Zimbabweans access to its content in different formats and not always requiring internet connection.

4.2. The Challenges of Anticipating the Use of Technological Innovations in Newsrooms

Our respondents highlighted several challenges to implementing technological innovations in their countries. A common hurdle among our respondents is the lack of skills to develop these emerging technologies in their countries. Stears (Nigeria) used the grant to create a billing infrastructure to collect regular payments from readers, mainly focusing on "integrating with African payment gateways and receiving payment from African audiences" (R5). Therefore, it is essential to find "the right talent," which is not an easy task, as these professionals need "to complement particular parts of the [media] business" (R5).

Our interviewee from the Lebanese organisation Daraj Media shared the same feeling. R9 told us that the news outlet "took literally more than half of the grant duration to figure out a team and who we are going to work with." LIFT-im is an innovation lab based at Daraj Media, aiming to bring emerging technologies to newsrooms. It is a project that requires high knowledge of advanced technologies such as AI:

> We came across a company in Jordan that is doing outstanding work. They proposed to work with us on deploying AI solutions that we imagined, but it was costly. We would have to pay much more than what we had received from Google. (R9)

An in-house team was also expensive for the organisation, as salaries for tech professionals tend to be higher. Besides that, there is intense competition from foreign companies and organisations outside the news industry for technologists.

As a result, it took a long time for the organisation to figure out whom to work with, as it required the team to know about AI, be native English and Arabic speakers, and have a genuine interest in media: "You need to tick too many boxes. If they exist, they are already employed and we can't afford them," explained R9. The solution came up through a collaboration between academia and Daraj. The news organization found a professor based in Paris (France) with a team in Beirut who agreed to become part of the LIFT-im lab and work together on these projects. The team has developed an AI-driven tool to look at Twitter in real-time to detect tweets generated by bots from the ones created by actual people, helping journalists to analyse how bots lead the conversation and what kind of impact they might have in the public discourse in Lebanon.

Other organisations relied on third-party companies to support them in developing their proposals. The Lebanese WhiteBeard is a team of engineers, designers, and managers who offer software solutions and insightful guidance for companies, mainly specialised in the news industry, as one of its co-founders worked for many years for the newspaper *L'Orient-Le Jour* (Lebanon). This tech company was responsible for helping other news organisations to deploy their innovative solutions using their GNI grants, such as *L'Orient-Le Jour*, Nida al Watan, and Rasseff22. Whitebeard also received a GNI grant to develop a Customer Relationship Management solution for smaller news outlets, allowing these newsrooms to manage subscriptions better using a tool that combines a metered paywall and a locked system based on data signals from audiences.

The lack of technological knowledge in the region might not have been anticipated by these organisations or Google, which did not work with them to develop these projects. For example, Citizen Bulletin also relied on other organisations to build the project. Once the existing funding was over, they folded.

All respondents bemoaned that their GNI were developed independently without help from Google. We learned that the tech giant does not support these news organisations with the needed skills for deploying these technologies, and the meetings are limited

to checking status. Some respondents mentioned that Google outsourced the whole process to third-party companies. It must be said that others noted that, in cases when they had to deal with former journalists working at Google, they had a more positive experience as former journalists, "they understood our reality better" (R9). Additionally, a few respondents noted that the process was very smooth compared to other funds they got.

4.3. The Inclusiveness of Technological Innovations and Its Limitations

By creating more diverse project teams or involving different stakeholders, some projects brought this inclusiveness approach to their projects. Outsourcing is a tool that helps in this process. However, specific organisations, particularly those with headquarters abroad, ended up developing their projects in other countries outside Africa or the Middle East. For example, Legit.ng (Nigeria) did not have information about the ReCo project, a content recommendation tool, as "it was made overseas," in its office in Ukraine, according to a representative who did not agree to be interviewed by us. Pulse.ng (Nigeria) is owned by the parent company, Ringier, based in Switzerland. We contacted a representative who also knew nothing about the project. In this aspect, the inclusiveness of these projects reveals to be poor. In some respects, being led and developed by organisations in Western countries, some projects limit their possibilities to contribute to local development and mitigate the low level of technological development in the region.

Other organisations brought an inclusive spirit to their projects by engaging new publics or providing an open space to create solutions for their problems, such as Africa Uncensored (Kenya), 263Chat (Zimbabwe), and Egab (Egypt). Similarly, EcoNai+ from Ripples Nigeria promises to track and mark changes to environmental phenomena using geo-journalism and crowdsourcing data. This platform allows users to collect, visualise, and report on data from communities impacted by climate change. According to R4, EcoNai+ is an ecosystem of:

> A couple of tools to help across the value chain of the environmental report, tracking, data capture or unreported to help journalists, researchers, scientists, policy formulators, as well as community members, convert their fears and their worries and anxieties about the environment into actual data that can help to drive and attract intervention for change. (R4)

By looking at "the most disadvantaged and underserved communities in a country" (R4), EcoNai+ brings this aspect of inclusiveness of technological innovations. However, it comes with limitations: "Geo-journalism involves a lot of technical training, acquisition of skills, and some technical tools," requiring training for people to learn how to use these tools. Furthermore, technological structures involve multiple owners, actors, and stakeholders that embroil the datafication and technical processes (Parks & Starosielski, 2015). To avoid it, GeoViz+, a tool to visualise data, relies on an easy-to-use approach of out-of-the-box tools, such as Flourish and DataWrapper (see de-Lima-Santos et al., 2021).

Equally important is to think about the technologies available to citizens. The reliance on SMS by Africa Uncensored helps to reach wider audiences in Kenya, as it "is very cheap, and a lot of people have in their mobile plans for free" (R2). Food For Mzansi targeted three agricultural communities as the organisation sought to create "community impact" (R6) by giving voice to these African youth and young small-scale farmers in these regions. The same goal had TelQuel Digital using the grant to "create podcasts focused on Moroccans living abroad" (R3).

4.4. The Responsiveness and Continuity of These Projects

All these grants are co-funded, meaning these news outlets still need to invest money into developing these projects. While some organisations use staff hours as part of the co-funding scheme, others had to invest money to develop these projects, as the grants were insufficient to cover them. The interviewees did not precisely describe the co-funding mechanism. While some mentioned that Google sponsored 70% of the project, others said it was 60% or 80% for them. Given this lack of common standards, it is hard to understand how Google decides on the grant's value. Some respondents bemoaned that they requested more funds in their application, but Google decided to give them less.

Some technologies were developed to a broader scope. EcoNai+, for instance, has the mission to contribute to solving the problem of climate change using technological media innovation. For R4, this solution allows Ripplers to become a "media tech company," preparing for the industry's future and potentially touting new revenue streams. Africa Uncensored also saw the potential of its tool during the Covid-19 pandemic for health function. Therefore, the team decided to explore this function to fundraise it for then expand it to other scopes.

Conversely, other organisations have yet to finalise their development even with the end of the grant period. For example, Community Media Network has not yet made Rabet available. According to R10, the political scenario in Jordan hampered its release:

> I don't know how successful it will be. We're going through two different problems. First, the space for civil societies is shrinking in our country. We are under a lot of pressure, myself and my organisation, and the tax people suddenly start to be interested in us, and we're facing a lot of bureaucracy. We're worried that if we put this up online publicly, we will get in further trouble. We're trying to limit the problem.

> On the other hand, we might have a new government. We want this to be available, at least to key editors and journalists, so they can use the information. (R10)

As technologies require constant reassessments and upgrades to keep pace with changing circumstances and the evolution of the industry, it is also essential that news outlets can respond adequately and timely to them. However, most organisations do not seem to have a clear path to continue developing these projects. Half of the organisations interviewed mentioned that they have applied or will apply for further funding from GNI Innovation Challenge. Some respondents said they would apply for other GNI grants, such as Equity Fund, to continue their projects. These answers stress the technological innovations' pathologies emerging in the news industry, such as path dependency (David, 2007) and technological lock-in mechanisms (Arthur, 1989).

Equally problematic was that some organisations mentioned the need to fundraise money from philanthropic institutions to continue developing these projects. This clearly shows that some projects are not yet sustainable. As a result, news outlets are not prepared to carry on these projects, putting at risk their continuity and clearly showing that Google did not help these organisations to mitigate the costs of these technological innovations. R2 clearly stated: "In the end, we realised it would be costly running this project after the grant is over. So, we decided we are going to continue it for a little bit."

5. Discussion and Conclusions

Our study shows a discrepancy in the deployment of emerging technological innovations in Africa and the Middle East. While North and West African news outlets rely primarily on the traditional use of technologies to create or develop tools to support their organisations' business models, in the Middle East, more emerging technologies are being deployed with GNI Innovation Challenge grant, particularly in Israel, Jordan, and Lebanon. In Southern Africa, the focus is on building new audiences. These different approaches reflect on the level of technological development in these regions and show how these innovations have a diverse range of solutions, from the most advanced (e.g., AI and immersive technologies) to the simplest ones (e.g., SMS integration systems and online portals). This shows how diverse and complex this geographical region is.

In the Global South, philanthropic routines lead news organisations, notably smaller and independent ones, to concentrate almost exclusively on funds provided by these institutions to sustain their business (Lugo-Ocando, 2020). While news organisations fail to generate enough revenue streams to create sustainable business models, the reliance on "Silicon Valley for funding and organisational imperatives" (Poell et al., 2022, p. 12) initially appeared as a promising path for sustainability. By giving this one-year grant, Google expects news organisations to solve their long financial sustainability problem and adopt technological innovation that will disrupt their business models and put them on the path to sustainability. Google seems to frame journalistic innovation as achievable only through its proprietary and technological capacities. However, as some of these organisations highlighted, there are limitations to developing these projects in the region due to the lack of knowledgeable IT personnel, high hiring costs, and reliance on third-party vendors. As a result, most projects ended up being minimum viable products of their original idea.

Furthermore, Google expects news outlets to co-fund these projects for one year. Many news organisations do not have the resources to co-finance these projects, which could indicate the lack of grantees in many African countries. Conversely, as some respondents mentioned, these GNI grants leave a feeling of validation for these pitches, giving the hope that these technological innovations will help these organisations in the future.

Thus, tech companies set the terms and conditions, leading news organisations to adapt incessantly to their needs. This shows how the power and functions provided by "Big Tech" platforms, such as Google and Facebook, continue to pervade news organisations, sustaining power asymmetries. As a result, to develop journalistic innovation in their newsrooms, grants such as the GNI Innovation Challenge seem to be the path of least resistance. However, what Google does with the GNI Innovation Challenge is an extension of "philanthrocapitalism," which Bishop and Green (2008) define as private wealth that "can advance the public good by applying entrepreneurial skills, speed, and score-keeping to our most persistent challenges" (p. ix). Similarly, these distinct regional characteristics demonstrate that philanthrocapitalism does not solve the inherent journalistic institutional crisis (Reese, 2021).

We believe that these grants could have the potential to become a global benchmark. For this, it is necessary not simply to offer money for the development of digital infrastructures, but also to provide support for these organisations during and beyond the development process, allowing them to fully launch a solution that offers maximum value to boost their overall presence in the digital news ecosystem. In other words, the design and deployment of the awarded projects should have decision support from Google, which could help news outlets to develop their technological innovations in good faith and with careful approaches, following RI principles (Voegtlin et al., 2022). Thus, these projects could guarantee new revenue streams for these organisations, while also bringing them to a sustainable path.

However, the examples presented in this study, such as Al Bawaba's proposal to utilise Google Cloud for its archive, demonstrate how news outlets increasingly rely on platforms' infrastructures to build technological

solutions (de-Lima-Santos & Salaverría, 2021; Parks & Starosielski, 2015). This situation further reinforces the infrastructural dependency of journalism on tech giants (Nechushtai, 2018). Additionally, the beneficiaries' decision to outsource critical parts of the projects to third-party companies did not allow them to acquire valuable knowledge that could help them become more independent. We contend that it further fuels the underlying power asymmetries between news media and platforms. As a result, dependency is not limited to infrastructure and resources but also knowledge, networks, and expertise.

Consequently, this article contributes to the critical discussion concerning the issue of news media organisations' capacity to innovate in an environment where large tech companies effectively control many of the tools and processes required to do so. Our study focuses on overlooked regions, namely Africa and the Middle East, where journalists are often in dire need of funding, thus risking a further entrenchment of contingency on platforms. Last, we wish to contribute to the ongoing dialogue about how platforms' programs could be shaped and executed to empower newsrooms. Even if it might not always be possible to prevent adverse effects from occurring, it is at least helpful to anticipate them, be responsive, and attempt to mitigate their impacts as much as possible.

Due to the limitations of language and vast scope, we could not reach out to every Google-funded organisation in the region. Although this limits our study, as it does not represent the entire variety of technological innovations adopted by these grantees, we combined methods to achieve a representative model that depicts the essential standard features to understand the influence of the GNI Innovation Challenge across the region. Similarly, some organisations were afraid of sharing data about these projects, as they had signed non-disclosure agreements with Google, restricting what they could share. Future studies could explore how these news outlets sustain some of these projects after the grant period and how these organisations continue developing emergent technologies in their newsrooms. A comparative analysis of the GNI Innovation Challenge between the Global South and a more privileged market, such as North America and the European Union, could also illuminate particular beneficiaries' treatments by Google. Despite these limitations, our study adds to the existing literature by demonstrating the power asymmetries between those who serve as the backbone for technological innovation processes and those dependent on these institutions.

In conclusion, these projects help us understand the challenges news outlets experienced in Africa and the Middle East. We can also perceive the processes involved in developing emerging technological innovations in a diverse geography area, contributing to broader stakeholders' visions of RI and helping them to adopt best practices that could empower them to create better solutions. Furthermore, the power dynamics embedded in these projects cannot be ignored as they influence the levels and trajectories of innovation dependency that bind Google and the project beneficiaries. Overall, this study demonstrated that dependence and power imbalance might negatively affect RI in the news industry.

Acknowledgments

The authors would like to thank Dr Arwa Kooli for assisting us during the data collection and all the interviewees who took the time to share their knowledge and experiences with us. Furthermore, this study was partially funded by the University of Amsterdam's RPA Human(e) AI and by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska Curie grant agreement No 765140.

Conflict of Interests

The authors declare no conflict of interests.

References

Aguilera, R. V., Rupp, D. E., Williams, C. A., & Ganapathi, J. (2007). Putting the S back in corporate social responsibility: A multilevel theory of social change in organizations. *Academy of Management Review*, *32*(3), 836–863. https://doi.org/10.5465/AMR.2007.25275678

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, *99*(394), 116–131. https://doi.org/10.2307/2234208

Barratt, E., & Berger, G. (Eds.). (2007). *African media since Ghana's independence*. Paarl Print. https://guyberger.ru.ac.za/fulltext/50years.pdf

Belair-Gagnon, V., Nelson, J. L., & Lewis, S. C. (2019). Audience engagement, reciprocity, and the pursuit of community connectedness in public media journalism. *Journalism Practice*, *13*(5), 558–575. https://doi.org/10.1080/17512786.2018.1542975

Bishop, M., & Green, M. (2008). *Philanthrocapitalism: How giving can save the world*. Bloomsbury Publishing USA.

Blaskó, B., Lukovics, M., & Buzás, N. (2014). Good practices in responsible innovation. In N. Buzás & M. Lukovics (Eds.), *Responsible innovation* (pp. 179–191). SZTE GTK.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Capobianco, A. (2021). *News media and digital platforms—Note by BIAC*. OECD. https://one.oecd.org/document/DAF/COMP/WD

Chua, S., & Westlund, O. (2022). Platform configuration: A longitudinal study and conceptualization of

a legacy news publisher's platform-related innovation practices. *Online Media and Global Communication*, *1*(1), 60–89. https://doi.org/10.1515/omgc-2022-0003

Collingridge, D. (1980). *The social control of technology*. St. Martin's Press.

David, P. A. (2007). Path dependence: A foundational concept for historical social science. *Cliometrica*, *1*(2), 91–114. https://doi.org/10.1007/s11698-006-0005-x

de Hoop, E., Pols, A., & Romijn, H. (2016). Limits to responsible innovation. *Journal of Responsible Innovation*, *3*(2), 110–134. https://doi.org/10.1080/23299460.2016.1231396

de-Lima-Santos, M. F., & Mesquita, L. (2021a). In a search for sustainability: Digitalization and its influence on business models in Latin America. In R. Salaverría & M. F. de-Lima-Santos (Eds.), *Journalism, data and technology in Latin America* (1st ed., pp. 55–96). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-65860-1_3

de-Lima-Santos, M. F., & Mesquita, L. (2021b). Data journalism beyond technological determinism. *Journalism Studies*, *22*(11), 1416–1435. https://doi.org/10.1080/1461670X.2021.1944279

de-Lima-Santos, M. F., & Mesquita, L. (2023). *Google News Initiative Innovation Challenge in Latin America: Business models between path dependence and power relations*. [Manuscript submitted for publication]. University of Amsterdam; Federal University of São Paulo; Dublin City Univerity.

de-Lima-Santos, M. F., Mesquita, L., de Melo Peixoto, J. G., & Camargo, I. (2022). Digital news business models in the age of industry 4.0: Digital Brazilian news players find in technology new ways to bring revenue and competitive advantage. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2037444

de-Lima-Santos, M. F., & Salaverría, R. (2021). From data journalism to artificial intelligence: Challenges faced by La Nación in implementing computer vision in news reporting. *Palabra Clave*, *24*(3), Article e2437. https://doi.org/10.5294/pacla.2021.24.3.7

de-Lima-Santos, M. F., Schapals, A. K., & Bruns, A. (2021). Out-of-the-box versus in-house tools: How are they affecting data journalism in Australia? *Media International Australia*, *181*(1), 152–166. https://doi.org/10.1177/1329878X20961569

Francis, D., & Bessant, J. (2005). Targeting innovation and implications for capability development. *Technovation*, *25*(3), 171–183. https://doi.org/10.1016/j.technovation.2004.03.004

Google. (2021). *How Google works with the news ecosystem*. https://blog.google/documents/105/How_Google_Works_with_the_News_Ecosystem_July_2021.pdf

Google News Initiative. (2022). 2021 *impact report*. https://newsinitiative.withgoogle.com/impact2021

Hendrickx, J., & Picone, I. (2020). Innovation beyond the buzzwords: The rocky road towardsa digital first-based newsroom. *Journalism Studies*, *21*(14), 2025–2041. https://doi.org/10.1080/1461670X.2020.1809494

Hodgkinson, G. P., & Healey, M. P. (2011). Psychological foundations of dynamic capabilities: Reflexion and reflection in strategic management. *Strategic Management Journal*, *32*(13), 1500–1516. https://doi.org/10.1002/SMJ.964

Kuai, J., Ferrer-Conill, R., & Karlsson, M. (2022). AI ≥ journalism: How the Chinese copyright law protects tech giants' AI innovations and disrupts the journalistic institution. *Digital Journalism*, *10*(10), 1893–1912. https://doi.org/10.1080/21670811.2022.2120032

Küng, L. (2013). Innovation, technology and organisational change: Legacy media's big challenges: An introduction. In T. Storsul & A. H. Krumsvik (Eds.), *Media innovations: A multidisciplinary study of change* (1st ed., pp. 9–13). Nordicom.

Lokot, T., & Diakopoulos, N. (2016). News bots: Automating news and information dissemination on Twitter. *Digital Journalism*, *4*(6), 682–699. https://doi.org/10.1080/21670811.2015.1081822

Lugo-Ocando, J. (2020). *Foreign aid and journalism in the Global South: A mouthpiece for truth* (1st ed.). Rowman & Littlefield.

Meese, J., & Hurcombe, E. (2021). Facebook, news media and platform dependency: The institutional impacts of news distribution on social platforms. *New Media and Society*, *23*(8), 2367–2384. https://doi.org/10.1177/1461444820926472

Morlandstø, L. (2017). Innovation and value creation in local media. *The Journal of Media Innovations*, *5*(1), 17–30. https://doi.org/10.5617/jomi.4350

Munoriyarwa, A., Chiumbu, S., & Motsaathebe, G. (2021). Artificial intelligence practices in everyday news production: The case of South Africa's mainstream newsrooms. *Journalism Practice*. Advance online publication. https://doi.org/10.1080/17512786.2021.1984976

Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism*, *19*(8), 1043–1058. https://doi.org/10.1177/1464884917725163

Nielsen, R. K., & Ganter, S. A. (2022). *The power of platforms: Shaping media and society* (1st ed.). Oxford University Press.

Papaevangelou, C. (2023). Funding intermediaries: Google and Facebook's strategy to capture journalism. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2155206

Parks, L., & Starosielski, N. (2015). Introduction. In *Signal traffic: Critical studies of media infrastructures* (1st ed., pp. 1–27). University of Illinois Press.

Paulussen, S. (2016). Innovation in the newsroom. In T. Witschge, C. W. Anderson, D. Domingo, & A. Hermida (Eds.), *The SAGE handbook of digital journal-*

ism (1st ed., pp. 192–206). SAGE. https://doi.org/10.4135/9781473957909.n13

Plantin, J. C., & Punathambekar, A. (2019). Digital media infrastructures: Pipes, platforms, and politics. *Media, Culture and Society*, *41*(2), 163–174. https://doi.org/10.1177/0163443718818376

Poell, T., Nieborg, D. B., & Duffy, B. E. (2022). Spaces of negotiation: Analyzing platform power in the news industry. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2103011

Poell, T., Nieborg, D., & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, *8*(4), 1–13. https://doi.org/10.14763/2019.4.1425

Posetti, J. (2018). *Time to step away from the "bright, shiny things"? Towards a sustainable model of journalism innovation in an era of perpetual change*. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-11/Posetti_Towards_a_Sustainable_model_of_Journalism_FINAL.pdf

Rashidian, N., Brown, P. D., Hansen, E., Bell, E. J., & Albright, J. R. (2018). *Friend and foe: The platform press at the heart of journalism*. Columbia Journalism Review. https://doi.org/10.7916/D8-15PQ-X415

Reese, S. D. (2021). *The crisis of the institutional press* (1st ed.). John Wiley & Sons.

Rodny-Gumede, Y. M. (2014). South African journalists' conceptualisation of professionalism and deviations from normative liberal values. *Communicare*, *33*(2), 54–69. https://doi.org/10.10520/EJC170577

Schiffrin, A. (2021). *Media capture: How money, digital platforms, and governments control the news* (1st ed.). Columbia University Press.

Schindler, P. (2018, March 20). *The Google News Initiative: Building a stronger future for news*. The Keyword. https://blog.google/outreach-initiatives/google-news-initiative/announcing-google-news-initiative

Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, *10*(10), 1832–1854. https://doi.org/10.1080/21670811.2022.2063150

Skowronski, W. (2009). Circulation boost? Newspapers explore delivery via electronic reader. *American Journalism Review*, *31*(3), 12–14.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. https://doi.org/10.1016/J.RESPOL.2013.05.008

Storsul, T., & Krumsvik, A. H. (2013). What is media innovation? In T. Storsul & A. H. Krumsvik (Eds.), *Media innovations: A multidisciplinary study of change* (1st ed., pp. 13–26). Nordicom. https://doi.org/10.13140/2.1.2082.5929

Taebi, B., Correljé, A., Cuppen, E., Dignum, M., & Pesch, U. (2014). Responsible innovation as an endorsement of public values: The need for interdisciplinary research. *Journal of Responsible Innovation*, *1*(1), 118–124. https://doi.org/10.1080/23299460.2014.882072

van de Poel, I. (2009). Values in engineering design. In A. Meijers (Ed.), *Philosophy of technology and engineering sciences* (pp. 973–1006). Elsevier. https://doi.org/10.1016/B978-0-444-51667-1.50040-9

Voegtlin, C., Scherer, A. G., Stahl, G. K., & Hawn, O. (2022). Grand societal challenges and responsible innovation. *Journal of Management Studies*, *59*(1), 1–28. https://doi.org/10.1111/joms.12785

Von Schomberg, R. (2011). *Towards responsible research and innovation in the information and communication technologies and security technologies fields*. Publications Office of the European Union. https://doi.org/10.2777/58723

About the Authors

Mathias-Felipe de-Lima-Santos (PhD) is a postdoctoral researcher in the Human(e) AI project at the University of Amsterdam and a research associate in the Digital Media and Society Observatory at the Federal University of São Paulo (Unifesp). Previously, he was a Marie Skłodowska-Curie fellow at the University of Navarra and an Erasmus visiting researcher at the Queensland University of Technology. Mathias-Felipe co-edited the book *Journalism, Data and Technology in Latin America* published by Palgrave Macmillan in 2021. His research focuses on the impact of technology on journalism, media, and online social networks.

Allen Munoriyarwa (PhD) is a senior lecturer in the Department of Media Studies at the University of Botswana. His research interests are in journalism, news production practices, and platforms and social media. He has also researched widely on data journalism, big data, and digital surveillance. His research employs different qualitative and quantitative methodologies.

Adeola Abdulateef Elega (PhD) is a lecturer at Nile University of Nigeria. He holds a PhD in communication and media studies from Eastern Mediterranean University, Turkey. Adeola previously served as editor-in-chief of *Gundem Newspaper*. His articles have been published in prestigious journals such as *International Journal of Communication*, *Journalism Practice*, *Environmental Communication*, *African Journalism Studies*, and *Public Relations Review*. Adeola's research interests include new media, journalism, interpersonal and intercultural communication, and meta-analysis in media studies.

Charis Papaevangelou is a PhD candidate at the Laboratory for Applied Social Sciences of the University of Toulouse. His PhD project is part of the European Training Network JOLT, which was funded by the European Union's Horizon 2020 program. His work concerns the political economy of online platform governance within the EU, primarily, focusing on online content governance. His work is situated at the nexus of media, political, cultural, and social sciences, and has a distinct interest in the relationship between platforms and news media organisations.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 344–354
https://doi.org/10.17645/mac.v11i2.6510

Article

"I Think Quality is More Important Than a Lot of Data" in Cities Datafication

Carl Chineme Okafor

Department of Media and Social Sciences, University of Stavanger, Norway; carl.c.okafor@uis.no

Submitted: 18 November 2022 | Accepted: 14 June 2023 | Published: 28 June 2023

Abstract
This article studies how the decision to connect data volumes to value is made by technologists and governance people in smart cities' datafication process. Its entry point is that datafication promises to use data to make cities liveable domains. Cities on the back of this promise presuppose that more data produce value and therefore fixate on exhaustive datafication. But datafication does not appear self-evident, and knowledge of how technologists and governance people connect data volumes to data value is quite unclear in media and communication literature. Using evidence from interviews (*n* = 6), datafication policy documents (*n* = 4), and a diverse dataset of city activities (*n* = 299) in the open data portal of a situated datafication site, the Stavanger Smart City, Norway, and with the theoretical support of critical data studies, this article responds to the question: How does data volume connect to data value in smart cities datafication? Its findings put data quality as the intermediary that makes this connection.

Keywords
data quality; data value; data volumes; smart city datafication; Stavanger Smart City

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

Exhaustive datafication ensures that cities' data repositories possess and express knowledge, capital, and power (Eubanks, 2017; Kitchin, 2013, 2021; Ma et al., 2020; Rose, 2020); and this prompts the interrogation of datafication by a critical group of media and communications scholars (e.g., Cardullo & Kitchin, 2019; Zuboff, 2019, pp. 8–14), believing that it is a capitalistic form of control and rationalisation of human mobility (Kitchin, 2021; Sadowski, 2019; Sadowski & Pasquale, 2015).

The crux of the debate is datafication's promise of a better outcome for modern cities using data, and cities' huge collection and valuation of data for governance decisions (Beaulieu & Leonelli, 2022, p. 6). Nowadays, cities datafy opinion formation, political debates, and distribution of services and opportunities (Lycett, 2013; Mejias & Couldry, 2019), making them collectors of vast volumes of data. Data is thus the mechanism to operationalise cities' communication technologies, but also to redefine engagements therein.

Data indeed possess value, but this article advances that volumes alone do not generate value or sustain cities' appetite for exhaustive datafication. Often, media and communications scholars focus on the value that volumes of data generate to cities in datafication (Mechant & Walravens, 2018; van der Graaf, 2018) but not specifically on how the connection between volume and value is made by the people who decide on data collection and valuation.

Thus, this article investigates cities' data collection appetite which is buoyed by the belief that more data will make them better places to live or more efficient to govern. Concretely, it examines if vast volumes of data alone are sufficient to create such value for cities, positing that data needs to be of a certain quality for precise analysis to generate value. By mapping data volumes and data quality, this article sheds light on the potential of

connecting data volumes and data quality in ways that create value.

To understand how value accrues from volumes and cities' propensity for exhaustive datafication, this article studies how decisions on data volumes and value are made by technologists and governance bodies who prioritise data collection and valuation in cities. The main research question in this regard is: How does data volume connect to data value in smart cities' datafication? Findings indicate that data quality enables this connection, contributing to our understanding of the connection between cities' intensive datafication and their appetite for data volume.

2. Literature Review

Smart cities' datafication is popular among governments and technologists (Karvonen et al., 2018; Kitchin, 2016), and scholars try to define smart cities with no common definition yet (Zhao et al., 2021). However, ideas about what a smart city is, can be, and/or should do are not in shortage (Csukás & Szabó, 2021). In this sense, smart cities are conceived, designed, and implemented by governments and technologists with a focus on city efficiency, allowing for larger control on mobility and use of resources, but also generating more data and providing an image of modernity (Al Nuaimi et al., 2015; Hashem et al., 2016)

Datafication is the foundation of smart cities, providing the raw material—data—upon which smart cities initiatives operate (Cardullo & Kitchin, 2019; Kummitha & Crutzen, 2017; Löfgren & Webster, 2020). Datafication claims to make cities smart when technical objects which are previously considered lifeless become cognitively conscious (Akhilesh, 2020). It is a longstanding quantification practice supported by digitisation (Lycett, 2013; (Mayer-Schönberger & Cukier, 2013, pp. 74–96; Mejias & Couldry, 2019). Its use in smart cities is extensive.

To "datafy" is to place phenomena into quantified formats, tabulate and analyse for decision-making objectives. van Dijck (2014, p. 198) explains that datafication is based on "a widespread belief in the objective quantification and potential tracking of all kinds of human behavior and sociality through online media technologies." In other words, everything that can be measured—relationships, experiences, moods—will be turned into a data point and added to a dataset (Ruckenstein & Schüll, 2017, p. 262). But datafication is also met with criticism from scholars (e.g., Chan et al., 2022; Lycett, 2013; Sadowski, 2019; Sadowski & Pasquale, 2015; Zuboff, 2019) who claim that it carries intrusive properties, including increasing citizen surveillance and nudging actions.

The data infrastructures of smart cities rely heavily on digital communication technologies and systems that enable datafication (Mohanty et al., 2016; Rose, 2020). This embeds smart cities' datafication to questions related to the role of communication infrastructures in modern societies, as they constitute "computing and network resources that allow multiple stakeholders to orchestrate their services and content needs" (Constantinides et al., 2018, p. 381). Datafication practices become intrinsic to decision-making, and thus to power (Sjøvaag & Ferrer-Conill, 2023). To that end, this article adopts Harrison et al. (2010, p. 2) definition of the smart city as "connecting the physical infrastructure, the IT infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city." This conceptualisation underlines the constitutive description of the so-called modern city as a convener of economies of scale and facilitator of agglomeration. The definition also approaches the smart city as a socio-technical construct (cf. Edwards et al., 2007) which understands humans to inscribe biases in technology and how these biases replicate themselves in the organisation of cities, thus questioning the notion of "smart" but also bringing into dialogue how data is generated and interpreted for use, i.e., data volumes, value, and provenance.

2.1. Data Volumes

Access to data is important to smart cities (Möller & Von Rimscha, 2017); the velocity, volume, and variety of data collection demonstrate this, but also the appetite for data (Beaulieu & Leonellli, 2022, pp. 6–10; Lycett, 2013). This article argues that volume is a derivative of velocity and variety, and contextually connects velocity, volume, and variety to advance that the velocious and varied data cities collect makeup data volumes. Volumes in cities connote the size of data that is collected from all available sources for decision-making (Al Nuaimi et al., 2015). Scholars (e.g., van Dijck, 2014; Zuboff, 2019), in their criticism of cities' conviction in governance through aggregated data, highlight that data volumes in cities' datafication process are often seen in the assumption of a manifest relationship between data and efficient cities. In this regard, dialogues on data volumes are often not without thoughts on how and where data comes from, i.e., data provenance, which is important to technologists and governance people in cities' datafication process (Beaulieu & Leonelli, 2022, pp. 7–8).

2.2. Data Value

Data value is the possible advantage cities generate from data volume through analyses and is often projected as the end goal of datafication (Al Nuaimi et al., 2015; Beaulieu & Leonelli, 2022, pp. 24, 39). Value is recognised in how cities mobilise data across contexts, space, and time, and its traces are evident in the technocratic, algorithmic, automated, and anticipatory behaviours cities exhibit (Al Nuaimi et al., 2015; Bibri, 2021; Löfgren & Webster, 2020). Scholars infer that cities recognise data to possess value both in the present and future (Bibri, 2021; Lycett, 2013; Sadowski, 2019; Sadowski & Pasquale, 2015; Rijshouwer

et al., 2022; van Dijck, 2014). In particular, the trust in algorithms to predict and pattern events in cities, but also to automate decisions, demonstrates this inference (Mayer-Schönberger & Cukier, 2013; O'Neil, 2016). Furthermore, cities' reliance on data to produce and evaluate social knowledge—e.g., "how to parent, police, govern, be healthy and put together a good soccer team" (Beaulieu & Leonelli, 2022, pp. 13–14)—or reinvent governance, are examples of the framings of data value in datafication studies (Kummitha & Crutzen, 2017; Noveck, 2018, pp. 123–126). Likewise, when cities consider data as economic, political, and social means, they, in other words, demonstrate the agency as value that data gain to, for instance, nudge, steer, and control behavioural boundaries or patterns therein (Beaulieu & Leonelli, 2022b; Cardullo & Kitchin, 2019; Kitchin, 2021; Mejias & Couldry, 2019; O'Neil, 2016, p. 191; Sadowski, 2019).

2.3. Data Provenance

In critical media and communication studies on datafication, data volume, and value are expansively discussed as important subjects, but not data provenance. Data provenance defines as the origin of data. It is also described as the conditions under which data is generated and disseminated (Beaulieu & Leonellli, 2022, p. 23). Beaulieu and Leonellli (2022) claim that focusing only on data volumes, without accounting for other important elements (e.g., data provenance) could result in risky data analyses and interpretation. Their contention underscores the need to account for and recognise what happens to data between acquisition and use in generating datafication outcomes.

Scholars in scientific knowledge production often question data collection methods, mostly to account for biases, which involve the properties of data. Similarly, when cities fail to account for the origin of the data they engage with, they risk flawed analysis and claims in decision-making. Data provenance shows the circumstances of data generation to support data use. Like volume and value, it should be a considerable subject in the interrogation of cities' datafication (boyd & Crawford, 2012). Studying the provenance of data helps to respond to questions of why, how, where, when, and whom of datafication. It supports technologists and governance groups with insights into making decisions on volumes and value. Connecting with volumes and value, the provenance of data articulates the constitutive character of cities' data assemblage which, as Kitchin and Lauriault (2014) claim, is rooted in conventions, traditions, and infrastructures.

To this end, this article focuses on analysing Data Volumes, Value, and Provenance, in response to its main research question which further opens out to:

- RQ1: What do the technical and governance people in the Stavanger Smart City want from their datafication practice and process?
- RQ2: What is important to the technical and governance people in the Stavanger Smart City datafication practice and process?
- RQ3: How do the technical and governance people in the Stavanger Smart City connect data volume and value to complete the datafication process?

3. Theoretical Framework

Cities' excitement about big data parallels that of the emergence of statistics in the late 18th century, but critical data studies (CDS) invite media and communications scholars to pay deeper attention to this exciting description of data and the cultures around it (Dalton et al., 2016). Based on datafication's promises of a better outcome for cities—a claim that influences its ubiquity—this invitation is to interrogate the datafication process beyond how and what cities use data for, to data collection and analysis.

Datafication provides data to smart cities as the raw material to operate, but also as the mechanism to understand their sociality (Dalton et al., 2016; van Dijck, 2014), hence CDS's call for a critical interrogation of data, their generation, and analysis. Scholars of diverse fields, including media and communications, engage with the assumptions of CDS to respond to the technical and organisational issues that data-intensive practices generate. Similarly, CDS substantially attends to the normative and privacy concerns that cities' datafication process generates, providing scholars with the theoretical tools to study the widespread consequences of big data in the social arena. From the point of view of data production and analysis, Dalton and Thatcher (2014), for instance, probe the manipulation of big data, including the motives and imperatives that often drive such data work. Other scholars have equally continued to count on CDS to interrogate how cities relate to data volumes and value, but bringing into the conversation how multiple relevant influences embed in the cultures of technological infrastructure political orientations, business, and economic plans or agendas to jointly frame datafication (Iliadis & Russo, 2016; Sadowski, 2019).

In this article, I engage the propositions of the CDS to study the socio-technological process of collecting volumes of data and extracting value from data, but also to know how this process manifests in data assemblages which consist of data systems of technological, political, social, and economic arrangements (Kitchin & Lauriault, 2014). Clarifying how data is consciously and unconsciously created, the means of collecting data, and what informs their analysis in smart cities datafication offers this article the space to bring attention to the entire process of interpreting data for value and how this is decided (boyd & Crawford, 2012).

Of interest to this article is the process of "cooking" data into context-dependent decisions. The "cooking" of data is a key theoretical assumption of CDS and is regarded in this article as the analytical process

of turning data into value. It highlights the misleading idea that data is neutral (Beer, 2016; Gitelman, 2013, pp. 167–171; Räsänen & Nyce, 2013), when they are not because they are often rooted in values, norms, epistemological claims, and philosophical outlooks (boyd & Crawford, 2012) which technologists and governance people in smart cities possess. In other words, all data are produced and, in the process, subject to choices about what to collect and how to analyse them for value (Kitchin, 2014; Sadowski & Bendor, 2019).

Therefore, I operationalise this assumption in studying the connection between data volume (which contextually embodies vast data collection) and value. This way, I investigate the belief that collecting more data is always better for cities to function. More concretely, in engaging the process of "cooking" data, I focus on how technologists and governance bodies who make decisions to aggregate data decide on their values; I also highlight the lack of clarity in scholarship and the social arena about how decisions on data collection and valuation are made (Andrejevic, 2014). This approach underlines the potentiality of datafication to mutate with new ideas and knowledge in cities, for example, when technologies are (re)invented, organisations change, business models are created and recreated, and political systems and economies get altered by new or old orders (Kitchin & Lauriault, 2014).

4. Data and Method

This article adopts a case study approach (Yin, 2018), using the Stavanger Smart City as a situated case. Stavanger is a purposive choice because it is diversly connected with the University of Stavanger as a research hub and provides this research access that would not be possible anywhere else. Stavanger is located in the southwest of Norway; it is both its fourth largest city with approximately 250,000 inhabitants and an energy hub. Stavanger has a smart city operation that dates back to 2016, with goals structured in five priority areas: health and welfare; education and knowledge; energy, climate, and environment; urban art; and governance and democracy (Stavanger City Council, 2016). Stavanger has also achieved smart city goals that include operating an open data portal, weed-control sensors, auto-measuring waste accumulation and disposal, digitising public services fault reporting, and managing city mobility. It has benefitted from the EU-funded programmes Horizon 2020 (2015–2020) for smart cities and communities in Europe and AI4Cities to support climate change goals with artificial intelligence.

Case studies are useful because they respond to the "how" or "why" questions of research. They also come in handy when researchers have limited control over events and the object of study is a contemporary phenomenon with complexities and contextual conditions that need to be studied closely and robustly. In addition, case studies often provide researchers with robust methodological tools to investigate cases of "decisions," "individuals," and "processes" (Yin, 2018). The case study method suits this research as it seeks to understand how technologists and governance people make decisions on data volume and value in a city's datafication process.

The data collection methods employed in the case study include interviews (Edwards & Holland, 2013), document analysis (Bowen, 2009; Dalglish et al., 2021; Grant & Kara, 2022), and a content analysis of an open data portal (Krippendorff, 2018, pp. 89–124). The methods give access to policy decisions that guide datafication, technologists, and governance people who implement datafication in Stavanger.

I conducted six semi-structured interviews with technologists and governance people in August and October 2022 (see Table 1 for informants' descriptions). The informants were chosen using maximum variation sampling to get a wide range of interactions from two distinct profiles, experts and practitioners who implement datafication in the smart city (Sandelowski, 1995 , 2000). They were recruited from two relevant sources—the Nordic Edge, a smart city and Internet of Things industry cluster associated with Stavanger, with members as vendors and participants in the smart city project, and the Stavanger Smart City Department, which is chiefly responsible for stakeholders' engagements and evaluation of datafication implementation in the case. These include face-to-face interviews with five participants and one online session on Microsoft Teams with a respondent who was physically unavailable. The interviews were conducted freely in the English language, lasting an average of 38 minutes per session. Respondents' privacy and ethical use of interview data were guaranteed, as required by the Norwegian Centre for Research Data (Norsk Sentre for Forskningsdata). Consent for this was fully obtained. Interviews covered demographic (name, age, professional status, and work affiliations) and research-related questions. The responses were comprehensive and a 50:50 male and female gender parity was observed.

I transcribed and used the thematic analyses (Braun & Clarke, 2006) to organise themes around my three analytical groupings of Data Volume, Value, and Provenance. The flexibility of thematic analysis allows me to informally determine themes' prevalence in the interview data, i.e., the prevalence of themes came from repeated data analyses. The theoretical freedom that thematic analysis enjoys also gives it the flexibility to richly account for my data as it reflects and explains the reality and the surface of my data, i.e., making its character transparent. This strategy also captures important themes from the data, with the "keyness" of themes not necessarily depending on quantifiable measures but on relevance to my research questions.

Using the thematic analysis as well, I analysed the four datafication policy papers: (a) the *ICT Strategy for Stavanger Municipality*, (b) the *Joint Social Element of the Municipal Masterplan for New Stavanger 2020–2034* (both were translated from Norwegian to English), (c) *the*

Table 1. Description of the study's sample.

Study sample: Interview respondents (IRs) and datafication documents (DDs)	Description
IR01	Governance/technical actor: Data and network engineer, with more than 20 years of experience in digitalisation and innovation; huge involvement in datafication decisions and implementation in Stavanger
IR02	Technical actor: Geodetic engineer, with more than eight years of experience in specialised spatial data management
IR03	Technical actor: Social scientist and IT expert, with more than 10 years of experience in spatial mapping research and innovation
IR04	Governance/technical actor: Technologist, with expertise in cyber security, open data, micro-mobility, citizen involvement, and sustainability in smart cities; more than four years of experience with Stavanger Smart City
IR05	Technical actor: Expertise in smart city systems and applications with more than 10 years of experience in public sensor systems management
IR06	Governance actor: IT innovation strategists with specialisation in digital business, service design, and innovation; more than six years of experience with Stavanger Smart City
DD01	*ICT strategy for Stavanger Municipality*
DD02	*Joint Social Element of the Municipal Masterplan for New Stavanger 2020–2034*
DD03	*Stavanger Digital Strategy 2014–2029*
DD04	*Roadmap for Stavanger Smart City*

Stavanger Digital Strategy 2014–2029, and (d) the *Roadmap for the Smart City Stavanger*, to organise themes into my three analytical groups of Data Volume, Value, and Provenance. I used the READ approach: (a) ready your materials, (b) extract data, (c) analyse data, and (d) distil your findings (Dalglish et al., 2021) to do this and gained familiarity with the manifest contents of Stavanger's datafication plan, including how technologists and governance people understand the plan. I did a content analysis of Stavanger's open data portal to gain insight into the data that is collected (see Table 2) and identify from its metadata themes that fall into my analytical groupings. In organising my analysis of the data, I relied on the theoretical views of the CDS, specifically observing the "cooking" of data in the process.

5. Results

Two data structures, the open data portal (for the public) and data lake (for internal operations), illustrate Stavanger's data use practice. However, I did a content analysis of only the open data portal (n = 299) which is available to the public. First, I find that eight categories of datasets including an Expired category (n = 94) are specific to Stavanger and that there are datasets in duplicates and unusual groupings in contents. I then make new groupings of four (see Table 2) from their original groupings excluding the expired datasets, by uniting and or renaming three categories, i.e., merging Bicycle measurements (n = 43) and Transport (n = 31) datasets to form Transport and Mobility (n = 74); Livelihoods (n = 25),

Table 2. Analysis of open data.

Transport and Mobility (n = 74)	Maps, Emergency, and Public Safety (n = 51)	Culture and Livelihoods (n = 68)	Weather and Environment (n = 12)
For example, parking, city bike counter, and cycling/hiking routes	For example, mapping, road noise levels, pedestrian traffic, school routes, vehicle charging stations, and speed limits	For example, registered unemployment/disability pensioners, immigration, and operating expenses for museums and cinemas	For example, waste management/refuse containers, rain/manhole/bathing temperature sensors, and air quality measurement

Culture (n = 12), and Statistics (n = 31) to form Culture and Livelihoods (n = 68); and renaming Maps, Emergency, and Public Safety (n = 51). Weather and Environment (n = 12) retain their categorisation. This was to make my analysis transparent and understandable.

Excluding the 94 expired datasets, the portal contains 207 datasets that are actively specific to Stavanger. Their metadata provides insights on the dates of collection (starting from 2016), update schedules (every two and five minutes for parking and city bikes data respectively, hourly for air quality data, daily for cycling and rain data, annually and on-demand for hiking routes and trails data), the protocol that governs data use and reuse (the Norwegian Licence for Open Government Data) and data storage formats (CSV, JavaScript Object Notation, GPS Exchange, Microsoft Word text document, and PDF).

5.1. Thematic Analysis

I use the thematic analysis to analyse the six interview transcripts and four datafication policy papers which are part of my empirical materials. The documents are (a) the *ICT strategy for Stavanger Municipality*, (b) the *Joint Social Element of the Municipal Masterplan for New Stavanger 2020–2034* (both were translated from Norwegian to English), (c) the *Stavanger Digital Strategy 2014–2029*, and (d) the *Roadmap for the Smart City Stavanger*. In addition to providing me with historical insights into Stavanger's datafication goals and implementation context, I find in these materials 10 themes that fall into my three analytical groupings of Data Volume, Value, and Provenance (see Table 3).

Data Volume is discussed through five themes. Data Volume in the context of smart city datafication connotes the size and variety of data collected from all available collection sources. Data from documents and interview analysis describe Stavanger's data collection practice, but also the number of datasets and the varied formats they are stored in the open data portal. One of the prominent themes concerns managing large amounts of data for society and its citizens. As informant DD01 states:

> With control and an overview of the data, it is easier for the municipality to use data in new contexts such as artificial intelligence, data analysis, and big data. More data can be compiled and create new insights and improved services. The municipality will also share data [open data] with other players to contribute to innovation and service improvement for the benefit of citizens. The municipality must therefore have discretion, control, and access to all data that the municipality produces.

Another aspect of Data Volume in the themes concerns how new technologies will continue to produce more data, as informant DD01 again explains:

> The municipality manages large amounts of data on behalf of society and its citizens, and new technology will produce even more. The investment in open data [data sharing] will continue so that we facilitate innovation and reuse also outside the municipality.

Also, Volume is the theme that reflects on how data collection is based on legal, economic, and clerical mandates of Stavanger's seven service departments, and informant IR01 explains this thus:

> It is very individualistic, from department to department. It depends on the tasks and services that they have the mandate to operate. For instance, the garbage people [Waste Disposal Unit] have data on different garbage bins. To the point, that would be what kind of clerical mandate we are set under.

In addition, informant IR04 underlines the extent of this theme saying: "When it comes to ownership and maintenance of data, it is up to different departments that have different professional fields, and there will be

Table 3. Themes from analysis of empirical data.

Analytical groups	Themes from datafication policy papers and interview transcripts
Data Volume	1. Municipality manages large amounts of data for society and its citizens 2, New technology will produce more data 3. Data collection is based on legal, economic, and clerical mandates 4. Data architecture and infrastructure must be built for data exchange and communication to have a special focus on scalability, openness, and interoperability 5. Municipality has standard data archiving and reuse process
Data Value	1. Data as an important resource is used for governance 2. The municipality must ensure ownership and access to data to build predictive decision-making 3. Data will be used and reused in other contexts that can provide new and better services in the future
Data Provenance	1. The quality of data collected is important for value generation 2. Quality data is realised from iteration and standardisation of collection

people for owning such data. Although, in theory, the city administrator owns all the data."

Further in relation to Data Volume is a theme expressing how the city's data architecture and infrastructure must be built for data exchange and communication and configured towards scalability, openness, and interoperability as supported by the governing protocol data use in the open data portal. On this, DD01 explains:

> Solid architecture and well-thought-out infrastructure are pillars of forward-looking, good citizen services, and effective ICT tools for the employees. This is therefore the municipality's main focus area within ICT. The architecture and infrastructure must be built so that it is possible to exchange data and communicate between different ICT systems, also across administrative levels. New ICT acquisitions, further development, and resource requirements depend on the ICT architecture. It must be facilitated for an efficient, modular, flexible, and service-oriented architecture and have a special focus on scalability, openness, and interoperability.

And finally, Volume is a theme that relates to the city's standard data archiving and reuse process, which informant DD01 explains:

> All systems that store archive-worthy material must have a strategy for handing over electronic material to Stavanger City Archives. Systems that are not approved as electronic archives can be considered integrated directly with the case/archive system. Reuse of master data across subject systems greatly contributes to saving resources and streamlining work processes, we must therefore decide which system/service is responsible for which data. The information must be registered and maintained in only one place and then made available to all systems that need the information.

In summary, the themes agreeably represent Data Volume and how Stavanger conceives and deploys it in its datafication process.

In terms of Data Value, three themes manifest in the analysis. But this also manifests in the protocol that governs data use and reuse in the open data portal analysis. Data value is the advantage that the smart city stands to generate from data volume. It is usually the end goal of cities' datafication process. One theme expresses this in the analysis as data being an important resource for governance, and informant DD01 confirms: "Data is an important resource today but will grow both larger and more important in the digital future."

Informant IRP01 also notes in this same breadth that: "We are collecting the data, putting them together across all the different sources in kind of a dashboard and reports to give us a better governance insight…data that help us back up the goals that we try to achieve."

Another theme that speaks to Data Value is the municipality's ownership and access to data to build predictive decision-making. Informant DD01 underscores this: "The municipality must ensure ownership and easy access to this important resource. In the future, the data will be used and reused in other contexts that can provide new and better services."

Informant IRP01 also explains:

> We have a lot of data in the municipality; we have water data, sensor data, and for different solutions—It could be HR, economic data, [and] health department which has a lot of data regarding our citizens. The purpose is to collect all the data and then we have a toolbox that gives us the opportunity to use the data for different purposes. We are starting to use the data for these new technologies on machine learning and prediction.

In agreement, Informant IR04 states: "Most of the operational part of the municipality is data-driven…a lot of the data that is collected within the municipality is to create statistical models for projections."

Similarly, a theme that talks about Data Value in the analysis is the expectation that data will be used in different contexts and levels to generate competent judgements. Informant DD01 references this: "Through analyses and big data, we can produce good decision-making information at all levels."

In addition to this, informant DD04 states: "Technology will be part of the solution—Whether new technology is used or existing technology is further developed…technology in a smart city context is a tool for creating economic, social, and environmental improvements.''

Consequently, the themes correspond to the advantages Stavanger expects from the volumes of data it generates.

When it comes to Data Provenance, which is the origin of and conditions under which data is generated and disseminated, two themes describe it in the analysis. The data collection dates and update schedule in the open data portal likewise express Provenance. One of the themes relates to how the quality of data that the city collects is important for value generation. This is as informant IR06 explains:

> We don't need to collect more than we are using…I think that [the] quality of the data is more important than having a lot of data…I think [that] quality data is data that somebody finds useful to make their tasks better at optimising something. But also, that [it] is collected and measured in a good way.

In agreement, informant IR03 explains that:

> This might sound like it comes from the schoolbooks and maybe it does; the main objective is

> to have FAIR data—findable, accessible, interoperable, and reusable. Quality as such is more related than [the] traditional perspectives on scaling levels and magnitude.

Furthermore, Data Provenance manifests in how iteration and standardisation of data collection result in quality data, and informant IRP01 explains this:

> We need at least to standardise the way that we are using data so that they make sense for us. If the systems we use are used in different ways for different scenarios, it will not make any sense to us, [and] if we collect the data because they are not consistent. If we are not standardising the way that we put data in the systems, we don't have good quality data, it loses a lot of quality and it is not good enough to be used. I would say that one of the main struggles in the municipality is that we are not standardised in our work processes, so the quality of the data that we put in will be different and it becomes confusing data.

Informant IR03 shares the same perspective, saying:

> This is along the discussion always, and it is quality meaning more detail, more accuracy, and precision. We have another word which might be related to quality—authoritative data—and it is not the same [as quality], but it means that these are data that you can trust, or they come from authority-level processes. For instance, the property data is one of Norway's most known authoritative datasets, and that compared to crowd-sourced or sensor-based data that haven't actually been qualified [to be sure] if they are correct. That means that it is part of the quality sign that the data is authoritative, meaning that it can be trusted, that it has been collected correctly or according to the laws, instructions, or standards that specify the necessary requirements.

Quality as a component for assessing trust and validity in data thus summarises Data Provenance in the analysis. It is also a component of Provenance and indicates that, in Stavanger's datafication process, data quality is functional in connecting volume to value.

6. Discussion

The results presented earlier contribute to my understanding of the datafication process in the Stavanger Smart City, specifically how technologists and governance people make decisions on data volume and value. But before I discuss this further, I would like to follow the scientific virtue of clarifying to highlight how the two actors I engaged with view datafication and the smart city concepts in their everyday operations.

There is no shared definition of the smart city in principle and practice; instead, actors define it according to their respective needs. Stavanger has a selective approach to accomplishing its smart city goals, it's understanding of the smart city nevertheless aligns with this article's definition as a city engaging with data and digital communications infrastructure (Harrison et al., 2010, p. 2). While datafication appears to be theoretically defined and settled in scholarship (Mayer-Schönberger & Cukier, 2013; Mejias & Couldry, 2019; Lycett, 2013; van Dijck, 2014. p. 198), I find that this is not the case amongst actors in Stavanger who are unclear about the theoretical label despite their practice of it. This implies that datafication in smart cities is context-driven (Lycett, 2013; Mejias & Couldry, 2019; Micheli et al., 2020).

Following from this, I also consciously point out for clarity that Stavanger's data lake, which is exclusive and unavailable for this study's analysis, however, comprises data from multiple systems and sources that are plugged into the city's digital infrastructure network. Respondents claim that data from therein are internally used to modulate and model perspectives and insights for city management. The data lake and open data portal differ from each other in terms of content, access, and use, but my findings from the open portal analysis show that Stavanger, like most smart cities setup, is compulsively obsessed with data volume, and its value as a governance tool (Al Nuaimi et al., 2015; Ma et al., 2020; Sadowski & Bendor, 2019). The results also indicate that actors have an idea of the importance of data provenance in decisions about data use or engagement, especially as it concerns the open data portal (Beaulieu & Leonellli, 2022, p. 23).

While substantiating the need for this study on the basis of the deficit of empirical thoughts on how decisions to connect volume and value are made in datafication despite extensive literature on datafication, my results establish that Stavanger has a clear datafication program to deliberately collect and generate value from data, trusting it and digital communications infrastructure to drive governance and societal wellbeing (cf. Beaulieu & Leonellli, 2022; Lycett, 2013). In this regard, my analysis suggests that the "cooking" of data is usually required to connect volume to value in Stavanger and that the provenance of data possibly plays an essential role in making this connection (Beaulieu & Leonellli, 2022, pp. 7–8). As an influenced process (Iliadis & Russo, 2016; Sadowski, 2019), datafication in Stavanger involves the "cooking" of data to match goals. In this case, I take note of how respondents talk about data quality—a component of provenance (boyd & Crawford, 2012)—as the intermediary for translating volume into value.

My results show that data is a valuable resource for Stavanger to productively engage, create opportunities (e.g., through its open data portal; cf. Gilbert, 2021), and fix the city's challenges (Kummitha & Crutzen, 2017; Noveck, 2018; van Dijck, 2014; Zuboff, 2019), but the connection between volume and value is achieved through quality, suggesting thus that quality serves an evaluative role in linking value to volume.

In other words, the results imply that value from large volumes of data is accessed through the quality of data that is collected and available to the city but also that the absence of quality could result in no value from volume.

Also based on the results of this study, data quality appears to empower datafication actors in Stavanger to, in the search of value, generate and obsess for more data. This, as respondents imply, is an iterative process to attain data quality, i.e., actors relying on iteration to get quality-level data that support their goal. Simply defined as data that is fit for purpose (Fox et al., 1994), data quality is also labelled by respondents as good data, i.e., data that is of high quality and which can enable good decisions, a claim that further accounts for its grounding in evaluation, but also infers that, in context, it is an ongoing process of improving data for value. This iterative search for value through quality, is in fact, the "cooking" that scholars refer to for data to align with contexts and priority-informed decisions on datafication (Hacking, 2007; Kitchin, 2021; Kitchin & Lauriault, 2014; Löfgren & Webster, 2020; Zuboff, 2019).

7. Conclusion and Contributions

The intensifying dimensions of datafication signify that it is not a passive process but an actively defined practice wherein cities decide on data that affords them specified value. This has become much more sophisticated with time, technology, and agenda and deployed to distribute opportunities, secure societies, and manipulate and modify social actions. In fact, it would appear that its essence is to enable societies to forecast and control their affairs using tons of data and communication infrastructure, a notion that has for long placed data volume and value in front of debates that concerns its premises. In these debates though, less attention has gone into discussing how decisions to connect volume to value are made by active actors of datafication. Having set out to study this—how decisions on data volume and value are made—through two essential actors (technologists and governance people) in a situated datafication site, the Stavanger Smart City, I have in this article made efforts to provide an empirically grounded argument that positions data quality as the intermediary for translating volume into value. My findings do not exclusively make claims that data quality is the only intermediary to translate volume into value but posit that it enables this connection.

While contributing to our understanding of how technologists and governance people make decisions on data volume and value in datafication, I further argue that data quality may as well be instrumental to cities' appetite for volumes of data.

My results hint at how data quality as a component of data provenance accentuates the role that provenance impliedly plays in establishing the reliability or otherwise of data. I reckon however that the seeming limited engagement with data quality amongst media and communications studies scholars may have provided the grounds for the many questions that datafication of the social arena generates. These questions are nonetheless vital in contemplation of datafication promises of a better outcome for modern cities, yet scholars' understanding of data quality as an intermediary to translate volume into value in datafication provides additional knowledge for interrogating datafication, more so leveraging on the theoretical sagacity of CDS.

8. Limitations and Future Research Recommendations

I do not make claims that this study answers all the questions about how the datafication process in cities can be understood, but I have made efforts to study how data quality may contribute to understanding datafication and that scholars can approach future studies of datafication from this perspective. In this breadth, I thus recognise and highlight that there are obvious limitations in this study, one of which is its reliance on a single case study, and particularly the low number of IRs captured in this study. A higher interview sample size (n = 12) was planned for, but saturation was achieved midway, yet this shortcoming is compensated with the use of complementary data sources to ground the findings. Future research can concretely investigate the likely dimensions of data quality in cities' datafication, to establish how the technologists and governance people who decide on data interpret data quality from their different work areas and if they have similar or dissimilar understandings of data quality, as well as how they manage to reconcile potential differences in interpretations of data quality to achieve data interoperability and sharing to meet cities datafication goals.

Acknowledgments

This work was supported by the Norwegian Research Council (Project No. 314257).

Conflict of Interests

The author declares no conflict of interests.

References

Akhilesh, K. B. (2020). Smart technologies—Scope and applications. In K. B. Akhilesh & D. P. F. Möller (Eds.), *Smart technologies* (pp. 1–16). Springer. https://doi.org/10.1007/978-981-13-7139-4_1

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), Article 25. https://doi.org/10.1186/s13174-015-0041-5

Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, *8*(17), 1673–1689.

Beaulieu, A., & Leonelli, S. (2022). *Data and society: A critical introduction*. SAGE.

Beer, D. (2016). How should we do the history of big data? *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951716646135

Bibri, S. E. (2021). Data-driven smart sustainable cities of the future: Urban computing and intelligence for strategic, short-term, and joined-up planning. *Computational Urban Science*, *1*(1), Article 8. https://doi.org/10.1007/s43762-021-00008-9

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. https://doi.org/10.3316/QRJ0902027

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cardullo, P., & Kitchin, R. (2019). Smart urbanism and smart citizenship: The neoliberal logic of "citizen-focused" smart cities in Europe. *Environment and Planning C: Politics and Space*, *37*(5), 813–830. https://doi.org/10.1177/0263774X18806508

Chan, J., Sanders, C., Bennett Moses, L., & Blackmore, H. (2022). Datafication and the practice of intelligence production. *Big Data & Society*, *9*(1). https://doi.org/10.1177/20539517221089310

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). Introduction—Platforms and infrastructures in the digital age. *Information Systems Research*, *29*(2), 381–400. https://doi.org/10.1287/isre.2018.0794

Csukás, M. S., & Szabó, R. Z. (2021). The many faces of the smart city: Differing value propositions in the activity portfolios of nine cities. *Cities*, *112*, Article 103116. https://doi.org/10.1016/j.cities.2021.103116

Dalglish, S. L., Khalid, H., & McMahon, S. A. (2021). Document analysis in health policy research: The READ approach. *Health Policy and Planning*, *35*(10), 1424–1431. https://doi.org/10.1093/heapol/czaa064

Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical data studies: A dialog on data and space. *Big Data & Society*, *3*(1). https://doi.org/10.1177/2053951716648346

Dalton, C., & Thatcher, J. (2014). What does a critical data studies look like, and why do we care? Seven points for a critical approach to "big data". *Society and Space*, *29*.

Edwards, R., & Holland, J. (2013). *What is qualitative interviewing?* Bloomsbury Academic.

Edwards, P., Jackson, S., Bowker, G., Knobel, C. (2007). *Understanding infrastructure: Dynamics, tensions, and design*. Santa Clara University.

Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (1st ed.). St. Martin's Press.

Fox, C., Levitin, A., & Redman, T. (1994). The notion of data and its quality dimensions. *Information Processing & Management*, *30*(1), 9–19. https://doi.org/10.1016/0306-4573(94)90020-5

Gilbert, S. (2021). *Good data: Power, paranoia, and prosperity in the digital age*. Welbeck.

Gitelman, L. (2013). *"Raw data" is an oxymoron*. The MIT Press.

Grant, A., & Kara, H. (2022). *Doing your research project with documents: A step-by-step guide to take you from start to finish*. Policy Press.

Hacking, I. (2007). Kinds of people: Moving targets. In *Proceedings of the British Academy Lectures* (Vol. 151, pp. 285–318).

Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., & Williams, P. (2010). Foundations for smarter cities. *IBM Journal of Research and Development*, *54*(4), 1–16. https://doi.org/10.1147/JRD.2010.2048257

Hashem, I., Chang, V., Anuar, N., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748–758.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2). https://doi.org/10.1177/2053951716674238

Karvonen, A., Cugurullo, F., & Caprotti, F. (Eds.). (2018). *Inside smart cities: Place, politics and urban innovation* (1st ed.). Routledge. https://doi.org/10.4324/9781351166201

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges, and risks. *Dialogues in Human Geography*, *3*(3), 262–267.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1). https://doi.org/10.1177/2053951714528481

Kitchin, R. (2016). *Getting smarter about smart cities: Improving data privacy and data security*. Department of the Taoiseach.

Kitchin, R. (2021). *Data lives: How data are made and shape our world* (1st ed.). Bristol University Press.

Kitchin, R., & Lauriault, T. P. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work*. SSRN. https://ssrn.com/abstract=2474112

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.

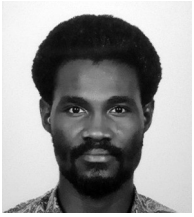
Kummitha, R. K. R., & Crutzen, N. (2017). How do we understand smart cities? An evolutionary perspective. *Cities*, *67*, 43–52.

Löfgren, K., & Webster, C. W. R. (2020). The value of big data in government: The case of "smart cities." *Big Data & Society*, *7*(1). https://doi.org/10.1177/2053951720912775

Lycett, M. (2013). "Datafication": Making sense of (big)

data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386.
Ma, M., Preum, S. M., Ahmed, M. Y., Tärneberg, W., Hendawi, A., & Stankovic, J. A. (2020). Data sets, modeling, and decision making in smart cities: A survey. *ACM Transactions on Cyber-Physical Systems*, *4*(2), 1–28. https://doi.org/10.1145/3355283
Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think* (1st ed.). Houghton Mifflin Harcourt.
Mechant, P., & Walravens, N. (2018). e-Government and smart cities: Theoretical reflections and case studies. *Media and Communication*, *6*(4), 119–122. https://doi.org/10.17645/mac.v6i4.1848
Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1428
Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, *7*(2). https://doi.org/10.1177/2053951720948087
Mohanty, S. P., Choppali, U., & Kougianos, E. (2016). Everything you wanted to know about smart cities: The Internet of Things is the backbone. *IEEE Consumer Electronics Magazine*, *5*(3), 60–70. https://doi.org/10.1109/MCE.2016.2556879
Möller, J., & Von Rimscha, M. B. (2017). (De)centralization of the global informational ecosystem. *Media and Communication*, *5*(3), 37–48. https://doi.org/10.17645/mac.v5i3.1067
Noveck, B. S. (2018). Open data: The future of transparency in the age of big data. In D. E. Pozen & M. Schudson (Eds.), *Troubling transparency* (pp. 206–225). Columbia University Press.
O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). Crown Books.
Räsänen, M., & Nyce, J. M. (2013). The raw is cooked: Data in intelligence practice. *Science, Technology, & Human Values*, *38*(5), 655–677.
Rijshouwer, E. A., Leclercq, E. M., & van Zoonen, L. (2022). Public views of the smart city: Towards the construction of a social problem. *Big Data & Society*, *9*(1). https://doi.org/10.1177/20539517211072190
Rose, G. (2020). Actually-existing sociality in a smart city: The social as sociological, neoliberal and cybernetic. *City*, *24*(3/4), 512–529.
Ruckenstein, M., & Schüll, N. D. (2017). The datafication of health. *Annual Review of Anthropology*, *46*(1), 261–278.
Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1). https://doi.org/10.1177/2053951718820549
Sadowski, J., & Bendor, R. (2019). Selling smartness: Corporate narratives and the smart city as a sociotechnical imaginary. *Science, Technology, & Human Values*, *44*(3), 540–563.
Sadowski, J., & Pasquale, F. (2015). The spectrum of control: A social theory of the smart city. *First Monday*, *20*(7). https://doi.org/10.5210/fm.v20i7.5903
Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing & Health*, *18*(2), 179–183.
Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, *23*(4), 334–340.
Sjøvaag, H., & Ferrer-Conill, R. (2023, March 14). Datasentre er en del av den digitale infrastrukturen—Og må reguleres deretter [Data centres are part of the digital infrastructure—And must be regulated accordingly]. *Stavanger Aftenblad*. https://www.aftenbladet.no/meninger/debatt/i/KnbjdM/datasentre-er-en-del-av-den-digitale-infrastrukturen-og-maa-reguleres-deretter
Stavanger City Council. (2016). *Roadmap for the Smart City Stavanger. Vision, goals and priority areas*.
van der Graaf, S. (2018). In Waze we trust: Algorithmic governance of the public sphere. *Media and Communication*, *6*(4), 153–162. https://doi.org/10.17645/mac.v6i4.1710
van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208.
Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.
Zhao, F., Fashola, O. I., Olarewaju, T. I., & Onwumere, I. (2021). Smart city research: A holistic and state-of-the-art literature review. *Cities*, *119*, Article 103406.
Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

About the Author



Carl Chineme Okafor is at this time a doctoral candidate in the Department of Media and Social Sciences at the University of Stavanger, Norway. He is a 2019 journalism fellow of the Reuters Institute for the Study of Journalism (RISJ)—the University of Oxford, UK. He has two MA degrees, in international diplomacy, and journalism and social communication. Datafication and journalism are his current research focus.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 355–366
https://doi.org/10.17645/mac.v11i2.6480

Article

Wellbeing Amid Digital Risks: Implications of Digital Risks, Threats, and Scams on Users' Wellbeing

Bindiya Dutt

Department of Media and Social Sciences, University of Stavanger, Norway; bindiya.dutt@uis.no

Submitted: 12 November 2022 | Accepted: 16 May 2023 | Published: 28 June 2023

Abstract
This study investigates how users perceive their wellbeing amid the risks associated with digital media use in Norway. According to the literature, some of these risks include digital dependence, online privacy, scams, thefts, information misuse, and harassment. To expand knowledge on how these and other digital risks are construed by users, this study addresses the following research questions: What implications do digital risks have on users' perceived sense of wellbeing? What are the solutions proposed by users to manage these risks? Methodologically, the inquiry is led through a qualitative approach comprising 17 semi-structured in-depth interviews of university students in Norway. The investigation centers on an interpretative phenomenological analysis. This study contributes to the existing literature by empirically evaluating the notion of digital wellbeing in the everyday choices of university students, thereby comprehending their safety concerns and how they manage online risks while exploring solutions to combat the risks of digital usage. The study adds value to the present literature on digital wellbeing by juxtaposing digital risks with the construct of wellbeing in digitalized societies.

Keywords
digital fraud; digital risks; digital wellbeing; mediatized risks; online scams; online theft

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

Whilst digital advances change communication behaviors and stimulate a social change in the way people use and access digital media in their day-to-day life, several risks emerge. These risks include, but are not limited to, online fraud, scams, stalking, information leakage, and data privacy (Balasubramanian, 2022; Masseno & Santos, 2018; Moore & Craciun, 2021). Though digitalization can be beneficial for instantaneous communication and connectivity, digital risks compromise users' online safety as well as autonomy. Likewise, with the personal data of users being registered electronically by digital platforms, social media apps, and mobile devices, data privacy becomes a serious concern (Gripsrud & Moe, 2010; Masseno & Santos, 2018), making users susceptible to digital risks while having ramifications on their perceived sense of wellbeing.

Even though previous research broadly depicts a connection between digital media usage and its adverse implications on wellbeing (Abeele, 2020; Baumer, 2013; Goodin, 2017; Moore & Craciun, 2021), it remains fragmented and does not explicitly link digital risks with all accounts of wellbeing—physical, social, psychological, and financial—thus creating a gap in the literature. Hence, this study aims to comprehend how users perceive their overall wellbeing amid the risks accompanied by digital usage. Studying users' perception of wellbeing considering the existing digital risks becomes important especially in a time when social interactions are increasingly moved online and it is often difficult to opt out. Thus, understanding digital wellbeing in the midst of digital risks becomes central in assessing digital daily life in a mediatized society. Further, through this investigation, the study intends to explore solutions that users propose to navigate risks

within the digital space that offers new trajectories for communication.

The investigation is carried out qualitatively including 17 in-depth interviews of university students in Stavanger, Bodø, and Oslo (Norway), to examine the implications of digital risks on their perceived wellbeing. This is because university students' wellbeing is currently at an all-time low (Baik et al., 2019) and digital risks are at an all-time high (Balasubramanian, 2022). Moreover, this query is pertinent because scholars contend that although digital use gives rise to several risks, less emphasis has been placed on citizens attitudes towards these risks (Grotto & Makridis, 2020).

Although this study acknowledges that digital media use can be beneficial for users, it focuses primarily on digital risks and the solutions to combat these risks. This is because the benefits of digital media use are accompanied by a myriad of risks that may create negative outcomes for users. Further, on a micro-level, this study discovers how users articulate their wellbeing around the risks of digitalization. In order to interpret the findings and propose relevant solutions to manage digital risks, the research rests on an interpretative phenomenological analysis (J. A. Smith & Osborn, 2004). An interpretative phenomenological analysis explains the crux of an existing phenomenon while eliciting responses from several individuals who share analogous experiences (Creswell & Poth, 2017). Here forth, the following sections present a literature review, methodology, research findings, and discussion.

2. Literature Review

The literature review begins by presenting a generic conceptualization of wellbeing while weaving in diverse theoretical perspectives. Thereafter, it delves into the construct of digital wellbeing and addresses existing risks within the digital media landscape. Additionally, it unveils the discourse on mediatization as well as on digital disconnection proposals in relation to wellbeing.

2.1. The Theoretical Paradigm of Wellbeing

Whilst the current paradigm of wellbeing described below addresses how literature theorizes "the good life," it does so without considering specific contexts which may encompass risk susceptibility, such as the digital arena. In fact, wellbeing within the digital context carries a whole new set of characteristics that may contradict the original idea of wellbeing.

Theories that focus on the generic concept of wellbeing attempt to identify things that are in due course good for an individual (Tiberius, 2020). For instance, list theories establish wellbeing with a list of items such as job, finances, work-life balance, social engagements, etc. (Haybron, 2008), whereas desire theory entails the fulfilment of one's desires (Haybron, 2008). However, these criteria-based wellbeing concepts do not reflect a comprehensive perspective on maintaining wellbeing especially when one's desires are not met, and the lists remain unchecked.

Likewise, activity theory "has popularized the idea that active involvement in activities causes happiness while enhancing wellbeing" (Diener, 1984, p. 558). Whereas flow theory proposes achieving a state of wellbeing through engaging in present-moment tasks, instead of being anxious about the future or obsessing over bygones (Csikszentmihalyi et al., 2014). Although activity and flow theories offer a framework to understand the processes that might lead to achieving a sense of wellbeing, a constant state of flow may neither be plausible nor make individuals risk averse. Besides, not all activities may augment wellbeing. In addition, these theoretical perspectives rule out the element of assessing and navigating risks.

Although the notion of wellbeing is embedded in creating a pleasant atmosphere in all spheres of life including physical, mental, emotional, and spiritual (M. Smith & Puczkó, 2008), the phenomenon of wellbeing remains ontologically subjective (Sumner, 1996). This is because wellbeing is based on one's personal assessment of life and remains grounded in how one perceives or appraises their experiences (Huppert, 2014). Based on this premise, wellbeing may not be entirely subject to generic pre-defined concepts but rather to individual experiences as well as one's state of mind at any given moment. Some philosophers take an antithetical stance on generalizing wellbeing. Instead, they assert that wellbeing must be looked at from the purview of specific contexts (Alexandrova, 2017). Therefore, context-specific inquiries centered on individual perception could augment wellbeing research.

2.2. Mediatization, Digital Risks, and Wellbeing

While traditional wellbeing theories have neglected the digital context, the theory of mediatization throws light on how media processes facilitate social change through digital communication (Hjarvard, 2013). In this study, I use Schulz's (2004) four components of the mediatization process which include *extension, substitution, amalgamation,* and *accommodation* as the analytical lenses to gauge how individuals understand digital media practices and their inherent risks. *Extension* suggests that media extends spatially and temporally wherein humans can receive digital messages easily. *Substitution* involves replacing face-to-face social interactions with digital communication. *Amalgamation* entails having mediatic involvement simultaneously coming together with non-media interactions. *Accommodation* centers on individuals accommodating to the way media operate, as they adapt to media logic.

Although these four components of mediatization explain how media processes bring on new approaches to communication, they do not reflect the risks associated with digital use. Although theorists acknowledge

that digital media's entry into different areas of life and its assimilation may pose challenges (Hjarvard, 2013), these challenges are neither explicitly nor cohesively spelled out, thus calling for deeper investigation.

Also, despite the challenges and risks, individuals are impelled into the mediatized arena. The mediatization of life may act as a push factor in facilitating communication through digital media channels. Owing to this, individuals are not only getting trapped in the digital sphere but they are also being set up for digital risks which may compromise their wellbeing. Furthermore, the negative outcomes of digital media use can create stress. The theory of stress and coping positions stress as a cognitive experience contingent upon how individuals appraise their association with a given environment (Lazarus, 1998). Research findings hint that some of the digital stress occurs due to connection and information overload (Andrejevic, 2013; LaRose et al., 2014). In this regard, digital wellbeing is interpreted as the balance that users may experience in being connected to digital platforms (Abeele, 2020). Though this insight may hold true to some extent, the concern remains whether digital wellbeing is simply a balancing act. Ruling out risks while comprehending digital wellbeing may constitute only a partial explanation.

The pervasiveness of media content (Couldry & Hepp, 2013) as well as the convergence of media technologies in the mundane (Deuze, 2011) give rise to several risks. While literature describes risk as perceived ambiguity (Holton, 2004), risk can be demarcated by three elements: probable loss, the consequence of loss, and the uncertainty of loss (Yates & Stone, 1992). Moreover, scholars contend that digital risks come entwined with the content, contact, and conduct of users, implying that risks may be subject to the online content that users are exposed to, whom they contact or communicate with, and the quality of communication which takes place digitally (Livingstone & Helsper, 2008).

As digital risks and threats are often used interchangeably in the literature, they can range from financial theft to cyberbullying and identity stealing. For example, cyberbullying literature finds little consonance in a standardized definition, and within the definitions presented, none comprise the word "threat" (Espelage & Hong, 2017). Research does not offer precision on how to conceptualize threats within the digital realm (Patton et al., 2019). However, a risk or a threat may be construed as harmful behavior intended to harass someone repeatedly using digital technology (König et al., 2010; P. K. Smith, 2009). Perhaps owing to several routine processes of life being online, from shopping to banking, users' reliance on digital devices can make the digital space lucrative for cybercriminals and online predators (Lallie et al., 2021). Not only that, but the digital world is also encapsulated by several other risks such as tracking users' data, threatening online privacy, stalking through location sharing, and harassment. The reasons why digital media use may pose risks for users are broadly highlighted below.

Firstly, participatory media practices make it much easier to track and gauge the users' attitudes (Ferrer-Conill, 2017). Due to digital metrics, companies know more about users than they knew before (Tandoc, 2014). Large enterprises, such as Google and Facebook, continually set up infrastructures around the world to store the users' data thus creating monopolistic trends and gaining power (Trittin-Ulbrich et al., 2021). Meanwhile, they employ available data to cater to the users by offering content as per their preferences (Arsenault, 2017). While companies seek to capitalize on users' data (Yoo et al., 2010) and advertisers push products to consumer segments by predicting patterns of online behavior (Shareef et al., 2018), apprehensions about data storage, protection, ownership, and privacy emerge.

Secondly, as opposed to traditional television and newspaper models, online media platforms are able to target wider audience segments (Fuchs, 2018). Users are deliberately targeted on social media platforms thus turning technical data into a socially covert influencer (Bolsover & Howard, 2017). For example, data analytics agencies such as Cambridge Analytics were accused of stealthily manipulating voters during the US political elections (Symeonidis et al., 2018). The problem of audiences' rights and privacy intensifies through such cases posing a threat to democracy as propaganda takes centerstage through digital media platforms.

Thirdly, the increasing number of online predators using the internet for harassment, sexual abuse, hacking, and theft poses new questions about how to cope with issues of personal security (Pawar et al., 2021). Research finds that cyberstalking is similar to offline traditional stalking behaviors that victimize and violate private space (Sheridan & Grant, 2007), thus risking a sense of wellbeing among cyber victims. Besides, in a quest for new connections and friendships, many users drift towards dating apps (Chen & Rahman, 2008). However, dating apps present various problems. For instance, location-sharing intensifies complications for users (Gillett, 2018). Sharing personal information can become a means to ease online stalking and harassment (Chugh & Guggisberg, 2022; Phan et al., 2021; Tokunaga & Aune, 2017). In this regard, digital systems can facilitate manipulation (Lee et al., 2019). Catfishing scams involve both financial and psychological risks for the victims, such as a loss of self-esteem, trauma due to experiencing deception, a state of shock, and feelings of distrust (Whitty & Buchanan, 2016). Such outcomes can leave adverse and even long-term negative associations for victims who may not only lose money during the online dating process but also lose a relationship which once appeared promising, thus compromising their sense of wellbeing. Although perpetual swiping may appear to offer wider options, its authenticity remains debatable. Further, the value of real-time human connection often gets compromised and replaced by online communication in the commodified app world (Krüger & Charlotte Spilde, 2020) thus

making deception easier in digital communication scenarios wherein one may not know who's hiding behind the screen.

Despite the aforementioned risks being present, users may either remain unaware of the risks concerning their privacy and security (Couch et al., 2012) or underrate the risks (Grotto & Makridis, 2020), while being oblivious to the repercussions of digitalization. Research shows that most users are not apprehensive about the likelihood of being a target of a scam or fraud as a result of their digital presence (Blank et al., 2019). Internet or digital addiction, including gaming and smartphone addiction (Almourad et al., 2020; Widyanto & Griffiths, 2006), as well as other issues such as aimless surfing leading to digital overuse and impulsive digital behaviors may be too compelling (Montag & Walla, 2016). Owing to this, users might downplay the risks associated with digital usage (Aboujaoude & Gega, 2021).

While digital risks have implications on different levels on users, they can be grouped into four broad categories: physical, financial, psychological, and social risks. For example, findings from a study show that excessive digital use accompanies physical problems, such as eye irritation and blurred vision (Gowrisankaran & Sheedy, 2015). Users prone to digital addiction explain having sleep difficulties, feelings of anxiety, and obsessions (Bakken et al., 2009). Whereas, from the standpoint of users' psychosocial wellbeing, dependence on digital devices may divert users' attention from forging real-time social connections (Dutt & Selstad, 2021). Other digital risks that may jeopardize a sense of wellbeing comprise online harassment, fraud, and deception. Research depicts that oppressive online exchanges result in reduced mental wellbeing (Festl et al., 2019).

Turning the focus to social media risks, scholars discover that unbalanced social media usage may relate to disorders, such as excessive selfies, self-obsession, self-promotion, and loss of interest in other hobbies (Gomez et al., 2022; Tang et al., 2022). Further, intimidation on social media platforms depicts adverse consequences on the wellbeing of online users, such as heightened stress, despair, anxiety, and behavioral problems (Kowalski et al., 2014). While studies find a damaging association between social media use and self-image (Faelens et al., 2021), online conflicts echo undesirable communication outcomes posing psychological and social risks. Research shows that social media engagement links to lower self-esteem among university students (Errasti et al., 2017). The fear of being excluded or missing out on information could constitute reasons for social media use. As social exclusion lessens a sense of wellbeing (Sjåstad et al., 2021), the need for inclusivity and connection may turn users towards social media platforms. Moreover, becoming influenced by a group may be another reason for mimicking social media behaviors (Aral, 2014; Macït et al., 2018).

In support of users' wellbeing, digital detox programs suggesting temporary or lasting digital disconnection emerge (Jorge, 2019; Syvertsen & Enli, 2020). However, these proposals put the entire onus of digital wellbeing on the users without reflecting on digital marketing strategies and other external factors that may pull users towards digital platforms. While academics indicate a linkage between digital disconnection and wellbeing (Baumer, 2013; Bélair-Gagnon et al., 2022; Bratsberg & Moen, 2015; Karppi et al., 2021; Syvertsen & Enli, 2020), whether or not users would opt to refrain from using digital media platforms remains contentious. The users' decisions may be guided by enforced digital usage (González-López et al., 2021), coerced digital usage (Barassi, 2019), or digital compulsions, as well as by the dopamine cycle which often centers on anxiety while anticipating rewards from digital activities (Macït et al., 2018). Therefore, whether disconnection and detox proposals are practical enough in the current mediatized scenario where digital use may not be an option but a necessity or compulsion requires further inquiry. To discover relevant answers pertaining to digital risks, this study proposes the following research questions:

- RQ1: What implications do digital risks have on users' perceived sense of wellbeing?
- RQ2: What are the solutions proposed by users to manage digital risks?

3. Methods

To gauge how informants in this study perceive wellbeing amid the risks associated with digital media usage and what remedies they imagine for the problems they identify with, the study relied on a qualitative inquiry. This was carried out through 17 in-depth semi-structured interviews, lasting between 45 and 75 minutes each, of university students in Stavanger, Bodø, and Oslo (Norway). In-depth interviews were used as they serve as processes that account for user experiences (Charmaz, 1990) and provide a detailed understanding of user perspectives while also offering a substantial description of their social environment (Silverman, 2016).

The interview guide included questions related to users' concerns about digital risks and the potential solutions to combat the risks. The risks were preclassified based on four broad categories including physical risks, financial risks, psychological risks, and social risks. The interviews were audio recorded, transcribed, and assembled into a corpus of textual data.

The study employed a purposive sampling method (Tongco, 2007) and the sample size was contingent upon the strategy of saturation (Mason, 2010), wherein new data ceased to offer any fresh information. Informant selection was based on their student status at various universities. Informants were recruited through the university library, student organizations, research schools, university housing, as well as campus sports clubs. To maintain ethical standards, all informants were made aware that their interview responses would be included

as part of a research project and that their identities would be kept confidential. Interviewees were offered a synopsis of the research at the beginning of the interviews. Additionally, they were assured that their data would be deleted upon the completion of the research. Participation in this research was purely voluntary and was not incentivized. Prior to the interviews, the study was granted ethical approval by the Norwegian Centre for Research Data (under Project No. 314257). Table 1 provides the informant sample profile.

The mode of analysis rested on an interpretative phenomenological analysis (J. A. Smith & Osborn, 2004) to place findings within the larger context of the role of digital media and its implications on wellbeing. Interpretative phenomenological analysis takes the direction of a reflective analysis wherein the researcher shows active engagement in comprehending the interviewees' account through initial note-taking and thereafter making a detailed analysis by looking at patterns that emerge from each interview (P. K. Smith et al., 2009). In this study, the analysis explored how study participants comprehended digital risks and perceived their wellbeing amid the challenges they experienced while capturing the vital elements of these experiences.

4. Findings

The findings from this research are arranged into the following sub-sections and are analyzed through four risk categories: physical risks, financial risks, psychological risks, and social risks. The implications of digital risks on informants' perceived wellbeing are synthesized into these sub-sections which present extracts from informant interviews along with the proposed solutions to manage digital risks.

4.1. Digital Risk Perception

One of the key findings of this study shows that informants feel helpless against having to engage in digital processes due to the forced digitalization of mundane services. They view digital use as a compulsion rather than an option. Since most of the services in Norway are digitalized, informants feel forced to get things done online even if they do not wish to. They also concur that coerced digitalization can become problematic, especially for those individuals that are not digitally savvy or do not wish to have their data shared on public platforms.

Informants agree that information sharing and excessive assimilation of digital media into daily life pose a threat to security while adversely influencing wellbeing. Several informants state that they would prefer using certain services non-digitally, for instance making doctors' appointments over a phone call or in-person to maintain privacy, rather than registering their health data online. This is exemplified by the following quote from a student in Stavanger:

> In Norway, everything is becoming digital. Oftentimes, I feel like I am being forced to share private information on digital platforms. For instance, if I wish to make a doctor's appointment, I must do it online. Whether I am comfortable sharing my personal health history online is not their concern. This kind of compulsion is not acceptable to me, and I strongly feel that there ought to be other non-digital options for users who do not consent to share personal details digitally. (P8)

The above finding echoes with the component of *substitution* within mediatization theory which affirms that mundane interactions are getting substituted with digital

Table 1. Informant sample profile.

Informants	Gender	Field of study	Study program	Nationality	University location
P1	Male	Computer science	Master's	Nepalese	Stavanger
P2	Male	Data science	Master's	Kenyan	Stavanger
P3	Male	Theology	Bachelor's	Norwegian	Oslo
P4	Male	Physics	Master's	British	Stavanger
P5	Male	Risk management	Post-doc	Colombian	Stavanger
P6	Female	Data science	PhD	Indian	Stavanger
P7	Male	Geology	Post-doc	American	Stavanger
P8	Male	Computer science	Bachelor's	Norwegian	Stavanger
P9	Male	Engineering	PhD	Pakistani	Stavanger
P10	Female	Theology	Master's	Norwegian	Oslo
P11	Male	Pedagogy	Bachelor's	Norwegian	Stavanger
P12	Male	Political science	PhD	British	Stavanger
P13	Male	Business	Bachelor's	Norwegian	Stavanger
P14	Female	Biology	PhD	Brazilian	Stavanger
P15	Male	Petroleum engineering	PhD	Iranian	Stavanger
P16	Male	Social work	PhD	Ethiopian	Bodø
P17	Female	Psychiatry	PhD	Norwegian	Bodø

interactions. This not only poses privacy risks for users but also threatens their sense of freedom especially when they prefer not to use digital platforms for information sharing. Besides, informants identify several other digital risks and believe that they cannot trust digital media platforms. Mainly, the risks of being hacked and financially scammed pose a top threat. In this regard, most informants describe feeling digitally unsafe and do not trust the government to protect them from digital risks. While those studying data or computer science notice the red flags in the digital context, other informants exhibit unawareness towards matters concerning privacy risks. However, when asked specifically about online privacy, most disclose not being comfortable having their private information, such as their address and phone numbers, displayed on digital platforms. Informants express feeling worried when they see their personal information flashed online and cannot have it removed at their discretion. Several informants also find digital media tracking, recommendations, and personalized advertising intrusive:

> Since I am studying data science, I watch out for red flags on digital platforms to ensure I am not scammed. But I know of many fellow students who have been victims of digital scams. There is too much risk in the digital environment—hacking, spam, tracking, fraud—and the list goes on. One way to avoid digital risks is to be cautious and increase digital literacy. (P2)

This implies that online privacy remains a serious concern, a breach of which may lead to digital misconduct. Implementing robust security measures may support privacy protection. Additionally, the removal of personal information from online platforms at the users' discretion would not only offer digital users a sense of autonomy but also strengthen feelings of safety and wellbeing.

4.2. Implications of Digital Risks

4.2.1. Physical Risks

Amongst the physical risks, informants report experiencing tired eyes, shoulder stiffness, wrist pain, hip pain from sitting for many hours, and other postural problems after using digital devices at a stretch. One of the reasons that the informants present for being digitally dependant is having free access to wireless networking at Norwegian universities, which makes it easy to navigate the internet. Informants agree that not having free internet access 24/7 would limit their digital consumption. As a remedial measure, having to pay for digital use may curb digital dependence:

> Digital use has seeped into all areas of life. It has become overwhelming due to all assignments being digital. Interpersonal and social communication is also digital. I feel like I am constantly staring at a screen, even while commuting or waiting at a restaurant. This has affected my eyes, posture, and sleep cycle. (P3)

The component of *amalgamation*, within mediatization, which suggests that media activities come together with non-media interactions resonates with this finding. While informants reveal engaging in several mundane activities through digital media platforms, they also admit that they use technology while performing other tasks. Though having perpetual access to digital technology fills the communication gap, it tends to facilitate digital dependence, thus depleting a sense of wellbeing.

Solutions to cope with the physical problems that arise due to digital use include taking enough breaks as well as partaking in complementary wellness practices such as yoga, meditation and stretching to relax the body and relieve muscle stiffness. Likewise, participating in other non-digital activities, such as physical exercise, going for walks, being in nature, playing board games rather than staring at a screen, and meeting friends in person instead of chatting online can be beneficial for overall wellbeing. Also, being goal-oriented in one's digital use by deciding beforehand what needs to be searched for online can limit screen time. Turning off digital devices long before going to bed is another way to manage physiological risks.

4.2.2. Financial Risks

Despite the benefits of digital banking, the fear of online fraud and theft remains a serious concern among informants. Other financial risks entail buying compulsiveness and easy access to online shopping. Informants concur that digital scams resulting in financial losses leave a lasting impression on them:

> The Norwegian identity number allotted to individuals is used everywhere, which does not make things safer. I know a few students who have been victims of online theft in Norway. They have clicked on links appearing to be sent by their bank, given away their one-time passwords, and have fallen into the scammers' trap. The police could not track the scammers, leaving the victims distraught. (P1)

While financial scams create a sense of tangible loss, they also create mental and emotional friction. This finding resonates with the reflections in literature which assert that financial loss is not a standalone occurrence. Rather, it comes intertwined with psychological risks for the victims. For instance, victims of catfishing scams on dating websites often experience mixed emotions ranging from a state of shock, remorse, guilt, self-blame, and trauma due to being deceived (Whitty & Buchanan, 2016). Such outcomes can leave the victims with unpleasant memories. In the process, they may not only lose money but

also lose a sense of trust, thus disturbing their sense of wellbeing.

Suggestions to manage financial risks include being alert and aware of online scams. Furthermore, informants recommend not clicking unknown links, not giving one-time passwords to anyone online or over the phone, and not sending money to digital acquaintances on dating websites to avoid catfishing scams. Other steps involve securing one's financial details and passwords. Employing a strong verification system with the bank can help prevent financial fraud. Additionally, using reliable websites for online purchases could be worthwhile.

4.2.3. Psychological Risks

Informants report that digital devices can become addictive, posing a danger to their psychological wellbeing. Further, a technological breakdown can result in inconveniences and delays especially when alternative avenues are unavailable:

> I often feel a sense of remorse and regret after using digital devices for too many hours. It seems like a waste of time scrolling endlessly and gathering information which serves no purpose. I have decided to improve my digital hygiene and engage in non-digital activities. (P5)

Solutions to prevent feelings of regret due to overusing digital devices involve controlling the constant temptation to search and scroll online. This can be done by setting specific time frames for digital activities. Practicing digital hygiene by compartmentalizing time for digital and non-digital activities can lead to a balanced approach while preventing wastage of time or remorse thereafter.

More specifically in response to digital fraud, an informant presents a detailed account of the psychological impressions following a scam:

> Before coming to Norway, I registered on a home rental website where a homeowner offered to rent me a room in his house. He asked me to send him rent money in advance. Although skeptical, I was in desperate need of a place to stay, at least for the initial months. So, I sent the advance money. When I asked him for the contract, he dodged me and instead suggested that I could date his girlfriend. He kept sending me her pictures. When I insisted on getting the contract or money back, he blocked me on the chat and disappeared. I could not track him. This incident was highly unexpected, it made me feel cheated. The money gone was one thing, but it shook me up mentally. I questioned my smartness for a long time and blamed myself. (P16)

While informants victimized by digital scams tend to blame themselves, experts suggest practicing self-compassion. Additionally, controlling digital impulses could be a tool to manage falling into the risk zone. Waiting to respond to digital requests allows time for reflecting on whether the offer is authentic. Getting a second opinion and carefully contemplating the situation before reacting spontaneously or giving out money instantly can help avert fraud. Apart from this, developing coping strategies such as altering the problem or changing emotional responses to problems can assist in handling psychological stress (Lazarus, 1998).

4.2.4. Social Risks

Oscillating between social media rewards and risks seems to activate a conflicting stance amongst informants. They report that the need for attention and acknowledgment drives them towards social media platforms, whereas the fear of social isolation is one of the factors that pulls them into social media use. Apart from this, informants fear missing out on pertinent information as well as feeling excluded when not present on social media platforms:

> Not being on social media ends up inducing a fear of missing out. But then being present on social media platforms triggers online comparison and anxiety. When I see others glamourizing their persona, posting happy pictures on social media to project a certain image, or boasting about their accomplishments, I feel like their life is perfect and everyone else needs to measure up to the social media standards. Comparison doesn't feel good but social media facilitates it. (P12)

Although access to social media platforms may ease communication while removing constraints of time and space, as seen in the component of *extension* within mediatization, it also elicits feelings of comparability amongst users. Passively consuming others' social media feeds tends to trigger online comparison. While social media presence may offer a sense of temporary inclusivity and connection, informants describe feeling anxious due to lacking something in comparison to the mediatic lives of others. Moreover, informants reveal that their personal safety gets compromised while meeting digital acquaintances in person, for instance through dating websites which they construe as risky:

> People are posting everything on social media, even the uninteresting stuff. Private matters are now made public through social media platforms. It seems like people are constantly seeking acknowledgment and attention. Perhaps, they are desolate and need support from others. But digital life cannot be equated with real life; no amount of digital connection can fill the real-life gaps. (P17)

To relieve these concerns, suggestions to manage social risks include not disclosing private information on digital

platforms which may later backfire. Alternatively, finding ways to connect with people non-digitally could serve as a solution. Also, informants suggest steering away from online comparison by raising digital literacy and mindful self-training. Being cognizant that digital notifications are not rewards could also be helpful in coping with digitally induced social stressors.

5. Discussion

In response to the main research questions, findings indicate that digital risks have several implications on the users' wellbeing as they traverse the digital space. As these risks appear in varied areas of life including physical, psychological, financial, and social, they often linger unmanaged either due to obscurity or external threats that may not be in the users' domain. Even when digital risks can be managed by users to some extent through employing self-discretionary techniques, external threats to privacy continue to persist, actuating the dichotomy of control. Hence, the proposed solutions offered by informants to deflect negative outcomes of digital use not only include self-regulatory behaviors but also extend to interventions on diverse levels.

Although this study finds that while many informants downplay risks, when asked to conscientiously contemplate digital risks they opine that these risks must be managed by government organizations and digital policymakers leading to more secure measures for user protection. Leaving a digital trail behind raises privacy concerns among users. Therefore, to circumvent privacy risks, one of the propositions that the informants cohesively agree on is that their private data be easily removed from digital platforms upon request regardless of whether they previously consented to having it online.

Findings also depict that the approach to digital risks is contingent upon the experiences of individual users as well as whom they digitally encounter. For instance, those informants who have experienced digital financial scams are more wary than others that have not experienced them. Also, the construal of risks varies amongst informants; most informants remain apprehensive about tangible digital financial losses as opposed to psychological risks that may be difficult to identify instantly and take time to process.

Informants tend to overlook the risks associated with digital use on a day-to-day basis as they assert being dependent on their devices. This corroborates previous literature findings that users often underestimate risks (Aboujaoude & Gega, 2021; Grotto & Makridis, 2020). Likewise, in regard to assessing probable loss (Yates & Stone, 1992), informants unscathed by digital scams do not seem to anticipate the likelihood of incurring loss while engaged in digital exchanges. However, those that have either been victimized by digital scams or have been threatened by fraudsters respond differently having witnessed the consequence of loss. This shows that digital scams can leave a lasting impression on the victims.

The theory of mediatization shows itself in the findings through the components of *extension*, *substitution*, and *amalgamation*. Displaying semblance with *extension*, informants acknowledge that digital technology eases communication and removes constraints of time and space. However, easy access to social media platforms encompasses the risk of online comparisons. As established in the component of *substitution*, informants contend that digital technology replaces human connection with media activities. Swapping media activities with non-media activities, such as face-to-face communication and community building, tend to facilitate social disconnection thus hindering a sense of wellbeing. Additionally, informants admit to engaging in digital activity even while performing other tasks, which echoes with *amalgamation*. The component of *accommodation* does not reflect in the findings as informants do not explicitly articulate how they adapt to media logic. However, findings support the claim that excessive assimilation of digital media into daily life may accompany challenges (Hjarvard, 2013; Schulz, 2004). Informants express an aversion to excessive assimilation of digitalization into daily life processes, particularly in response to their privacy, in financial and health matters. Also, over-digitalization in Norway seems to deplete a sense of interpersonal and social connection due to digital dependence which alters the wellbeing perception among informants.

Conversely, differing from the premise of activity theory, findings show that not all activity may offer a sense of wellbeing. An over-indulgence in digital activity may deplete happiness by triggering feelings of remorse amongst users. While discussing proposals such as digital disconnection and digital detox (Jorge, 2019; Syvertsen & Enli, 2020), this study uncovers that such propositions may not be practically feasible because of coerced digitalization. Further, this particular finding supports previous literature which suggests that digital coercion or enforcement (Barassi, 2019; González-López et al., 2021) serve as a deterrent to withdrawal. Additionally, digital compulsions that push the reward-seeking dopamine cycle through digital activities may dissuade users from digital disconnection (Macït et al., 2018).

6. Conclusion

This study contributes to the current literature by empirically assessing the notion of digital risks and wellbeing in cohesion so that digital wellbeing is not an afterthought. Conclusively, findings depict that although digital wellbeing entails striking a balance during digital activities, it also involves assessing, preventing, and managing risks to inhibit negative outcomes that may otherwise emerge from digital use. Accordingly, expanding on existing definitions, digital wellbeing can be construed as feeling safe and equipped to manage risks in all areas including physical, psychological, financial, and social, within the digital environment.

While informants in this study show how mediatization materializes in their daily life as they combine digital and non-digital activities seamlessly, they also articulate their feelings around mediatization. As seen in the component of *substitution* within mediatization, due to untethered digital access informants often tend to substitute face-to-face communication with digital communication. However, relying predominantly on digital communication to connect with others entails social risks, such as experiencing social dissatisfaction. Moreover, echoing the component of *amalgamation* within mediatization, informants engaged in digital activity while simultaneously performing other activities express having a shorter attention span, barring them from focusing on a singular task. Similarly, resonating with the component of *extension* within mediatization, despite the ease of communication and perpetual access to a wider digital network, social media platforms enclose the risk of online comparisons which pose psychological risks. These revelations not only provide a glimpse into the ways in which different components of mediatization manifest in the mundane, but it also pushes the theoretical discourse further by offering insights into how informants interpret mediatization.

Further, the key takeaways of the study encompass conscientious risk cognizance, risk management skills as well as effective coping mechanisms to deal with undesired outcomes of digital use. While findings illustrate that most users are aware of digital risks and desire stronger external measures for risk management, digital dependencies and risk denial may tend to outweigh risk cognizance. Also, different digital risk categories carry distinct implications on the users' perceived sense of wellbeing. For instance, most users reiterate feeling threatened by tangible financial scams and data privacy issues yet display risk aversion and tolerance towards social or physical risks that they may experience within the digital context. This could hint that definite and noticeable risks may be regarded as remarkably intense and easier to register than abstract risks.

To augment a sense of wellbeing amid existing risks, some adaptive tools to cope with digitally induced distress include removing stressful triggers, reappraising tense situations, taking breaks from technology, journaling, practicing relaxation techniques such as yoga, finding humor to engage in positive emotions, seeking support from others, and offering gratitude for the good things in life. These solutions are feasible to the extent wherein users are free from external constraints and can exercise autonomy in their own digital behaviors. However, concerns continue to remain at the government as well as the digital policy level that users have no control over.

Though the ubiquitous digital technology offers conveniences, it poses several risks, as revealed in the narrative, thus calling for substantial interventions to mitigate these risks. Digital risk management requires intervention on various levels, including the government, digital policymakers, digital platform creators, university organizations, and end users. These interventions could include updated regulations to protect users' data privacy, the development of user-centric digital policies, national-level programs to raise awareness of digital risks, and educational initiatives for digital risk assessment and digital risk management. Additionally, offering coping resources to scam victims as well as implementing practical measures such as removing the private data of users from digital platforms upon request while supporting users' right to withdraw consent can foster a safer digital environment. Likewise, creating robust digital protection systems could help relieve digital risks.

Although using self-discretion to protect oneself from digital risks is a key factor in staying digitally safeguarded, it may not be enough to restrict negative outcomes. Thus, future research could be directed towards digital policies and the government's role in protecting users from experiencing digital risks. Only when digital platforms, digital policies, and government bodies are collectively in sync with users' rights to digital security can the users experience complete autonomy, safety, as well as wellbeing in the digital realm.

As a limitation, this study includes a niche sample, university students who are digitally savvy. However, the sample is heterogeneous. Although the informants in the study are university students (the only common factor amongst them), they are from diverse nationalities, study programs, and study levels. Most of them work part-time and some have previously worked in full-time jobs, enabling them to bring a varied as well as a multi-cultural perspective in response to the inquiry. Still, future studies seeking the viewpoint of other digital user segments such as high school students or full-time employees in various industries could offer further insights into how they might perceive the wellbeing concept amid digital risks.

Acknowledgments

This work was supported by the Norwegian Research Council (Project No. 314257). Special thanks to Dr. Helle Sjøvaag and Dr. Raul Ferrer-Conill.

Conflict of Interests

The author declares no conflict of interests.

References

Abeele, M. V. (2020). *Digital wellbeing as a dynamic construct*. *Communication Theory*, *31*(4), 932–955. https://doi.org/10.1093/ct/qtaa024

Aboujaoude, E., & Gega, L. (2021). Editorial Perspective: Missing the forest for the trees—How the focus on digital addiction and gaming diverted attention away from wider online risks. *Child and Adolescent Mental Health*, *26*(4), 369–371. https://doi.org/10.1111/camh.12503

Alexandrova, A. (2017). *A philosophy for the science of well-being*. Oxford University Press.

Almourad, M. B., McAlaney, J., Skinner, T., Pleya, M., & Ali, R. (2020). Defining digital addiction: Key features from the literature. *Psihologija*, *53*(3), 237–253.

Andrejevic, M. (2013). *Infoglut: How too much information is changing the way we think and know*. Routledge.

Aral, S. (2014). The problem with online ratings. *MIT Sloan Management Review*, *55*(2), 47–52.

Arsenault, A. H. (2017). The datafication of media: Big data and the media industries. *International Journal of Media & Cultural Politics*, *13*(1/2), 7–24.

Baik, C., Larcombe, W., & Brooker, A. (2019). How universities can enhance student mental wellbeing: The student perspective. *Higher Education Research & Development*, *38*(4), 674–687.

Bakken, I. J., Wenzel, H. G., Götestam, K. G., Johansson, A., & Øren, A. (2009). Internet addiction among Norwegian adults: A stratified probability sample study. *Scandinavian Journal of Psychology*, *50*(2), 121–127.

Balasubramanian, A. (2022). *Data, digital risks, and financial markets* [Unpublished doctoral dissertation]. Stanford University.

Barassi, V. (2019). Datafied citizens in the age of coerced digital participation. *Sociological Research Online*, *24*(3), 414–429.

Baumer, S. (2013). Social media, human connectivity and psychological well-being. In S. Price, C. Jewitt, & B. Brown (Eds.), *The SAGE handbook of digital technology research* (pp. 71–87). SAGE.

Bélair-Gagnon, V., Bossio, D., Holton, A. E., & Molyneux, L. (2022). Disconnection: How measured separations from journalistic norms and labor can help sustain journalism. *Social Media + Society*, *8*(1). https://doi.org/10.1177/20563051221077217

Blank, G., Dutton, W. H., & Lefkowitz, J. (2019). *Perceived threats to privacy online: The internet in Britain. Oxford internet survey 2019*. Oxford Internet Institute.

Bolsover, G., & Howard, P. (2017). Computational propaganda and political big data: Moving toward a more critical research agenda. *Big Data*, *5*(4), 273–276. https://doi.org/10.1089/big.2017.29024.cpr

Bratsberg, L., & Moen, T. (2015). *Logg av* [Log off]. Cappelen Damm.

Charmaz, K. (1990). "Discovering" chronic illness: Using grounded theory. *Social Science Medicine*, *30*(11), 1161–1172.

Chen, G., & Rahman, F. (2008). Analyzing privacy designs of mobile social networking applications. In M. Guo, Z. Wang, F. Tang, & C.-Z. Xu (Eds.), *2008 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing* (pp. 83–88). IEEE.

Chugh, R., & Guggisberg, M. (2022). Stalking and other forms of dating violence: Lessons learned from you in relation to cyber safety. *Journal of Interpersonal Violence*, *37*(9/10), NP6760–NP6784.

Couch, D., Liamputtong, P., & Pitts, M. (2012). What are the real and perceived risks and dangers of online dating? Perspectives from online daters: Health risks in the media. *Health, Risk & Society*, *14*(7/8), 697–714.

Couldry, N., & Hepp, A. (2013). Conceptualizing mediatization: Contexts, traditions, arguments. *Communication Theory*, *23*(3), 191–202. https://doi.org/10.1111/comt.12019

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE.

Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2014). Flow. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology* (pp. 227–238). Springer.

Deuze, M. (2011). Media life. *Media, Culture & Society*, *33*(1), 137–148.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575.

Dutt, B., & Selstad, L. (2021). The wellness modification of yoga in Norway. *International Journal of Spa and Wellness*, *5*(1), 33–49. https://doi.org/10.1080/24721735.2021.1948274

Errasti, J., Amigo, I., & Villadangos, M. (2017). Emotional uses of Facebook and Twitter: Its relation with empathy, narcissism, and self-esteem in adolescence. *Psychological Reports*, *120*(6), 997–1018.

Espelage, D. L., & Hong, J. S. (2017). Cyberbullying prevention and intervention efforts: Current knowledge and future directions. *The Canadian Journal of Psychiatry*, *62*(6), 374–380.

Faelens, L., Hoorelbeke, K., Cambier, R., van Put, J., Van de Putte, E., De Raedt, R., & Koster, E. H. W. (2021). The relationship between Instagram use and indicators of mental health: A systematic review. *Computers in Human Behavior Reports*, *4*, Article 100121. https://doi.org/10.1016/j.chbr.2021.100121

Ferrer-Conill, R. (2017). Quantifying journalism? A study on the use of data and gamification to motivate journalists. *Television & New Media*, *18*(8), 706–720.

Festl, R., Reer, F., & Quandt, T. (2019). Online sexual engagement and psychosocial well-being: The mediating role of sexual victimization experiences. *Computers in Human Behavior*, *98*, 102–110.

Fuchs, C. (2018). Propaganda 2.0: Herman and Chomsky's propaganda model in the age of the internet, big data and social media. In J. Pedro-Carañana, D. Broudy, & J. Klaehn (Eds.), *The propaganda model today: Filtering perception and awareness* (pp. 71–91). University of Westminster Press.

Gillett, R. (2018). Intimate intrusions online: Studying the normalisation of abuse in dating apps. *Women's Studies International Forum*, *69*, 212–219. https://doi.org/10.1016/j.wsif.2018.04.005

Gomez, M., Klare, D., Ceballos, N., Dailey, S., Kaiser, S., & Howard, K. (2022). Do you dare to compare?: The key characteristics of social media users who

frequently make online upward social comparisons. *International Journal of Human–Computer Interaction*, *38*(10), 938–948.

González-López, Ó. R., Buenadicha-Mateos, M., & Sánchez-Hernández, M. I. (2021). Overwhelmed by technostress? Sensitive archetypes and effects in times of forced digitalization. *International Journal of Environmental Research and Public Health*, *18*(8), Article 4216.

Goodin, T. (2017). *Off: Your digital detox for a better life*. Abrams Image.

Gowrisankaran, S., & Sheedy, J. E. (2015). Computer vision syndrome: A review. *Work*, *52*(2), 303–314.

Gripsrud, J., & Moe, H. (2010). *The digital public sphere: Challenges for media policy*. Nordicom.

Grotto, A. J., & Makridis, C. (2020). *Perception of digital risks: Evidence from 54 countries*. SSRN. https://dx.doi.org/10.2139/ssrn.3711862

Haybron, D. M. (2008). Philosophy and the science of subjective well-being. In M. Eid & R. J. Larsen (Eds.), *The science of subjective well-being* (pp. 17–43). The Guilford Press.

Hjarvard, S. (2013). *The mediatization of culture and society*. Routledge.

Holton, G. A. (2004). Defining risk. *Financial Analysts Journal*, *60*(6), 19–25.

Huppert, F. A. (2014). *The state of wellbeing science: Concepts, measures, interventions, and policies*. Wiley.

Jorge, A. (2019). Social media, interrupted: Users recounting temporary disconnection on Instagram. *Social Media + Society*, *5*(4). https://doi.org/10.1177/2056305119881691

Karppi, T., Chia, A., & Jorge, A. (2021). In the mood for disconnection. *Convergence*, *27*(6), 1599–1614.

König, A., Gollwitzer, M., & Steffgen, G. (2010). Cyberbullying as an act of revenge? *Journal of Psychologists and Counsellors in Schools*, *20*(2), 210–224.

Kowalski, R., Giumetti, G., Schroeder, A., & Lattanner, M. (2014). A meta-analysis of factors predicting cyberbullying perpetration and victimization: From the social cognitive and media effects approach. *Psychological Bulletin*, *140*(4), 1073–1137.

Krüger, S., & Charlotte Spilde, A. (2020). Judging books by their covers—Tinder interface, usage and sociocultural implications. *Information, Communication & Society*, *23*(10), 1395–1410.

Lallie, H. S., Shepherd, L. A., Nurse, J. R., Erola, A., Epiphaniou, G., Maple, C., & Bellekens, X. (2021). Cyber security in the age of Covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Computers & Security*, *105*, Article 102248.

LaRose, R., Connolly, R., Lee, H., Li, K., & Hales, K. D. (2014). Connection overload? A cross cultural study of the consequences of social media connection. *Information Systems Management*, *31*(1), 59–73.

Lazarus, R. S. (1998). The stress and coping paradigm. In R. S. Lazarus (Ed.), *Fifty years of the research and theory of R. S. Lazarus: An analysis of historical and perennial issues* (pp. 182–220). Routledge.

Lee, U., Han, K., Cho, H., Chung, K.-M., Hong, H., Lee, S.-J., Noh, Y., Park, S., & Carroll, J. M. (2019). Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions. *Ad Hoc Networks*, *83*, 8–24.

Livingstone, S., & Helsper, E. J. (2008). Parental mediation of children's internet use. *Journal of Broadcasting & Electronic Media*, *52*(4), 581–599.

Macït, H. B., Macït, G., & Güngör, O. (2018). A research on social media addiction and dopamine driven feedback. *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, *5*(3), 882–897.

Mason, M. (2010). Sample size and saturation in phd studies using qualitative interviews. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *11*(3). https://doi.org/10.17169/fqs-11.3.1428

Masseno, M. D., & Santos, C. T. (2018). Assuring privacy and data protection within the framework of smart tourism destinations. *Media Laws—Rivista di Diritto dei Media*, *2018*(2), 251–266.

Montag, C., & Walla, P. (2016). Carpe diem instead of losing your social mind: Beyond digital addiction and why we all suffer from digital overuse. *Cogent Psychology*, *3*(1), Article 1157281.

Moore, K., & Craciun, G. (2021). Fear of missing out and personality as predictors of social networking sites usage: The Instagram case. *Psychological Reports*, *124*(4), 1761–1787.

Patton, D. U., Leonard, P., Elaesser, C., Eschmann, R. D., Patel, S., & Crosby, S. (2019). What's a threat on social media? How Black and Latino Chicago young men define and navigate threats online. *Youth & Society*, *51*(6), 756–772.

Pawar, S. C., Mente, R., & Chendage, B. D. (2021). Cyber crime, cyber space and effects of cyber crime. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *7*(1), 210–214.

Phan, A., Seigfried-Spellar, K., & Choo, K.-K. R. (2021). Threaten me softly: A review of potential dating app risks. *Computers in Human Behavior Reports*, *3*, Article 100055. https://doi.org/10.1016/j.chbr.2021.100055

Schulz, W. (2004). Reconstructing mediatization as an analytical concept. *European Journal of Communication*, *19*(1), 87–101.

Shareef, M. A., Mukerji, B., Alryalat, M. A. A., Wright, A., & Dwivedi, Y. K. (2018). Advertisements on Facebook: Identifying the persuasive elements in the development of positive attitudes in consumers. *Journal of Retailing and Consumer Services*, *43*, 258–268.

Sheridan, L. P., & Grant, T. (2007). Is cyberstalking different? *Psychology, Crime & Law*, *13*(6), 627–640.

Silverman, D. (2016). *Qualitative research*. SAGE.

Sjåstad, H., Zhang, M., Masvie, A. E., & Baumeister, R. (2021). Social exclusion reduces happiness by creat-

ing expectations of future rejection. *Self and Identity*, *20*(1), 116–125.
Smith, J. A., Larkin, M., & Flowers, P. (2009). *Interpretative phenomenological analysis: Theory, method and research*. SAGE.
Smith, J. A., & Osborn, M. (2004). Interpretative phenomenological analysis. In G. M. Breakwell (Ed.), *Doing social psychology research* (pp. 229–254). The British Psychological Society; Blackwell.
Smith, M., & Puczkó, L. (2008). *Health and wellness tourism*. Routledge.
Smith, P. K. (Ed.). (2009). *Cyberbullying: Abusive relationships in cyberspace*. Hogrefe.
Sumner, L. W. (1996). *Welfare, happiness, and ethics*. Clarendon Press.
Symeonidis, I., Biczók, G., Shirazi, F., Pérez-Solà, C., Schroers, J., & Preneel, B. (2018). Collateral damage of Facebook third-party applications: A comprehensive study. *Computers & Security*, *77*, 179–208.
Syvertsen, T., & Enli, G. (2020). Digital detox: Media resistance and the promise of authenticity. *Convergence*, *26*(5/6), 1269–1283. https://doi.org/10.1177/1354856519847325
Tandoc, E. C., Jr. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society*, *16*(4), 559–575.
Tang, W. Y., Reer, F., & Quandt, T. (2022). The interplay of the dark triad and social media use motives to social media disorder. *Personality and Individual Differences*, *187*, Article 111402. https://doi.org/10.1016/j.paid.2021.111402
Tiberius, V. (2020). Well-being, philosophical theories of. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 7110–7113). Springer.
Tokunaga, R. S., & Aune, K. S. (2017). Cyber-defense: A taxonomy of tactics for managing cyberstalking. *Journal of Interpersonal Violence*, *32*(10), 1451–1475.
Tongco, M. D. (2007). Purposive sampling as a tool for informant selection. *Ethnobotany Research Applications*, *5*, 147–158.
Trittin-Ulbrich, H., Scherer, A. G., Munro, I., & Whelan, G. (2021). Exploring the dark and unexpected sides of digitalization: Toward a critical agenda. *Organization*, *28*(1), 8–25.
Whitty, M. T., & Buchanan, T. (2016). The online dating romance scam: The psychological impact on victims—Both financial and non-financial. *Criminology & Criminal Justice*, *16*(2), 176–194.
Widyanto, L., & Griffiths, M. (2006). "Internet addiction": A critical review. *International Journal of Mental Health and Addiction*, *4*, 31–51.
Yates, J. F., & Stone, E. R. (1992). *The risk construct*. In J. F. Yates (Ed.), *Risk-taking behavior* (pp. 1–25). Wiley.
Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary—The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research*, *21*(4), 724–735.

About the Author

Bindiya Dutt is a student of media and communication in the Department of Social Sciences at the University of Stavanger.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 367–378
https://doi.org/10.17645/mac.v11i2.6486

Article

Post-Publication Gatekeeping Factors and Practices: Data, Platforms, and Regulations in News Work

Margareta Salonen *, Veera Ehrlén, Minna Koivula, and Karoliina Talvitie-Lamberg

Department of Language and Communication Studies, University of Jyväskylä, Finland

* Corresponding author (margareta.salonen@jyu.fi)

Submitted: 14 November 2022 | Accepted: 6 March 2023 | Published: 28 June 2023

Abstract
The gatekeeping literature has turned to look at the factors and practices that shape gatekeeping in the post-publication environment, i.e., after news has entered circulation. This article adds to the discussion and argues that news workers share gatekeeping power in the post-publication environment with audiences, platforms, and regulations. Further, this study extends the post-publication gatekeeping framework and considers it in the context of datafication. The article aims to broadly understand how (audience) data is part of editorial decision-making in news media from news workers' perceptions. The current study was conducted by interviewing news workers from three Finnish news organisations. The interview data was analysed utilising qualitative iterative content analysis. Our analysis revealed that the use of (audience) data in news organisations increasingly shapes news workers' journalistic decision-making processes. We found that news workers were ambivalent toward data (use) and that their reliance on platform data depended on the particular platform. Furthermore, when interviewed about journalism ethics, news workers only connected it with legislative issues, such as General Data Protection Regulation. Lastly, we could see that regulatory factors of data, i.e., legislation and media self-regulation, have power over news production and distribution. This study reflects how journalism (research) is shifting from an audience-centric view to a data-driven one, i.e., it is experiencing a *data turn*.

Keywords
audiences; datafication; gatekeeping; news workers; platforms; regulations

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

The growing use of audience data in newsrooms has impacted the power dynamics between news organisations, platforms, and audiences. The "digital trinity" of datafication, algorithmisation, and platformisation (Latzer, 2021) has led to a situation where journalists hold less gatekeeping power, i.e., a growing number of players have gained influence over news distribution, circulation, and the business itself (Salonen et al., 2022; Seuri & Ikäheimo, 2022). Through their digital footprints, such as interacting with news on social media, audiences are influencing editorial decision-making (Tandoc & Vos, 2016), and, in turn, social media platforms' algorithms tangle with news distribution and visibility, affecting the way news is shown to audiences (van Dijck et al., 2018). Algorithmic platforms and their users have entered the news ecosystem and currently intertwine with news processes, especially in the post-publication context, i.e., after news has entered circulation (Hermida, 2020; Salonen et al., 2022).

In this study, we examine datafied news work from the perspective of post-publication gatekeeping. Our aim is to broadly understand how (audience) data is part of editorial decision-making in news media from news workers' perceptions. More specifically, we contribute

to the literature by discussing what role regulations play in the datafied news environment. From the viewpoint of media accountability, news media consider and reflect the external control (such as laws of the country/region), the internal control (such as journalistic values, ethics, and press councils) as well as audiences and other members of civil society (Eberwein et al., 2019) when deciding what is newsworthy to publish. In the current study, we focus on media regulation from a twofold perspective: Firstly, legislation in the European Union, for example, the General Data Protection Regulation (GDPR), ensures and controls users, media organisations, and platform companies' rights over data (Meier & Trappel, 2022); and secondly, news organisations consider media self-regulation in editorial decision-making processes. This study addresses the suggestion of Porlezza and Eberwein (2022), Seuri and Ikäheimo (2022), and Seuri et al. (2022) that media regulation (the external control) and self-regulation (the internal control) should be investigated in the era of datafication.

Datafication is, in this study, understood as digital media's capacity to turn all human action and interaction into measurable digital traces (Breiter & Hepp, 2018). Essentially, knowledge of user characteristics and behaviours has become the main currency in the current media ecosystem (Ohlsson & Facht, 2017): In the process of datafication, (individual) data is monetised for its business potential (Couldry & Yu, 2018). Power over audience data is exercised by those with access to databases and expertise in processing and data mining (Andrejevic, 2014), such as news media and platform companies. Through digital profiling, platform companies can define which content users are exposed to and predict their (future) behaviour. News media depend on these infrastructural services as platform companies have developed considerable market and gatekeeping power (Meier & Trappel, 2022).

Ownership and control over news-related audience data have become complex, and therefore, it is important to understand how news media utilise and make decisions concerning data. In this research, we highlight the (potential) role of audience data from self and/or third-party-governed platforms in news workers' editorial decision-making. We do this by reflecting news workers' perceptions of their datafied working practices through the lens of post-publication gatekeeping theory, which looks at the factors that shape news after its publication. The current study adds to previous research on post-publication factors, practices, and the environment itself (Hermida, 2020; Salonen et al., 2022). Further, this study connects regulatory factors (legislation and media self-regulation) theoretically to the post-publication gatekeeping framework (Hermida, 2020) and provides evidence by applying the suggested extended framework empirically. Finally, this study takes a stance on the well-needed discussion of regulatory factors in the era of datafication.

2. Theoretical Framework of (Post-Publication) Gatekeeping in the Context of Datafied News Environment

In the 2020s, the context where news circulates is increasingly digital and datafied. Datafication of news has meant, for example, that newsrooms' editorial decisions are increasingly data-oriented (Vu, 2014). Datafication has also impacted economic models in journalism and challenged journalistic autonomy (Hanusch, 2017). This has led to the datafication of the digital news environment, and hence, the gatekeeping processes have also changed. Gatekeeping has been defined as "the process of culling and crafting countless bits of information into the limited number of messages that reach people each day" (Shoemaker & Vos, 2009, p. 1). The definition still holds up even though the theory has transformed considerably since its birth in the 1940s (Vos, 2019). This section first introduces gatekeeping theory in the digital age, then reviews studies focused on post-publication gatekeeping, and finally, discusses the role of regulations as a post-publication gatekeeping factor in datafied news work.

The digital news environment has brought changes to the ways gatekeeping theory has been utilised and developed for research in the past. In 1989, Abbott and Brassfield (1989) compared print and electronic media, and in 1990, Berkowitz (1990) studied the gatekeeping processes of local television news. More recently, the theory has been applied to online contexts. These studies have looked at the rise of news events on social media (Meraz & Papacharissi, 2013), user-generated visibility on media websites (Singer, 2014), visual gatekeeping practices (Pantti, 2015), and social media editors' impact on news diffusion (Welbers & Opgenhaffen, 2018), for example. Gatekeeping has also been researched from the viewpoint of data and metrics. Tandoc (2014) pointed out how web analytics is changing the gatekeeping process after decades of journalists paying little attention to audience opinion in their decision-making. Consequently, information gained from (audience) data is part of the journalistic gatekeeping process. Nowadays, the question is not about whether data affects newsroom decision-making but, rather, how much. Further, the theory field has started to turn towards the post-publication viewpoint, i.e., the context where news is constantly circulated in the datafied digital news environment.

Some previous studies have looked at the post-publication side of gatekeeping, even though they do not specifically talk about post-publication gatekeeping. Singer (2014) introduced the term secondary gatekeeping, which refers to how users can up or downgrade the visibility of an online news item, while Bruns (2018) talks about how users and journalists can act as gatewatchers to/of information that is relevant to be distributed further online. Wallace (2018), in turn, points out different types of digital gatekeepers in contemporary society: journalists, individual amateurs, strategic professionals,

and algorithms. These studies can be categorised as post-publication gatekeeping studies because they look at the gatekeeping factors (e.g., audiences/users and platforms) and practices (e.g., users' ability to up or downgrade or gatewatch) in the environment after news has been published.

Post-publication gatekeeping specifically has been previously researched by Hermida (2020) and Salonen et al. (2022). Salonen et al. (2022) introduced the concept of conversational gatekeeping, highlighting the social interactional nature of gatekeeping in the post-publication context. Further, they see gatekeeping materialising as and in social interaction: Journalists and audiences negotiate and create the conversational norms together and decide on the accepted content for the particular online platform. Further, Hermida (2020) introduced the framework of four factors (4Ps)—publics, platforms, paraphernalia, and practices—through which post-publication gatekeeping can be viewed and which shape the processes of gatekeeping. Publics refers to the news audiences, reaching from the members of the public to politicians, businesspeople, and journalists themselves. Platforms concern the platform giants such as Google, Meta, and Twitter. Paraphernalia refers to the materiality of gatekeeping, such as mobile devices, smart speakers, and software. Practices refers to social (spatial and temporal) practices around how users engage with the news; for example, whether the news is consumed on a bus or while lying in bed.

In their study, Hermida (2020, p. 16) calls for studies that consider digital metrics' influence on editorial decision-making by considering "how all or some of the four Ps impact flows of news and information post-publication." Specifically, regulation issues related to gatekeeping are under-researched even though regulation increasingly shapes the formation of news. With the term regulations, we refer to the legislative factors, i.e., laws of the country and/or region, and to the self-regulatory factors, namely, journalistic values and ethics. Academics have called for collective regulation of data-driven systems (Steedman et al., 2020). In the EU, legislation (the GDPR and several digital acts, for example) ensures and controls users, media companies, and platform companies' rights over data (Meier & Trappel, 2022). For example, some European data protection authorities have found the use of US-based Google Analytics unlawful due to the increased risk of being in breach of the GDPR that regulates the use of personal data (Roosa et al., 2022). These new digital laws will challenge platform companies' data monopolies, and as a result, companies such as Meta have threatened to withdraw from the EU market (Burgess, 2022). The new regulations will also affect data collected by third parties and used by the media. As a result, audiences may gain more control over their data, and the monetisation model of platforms and media companies could be undermined.

Further, Seuri et al. (2022) discuss the new gatekeeping regime and highlight the role and need to regulate platform giants and big media companies. In their future scenarios for the platform society, Seuri et al. (2022) see that regulation is needed to gain positive outcomes for the information environment and to counterbalance the network effects that create platform monopolies. Furthermore, Aral (2020) has called for structural reform: Data flows between platforms should be enforced to ensure sustainable interoperability between platforms. This means that user data and actions should no longer be monopolised by platform companies. The Digital Services Act is designed to combat this in the EU, which could mean that in the future, news companies will have better access to databases collected by platform companies.

From the viewpoint of media accountability and journalism ethics, news media are currently trying to balance the journalistic ethos of reporting what the audience needs to know and the data-driven view of what the audience wants to know (Hanusch, 2017). This notion raises ethical questions as the journalistic field also has its norms and ethics on which editorial decisions are based. In Finland, the context of our empirical study, the Finnish press council governs journalistic (self-regulation) guidelines and processes complaints related to them. In relation to media accountability, the press council is an important part of the journalistic institution and has been designed to oversee the media's responsibility to society, citizens, and the journalistic institution itself (Eberwein et al., 2019). However, in digital journalism driven by datafication and algorithms, there is still a lack of normative standards and regulations across Europe (Porlezza & Eberwein, 2022). Journalism scholars (e.g., Porlezza & Eberwein, 2022; Rydenfelt et al., 2022) have argued that journalistic self-regulation needs to be adapted to the era of datafication, news automation, and personalisation.

From these theoretical premises, we answer Hermida's call to apply the framework of the 4Ps (publics, platforms, paraphernalia, and practices) and tackle the question of digital metrics' connectedness to editorial decision-making. Thus, we propose the framework of "Post-publication gatekeeping factors in datafied news work" (Figure 1) that has been built on the premises of the 4Ps and the previous theoretical discussion of regulations and datafication of journalism. In the extended framework, we add the factor of regulations, i.e., legislation and self-regulation, to extend the 4Ps framework and see that all this is taking place in the context of a datafied news environment.

In this study, we use the term audiences instead of publics as it is more descriptive in depicting the people for whom the news is created. Further, we see that the factor of platforms includes all kinds of platforms used by the news media, self-governed and third-party governed. Furthermore, we see that the factor of paraphernalia—the materiality of the object—is omnipresent in the factor of platforms, as platforms are technological constellations. Therefore, this study considers paraphernalia (e.g., software) as embedded in the platform factor.

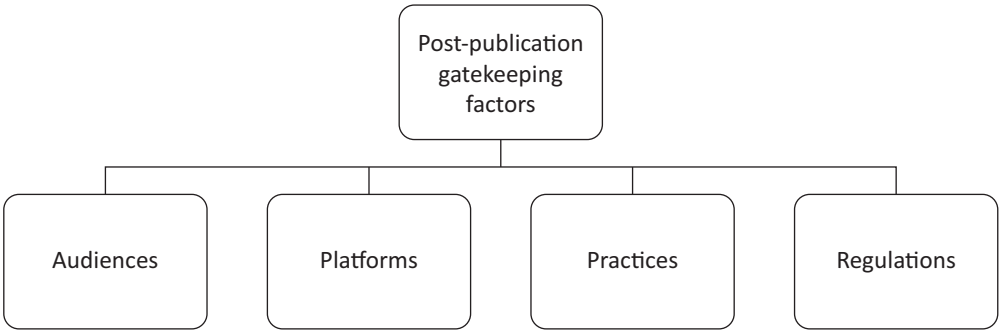


Figure 1. Post-publication gatekeeping factors in datafied news work.

By practices, we refer to the concept beyond audiences' social practices and include all kinds of journalistic practices that news workers can perform during their daily working routines. Thus, in this study, we have adapted from Hermida's (2020) study, the factors of audiences, platforms and practices, and added the factor of regulations.

Against this background, this study asks:

RQ: How do news workers perceive the datafied factors of audiences, platforms, and regulations, and how do these factors shape their working practices from the viewpoint of post-publication gatekeeping?

3. Methods and Materials

The material for the study was gathered by interviewing Finnish news workers (*N* = 9). Finland makes for a particularly interesting research context concerning audience data use because news reach and trust in news media are generally high (Newman et al., 2022). In global comparison, Finns trust news sites to use their personal data fairly more than people in other Western countries (Newman et al., 2022). This context may also be reflected in news workers' attitudes toward data use. We adopted a qualitative approach to comprehend news workers' perceptions and understandings of data-related issues and practices and interviewed nine news workers from three Finnish newspaper organisations. The first organisation is a large national daily newspaper, and the other two are sizable regional newspapers, both publishing the largest newspaper in their geographical area in terms of circulation.

Our study participants are personnel who deal with data-related questions in their everyday work in news organisations. Their views on data use in news organisations vary somewhat based on their titles, as some worked in content production roles (e.g., social media manager) while others had more managerial tasks (e.g., head of technological development). Importantly, however, all were experts in how their organisation utilises audience data. Hence, we refer to them collectively as news workers. Seven participants identified as male, one as female, and one as non-binary. Participants' job titles and years of experience can be found in Table 1.

The interviewees were selected for the study by peer recommendations within the organisations. Prior to the interviews, the participants were informed about the research project and the interview themes. They were also asked to review and approve a consent form guaranteeing the voluntary and anonymous nature of participation and the confidentiality of the interview material. Semi-structured interviews were conducted by the first author via Zoom and face-to-face between May and August of 2022. On average, interviews lasted 96 minutes (a total of 866 minutes). The interview themes were: (a) Collection and use of data from news media sites, (b) collection and use of data from social media

Table 1. Interviewees' job titles and work experience in years.

Interviewee	Work title	Years of experience
NW1	Producer	15
NW2	Manager	20
NW3	Lead Developer	2,5
NW4	Social Media Producer	6
NW5	Head of Business Development	8
NW6	Web Manager	12
NW7	Web Analyst	20
NW8	Head of News	13
NW9	Head of Technological Development	10

platforms, (c) ethics and responsibility in data collection and use, (d) news media's approach to social media platforms, (e) news media's approach to legislation dealing with news content and data practices, and (f) moderation of news content and discussions.

The interview material was analysed using qualitative iterative content analysis (Tracy, 2018). This method combines and alternates between deductive and inductive research approaches, i.e., it iteratively combines data and previous theory. This method was chosen because it "focuses on more narrow aspects of the data that have potential to extend specific theories or address practical problems" (Tracy, 2018, p. 63). In this study, the post-publication gatekeeping theory was used to gain an understanding of the current data-related practices and perceptions of news workers. At the same time, the qualitative data provided empirical evidence to support the proposed extension of Hermida's (2020) theory. Further, the iterative approach brought forward news workers' practices related to data use.

The interview material was analysed and coded by the first author. In the first round of analysis, the material was read and viewed in light of *what* was present. This reading highlighted the salience of the post-publication gatekeeping environment. Thus, throughout the following rounds of coding, the interviews were reflected in the light of the post-publication gatekeeping theory to ensure that all the factors based on previous literature (audiences, platforms, and regulations) were taken into account in the analysis. In addition, to ensure that all material discussing data-related practices was thoroughly reviewed, the parts mentioning data generally were coded as a fifth category. By doing so, the authors gained a comprehensive understanding of the interview material and advanced their conception of how post-publication gatekeeping factors are connected to data use in news media. In the final phase of the analysis, the authors jointly evaluated the fit between the formulated categories and coded content through a discussion on data excerpts relating to post-publication gatekeeping factors and different kinds of data-related practices evident in the interview data.

4. Findings: Post-Publication Gatekeeping Factors and Practices

In this section, we examine our data through the lenses of post-publication gatekeeping factors and practices in datafied news work, as Hermida (2020) suggested. We apply the extended framework previously suggested in the theory section (see Figure 1) to our empirical data. In the following, we analyse and provide empirical evidence of how news workers perceive the datafied factors of audiences, platforms, and regulations and how these factors may shape their working practices. First, we explain the significant and ambivalent role of audience data in journalistic decision-making processes; second, we open up the news workers' perspectives about self- and third-party governed platforms they use in their organisations; and third, we highlight the pivotal role that regulations (legislation and media self-regulation) play in news workers' decision-making processes.

4.1. Audience Factor and its Related Practices

News media workers' emphasis on audience data varied among organisations and individuals. This highlights the ambivalent attitudes toward data practices in news media. For some, audience data significantly impacted their decision-making processes, which the following excerpt illustrates:

> We base decisions less and less on emotions and feelings….Whether it's about planning journalism—what kinds of stories we want to make—or business decisions or designing our website—what kinds of functions we want to place there. In both cases, data is used diligently. (NW5)

However, not all interviewees put as much emphasis on audience data. They felt they had a responsibility to society to tell the most important news, no matter what metrics indicated, i.e., journalistic values guided their decision-making regarding the news. Audience data could guide the form of news stories, but newsworthiness and the story's content were news workers' decisions. In contradiction, some understood audience data as a means to emphasise particular kinds of content. For example, one interviewee said that they closely follow their audiences' actions by age group: "We follow what kind of content interests different aged audiences so that we can produce better content and are able to offer our subscribers the content they wish for" (NW9).

Interviewees' perceptions of audience data were indeed ambivalent: They were uncertain about how much data should and does affect their decisions and actions. Thoughts and perceptions about the influence of audience data also varied throughout the interviews, as news workers seemed to be weighing what they could say about data. Illustratively, one interviewee raised critical views of their sites' audience data use and presumed that it was unreliable regarding age, gender, and reading time statistics: "Reading time analytics is so unreliable that we can't base any decisions on it" (NW8). Later the same interviewee described that they could still partly rely on their (audience) data in cases such as planning the front page. They further described that analytics was understood as helping them to make decisions regarding their publication format or content:

> We've got new kinds of formats such as live broadcasting and radio shows, and as the number of staff stays the same, we need to let go of something. This is how long-term analytics can help us decide which [formats/content] are not so important for us to do. (NW8)

Data-driven news culture was also visible in organisations' ways of testing news headlines and how different kinds of headlines allured their audiences. Some had even predicted, for example, how particular headlines performed for different aged audiences.

The origins of the data also mattered. Most importantly, the data of social media audiences raised concerns. Some saw a difference between the use of social media data and data that was collected from the organisation's sites—audience data guides more social media activities than activities on their own sites. They described this as a way to keep journalistic decision-making in their hands. In addition, audience data was utilised to compare different social media platforms' performance. As one participant observed, "It's interesting to compare how the same video performs and interests on TikTok and Instagram and what kinds of audiences it allures" (NW4).

Also, audience data is considered when practices relating to paywalls are decided upon. News media aims to increasingly understand their audiences' movements on platforms and ways to monetise them. In our study, interviewees from all three organisations said they employ hard paywalls. While the previous literature considers a "hard paywall" to be "no access to content without subscription" (Myllylahti, 2014) or that they "allow no free content" (Pickard & Williams, 2014), our interviewees seemed to connect hard paywalls to granting access to content that is only available to subscribers. Thus, they distinguished between some content that is freely accessible to all and clear-cut content that is unreachable unless readers subscribe. With that in mind, in this study, we use the concept of a "hard paywall," as our interviewees presented it. Further, in our study, the number of news stories placed behind the paywall varied between organisations. The first newspaper distributes a large amount of free content as they see access to accurate information as their journalistic responsibility in building a democratic society. The two other newspapers had a different approach to paywall practices, and most of their content was behind a paywall. They further described how the content is usually distributed for free in situations when the news originates from the Finnish News Agency (STT) or deals with global, national, or local security:

> All the stories that we produce are primarily behind the hard paywall no matter where one enters the story, from social media or elsewhere….For free of charge, we offer things such as our columns, STT news, and for example, national instructions during Covid-19. (NW8)

The above-mentioned factors become apparent through newsroom practices—such as the paywall example illustrates—and manifest the central role of audience data in news workers' decision-making processes. The interviewees' statements reflect the contradictory views of audience data's role in news work. The ambivalence and uncertainty toward audience data were also visible inside single organisations: Interviewees from the same news organisation shared strikingly different views on the impact of audience data in their decision-making processes. In addition, the statements portray how audience data is a post-publication factor that shapes new workers' decision-making which makes it an issue of (post-publication) gatekeeping—what kind of content news workers publish or should publish, and how they frame their journalistic content.

4.2. Platform Factor and Its Related Practices

News organisations use a range of platforms to reach their audiences and distribute their content. Each platform has its unique mechanisms, and many are guided by algorithmic recommendations. When platforms were looked at through the lens of (post-publication) gatekeeping, the question of governance became evident—how much decision-making power the platforms afford to news workers. The interviewees described how organisations employed three different kinds of platform systems: self-governed, third-party governed, or a mix of these. On a self-governed platform (e.g., company websites and applications), the news organisations are fully in charge of the published content (what and when) and in control of dataflows (collection, storage, and handling; e.g., self-made analytics tool). From the viewpoint of paraphernalia, the materiality of technological objects is embedded in the platforms' software. That is, power over platforms is tied to the software being used: If a news organisation owns the software, it has more power to manage the platform it operates on. An effective way to gain control over data is to develop one's own analytics tools for the news organisation, as one interviewee states:

> It's a platform we've developed for GDPR reasons. Because if we used another [third-party] platform, the data would go just somewhere….We've protected [the data] so that people can reply anonymously, and those answers are not connected to any other information within the organisation. (NW3)

Self-governed platforms were understood as means to control data, but third-party governed platforms were seen as governed by the technology giants. This means that news media have little power over the affordances that guide the use of the latter ones. The most often mentioned platforms were Google and Meta and their products: Google Analytics for following newsroom metrics; Facebook, Instagram, Twitter, and Tiktok for social media activities. Interviewees from each organisation described how much traffic in news organisations' sites originates from social media platforms. They were unified in that the main function of social media is to promote journalistic content, i.e., making their brand known.

However, understandings of platform governance varied. Some interviewees stated that they are not particularly utilising social media (data) since the organisation prioritises its own products, whereas some organisations put much effort into social media publishing. News workers also pay attention to the vernaculars of the platform. As one interviewee states: "We bring it [the news story] in a format that is easy to understand, e.g., to Instagram. So that it also supports visual storytelling. And that's also shareable" (NW4). This highlights how the platforms' affordances guide journalistic decisions. Interestingly though, some interviewees also highlighted social media as a tool for building a democratic society: "We fulfil our democratic duty by reaching young readers and by sharing trustworthy information with them about the regional elections. Social media has, in this way, a pure journalistic and noble purpose to serve democracy'' (NW1).

As governance of platforms proved important, it also raised concerns about data reliability. This was particularly evident in the context of social media platforms. In some situations, social media companies' data (e.g., Meta's demographics) were deemed more trustworthy than the data of news organisations' platforms, as this type of data was harder to collect from their own sites in a trustworthy manner. As one interviewee explains: "I trust age distribution [data] more on social media because I know that knowing their users is their [platforms'] main business" (NW8). Further, the reliance between social media platforms varied as one interviewee explains: "Google Analytics offers its own view about the visitors, but when you compare that to social media [Meta's user data], it's good to remember that the [Google] data only gives some ideas" (NW7). Data reliability was also deemed important when news organisations use platforms that are a mix of self—and third-party governed platforms. For example, news organisations have built their tools on top of Google's infrastructure. As one interviewee explains:

> Within our conglomerate, we've previously used a company which has built us analytics [tools]….They have utilised the data that comes through Google Analytics….But now we use an external company that tries to identify problems related to data reliability….They have found errors and problems in data collection and are now helping us fix them. (NW8)

The role of the platform factor becomes even more evident in news organisations' moderation practices. News organisations seem to consider what the platforms afford them to do, i.e., do news workers need to moderate comments or does the platform or another service provider do it for them? The level of content moderation differed between newspapers. In the first news organisation, moderation for their website and Meta's social media platforms was bought as an outside service and was conducted by humans and machine learning software. Occasionally news workers moderate by themselves, e.g., on TikTok. News workers from the second news organisation said they had outsourced their website moderation, but they moderate their social media comments as part of their daily practices. News workers from the third news organisation explained that they do not currently use website commenting and that they rely solely on social media commenting in audience interaction. Social media moderation was seen as part of their daily practices: "Journalists who work online are in charge of social media moderation. You go there systematically to check the comments, but because you do that besides all other tasks, you can't fully concentrate on it'' (NW8).

The above-mentioned practices demonstrate that news workers base their decisions on platform data and vernaculars to some degree. Platforms can thus be seen as a post-publication gatekeeping factor that is a part of their decision-making process when deciding what to publish. Further, from the interviewees, it became evident that there is a hierarchy of trust towards different platforms. Organisations (and individuals inside them) value them differently based on the trustworthiness of the particular platform and its data practices.

4.3. Regulatory Factor and Its Related Practices

As our theory-based extension to Hermida's (2020) post-publication gatekeeping framework suggests, regulatory factors are part of the decision-making process of new workers, i.e., post-publication gatekeeping factors. The interviewees described how GDPR has extensively shaped their work practices. However, they gave considerably less attention to the aspects of journalism ethics and media self-regulation, which can also be regarded as part of post-publication gatekeeping. In the interviews, we asked direct questions, such as "how ethical and responsible do you think you are in your work." In these situations, none of the interviewees mentioned the journalistic guidelines, the ethical codebook for mass media set by the Finnish press council, but rather discussed the effects of GDPR on their work. This is notable because media self-regulation often takes place *after* the news is published, for example, when a story needs revising based on feedback from sources. Hence, the press council also has a pivotal role in the media self-regulation process as it oversees how news media follow the journalistic guidelines *after* a news item has been published. The interviewees' focus on GDPR, on the other hand, suggests that the law is an effective regulatory tool and part of news workers' everyday work. This is illustrated by the following quote from an organisation that received feedback concerning the use of Google Forms from their audiences: "Yes, I feel we're [ethical]. Earlier, we might have collected something through Google Forms but then gave up on it as it raises questions about data protection….[The] GDPR becoming effective was a turning point here" (NW2).

The quote also demonstrates that some audience members were concerned about regulatory issues. Furthermore, interviewees described how following GDPR has affected their working routines and increased audiences' rights over their data. This can be interpreted as a means for audiences to become part of the gatekeeping process. GDPR allows them to withhold the release of their personal data even after it has been handed over to news organisations and used, for example, in audience metrics. However, not all audience members seem to be equally aware of their rights, as one participant explains:

> Users can ask us to empty all their data from our records….These requests might come once or twice a week. They are usually people who work in the IT field, experts that are interested in data. An average user isn't usually so interested in their data as long as it's in somewhat good hands. (NW5)

While interviewees brought up the theme of GDPR and how it has shaped news media work, they also described that GDPR and related legal issues were not their areas of responsibility. They repeated that GDPR was something that they did not need to be responsible for because their organisation has specific professionals for GDPR and legal matters. As one participant explains: "I'm not sure how to answer except that GDPR has brought along strictness….These sorts of issues are not our team's responsibility. I don't work with these issues" (NW2). This is noteworthy given that our interviewees were the ones that are dealing with data-related issues in their daily work.

Regulation's shaping of practices can further be exemplified through cookies that also shape the consumption of news after its publication. Due to GDPR, audience members can refuse to share their cookie data with news organisations and third parties who have made agreements with the organisations. Some interviewees believed that data will be regulated increasingly in the future and that regulation will bring changes to data use, both to social media companies and to news organisations themselves. They were worried about the demise of third-party cookies. This would make it difficult to access audience data and would thus challenge media organisations' business logic: "Google will close its support to third-party cookies, which means operating them will become more difficult….Also, in Finland, they [data protection officers] regulate the way we and other publishers operate now and in the future" (NW7).

At the same time, interviewees were after stricter regulation for the major platforms such as Google and Meta. They called for compensation for news content they had created, which now circulates for free on social media platforms in the post-publication context: "Many working in the industry hope that Google and Facebook would share their revenue with us, just as has happened in Australia and France in the past" (NW7). Interviewees hoped that future regulation would help their businesses, for example, news media could have access to cross-platform data to see how their audiences migrate across their own and third-party platforms. At the moment, due to regulation and social media platforms' unwillingness to share their user data, news media organisations base their decisions on a single social media platform or their own platforms' data flows. This further demonstrates the gatekeeping power that these social media giants and regulations possess over news organisations.

5. Discussion and Conclusion

In the previous section, we have demonstrated how the post-publication factors of audiences, platforms, and regulations, as well as their intertwined practices, iteratively shape news workers' editorial decision-making processes in datafied news work. We have highlighted the role of regulatory factors in the current (post-publication) gatekeeping processes and extended the framework of "Post-publication gatekeeping" by Hermida (2020) to include regulatory factors and practices in datafied news work. Our empirical findings support previous studies (e.g., Hermida, 2020; Salonen et al., 2022; Seuri & Ikäheimo, 2022) claiming that gatekeeping power is shared between multiple (f)actors in the current datafied news environment. To illustrate how the factors and practices are situated in the traditional journalistic gatekeeping process, we present Figure 2.

The figure illustrates how post-publication gatekeeping factors and practices iteratively shape the traditional journalistic gatekeeping process, which includes all the culling and crafting of information that takes place in newsrooms before a news item is published. After the news item has been published, it circulates in the datafied news environment. From there on, the news item interacts with audiences, platforms, and regulatory gatekeeping factors. These create new practices and shape traditional journalistic gatekeeping in news organisations. For example, the audience factor is present in the ways audience data guide decision-making over news headlines. The platform factor is present, for example, in cases where social media metrics guide what kind of content is published on a particular platform. Further, regulatory factors outline how audience data can be utilised. For example, audiences have the right to withdraw their data from news organisations' databases. These kinds of withdrawals shape audience metrics and can, therefore, also shape decision-making processes. For future studies, we recommend that researchers apply the suggested framework empirically to validate it further and more extensively map the factors and practices that shape post-publication gatekeeping processes in the datafied news environment.

The current study further contributes theoretically to the field of journalistic gatekeeping studies—it brings together the previous discussions of gatekeeping factors (e.g., Salonen et al., 2022; Wallace, 2018) and practices (e.g., Bruns, 2018; Singer, 2014) that shape the

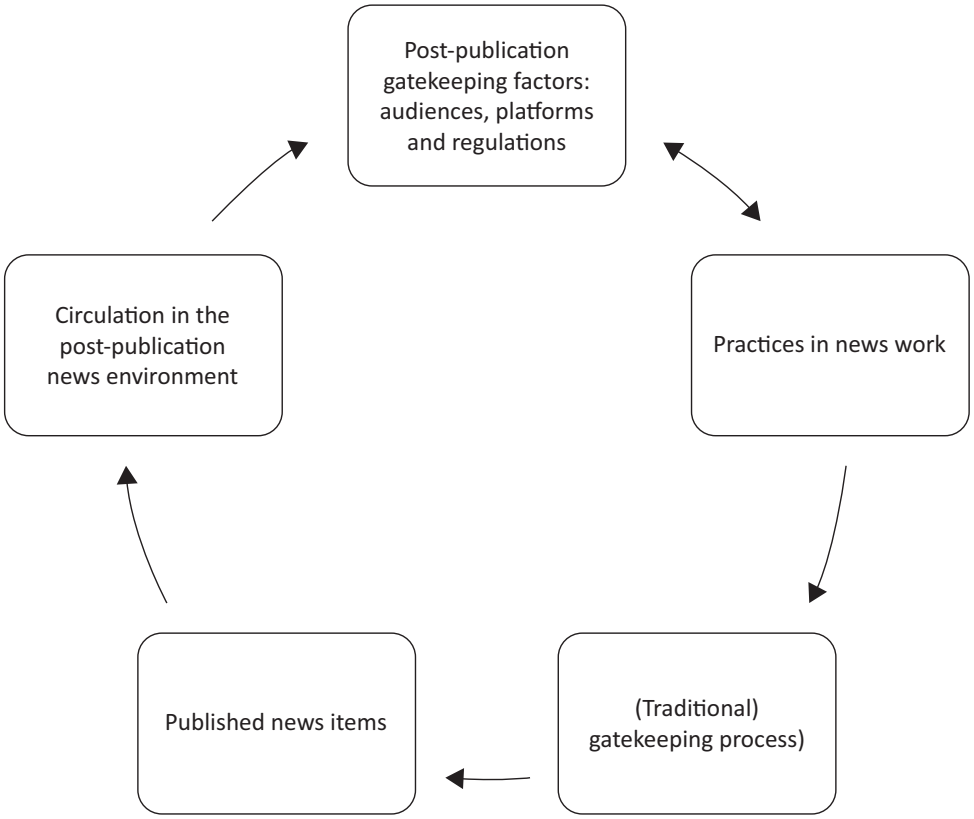


Figure 2. (Post-publication) gatekeeping processes in datafied news work.

production and distribution of news. Furthermore, the study answers previous research's (Aral, 2020; Porlezza & Eberwein, 2022; Seuri et al., 2022) call to examine regulations in the datafied New Gatekeeping Regime. We have done this by not only suggesting the extension of regulatory factors to Hermida's (2020) framework but also by unravelling the news workers' contradicting perceptions concerning legal and ethical matters of their work. Our findings somewhat align with previous research (e.g., Ekström et al., 2022; Rydenfelt et al., 2022) on how news workers work with data; its use is negotiated with other, often journalistically-driven values. However, our study participants highlighted aspects of the law (i.e., the GDPR) more than media self-regulation, even when asked about the ethics of news work. It is noteworthy that while the Finnish press council is highly valued among news workers in Finland, it was not discussed in the interviews. This could be due to uncertainty arising from datafied news work. Several of our participants stated that they were unsure how data is used and that other non-editorial personnel might be better equipped to discuss legal aspects of their work. Data and knowledge were understood as being "out there."

Further, part of this confusion between law and ethics could be explained by the idea that the law (GDPR) limits the misuse of data collection and use and, therefore, forces more ethical approaches. Ethics is written into the GDPR, at least at the minimum. As the EU's new set of Digital Acts comes into force, the European media market will face changes in business practices and models (Newman et al., 2022). These changes may also impact the future of media self-regulation. It is, therefore, crucial for news organisations to prepare for these changes and consider ethical aspects when planning their future. The interconnected nature of legislation and media self-regulation should also be the focus of further research.

Furthermore, our study confirmed that web analytics and the knowledge generated through audience data shape the journalistic gatekeeping process (cf. Tandoc, 2014), evident in practices such as personalising headlines and decisions relating to the front page. This supports the idea that journalism is shifting from an audience-centric view to a data-driven one, i.e., journalism (research) is experiencing a *data turn*. This is also connected to reliance on data and platforms. In our empirical data, the significant role of data processing technologies and, most notably, the various third-party actors who provide these technologies sheds light on what the digital trinity of algorithmisation, datafication, and platformisation (Latzer, 2021) looks like in everyday news work and how it informs the understanding of not only the audiences but also that of news workers concerning newsworthiness. This was visible in our findings, for example, as reliance on social media data.

However, reliance on data raises some critical concerns. First, our focus on news media organisations and their everyday working practices shows that the

understanding of the audience generated through data does not always reflect reality. This mirrors recent critical data and media research arguments on how knowledge created through data is, in fact, often ambiguous (Pink et al., 2018). Consequently, datafication may produce a fundamental misfit between data-generated understanding of an individual and their own experience (Talvitie-Lamberg et al., 2022)—to which the data double concept also refers (see e.g., Ruckenstein, 2014). Second, and as our study demonstrated, there is an ambivalence in how the news workers spoke of their data use, even inside a single organisation. This illustrates the controversial role of audience data in news work. News workers aim to base their decisions on platform metrics, but the knowledge of audiences is highly dependent on the (single) platform-generated data and the particular data processing tools (Aral, 2020). This means that journalistic decision-making becomes increasingly dependent on the platforms and their data processing practices. Further, this could lead to an undesirable situation where third parties are given too much power over journalistic decision-making (Salonen et al., 2022).

The knowledge of how data is used and exploited across different units in news organisations is fragmented. With further overall discussion on data-related practices, news organisations could form a less ambivalent relationship with the data they possess. As our results indicate, a move in this direction seems salient, given that audience data increasingly guides decision-making and newsworthiness in the news media. Therefore, involving news workers from different units and positions in discussions on data use and its regulatory aspects would help create a more holistic understanding of data and give individual news workers confidence in managing their responsibilities and expertise. We have begun to unravel this topic but also recognise the limitation of the size and nature of our dataset. Therefore, we invite future scholars to dig deeper into the ambivalence of data use in news workers' daily practices.

Acknowledgments

The authors would like to thank the Media Industry Foundation of Finland, which funded the research, and the editors and reviewers for their time and comments that helped to improve the manuscript.

Conflict of Interests

The authors declare no conflict of interest.

References

Abbott, E. A., & Brassfield, L. T. (1989). Comparing decisions on releases by TV and newspaper gatekeepers. *Journalism Quarterly*, *66*(4), 853–856. https://doi.org/10.1177/107769908906600411

Andrejevic, M. (2014). Big data, big questions: The big data divide. *International Journal of Communication*, *8*, 1673–1689. https://ijoc.org/index.php/ijoc/article/view/2161/1163

Aral, S. (2020). *The hype machine: How social media disrupts our elections, our economy and our health—And how we must adapt*. HarperCollins.

Berkowitz, D. (1990). Refining the gatekeeping metaphor for local television news. *Journal of Broadcasting & Electronic Media*, *34*(1), 55–68. https://doi.org/10.1080/08838159009386725

Breiter, A., & Hepp, A. (2018). The complexity of datafication: Putting digital traces in context. In A. Hepp, A. Breiter, & U. Hasebrink (Eds.), *Communicative figurations: Transforming communications in times of deep mediatization* (pp. 387–405). Palgrave Macmillan.

Bruns, A. (2018). *Gatewatching and news curation: Journalism, social media, and the public sphere*. Peter Lang.

Burgess, M. (2022, August 9). Will Europe force a Facebook blackout? *Wired*. https://www.wired.co.uk/article/facebook-eu-us-data-transfers

Couldry, N., & Yu, J. (2018). Deconstructing datafication's brave new world. *New Media & Society*, *20*(12), 4473–4491. https://doi.org/10.1177/1461444818775968

Eberwein, T., Fengler, S., & Karmasin, M. (2019). Theory and practice of media accountability in Europe: An introductory overview. In T. Eberwein, S. Fengler, & M. Karmasin (Eds.), *Media accountability in the era of post-truth politics: European challenges and perspectives* (pp. 3–18). Routledge.

Ekström, M., Ramsälv, A., & Westlund, O. (2022). Data-driven news work culture: Reconciling tensions in epistemic values and practices of news journalism. *Journalism*, *23*(4), 755–772. https://doi.org/10.1177/14648849211052419

Hanusch, F. (2017). Web analytics and the functional differentiation of journalism cultures: Individual, organizational and platform-specific influences on newswork. *Information, Communication & Society*, *20*(10), 1571–1586. https://doi.org/10.1080/1369118X.2016.1241294

Hermida, A. (2020). Post-publication gatekeeping: The interplay of publics, platforms, paraphernalia, and practices in the circulation of news. *Journalism & Mass Communication Quarterly*, *97*(2), 469–491. https://doi.org/10.1177/1077699020911882

Latzer, M. (2021). *Digital trinity—Controllable evolution—Everyday religion: Characteristics of the socio-technical transformation of digitalization*. Department of Communication and Media Research. https://www.zora.uzh.ch/id/eprint/204675/1/Latzer_Digital_trinity_controllable_evolution_everyday_religion.pdf

Meier, W. A., & Trappel, J. (2022). Media transparency: Comparing how leading news media balance the

need for transparency with professional ethics. In J. Trappel & T. Tomaz (Eds.), *Success and failure in news media performance: Comparative analysis in the media for democracy monitor 2021* (pp. 255–273). Nordicom.

Meraz, S., & Papacharissi, Z. (2013). Networked gatekeeping and networked framing on #Egypt. *The International Journal of Press/Politics*, *18*(2), 138–166. https://doi.org/10.1177/1940161212474472

Myllylahti, M. (2014). Newspaper paywalls—The hype and the reality: A study of how paid news content impacts on media corporation revenues. *Digital Journalism*, *2*(2), 179–194. https://doi.org/10.1080/21670811.2013.813214

Newman, N., Fletcher, R., Robertson, C., Eddy, K., & Nielsen, R. (2022). *Reuters Institute digital news report 2022*. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022

Ohlsson, J., & Facht, U. (2017). *Ad wars: Digital challenges for ad-financed news media in the Nordic countries*. Nordicom. http://hdl.handle.net/2077/59562

Pantti, M. (2015). Visual gatekeeping in the era of networked images: A cross-cultural comparison of the Syrian conflict. In T. P. Vos & F. Heinderyckx (Eds.), *Gatekeeping in transition* (pp. 203–223). Routledge.

Pickard, V., & Williams, A. T. (2014). Salvation or folly? The promises and perils of digital paywalls. *Digital Journalism*, *2*(2), 195–213. https://doi.org/10.1080/21670811.2013.865967

Pink, S., Lanzeni, D., & Horst, H. (2018). Data anxieties: Finding trust in everyday digital mess. *Big Data & Society*, *5*(1), 1–14. https://doi.org/10.1177/2053951718756685

Porlezza, C., & Eberwein, T. (2022). Uncharted territory: Datafication as a challenge for journalism ethics. In M. Karmasin, S. Diehl, & I. Koinig (Eds.), *Media and change management* (pp. 343–361). Springer.

Roosa, S., Ritzer, C., Martin, N., Jansen, J., Rosenzweig, D., Schuitema, N., Filkina, N., & Ghebali, B. (2022). *European rulings on the use of Google Analytics and how it may affect your business*. Norton Rose Fulbright. https://www.dataprotectionreport.com/2022/02/european-rulings-on-the-use-of-google-analytics-and-how-it-may-affect-your-business

Ruckenstein, M. (2014). Visualized and interacted life: Personal analytics and engagements with data doubles. *Societies*, *4*(1), 68–84. https://doi.org/10.3390/soc4010068

Rydenfelt, H., Haapanen, L., Haapoja, J., & Lehtiniemi, T. (2022). Personalisation in journalism: Ethical insights and blindspots in Finnish legacy media. *Journalism*. Advance online publication. https://doi.org/10.1177/14648849221138424

Salonen, M., Olbertz-Siitonen, M., Uskali, T., & Laaksonen, S. M. (2022). Conversational gatekeeping—Social interactional practices of post-publication gatekeeping on newspapers' Facebook pages. *Journalism Practice*. Advance online publication. https://doi.org/10.1080/17512786.2022.2034520

Seuri, O., & Ikäheimo, H. P. (2022). *Gatekeeping in the digital age*. Sitra. https://www.sitra.fi/en/publications/gatekeeping-in-the-digital-age

Seuri, O., Ikäheimo, H. P., & Huhtamäki, J. (2022). What happens when platforms mediate the audience–journalism relationship? In V. Manninen, M. Niemi, & A. Ridge-Newman (Eds.), *Futures of journalism* (pp. 227–243). Palgrave Macmillan.

Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping theory*. Routledge.

Singer, J. B. (2014). User-generated visibility: Secondary gatekeeping in a shared media space. *New Media & Society*, *16*(1), 55–73. https://doi.org/10.1177/1461444813477833

Steedman, R., Kennedy, H., & Jones, R. (2020). Complex ecologies of trust in data practices and data-driven systems. *Information, Communication & Society*, *23*(6), 817–832. https://doi.org/10.1080/1369118X.2020.1748090

Talvitie-Lamberg, K., Lehtinen, V., & Valtonen, S. (2022). Tactics of invisibility: How people in vulnerable positions make datafied everyday life livable. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448221136077

Tandoc, E. C., Jr. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society*, *16*(4), 559–575. https://doi.org/10.1177/1461444814530541

Tandoc, E. C., Jr., & Vos, T. P. (2016). The journalist is marketing the news: Social media in the gatekeeping process. *Journalism Practice*, *10*(8), 950–966. https://doi.org/10.1080/17512786.2015.1087811

Tracy, S. J. (2018). A phronetic iterative approach to data analysis in qualitative research. *Journal of Qualitative Research*, *19*(2), 61–76. https://doi.org/10.22284/qr.2018.19.2.61

van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Vos, T. P. (2019). Journalists as gatekeepers. In K. Wahl-Jorgensen & T. Hanitzsch (Eds.), *The handbook of journalism studies* (pp. 90–104). Routledge.

Vu, H. T. (2014). The online audience as gatekeeper: The influence of reader metrics on news editorial selection. *Journalism*, *15*(8), 1094–1110. https://doi.org/10.1177/1464884913504259

Wallace, J. (2018). Modelling contemporary gatekeeping: The rise of individuals, algorithms and platforms in digital news dissemination. *Digital Journalism*, *6*(3), 274–293. https://doi.org/10.1080/21670811.2017.1343648

Welbers, K., & Opgenhaffen, M. (2018). Social media gatekeeping: An analysis of the gatekeeping influence of newspapers' public Facebook pages. *New Media & Society*, *20*(12), 4728–4747. https://doi.org/10.1177/1461444818784302

About the Authors

Margareta Salonen (MA) is a doctoral researcher in journalism at the Department of Language and Communication Studies, University of Jyväskylä, Finland. Her work focuses on the triadic power relations between news media, audiences, and social media platforms. More precisely, she researches social interaction between news media and audiences, datafication of news work, and journalism ethics on social media platforms. She is a member of the Finnish press council and the Student and Early Career Representative of the ICA Journalism Studies Division.

Veera Ehrlén (PhD) is a postdoctoral researcher in journalism at the Department of Language and Communication Studies, University of Jyväskylä, Finland, where she also earned her doctorate in journalism in 2021. She has expertise in studying mediatisation of leisure cultures, digitalisation of social interaction, social media consumption and content distribution practices, and news consumption in data-driven digital environments.

Minna Koivula (MA) is a doctoral researcher in journalism at the Department of Language and Communication Studies, University of Jyväskylä, Finland. Her research examines journalism through the lens of organisational communication research and explores the shifting institutional context of work in the media field.

Karoliina Talvitie-Lamberg (PhD) is an assistant professor in media, communication and journalism at the Department of Language and Communication Studies, University of Jyväskylä, Finland. Her research interests include everyday datafication in media publics, inclusion and vulnerabilities, data ethics, digital disconnection, and AI and bot-enhanced communication. She is a specialist in digital inclusion issues in media and public service development. Her work has been published in several esteemed journals, including *New Media and Society*, *Computers in Human Behavior*, and *Surveillance and Society*.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 379–391
https://doi.org/10.17645/mac.v11i2.6401

Article

Platforms and Exposure Diversity: Towards a Framework to Assess Policies to Promote Exposure Diversity

Heritiana Ranaivoson * and Nino Domazetovikj

imec-SMIT, Vrije Universiteit Brussel, Belgium

* Corresponding author (hranaivo@vub.be)

Submitted: 30 October 2022 | Accepted: 23 March 2023 | Published: 28 June 2023

Abstract
The fragmentation of consumption and algorithms' increasing impact on how content is recommended and displayed makes it even more important to analyse and promote exposure diversity, i.e., the extent to which audiences are exposed to, discover, and engage with diverse content. Although there is a growing literature addressing how to define media diversity in the context of the challenges posed by platformisation, this article translates the normative dimensions into a framework for operationalising exposure diversity into a tangible policy goal, taking into account datafication and its consequences in terms of increasing data requirements towards platforms. The main objective of this study is to analyse initiatives to assess exposure diversity in the platform era and to discuss how such assessment could be improved, particularly for policy initiatives. This involves addressing several challenges of existing approaches for the assessment of exposure diversity related to defining an appropriate frame of reference, determining the degree of diversity required, dealing with data transparency issues, and promoting user autonomy. To achieve this, we propose a framework for analysing initiatives aimed at assessing and promoting exposure to media diversity. Our framework is composed of four key features: measures (type of initiative), metrics (quantifying exposure diversity), data collection methods, and data requirements. We apply this framework to a set of 13 initiatives and find that policy initiatives can benefit from adopting metrics based on distances and experimenting with data collection methods.

Keywords
audiovisual policy; data requirement; exposure diversity; media pluralism; news policy; online platforms; platform regulation; recommender systems

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

With the fast-paced development and adoption of digital technologies in the distribution and consumption of media content, assessing media diversity has become even more important. Media diversity is a significant objective (Helberger, 2011), which is formative for democratic societies. With a long tradition in free speech and democratic theory (Ash, 2016), it has become a central value in media law and policy (Helberger et al., 2020). Having diversity in cultural goods and services offered to citizens, so that they have access to a varied range of available views and ideas, is perceived as a desirable objective for policies addressing the cultural and media sectors. Following Napoli (1997), we distinguish, within media diversity, between source, content, and exposure diversity. Source diversity focuses on content producers and media outlets. Content diversity is concerned with the content features, e.g., the opinions and views expressed, the origin and the language, etc. Exposure diversity corresponds to the extent to which audiences are exposed to, discover, and engage with diverse content (Napoli, 2011).

Due to digitalisation, any content and the ways in which users interact with that content can be transformed into bits of data. Mayer-Schönberger and Cukier (2013) refer to datafication as the process of recording and computing data from different sources and the value that could be obtained from it. Combined, digitalisation and datafication contribute to distinct content types being accessible on the same platform while enabling providers access to unprecedented insights into human behaviour and new forms of creating and extracting value (Kennedy et al., 2015; Mayer-Schönberger & Cukier, 2013). Within this context, the boundaries between media sectors are blurred and there is increased competition between companies that were not competitors before—especially as they now compete for attention (Picard, 2011). Digital transformation and convergence have pushed media companies to extend their set of activities (e.g., newspapers producing podcasts or broadcasters developing on-demand services) resulting in an increase of available content and shifting the competitive landscape from sectoral competition to a situation where many companies and conglomerates have extended their activities beyond the boundaries of their native industries. Due to the proliferation of content across multiple delivery platforms, audiences can conveniently switch between reading news articles on social media and consuming multimedia content on streaming services from a single access point. Digital distribution of content has little effect on costs thus promoting ubiquitous availability across various distribution platforms. Therefore, content convergence can lead to distribution divergence (Dal Zotto & Lugmayr, 2016) as well as consumption fragmentation.

In a context where algorithms modulate how content is recommended and displayed, consumption fragmentation has rendered it more crucial to update the analysis and assessment of exposure diversity (Lambrecht et al., 2022). Considering the changing media landscape, platforms and digital media providers are operating under time and attention constraints, which has highlighted the role of intermediation in creating and capturing value (Kostovska, 2022). These circumstances have prompted the inception of novel approaches to curating and sorting content. There is a complete change in how exposure diversity is considered with the algorithmic influence on recommendation and display (Vrijenhoek et al., 2021) and the possibility to have access to in-depth data on consumption. The core challenge in examining exposure diversity in the platform era is related to algorithms' impact on diversity (Haim et al., 2018; Jürgens & Stark, 2022; Sørensen & Schmidt, 2016). Algorithms can be flawed by inherent biases and can produce unfavourable outcomes such as inequality of representation and coverage (Ranaivoson, 2019), which are associated with challenges identified in the literature as the "filter bubble" (Pariser, 2011) and the "echo chamber" (Colleoni et al., 2014). Colleoni et al. (2014) claim that users are in echo chambers when their prior political views are reinforced due to selective online exposure to political content. In a similar metaphor, in filter bubbles, users are never aware of what others think outside their bubble regarding political, moral, or scientific issues (Dündar & Ranaivoson, 2022). However, the filter bubble rationale has attracted some criticism. Bruns (2019, p. 8) calls it "the dumbest metaphor on the internet" since, "to the extent that it occurs," homophily results in the first place from users' agency. In addition, empirical evidence that warrants any strong worries about filter bubbles is lacking (Zuiderveen Borgesius et al., 2016). Haim et al. (2018) note that empirical evidence on the existence of the filter bubble and its effects, especially in the context of news, is limited. Of the existing handful of studies, none has been able to prove genuine negative effects of filter bubbles (Ranaivoson, 2019). Nonetheless, the sheer possibility of filter bubbles needs to consider exposure diversity in the context of algorithms.

It is suggested by van Dijck (2014) that it is not accurate to view data as solely a reflection of neutral human behaviour harnessed by platforms and that the role played by platform interventions and interpretations in creating and refining these data resources should be acknowledged. Therefore, established rationales for promoting diversity are being challenged not only by the novel ways of distributing and consuming content but also by the ways in which platforms utilise data to modulate intermediation. As normative notions, media diversity and pluralism are operationalised through measures that aim to increase source and content heterogeneity, with the goal of promoting variety (of information, views, content, and ownership; Ranaivoson, 2007). In contrast, exposure diversity looks at the audience dimension of media diversity (Helberger, 2012). It is therefore not enough to focus on source and content diversity, and it is necessary to consider the specificities of consumption diversity and the role played by all ways of modifying exposure.

Scholarly literature focused on approaches and methodologies for assessing exposure diversity is limited in terms of the volume of studies that incorporate consumption and algorithmic suggestions' diversity. Whereas the issue of defining media diversity in the context of platformisation has been discussed (Helberger, 2018; Helberger et al., 2018; Hendrickx et al., 2020; Joris et al., 2020; Ranaivoson, 2019), the question of how normative dimensions can be translated into tangible measures requires further consideration. Therefore, to advance our understanding of exposure diversity, it is important to adopt a more systematic approach to examine its conceptual features. To accomplish this, we develop a framework for assessing diversity of exposure in the platform era and apply it to a comparison of initiatives aimed at assessing and promoting exposure diversity. Our primary research question is: What characterises initiatives to assess exposure diversity in the platform era and how could such assessment be improved, in particular for policy initiatives?

To answer this question, in Section 2, we discuss the challenges of promoting exposure diversity, which are notably related to the fact that media policy has been mainly aimed at promoting source and content diversity. In Section 3, we present a framework to evaluate initiatives to assess and promote exposure diversity in a datafied society. The following main features of the framework are discussed: measures, metrics, methods, and data requirements. Section 4 presents an overview and compares such initiatives. The main results of the comparison are discussed following our framework. The conclusion focuses on policy implications and recommendations that can be derived from the analysis.

2. The Challenges of Promoting Exposure Diversity

In the EU, media policy has largely sought to promote pluralism by targeting the availability of diverse content. Media diversity has been primarily analysed from the point of view of production and distribution and their impact on consumption. Therefore, current media diversity policies remain rooted in scarcity logic (Helberger, 2019) even though source diversity alone does not secure diversity of the overall output (Helberger, 2012). With digitalisation and datafication, it is important to consider beyond what is available, what actually gets recommended and displayed to users (Ranaivoson, 2020), and the role data play in these processes.

Existing literature highlights several challenges that must be considered to provide a better understanding of the concept. Firstly, as outlined by Helberger et al. (2018), diversity of exposure is a means to an end rather than an aim in itself. Since providing more diversity might come at a cost (Stirling, 2007), it is key to understand how diversity of exposure can contribute to the overall goals. Helberger et al. (2018) highlight three perspectives for considering exposure diversity as a societal goal which can also contribute towards deriving normative positions. These are the individual autonomy perspective, the deliberative perspective, and the adversarial perspective (Helberger et al., 2018). Within this framework, exposure diversity can be utilised to respectively "extend individual choice" and provide individuals with "more opportunities to realize their interests," "promote rational public debate and the formation of a reasoned public opinion," or provide "a corrective to the tendency of public debates to be dominated by existing elites and powerful interests" (Helberger et al., 2018, p. 195).

From a more applied but still conceptual perspective, exposure applies to diverse types and formats of content in different contexts. It seems difficult to design a single approach for assessing and promoting exposure diversity, which would apply to cultural or linguistic diversity or representations of minorities in media outlets. A related critical question from a policy perspective is to decide what an adequate degree of diversity of exposure would be by determining specific benchmarks (Helberger, 2011). There is no consensus on how diversity should be quantified at the recommender system level (Kunaver & Požrl, 2017).

Moreover, difficulties arising from a lack of data transparency and data collection challenges add another layer of complexity to the aspiration of defining indicators to assess exposure diversity. Data are used by media companies to influence consumption behaviours, e.g., in the design and functioning of recommender systems. While such data and data about actual consumption are needed to assess exposure diversity, there is no unified, transparent way for external actors to access it. As proprietary audience data is a valuable resource for developing and maintaining competitive advantage, platforms are incentivised to refrain from data sharing, which can hinder attempts to evaluate exposure diversity (Hagiu & Wright, 2020; van Dijck et al., 2018). Ultimately, data is an asset and firms have an incentive to keep as much as possible of it siloed. For instance, even though Netflix has access to abundant and granular data about the viewing habits and preferences of their users, they only selectively release data on some aspects of their service (Wayne, 2022).

The growing importance of algorithms coupled with the difficulties of predicting their outcomes calls forth for algorithm auditing as an early bias and problem detection tool when new versions are released. However, such an approach can be costly for regulators and those audited and challenging concerning the quantity and quality of data that can be collected as well as the legal soundness of the approach in general (PEReN & Regalia, 2021).

Finally, when discussing the question of data and exposure diversity, a challenge to consider is related to users' agency. Ensuring that users' personal autonomy and privacy and how they envision them are respected (Helberger et al., 2018; Stasi, 2019) can conflict with what experts or policymakers could identify as the "socially-desired distribution of audience attention" (Napoli, 2011, p. 256). Transparency and privacy obligations are at the core of the discussions on designing monitoring mechanisms, as can be seen in the recently adopted EU Regulation Digital Services Act.

3. A Framework to Address Initiatives for Exposure Diversity

Since exposure diversity corresponds to the extent to which audiences are exposed to, discover, and engage with diverse content (Napoli, 2011), it also relates to the users' media diets. Few studies in academic literature incorporate the perspectives of consumption diversity and diversity of algorithmic suggestions when proposing ways for assessing exposure diversity. Furthermore, conceptual advances would be useful for policy purposes.

Therefore, we propose a framework to analyse and compare current or recent initiatives to assess and eventually promote exposure diversity. It distinguishes between measures, metrics, methods, and data

requirements as features of all initiatives included in our overview. Measures are initiatives to achieve more diversity, they correspond to the overall objective, *why* the assessment and promotion of exposure diversity take place. This relates to the discussion on exposure diversity as a societal goal (Helberger et al., 2018). Metrics are less normative and somehow more concrete as they define *what* is being assessed. We expect that the approach changes depending on the type of content and the context. Considering metrics is also rendered necessary by the lack of consensus regarding how to assess exposure diversity—e.g., at the recommender system level (Kunaver & Požrl, 2017). Methods correspond to *how* such assessment is made, notably regarding data collection. Finally, data requirements give one of the main *limits* of the initiatives in terms of Transparency, Parsimony, and Replicability. Actually, while considerable amounts of data are generated, as discussed before, there are large data collection and use challenges.

Our framework first appeared in the context of a study performed for the European Commission on media plurality and diversity online (Lambrecht et al., 2022). It is notably based on a previous study for UNESCO on suitable properties of indexes of media and cultural diversity (Ranaivoson, 2007), which was dealing with *metrics*. The importance of *measures* is highlighted by Helberger et al. (2018), in particular for diversity-by-design measures. The term "framework" itself derives from the Australian Communications and Media Authority (ACMA) study on news measurement framework, which was focused on developing new approaches to assess media diversity in a context of declining local news and a need for better understanding and quantifying public interest journalism. Finally, an original contribution of this framework, compared to the version in Lambrecht et al. (2022), is the emphasis put on data requirements. This is to allow us to take into account how datafication is impacting exposure diversity, i.e., how audiences are exposed to, discover, and engage with diverse content.

3.1. Measures

Measures are initiatives to achieve more diversity. We distinguish between Industry, Policy, and Research initiatives. Industry initiatives are led by one industry partner or a consortium of industry partners to develop solutions to improve exposure diversity. Policy initiatives include law-making and policies aimed at monitoring or improving exposure diversity. Research initiatives are led by universities or academic institutions to assess exposure diversity. Among the latter, we have tried to exclude Research projects of a smaller scale. There are indeed a great number of studies that aim at increasing exposure diversity—see, e.g., Latha and Nadarajan's (2019) mapping of approaches to improve recommendation diversity). They can be focused on software solutions. For example, Kamishima et al.'s (2012) recommender system aims at providing neutral recommendations to users in relation to a specific viewpoint. Latha and Nadarajan's (2019) approach incorporates diversity into its recommendations and is applied to movies and news. Helberger (2018) recommends the use of tools for alternative recommendation settings and technologies that make users aware of their filter bubble.

3.2. Metrics

Metrics refer to what is being assessed. This includes of course assessing the diversity of exposure, but also identifying the barriers to diversity. Recent literature overviews point to the too high number of conceptualisations of news and media diversity (Hendrickx et al., 2020; Joris et al., 2020). Besides, media diversity is assessed by a broad spectrum of scholars, from social to computer sciences. Despite this, the initiatives to promote and assess exposure diversity use indicators that derive from and sometimes combine the four following basic metrics: Count, Percentages, Dual Indexes, and Distances (Lambrecht et al., 2022). These basic metrics can be aggregated or combined to assess the diversity of exposure.

Count represents the number of elements that are part of a set. The elements can be various units, which can be part of a subset or grouped in a category. Count can be used to quantify to what extent units of various subsets or categories are represented in a set. Percentages correspond to all indicators consisting in measuring the relative share of a category within a system, e.g., the shares of films in Netflix's catalogue per country of origin. Dual Indexes (Stirling, 2007) notably include both the entropy index and the Herfindahl-Hirschman index. As Stirling's (2007) denomination alludes, both indices combine in their quantification what Count and Percentages assess respectively.

Distances correspond to the level of differences, or dissimilarity, between every pair of items consumed, for example. Distances are commonly used by computer scientists to assess diversity (Kunaver & Požrl, 2017). Using cosine similarity, they assess how similar two items are. Mathematically, cosine similarity corresponds to the normalised (between 0 for *maximum similarity* and 1 for *maximum dissimilarity*) cosine of the angle between two vectors that are projected in a multidimensional space (Prabhakaran, 2018). Distances can be calculated for any content, assuming that the data is labelled (e.g., using metadata) or can be categorised (Lambrecht et al., 2022). For example, a common computational approach to assess distances is the bag-of-words model, which compares texts according to the words they use. Distances are often used to assess recommender systems' diversity—especially in comparison to Count or Percentages. In practice, the diversity of recommendations will be increased by suggesting items that are further from a user's preferences, although always within certain limits (Lambrecht et al., 2022). For example, someone who has watched a video about football

could be proposed other videos about other sports, because football is a subcategory of sport or because the recommender system has found that those who have watched videos about football often also watch videos about other sports.

3.3. Methods

Methods are ways to collect data that can be used to assess the diversity of exposure. Methods include Surveys, Online Data Scraping, and Experiments. Surveys consist in gathering information from a sample of individuals, usually through a questionnaire but this also includes interviews. An example is the Office of Communications's (Ofcom) collection of data via consumer research for their Measurement Framework for Media Plurality (MFMP). Collected information is therefore declarative. In contrast, Online Data Scraping consists in extracting data from human-readable output or content available online, which provides insights into actual (consumption) behaviour at aggregate and, in some cases, individual levels. For example, for the reports on "Australian content," the researchers entered into a database all titles identified on Stan, Netflix, and Amazon Prime Video when inspecting content included in relevant categories on each service (such as Australian movies and Australian TV) and using search terms ("Australia," "Australian," "Austral*") to identify relevant content not included in these categories (Lobato & Scarlata, 2019). Experiments also aim at analysing the actual behaviour of individuals but by putting them in a setting where their behaviours can be directly observed and hence more easily explained. Experiments can also consist of simulations to test features such as algorithmic systems. For example, in the discoverability index Research initiative, Tétu and Dubois-Paradis (2020) built two profiles on Netflix to observe which films get recommended to these profiles, based on their different viewing behaviours.

3.4. Data Requirements

Finally, the initiatives can be compared based on data requirements. The emphasis is put upon which data would be required for the methodologies to be applied. The fourth feature of our framework considers the profound changes in media consumption resulting from datafication. Huge amounts of real-time data are produced but this is also challenging for initiatives aiming at promoting and/or assessing exposure diversity: Given the available data, a suitable initiative should not be too demanding in terms of data necessary to build metrics. Within this feature, we distinguish between Transparency, Parsimony, and Replicability. *Transparency* means how easy the measures, metrics, and methods are to understand. The initiative is Transparent when the assumptions are explicit, including regarding the data used to compute the metrics and more generally apply the initiatives. *Parsimony* corresponds to the costs underlying the use of data by the initiatives. This includes the financial cost to access the data and, beyond, how simple it is to assess exposure diversity (e.g., required computing power and conversely the speed at which calculations can be done). *Replicability* means to what extent the indicators can repeatedly be assessed over time in the same context. This notably concerns the type of data the frameworks rely on. Questions here include the type of data, their periodicity, and how they are collected. Besides, the access to data is considered: To what extent are data available publicly? Who owns such data? How is access to these data ensured? And to whom?

4. A Review of Initiatives for Exposure Diversity

4.1. Methodology

We employ a multi-method approach to investigate a nascent aspect of media and communications policy including document analysis (Karppinen & Moe, 2012) and expert interviews (Van Audenhove & Donders, 2019). Expert interviews are suitable for gathering non-codified knowledge and insights about emerging issues in media and communications policy research, whereas the selection of interviewees is identified as a critical aspect of the research design (Hammersley & Atkinson, 1995). We conducted 15 semi-structured expert interviews with professionals, academics, and technical experts from various geographic locations such as Belgium, the Netherlands, France, the UK, Australia, Canada, and Switzerland (see Table 1). A snowball effect was created by encouraging the interviewees to suggest other potentially relevant interviewees. The information collected through the interviews allowed us to complement and contextualise the document analysis of 13 initiatives aimed at assessing exposure diversity (see Table 2). The selection of initiatives was based on a snowball process as well. The document analysis involved a systematic comparison of the initiatives, with the coding being informed by both publicly available information and undisclosed information gathered during the interviews. The framework used for the analysis enabled the initiatives to be compared in terms of why, what, how, and to what extent they allow for the assessment of exposure diversity.

Regarding measures, initiatives can be either Policy, Industry, or Research. Metrics can include and sometimes cumulate Count, Percentages, Distances, and Dual (for Dual Indexes). Similarly, there can be one or more methods among Surveys, Scraping, and Experiments. Data requirements for each of these initiatives are more difficult to assess in a comparative manner. This is why we propose a more granular approach and distinguish between Transparency, Parsimony, and Replicability. The assessment though remains a bit subjective and qualitative, especially in comparison to our other features. Transparency and Replicability refer to easiness, respectively to understand and to repeat, which are difficult

Table 1. Interviews conducted in the frame of the project.

Name	Position	Organisation	Date of interview
Glen Joris	PhD student	Ghent University	30/11/2020
Joris Mattheijssens	Data scientist	Vlaamse Radio- en Televisieomroeporganisatie	30/11/2020
Sanne Vrijenhoek	Researcher	University of Amsterdam	02/12/2020
Johan Loeckx	Researcher	Vrije Universiteit Brussel	02/12/2020
Nicolas Rolin Lucas Verney Benoît Rottembourg	Researchers	Pôle d'Expertise de la Régulation Numérique Inria	03/12/2020 16/02/2021
Nava Tintarev	Professor	TU Delft	04/12/2020
Olaf Steenfadt	Global project director	Media Ownership Monitor	10/12/2020
Michèle Rioux	Director	Université du Québec à Montreal	10/12/2020
Eleonora Mazzoli	PhD student	London School of Economics and Political Science	08/01/2021
Ramon Lobato	Professor	Royal Melbourne Institute of Technology	21/01/2021
Nanao Kachi	Director of social and consumer policy	Canadian Radio-Television and Telecommunications Commission	22/01/2021
Véronique Guèvremont	Professor	Laval University	22/01/2021
Sébastien Noir	Head of software engineering, technology, and innovation	European Broadcasting Union	25/01/2021
Catherine Johnson	Professor	University of Huddersfield	27/01/2021

to objectively assess. Parsimony refers to costs but all data was not available to extensively compare initiatives on that aspect. For all these reasons, that feature of the framework should be taken cautiously. Similarly, the information through our document analysis and our interviews could not always allow us to properly assess this feature, hence the note that data is in that case not available (n/a).

4.2. The Rise of Policy Initiatives Towards Exposure Diversity

Considering measures, the comparison first shows that Policy initiatives have appeared that take exposure diversity into account. Napoli (2011) deemed that exposure diversity, despite being increasingly regarded as a critical component when assessing diversity within online environments, is very seldom the object of policy interventions. This seems to be changing now. While media diversity remains usually assessed by regulatory authorities by considering source diversity (ACMA, 2020), our sample includes four Policy initiatives: two established ones (Ofcom's MFMP and the EAO's yearly publication on the visibility of audiovisual works on transactional video-on-demand services) and two more in the making. For the latter, two reports published respectively in Australia (ACMA, 2020) and in Canada (Parliament of Canada, 2021) could lead to policies to promote exposure diversity among others. Bill C-11 is the second attempt to amend Canada's Broadcasting Act after a bid to modernise the Act with Bill C-10 ended unsuccessfully with the dissolution of the 43rd Canadian Parliament in 2021.

Besides one Industry initiative, most other initiatives are Research initiatives. As algorithms (notably recommender systems) and their impact on diversity play a significant role in the analysis of exposure diversity, most research revolves around computer science. However, some projects like PersoNews or DIAMOND rely on an interdisciplinary approach, involving social sciences, e.g., to analyse the impact of algorithms on users, to understand how they fit in media business strategies, or to devise policy recommendations.

4.3. A Focus on One or Two Metrics

All metrics appear in our sample of initiatives: Percentages in eight of them, Distances in seven of them, Count in six of them, and Dual in three of them. Only two Research initiatives rely on all these metrics, PersoNews and ENSURE, which reflects the broad scope of research they involve. Both include several more focused projects, which can have different emphases in terms of exposure diversity.

Another interesting specificity concerns the use of Distances, which has so far never been used in Policy initiatives but exclusively in Research initiatives (the European Broadcasting Union PEACH Industry initiative being the exception). Moreover, these Research initiatives always involve computer scientists, which reflects

Table 2. A comparison of initiatives to assess/promote exposure diversity based on our framework to address initiatives for exposure diversity.

Abbreviation	Media sector	Measure	Metrics	Methods	Data requirement	Brief description	Lead organisation	Country
EAO Visibility	Audiovisual	Policy	Percentages and Count	Scraping	n/a	Yearly analysis by the EAO (European Audiovisual Observatory) of the visibility (promotion) of audiovisual works on transactional video-on-demand services	EAO	Europe
Australian content	Audiovisual	Research	Count and Percentages	Scraping	Transparent and Replicable	Yearly reports examining the availability and discoverability of Australian screen content on subscription video-on-demand services	Royal Melbourne Institute of Technology	Australia
Broadcasting Act	Broadcasting	Policy	n/a	n/a	n/a	A 2020 report reviewed the regulatory and legislative frameworks for broadcasting and telecommunications in Canada; the legislation (Bill C-11, currently under final consideration in the Canadian Senate) could lead to some recommendations being applied	Canadian Radio-Television and Telecommunications Commission	Canada
PEACH	Broadcasting	Industry	Distances	Scraping	Transparent and Parsimonious	Personalisation for EACH (PEACH) is a "personalisation and recommendation ecosystem developed by broadcasters for broadcasters"; the recommended content should broaden a user's horizon	European Broadcasting Union	Switzerland
Discoverability index	Music Audiovisual Books	Research	Percentages	Experiments	Transparent and Replicable	Index considering the presence, visibility, and recommendation of content (music, audiovisual, book), developed by the LATICCE lab	Université du Québec à Montréal	Canada
Recoloco	n/a	Research	Distances	Scraping	n/a	REcommanding personalized COntent for LOcal Communities (Recoloco) has developed software and approaches to personalise Postbuzz's content and user experience, a digital replica of your physical mailbox; this has included automatically identifying new content tags, recommending content tags, and profiling users	Vrije Universiteit Brussel	Belgium

Table 2. (Cont.) A comparison of initiatives to assess/promote exposure diversity based on our framework to address initiatives for exposure diversity.

Abbreviation	Media sector	Measure	Metrics	Methods	Data requirement	Brief description	Lead organisation	Country
MFMP	News	Policy	Count, Percentages, and Dual Indexes (Herfindahl-Hirschman index)	Scraping and Surveys	Transparent	The MFMP was developed by the UK Ofcom to measure media diversity via availability, consumption, impact, and contextual factors	Ofcom	UK
PersoNews	News	Research	All	Scraping and Experiments	n/a	The European Research Council PersoNews (profiling and targeting news readers and implications for the democratic role of the digital media, user rights, and public information policy) project investigated the impact the trend for personalisation has on the role of digital media in society and how that can be assessed; it has been followed by several projects such as the Research Priority Area Human(e) AI	University of Amsterdam	Netherlands
NewsDNA	News	Research	Percentages and Distances	Experiments	Transparent and Replicable	Diversity in the News Through Algorithmization (NewsDNA) was an interdisciplinary four-year research project (2018–2022) to develop and test an algorithm that uses news diversity as a key driver for personalised news recommendations	Ghent University	Belgium
DIAMOND	News	Research	Distances	Surveys	Transparent and Replicable	Diversity and Information Media: New Tools for a Multifaceted Public Debate (DIAMOND) was a Flemish (Belgian) interdisciplinary four-year research and valorisation project (2017–2021) on news diversity	KU Leuven	Belgium

Table 2. (Cont.) A comparison of initiatives to assess/promote exposure diversity based on our framework to address initiatives for exposure diversity.

Abbreviation	Media sector	Measure	Metrics	Methods	Data requirement	Brief description	Lead organisation	Country
CPN	News	Research	Distances	Surveys and Experiments	Transparent and Parsimonious	The EU H2020 Content Personalisation Network (CPN)project has offered news organisations transparent and easily integrated software to personalise their content ("bring your audience the right stories at the right time'')	Vlaamse Radio- en Televisieomroe-porganisatie	Belgium
ACMA report	News	Policy	Count and Percentages	Survey and Scraping	Transparent and Replicable	The ACMA 2020 report (*News in Australia: Diversity and Localism. News Measurement Framework*) sets out an alternative framework to measure the current levels of news diversity and the availability of local news throughout Australia	ACMA	Australia
ENSURE	Various	Research	All	Surveys and Experiments	Transparent and Replicable	The ExplaiNing SeqUences in REcommendations (ENSURE) project looked at ways of improving the transparency and decision support for recommender systems; it is linked to other research projects on diversity (recommender systems, viewpoint diversity, etc.)	TU Delft	Netherlands

the fact that computer scientists tend to assess diversity by using Distances (Kunaver & Požrl, 2017).

Conversely, initiatives that make use of Count and/or Percentages seldom make use of Distances. The problem is that these metrics consider different, complementary aspects of diversity, as the Stirling (2007) model shows. For example, increasing the diversity of recommendations on a music streaming platform can be done by proposing songs by more different artists, instead of always by the same artist (which can be assessed by combining Count and Percentage) as well as by proposing songs in less familiar genres (which can be assessed in terms of Distances).

4.4. Methods: Main Results

The three types of methods (Surveys, Scraping, and Experiments) appear well-represented in our sample. The most striking result is the fact that Policy initiatives never rely on Experiments. One explanation could be that such Experiments better fit exploratory analysis whose results still need to be generalised after, which makes them probably less fit in the context of a Policy initiative.

The analysis of the initiatives also shows that most Research initiatives on exposure diversity involve computer science. This seems logical in a context where algorithms (notably recommender systems) and their impact on diversity play a significant role in exposure diversity. Some Research initiatives (and in our view, the most interesting ones) rely on an interdisciplinary approach, involving social sciences, e.g., to analyse the impact of algorithms on users, to understand how they fit in media business strategies, or to devise policy recommendations.

4.5. Transparent but not Parsimonious Initiatives

Finally, regarding data requirements, most initiatives are Transparent (nine vs. three). The assessment was made based on the information provided about data used to assess and eventually improve exposure diversity, as well as on the metrics actually used. Neither Recoloco nor EAO Visibility provided information on the data used (Recoloco notably provided no information on the data used to profile users). As for PersoNews, the lack of Transparency rather stemmed from the lack of information on the metrics themselves.

Only two initiatives are Parsimonious: PEACH and CPN. PEACH provides tools aimed at editors and journalists. It is developing algorithms to recommend content which will broaden a user's horizon (i.e., to educate them). Data used are the ones collected already by media providers, it is rather the functioning of the algorithms that is tweaked in another direction. In the CPN Research initiative, exposure diversity was assessed based on partner public service media's algorithmic systems and a rather classical approach, hence Parsimony, although user research is more costly in the process. Interestingly, a common point of PEACH and CPN is that they are led by media service providers.

Finally, six initiatives are Replicable. Replicability depends on data type and ownership. A difference is to be made notably between whether data are available publicly or not. Hence, EAO Visibility is relying on a partnership with Ampere Analysis, the EAO is therefore not owning data. In comparison, for the discoverability index, the LATICCE team has set up profiles on streaming services to compare received recommendations, which could be replicated by anyone with internet access and a subscription to the analysed services.

5. Discussion

The increasing importance of platforms in modern society has contributed to a growing apprehension from governments and advocacy groups regarding issues such as privacy and surveillance concerns, potential threats to freedom of expression, as well as the possibility of technological and infrastructure domination (Gillespie, 2018). Data is increasingly important but is often controlled by a few powerful actors. The resulting imbalance of power between those who provide the data and those who control it can impact how institutions and public discussions are governed (Kennedy et al., 2015; van Dijck et al., 2018). Diversity, as a crucial objective of media policy, needs to be readdressed in a context where online platforms' domination over access to content has altered the way citizens are exposed to media. The aim of this article is to provide a framework to analyse initiatives meant to assess and promote exposure diversity in the platform context. We develop and apply such analytical framework to a sample of initiatives and highlight drawbacks as to the preparedness of current policy in considering datafication when assessing exposure diversity across media landscapes. We highlight the importance of considering exposure diversity as a policy goal and propose approaches for defining and assessing it.

Our main contribution is the framework to assess and compare initiatives towards exposure diversity. This framework consists of four features: measures, metrics, methods and data requirements. Measures are initiatives to achieve more diversity and they correspond to the overall objective, why the assessment and promotion of exposure diversity take place. Metrics are less normative and somehow more concrete as they define what is being assessed. Methods correspond to how such assessment is made, notably regarding data collection. Finally, data requirements consider the changes that datafication has brought to the whole media consumption process, assessing the initiatives in terms of Transparency, Parsimony, and Replicability. The framework is applied to a set of 13 initiatives aimed at assessing and eventually promoting exposure diversity. Applying the framework allows us to compare in a systematic way these initiatives according to why, what,

how, and to what extent these initiatives allow us to assess exposure diversity. Besides, the features of the frameworks can be combined in the analysis.

All this allows us to respond to the research question: What characterises initiatives to assess exposure diversity in the platform era? How could such assessment be improved, in particular for Policy initiatives?

Firstly, as exposure diversity is recognised in the policy sphere as an important objective that requires policies to promote and foster it, there is a growing number of Policy initiatives aimed at assessing and promoting exposure diversity. Secondly, based on our analysis, it may be argued that future designs of policy measures need to include Distances as metrics, as direct applications of research (in computer science). The focus on Percentages and Count leads to a reductive view of only one side of the problem. Distances have been applied only rarely as metrics to assess diversity in the cultural and media sectors (for example, see Farchy & Ranaivoson, 2011). One reason may be that measuring Distances has proved difficult for a long time for social scientists, either conceptually or technically. However, as computer scientists are doing it on a regular basis, there should be fewer obstacles now to such an endeavour. Furthermore, we recommend more generally for future Policy initiatives to develop a set of indicators instead of very synthetic indexes. This would enable adopting a multifaceted approach in grasping challenges related to diversity and a more comprehensive operationalisation of diversity.

Thirdly, Policy initiatives never rely on Experiments. One explanation could be that such Experiments are better suited for exploratory analysis and their results still need to be generalised after, which likely makes them less fitting in a policy context. More general considerations in terms of methods are about the importance of an interdisciplinary perspective for Policy initiatives and more generally for all initiatives. Social sciences and computer sciences remain too separated while some of the most interesting frameworks rely on interdisciplinary approaches. In a similar way, it is also crucial to involve experts and practitioners in the design and piloting of measurement frameworks, such as media professionals, algorithm developers, representatives from non-profit organisations and civil society, etc. Fourthly, for Policy initiatives as for other initiatives, it is important to assess constraints posed by data requirements. Transparency, Parsimony, and Replicability are important for citizens to understand these Policy initiatives and for such initiatives to be efficient and applied in the long term.

6. Conclusion

The proposed framework to analyse initiatives aimed at assessing and promoting exposure diversity in the platform context could also be used to assess future initiatives. Several important measurement frameworks are under discussion and should be followed up. In Canada, the proposed Bill C-11 is under discussion that would regulate the outcome of recommender systems of platforms and digital media providers serving the Canadian market with the aim to improve the prominence and discoverability of local content, which could likely lead to adopting new or adapting existing measurement frameworks. In Australia, the ACMA's report would still need to be validated and implemented before becoming an actual measurement framework.

A major difficulty lies in how to best involve—or address—online platforms. Online platforms impact all media sectors through algorithmic gatekeeping (Helberger, 2019; Napoli, 2015), which they deploy in order to automatically filter, rank, and recommend content (Haim et al., 2018). Nonetheless, understanding their impact necessitates an ability to evaluate the scope of diversity they make available and recommend. Currently under debate is whether platforms should be obliged to share data with national regulatory authorities, given the risks related to technological and infrastructure domination and the fundamental imbalances between those that provide data and those who control it (Gillespie, 2018; Kennedy et al., 2015; van Dijck et al., 2018). Alternatively, researchers could be allowed to employ data-gathering methods that do not require the platforms' authorisation—e.g., by recruitment of internet users to install an apparatus that automatically records and reports their internet usage (see Kitchens et al., 2020) or by using bots to scrape and collect data. This could benefit from online platforms being incentivised to share data allowing to measure media diversity. In any case, doubts would remain regarding the collected data's reliability and the metrics to be devised to audit platforms.

A follow-up and more extensive overview of initiatives used to not only measure exposure diversity but, beyond, promote prominence, discoverability, and serendipity, will be key to ensuring initiatives to assess and eventually promote exposure diversity remain relevant.

Acknowledgments

The authors acknowledge a Horizon2020 grant (Grant No. 952156).

Conflict of Interests

The authors declare no conflict of interests.

References

Ash, T. G. (2016). *Free speech*. Yale University Press.

Australian Communications and Media Authority. (2020). *News in Australia: Diversity and localism*. https://www.acma.gov.au/publications/2020-12/report/news-australia-diversity-and-localism

Bruns, A. (2019). Filter bubble. *Internet Policy Review*, *8*(4). https://doi.org/10.14763/2019.4.1426

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*(2), 317–332.

Dal Zotto, C., & Lugmayr, A. (2016). Media convergence as evolutionary process. In A. Lugmayr & C. Dal Zotto (Eds.), *Media convergence handbook: Firms and user perspectives* (Vol. 2, pp. 3–16). Springer.

Dündar, P., & Ranaivoson, H. R. (2022). Science by YouTube: An analysis of YouTube's recommendations on the climate change issue. *Observatorio (OBS*)*, *16*(3). https://doi.org/10.15847/obsobs16320222061

Farchy, J., & Ranaivoson, H. (2011). Do public television channels provide more diversity than private ones? *ENCATC Journal of Cultural Management and Policy*, *1*, 50–63.

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Hagiu, A., & Wright, J. (2020). When data creates competitive advantage. *Harvard Business Review*, *98*(1), 94–101.

Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, *6*(3), 330–343.

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in practice*. Routledge.

Helberger, N. (2011). Diversity by design. *Journal of Information Policy*, *1*, 441–469.

Helberger, N. (2012). Exposure diversity as a policy goal. *Journal of Media Law*, *4*(1), 65–92.

Helberger, N. (2018). Challenging diversity—Social media platforms and a new conception of media diversity. In M. Moore & D. Tambini (Eds.), *Digital dominance: The power of Google, Amazon, Facebook, and Apple* (pp. 153–175). Oxford University Press.

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207.

Helberger, N., Moeller, J., & Vrijenhoek, S. (2020). *Design pour la diversité—Diversité des contenus à l'ère numérique* [Design for diversity—Content diversity in the digital era]. Government of Canada. https://www.canada.ca/fr/patrimoine-canadien/services/diversite-contenus-ere-numerique/design-diversite.html

Hendrickx, J., Ballon, P., & Ranaivoson, H. (2020). Dissecting news diversity: An integrated conceptual framework. *Journalism*, *23*(8), 1751–1769.

Joris, G., De Grove, F., Van Damme, K., & De Marez, L. (2020). News diversity reconsidered: A systematic literature review unraveling the diversity in conceptualizations. *Journalism Studies*, *21*(13), 1893–1912.

Jürgens, P., & Stark, B. (2022). Mapping exposure diversity: The divergent effects of algorithmic curation on news consumption. *Journal of Communication*, *72*(3), 322–344.

Kamishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012). Enhancement of the neutrality in recommendation. In M. de Gemmis, A. Felfernig, P. Lops, F. Ricci, G. Semeraro, & M. Willemsen (Eds.), *Decisions@ RecSys'12* (pp. 8–14).

Karppinen, K., & Moe, H. (2012). What we talk about when we talk about document analysis. In M. Puppis & N. Just (Eds.), *Trends in communication policy research: New theories, methods and subjects* (pp. 177–193). Intellect.

Kennedy, H., Poell, T., & van Dijck, J. (2015). Data and agency. *Big Data & Society*, *2*(2). https://doi.org/10.1177/2053951715621569

Kitchens, B., Johnson, S. L., & Gray, P. (2020). Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption. *MIS Quarterly*, *44*(4), 1619–1649. https://doi.org/10.25300/misq/2020/16371

Kostovska, I. (2022). Blockchain ecosystems in the creative industries: Big dreams for micro-payments. In S. Baumann (Ed.), *Handbook on digital business ecosystems* (pp. 405–423). Edward Elgar.

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems—A survey. *Knowledge-Based Systems*, *123*, 154–162.

Lambrecht, I., Mazzoli, E. M., Ranaivoson, H., Domazetovikj, N., Tambini, D., & Valcke, P. (2022). An assessment of the effectiveness of measures related to prominence and the findability/discoverability of general interest content and services. In Directorate-General for Communications Networks, Content and Technology (Ed.), *Study on media plurality and diversity online: Final report* (pp. 95–157). Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/475bacb6-34a2-11ed-8b77-01aa75ed71a1/language-en/format-PDF/source-266738523

Latha, R., & Nadarajan, R. (2019). Analysing exposure diversity in collaborative recommender systems—Entropy fusion approach. *Physica A: Statistical Mechanics and its Applications*, *533*, Article 122052.

Lobato, R., & Scarlata, A. (2019). Australian content in SVOD catalogs: Availability and discoverability. Digital etnography Research Centre.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Napoli, P. M. (1997). Rethinking program diversity assessment: An audience-centered approach. *Journal of Media Economics*, *10*(4), 59–74.

Napoli, P. M. (2011). Exposure diversity reconsidered. *Journal of Information Policy*, *1*, 246–259.

Napoli, P. M. (2015). Social media and the public inter-

est: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy*, *39*(9), 751–760.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.

Parliament of Canada. (2021). *C-11: An act to amend the Broadcasting Act and to make related and consequential amendments to other Acts*. https://www.parl.ca/legisinfo/en/bill/44-1/c-11

PEReN, & Regalia. (2021). *Evaluation methods for content recommendation algorithms*. Direction Générale des Médias et des Industries Culturelles. https://www.peren.gouv.fr/rapports/2021-05-06%20-%20Evaluation%20Methods%20for%20Content%20Recommendation%20Algorithms%20-%20PEReN-Regalia.pdf

Picard, R. G. (2011). *The economics and financing of media companies*. Fordham Univ Press.

Prabhakaran, S. (2018). *Cosine similarity—Understanding the math and how it works?* machinelearningplus. https://www.machinelearningplus.com/nlp/cosine-similarity

Ranaivoson, H. (2007). *Measuring cultural diversity: A review of existing definitions*. UNESCO.

Ranaivoson, H. (2019). Online platforms and cultural diversity in the audiovisual sectors: A combined look at concentration and algorithms. In L. A. Albornoz & M. T. Garcia Leiva (Eds.), *Audiovisual industries and diversity economics and policies in the digital era* (pp. 100–118). Routledge.

Ranaivoson, H. (2020). Cultural diversity. In R. Towse & T. Navarrete Hernández (Eds.), *Handbook of cultural economics* (3rd ed., pp. 183–191). Edward Elgar.

Sørensen, J. K., & Schmidt, J. H. (2016). *An algorithmic diversity diet? Questioning assumptions behind a diversity recommendation system for PSM*. Unpublished manuscript.

Stasi, M. L. (2019). Social media platforms and content exposure: How to restore users' control. *Competition and Regulation in Network Industries*, *20*(1), 86–110.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, *4*(15), 707–719.

Tétu, M., & Dubois-Paradis, S. (2020). *Quelle place pour les produits audiovisuels québécois en ligne? Un modèle d'observation empirique de la découvrabilité* [What place for Quebecois audiovisual products online? An empirical observation model of discoverability]. LATTICE; CEIM.

Van Audenhove, L., & Donders, K. (2019). Talking to people III: Expert interviews and elite interviews. In H. Van den Bulck, M. Puppis, K. Donders, & L. Van Audenhove (Eds.), *The Palgrave handbook of methods for media policy research* (pp. 179–197). Palgrave Macmillan.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208.

van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.

Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D., & Helberger, N. (2021). Recommenders with a mission: Assessing diversity in news recommendations. In F. Scholer & P. Thomas (Eds.), *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 173–183). Association for Computing Machinery.

Wayne, M. L. (2022). Netflix audience data, streaming industry discourse, and the emerging realities of "popular" television. *Media, Culture & Society*, *44*(2), 193–209.

Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., de Vresse, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1). https://doi.org/10.14763/2016.1.401

About the Authors

Heritiana Ranaivoson is a senior researcher and project leader at Vrije Universiteit Brussel's imec-SMIT (Belgium) He holds an MSc in economics and management from the École Normale Supérieure de Cachan and a PhD in industrial economics from Université Paris 1, Panthéon-Sorbonne. He has led several research projects at international and national levels for the European Commission (H2020, study contracts), Unesco, Google, and others. He has published extensively in the fields of cultural and media diversity, media policy, audiovisual platforms, and recommender systems.

Nino Domazetovikj is a PhD researcher at Vrije Universiteit Brussel's imec-SMIT (Belgium) where he is part of the Media Economics and Policy unit. His research interests are situated at the crossroads of media economics, media management, and media policy studies. His current work is focused on the operations of global video-on-demand providers and how they alter financing and production practices in European audiovisual markets.

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 392–405
https://doi.org/10.17645/mac.v11i2.6393

Article

Media Concentration Law: Gaps and Promises in the Digital Age

Theresa Josephine Seipp

Institute for Information Law, University of Amsterdam, The Netherlands; t.j.seipp@uva.nl

Submitted: 30 October 2022 | Accepted: 13 April 2023 | Published: 28 June 2023

Abstract
Power concentrations are increasing in today's media landscape. Reasons for this include increasing structural and technological dependences on digital platform companies, as well as shifts in opinion power and control over news production, distribution, and consumption. Digital opinion power and platformised media markets have prompted the need for a re-evaluation of the current approach. This article critically revisits and analyses media concentration rules. To this end, I employ a normative conceptual framework that examines "opinion power in the platform world" at three distinct levels (individual citizen, institutional newsroom, and media ecosystem). At each level, I identify the existing legal tools and gaps in controlling power and concentration in the digital age. Based on that, I offer a unifying theoretical framework for a "digital media concentration law," along with core concepts and guiding principles. I highlight policy goals and fields that are outside the traditional scope yet are relevant for addressing issues relating to the digital age. Additionally, the emerging European Union regulatory framework—specifically the Digital Services Act, the Digital Markets Act, and the European Media Freedom Act—reflects an evolving approach regarding platforms and media concentration. On a final note, the analysis draws from the mapping and evaluation results of a Europe-wide study on media pluralism and diversity online, which examined (national) media concentration rules.

Keywords
digital platforms; editorial independence; European regulation; media concentration; media pluralism; opinion power; structural dependency

Issue
This article is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Introduction

Today's media environment is characterised by shifting opinion power, changes in news production, distribution, and consumption, and increased structural and technological dependences on digital platform companies. Throughout this article, I will refer to "digital platform companies" (hereafter "platforms") as (a) the services, platforms, and infrastructures of large platform companies, and (b) the "firm's" or "company's" corporate and business strategies. This is in line with the definitions of "platform" and "platform companies" brought forward by Gorwa (2019, p. 856) and Simon (2022, p. 1833).

It is observable that the power of platforms extends far beyond economic and data power, thereby affecting the entire media ecosystem, public sphere, and democracy (Helberger, 2020). These changing power dynamics in the media ecosystem and platforms' ability to influence public opinion formation contribute to growing media concentration trends that raise concerns about media pluralism, particularly as the shifting media landscape threatens local and independent journalism (Pickard, 2020; Seipp et al., 2023). Such developments might endanger not only a pluralistic media landscape but democracy as a whole (Helberger, 2020, p. 845). While these developments are relevant to many democracies around the world, this article focuses on Europe and its tradition of imposing positive obligations on states to prevent media concentration and promote pluralism and freedom (Tambini, 2021). Accordingly, in a well-functioning democracy, measures should be put in place to disperse "opinion power" (stemming from the

German notion "Meinungsmacht"), which is defined as "the ability of the media to influence processes of individual and public opinion formation" (Neuberger, 2018, p. 56; see also Baker, 2007; Seipp et al., 2023).

Media concentration law is the relevant legal tool for doing exactly that. Namely, addressing concentration and preventing dominant opinion power from accruing. However, existing tools are unable to capture increased concentration trends driven by datafication, digitalisation, and platformisation of the media ecosystem while (national) reform initiatives have not yet proven successful (KEK, 2018; Lobigs & Neuberger, 2018; Ofcom, 2021). This is what prompted the EU to urge a review of media concentration laws and measurements in its recently proposed European Media Freedom Act (European Commission, 2022a). Since creating concentration rules for the digital age is complicated, it is useful to revert to the normative foundations of traditional media concentration law, intersecting competition law, media law, and constitutional law. In this article, I propose a unifying theoretical framework for a novel "digital media concentration law" based on concentration rules' traditional normative foundations and incorporating relevant policy goals, concepts, and guidelines from multiple legal areas, including data protection and privacy law, consumer law, contract law, competition law, (tele)communication law, media law, platform regulation, and AI law (see Figure 1).

1.1. Methodology

A normative legal research method is used to investigate media concentration law and related policy responses. Because "normative choices have policy consequences" (Popiel, 2022, p. 33), understanding the underlying normative goals of legal tools (Cornils, 2020, p. 14; Ganter, 2022) is useful to assess long-term and institutional effects. Economic and competition-driven policies generally fail to effectively protect media pluralism, public values, and democracy (Baker, 2002, p. 30). Hence, to deal with media concentration, the normative foundations of the respective rules are best equipped to inform effective policy choices. In the words of Lin and Lewis (2022, p. 2), discussions about the digitalisation of the news media shall focus on what AI and technology *should* (rather than *could*) do for them and democracy. To that end, this article follows a normative and prescriptive approach. It discusses the normative foundations of European media concentration law, based on a (doctrinal) analysis of regulatory frameworks and documents, research reports and studies, and literature. Flowing from that, the normative findings are then woven together in a prescriptive narrative to propose potential remedies.

Of prime relevance for this analysis is the EU-wide "Media Pluralism and Diversity Online" study (Centre on Media Pluralism and Media Freedom [CMPF] et al., 2022), which maps and evaluates media concentration rules. The investigated elements of EU member states (including the UK) regulation are those aiming specifically at limiting media concentration and promoting media pluralism. I draw from this study's findings and national examples, in addition to other relevant literature, studies, and regulatory frameworks. These include findings from an EU study on digital advertising and publishers (Armitage et al., 2023) and other relevant national studies (KEK, 2018; Lobigs & Neuberger, 2018; Ofcom, 2021). To identify general approaches, several national examples are highlighted to support my arguments, though they are not compared in detail. Despite the complexity of this topic, due to space constraints, only a selection of examples is referred to. Furthermore, various gaps related to failures to assess concentration and opinion power, such as the inability of TV audiences to share measurements to genuinely represent news consumption and opinion formation, have previously been identified. Hence, my gap analysis as described in this article correlates with the significant power shifts in the media landscape at three levels: (a) the shifting impact over individual news consumption and exposure; (b) the shifting power dynamics inside automated, datafied, and platform-dependent newsrooms; and (c) the arrival of new players, particularly platforms and their systemic power and growing structural dependencies (Seipp et al., 2023). The three-level conceptualisation of opinion power further guides the analysis and, for each level, I outline normative goals, gaps, and potential remedies for the digital age.

2. Normative Foundations of European Media Concentration Law

The relationship between media concentration and economic, journalistic, and political power seems evident as journalistic power stems from the economic power of media companies, more specifically, the "capital owners of these companies" (Knoche, 2021, pp. 374–375). Those with journalistic power can "enforce information, opinion, legitimization, and ideology" that conforms with the interests and goals of those in power (Knoche, 2021, pp. 374–375). Those in the media with "economic, journalistic, and political power" can influence individual and public opinion formation and, hence, wield what I call "opinion power."

Media concentration law aims at ensuring the dispersal of "opinion power" by controlling and measuring the (economic) effects of media market concentration and through the promotion of public values, particularly media pluralism. As Helberger et al. (2017) stress, the public value(s) at stake depend on the context. Here, the relevant public values are the promotion of media pluralism, the safeguarding of equal opportunity to communicate and participate in the public sphere, of democratic power distribution, and of transparency (Baker, 2007; Karppinen, 2013; Schulz, 1998; Seipp et al., 2023). The dual goal of safeguarding competition and media pluralism encapsulates the intertwined nature of

concerns about economic sustainability and public values in the digital media ecosystem. Both aspects need to be addressed to achieve the normative goals of media concentration law: preventing concentrated power over public opinion and encouraging the wider distribution of power to participate in public discourse (Baker, 2002, 2007). To attain these goals, media law alone is insufficient; all policy fields relevant to the media and communication sphere are applicable (Figure 1). Hence, I recommend including the normative underpinnings and aspired public values of media concentration law as well as measures enabling fair competition in any efforts to govern the digital media ecosystem.

In previous years, there has been a strong push to revisit media concentration laws, as seen in the European Media Freedom Act and similar national initiatives (e.g., by Ofcom and KEK). As previously mentioned, this article focuses on Europe, where states have positive and negative obligations to protect free expression, media freedom, and media pluralism (Tambini, 2021). Individual and public freedom of expression safeguards, envisaged in Article 10 of the European Convention on Human Rights (Council of Europe, 1950) and Article 11 of the EU Charter (Charter of Fundamental Rights of the European Union, 2012), are needed to enable a free marketplace of ideas in which truth, self-government, and autonomy prevail. Media freedom is institutional in nature, and specific privileges and protections are indispensable due to their societal value in facilitating free speech and opinion formation (Tambini, 2021). Therefore, media freedom and pluralism are means to an end—truth, democracy, and individual autonomy—and not ends in themselves. In this respect, states are obliged to (proactively) guarantee a healthy media environment and prevent dominant opinion power and concentration as a democratic prerequisite.

The High-Level Group on Media Freedom and Pluralism (2013, pp. 15–16), convened by the European Commission, drafted one of the most comprehensive reports setting forth principles for "a free and pluralistic media to sustain democracy," highlighting the threats of concentration. Concentration jeopardises media freedom and pluralism and some form of global settlement for democracies to resolve doctrinal and constitutional differences is desired to address the issues coherently and globally (Tambini, 2021). Henceforth, I resort to and refer to the following list of concentration threats in the gap analysis that has previously guided reports worldwide (CMPF et al., 2022; Mendel et al., 2017):

- Threat 1: Excessive media ownership or advertising client influence;
- Threat 2: Commercial media ownership concentration;
- Threat 3: Changing business models and their consequences for the quality of journalism;
- Threat 4: Lack of media ownership transparency and opacity of funding sources;
- Threat 5: Potential conflicts of interest arising from journalists' closeness to business interests with implications for the political space.

The proposed remedies to address these threats are not exhaustive. Instead, broader realignments of future policy objectives are needed. The thinking here must go beyond existing rules and encompass policy fields not previously considered part of media concentration law. Although legal responses are necessary, as van Drunen and Fechner (2022) highlight, they are not the only way to achieve certain goals and establish norms. Professional ethical guidelines, internal organisational measures, and so on can contribute to tackling concentration threats, and thereby enabling media pluralism. Hence, I explore regulatory areas beyond the scope of existing media concentration law (Figure 1) and non-regulatory approaches to empower autonomous individuals and (news) institutions.

3. Analysis: Gaps and Promises for a Digital Media Concentration Law

This section presents the normative conceptual framework of shifting "opinion power" and explores how current media concentration laws are insufficient in addressing the increasing concentration threats in the digital media landscape at three levels: the individual citizen, institutional newsroom, and media ecosystem levels. At each level, I highlight potential remedies that can help fulfil the normative goals and protect public values and ensure fair competition.

3.1. Individual Citizen Level

3.1.1. Normative Conceptual Framework: Power Over News Consumption and Exposure

Individual news consumption and exposure are increasingly governed by algorithms controlled by platforms and informed by data collected on user behaviour, such as traffic to news websites, personal interests and beliefs, and location data (Diakopolous, 2019). Individual autonomy and freedom of choice are significant aspects of empowering users. Hence, the ability of profit-driven platforms to steer news exposure and control attention through algorithms, based on platform design choices (Viljoen et al., 2021, p. 2), may affect how autonomous and free choices in news consumption and opinion formation are. Digital media concentration rules at the individual level intend to limit interference over user choice and autonomy to protect free and independent opinion formation. Normatively, this goal stems from the principle of equal opportunity to communicate by enabling everyone to benefit from a "structurally possible, real and equal opportunity to actively or passively participate in the communication and public opinion formation process" (Schulz, 1998, p. 180).

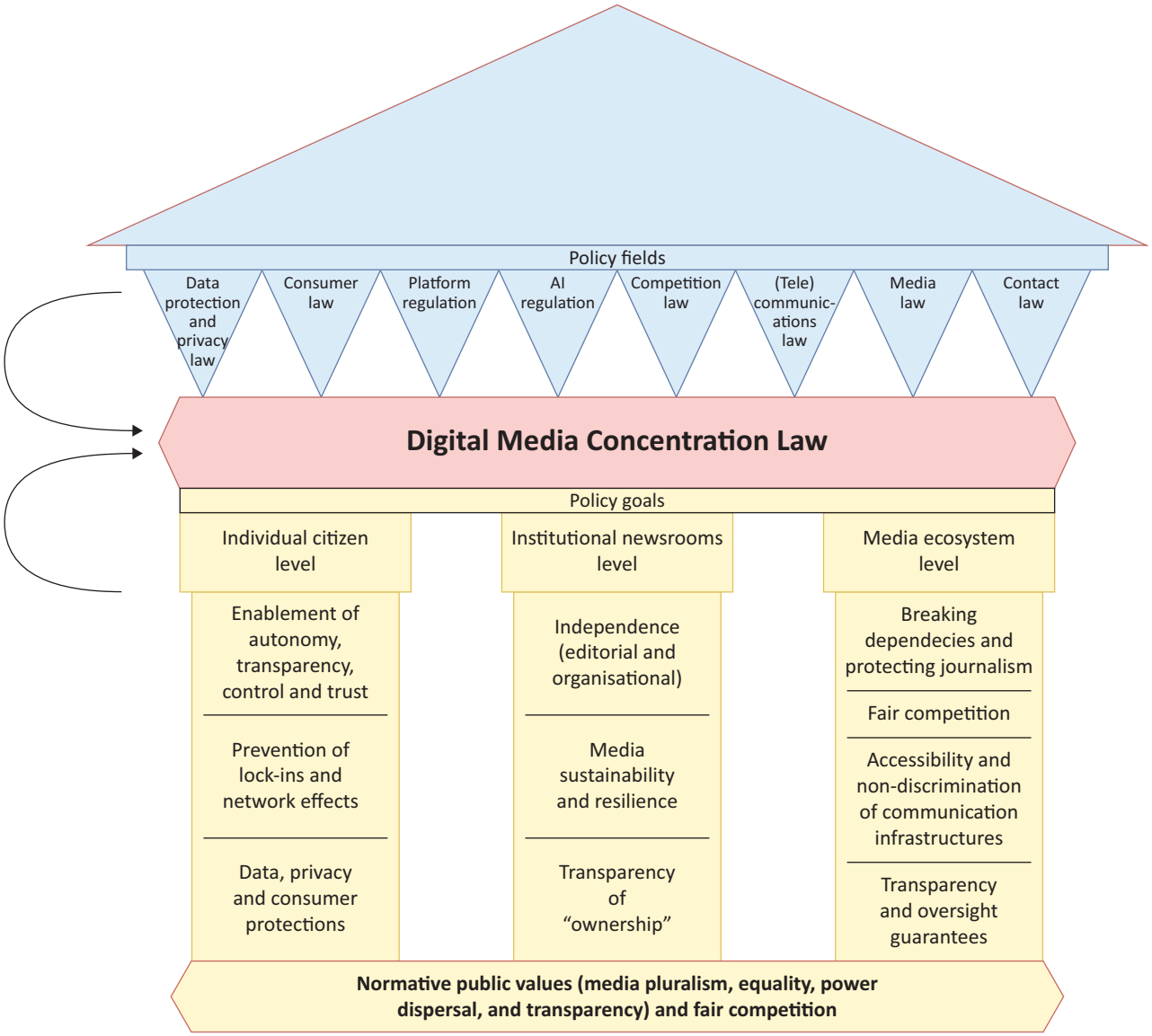


Figure 1. Illustration of a unifying theoretical framework for a "digital media concentration law" and the relevant policy goals and fields that feed into it.

3.1.2. Gaps

Several measures aim to minimise the influence of media owners or others on individual opinion formation (Threat 1). Media reach limitations focus on the supply, reach, and availability of one media source to the public. Audience shares measurements additionally aim to gauge "real consumption." Some EU member states have set thresholds for the allocation of broadcasting licences (CMPF et al., 2022, p. 206), while others limit capital rights shares and voting rights shares to confine the power of individual actors within a media company. Media reach limitations, however, neglect the influence of online media, new services, and platforms on opinion formation (CMPF et al., 2022, pp. 204–207). Similarly, audience share measures mostly target traditional media. Only six EU member states (e.g., Croatia, Italy, and Germany) address online media (CMPF et al., 2022, p. 208). However, here, online media refers to on-demand services offered by broadcasters, not online news services, let alone platforms. Some platforms, like Meta, focussed less on news and journalism. At the same time, the reliance on platforms, especially TikTok, Snapchat, and Instagram, amongst younger audiences keeps growing (Newman et al., 2022). Therefore, media reach constraints and audience share measurements focused on traditional media fail to capture the role of platforms in today's news consumption and exposure.

Furthermore, personalisation and recommendations are hugely relevant for individuals to navigate unprecedented amounts of information and news circulating in the digital public sphere. Changing business models and the effects on quality journalism (Threat 3) also affect consumption and exposure as does platforms' ability to algorithmically tailor news feeds, making "exposure diversity" gain importance. Intermediaries do not control

access to the medium as gatekeepers, instead they "control powerful transmission nodes and direct user flows by continuously engaging user attention" (Lobigs & Neuberger, 2018, p. 70, translation by the author; van der Vlist & Helmond, 2021). Controlling algorithmic infrastructures that manage user attention on platforms provides power over exposure and diversity, as online news consumption is not inherently pluralistic, despite the unprecedented availability and diversity of information sources (Napoli, 1999). The attention economy and changing audience behaviour online further render traditional audience share measurements obsolete. To gauge influence over opinion formation and set limits, one cannot ignore the far-reaching role of attention control and persuasive tools via algorithmic systems, as well as those new actors who have such direct and novel control over exposure. There is no easy way of doing so because by making effects part of the assessment, the threshold for application of media concentration law heightens. As Tambini (2021, p. 154) highlighted, "the key metric is not audience share on [the] national level, but data consolidation." Hence, it might be easier to identify control over choice architectures (including data and targeting algorithms) than the effects of selective exposure. The following subsection deals with potential solutions at the individual citizen level, focussing on empowering users by allowing true autonomy and agency, which is needed to meet the normative goal of sustaining free and independent opinion formation.

3.1.3. Policy Goals and Potential Remedies

3.1.3.1. Enablement of Autonomy, Transparency, Control, and Trust

Autonomous news users, enhanced user control, and trust can help counter the negative effects of changing business models (Threat 3). Transparency obligations have played a significant role in measures to increase trust. For instance, the proposed EU AI Act (European Commission, 2021) and the German "Medienstaatsvertrag" (Die Medienanstalten, §83, §93) both require that automatically generated content be labelled as such to enhance trust in automated tools. Yet, transparency obligations *alone* are not enough. Effective recommender transparency requires a certain level of AI literacy to comprehend the information and the technical interfaces used to control personalisation (Deuze & Beckett, 2022). Research shows that obligations to make transparent information about automated decision-making, like the Regulation of the European Parliament and of the Council of 27 April (2016; General Data Protection Regulation [GDPR]) requires or about the main parameters of platforms recommender systems (European Commission, 2022b, Article 27) are only a small (and slightly less relevant) portion of the information that is important to individuals' trust (van Drunen et al., 2022, p. 36). From a normative viewpoint, to protect free and independent opinion formation, merely making more information transparent is insufficient. Instead, enabling user choice and enhancing accountability may be considered as a relevant complementary avenue (van Drunen et al., 2019, 2022).

Connectedly, the European Media Freedom Act proposes a "right of customisation" (of audio-visual media offer; European Commission, 2022a, Article 19) and the Digital Services Act requires platforms that use recommenders to allow user choice, including enabling at least one option for recipients of the service to modify or influence those main parameters (European Commission, 2022b, Article 27). The latter is the first provision of its kind, demonstrating the importance of increasing not only transparency but also control to enable independence and autonomy. Previous research confirmed that control mechanisms over news recommendation algorithms are "extremely valued" by users (Harambam et al., 2019). Users need a certain level of trust in the quality of information to freely form opinions from a diverse pool of information and viewpoints. In fact, the media's ability to fulfil its role in society is predicated on citizens' ability to trust the media, while citizens cannot fulfil their role in the democratic process unless they can trust the media (van Drunen et al., 2022). While it seems indispensable to enhance transparency and user control, merely requiring "alternative options" (e.g., Digital Services Act) may not suffice if the design and adoption of such alternatives remain at the discretion of very large online platforms (Helberger, van Drunen, et al., 2021).

So far, media concentration rules do not fully encompass elements of transparency, control, and choice to empower users as "active agents." I envision those elements as being part of an extended revision of media concentration rules for the digital. We can observe that the first elements of this are already emerging in the EU framework (e.g., Digital Services Act). Some open questions remain, though, such as the extent to which the ability to turn off personalisation is sufficient in providing users with choice, how exactly it promotes diversity, as well as how autonomous are users' choices and how much control do they truly wield once they find themselves in technical and infrastructural lock-ins. The avoidance of lock-ins and network effects, as well as data, privacy, and consumer protections, are other elements to complete the puzzle for the individual citizen level.

3.1.3.2. Prevention of Lock-In and Network Effects

Preventing lock-ins helps to enable user autonomy, choice, and control, addressing the normative threat of concentration and influence of (commercial) power in the media (Threat 2). In Diakopolous' (2019, p. 183) words, "given their ability to influence attention, interaction, and communication, the choices made in the design of their interfaces and algorithms are anything but neutral," and end-user autonomy is curtailed if they depend on established infrastructures and platforms

(Napoli, 2015). Accordingly, measures to avoid lock-ins and prevent network effects seem indispensable to fulfil the normative goals and to truly enable competition. The Digital Markets Act also recognises that "core platform services have very strong network effects, the ability to connect many business users with many end users through the multi-sidedness of these services, a significant degree of dependence of both business users and end users, lock-in effects" (European Commission, 2022c, preamble), which reduce end users' choice in practice, affects fair competition, and threatens user rights. Preventing large platforms' consumer profiling and enhancing contestability and transparency seem to be relevant initial steps to counter lock-ins and avoid dependences.

3.1.3.3. Data, Privacy, and Consumer Protections

Platforms "collect a massive amount of personal data from consumers, who are not capable of making sufficiently voluntary and informed decisions about the collection and use of 'their' personal data…which endangers their informational self-determination and privacy" (Kerber & Specht-Riemenschneider, 2021, p. 4). Protective measures matter for media pluralism and freedom, as privacy concerns may have a chilling effect on free speech (Cohen, 2013). As I will elaborate below, the protections of data rights and privacy envisaged in legal instruments like the GDPR and protected under Article 8 of the European Convention on Human Rights (Council of Europe, 1950) and Articles 7 and 8 of the EU Charter (Charter of Fundamental Rights of the European Union, 2012), are, despite some flaws, significant initial steps towards constraining the collection, processing, and use of personal data.

Structural digital vulnerabilities are architectural, relational, and data-driven and can be triggered by asymmetric power relations (Helberger, Micklitz, et al., 2021, p. 145). News consumption triggers vulnerability because data is analysed to target audiences and to capture attention through personalised news exposure, with significant ramifications for individual rights, public opinion, and democracy. As news travels via platforms, and personalised news exposure without user control may allow platforms to manipulate opinion, digital architectures must be designed in a way to not exclude or disadvantage news consumers' free and autonomous choices. Currently, architectural "dark patterns," namely "user interface design choices that benefit an online service by coercing, steering, or deceiving users into making unintended and potentially harmful decisions" may interfere with free and independent opinion formation (Helberger, Micklitz, et al., 2021, p. 6).

The Digital Markets Act (European Commission, 2022c, Article 5) imposes limits on "gatekeepers," such as the prohibition to combine and cross-use personal data with third-party personal data, to limit data-driven competitive advantages. This is a welcome approach to limit private data power; however, under the GDPR, users can simply provide consent. Therefore, allowing end-user consent is not enough to make data-driven marketplaces competitive (Graef, 2021) nor to protect consumers, as architectures could exploit user vulnerabilities and nudge towards consenting (Helberger, Micklitz, et al., 2021). Furthermore, Helberger et al. (2017) argue that realising public values in platform-based public activities requires "cooperative responsibility." Consequently, rules ought to not only allocate institutional accountability on and data collection constraints on platforms but also concentrate on architectural design decisions, such as the configuration of recommendation or sorting algorithms (Helberger et al., 2017, p. 2).

In sum, existing legal tools cannot effectively limit power or measure concentration because they are limited to traditional media and ignore how attention control influences news consumption and exposure. To deal with concentration and opinion power at the individual citizen level, we need to look beyond traditional, media-centred tools and expand the toolbox with more audience- and user-centric remedies.

3.2. Institutional Newsroom Level

3.2.1. Normative Conceptual Framework: Power Over Editorial Decision-Making and Agendas

At the institutional newsroom level, growing automation, datafication, digitalisation, and platformisation of newsrooms have implications for the media's normative role and editorial independence (Seipp et al., 2023; van Drunen & Fechner, 2022). Traditionally, opinion power is based on an editor's ability to curate and set an agenda in combination with the means to reach an audience (Jarren, 2018). As control over audience connection shifts from news media to platforms (Nielsen & Ganter, 2022; Simon, 2022), opinion power and control inside newsrooms also shift. Henceforth, policy goals at this level aim at protecting editorial independence, enabling media resilience and sustainability, and imposing transparency obligations. This is informed by the normative objective to control those with power over the media and public opinion. Potential remedies should aim at promoting media pluralism, and specifically, measures to control the structural power relations within the media ecosystem, including the "struggles over the framing and agenda of public discussions, and political and corporate decisions about the architecture and ownership of media systems" (Karppinen, 2013, p. 80).

3.2.2. Gaps

Ownership restrictions assume that "many owners" equal "plurality," which has traditionally been justified by "the normative assumption…that emphasises the importance of diverse ownership to guarantee equal distribution of communicative power" (Just, 2022, p. 188).

Traditional ownership constraints, however, ignore new sources of power and control in the media. In fact, current rules focus on traditional media actors, whereas in the digital age, new and extremely powerful actors have entered the stage. While limitations on media ownership remain important, its definition must be rethought for the digital age as objects and characteristics of control change. Baker (2002, p. 57) observed decades ago that "the fundamental issue is control, for which ownership is a loose but poor proxy." The growing importance of control over data, skills, and knowledge is eroding traditional understandings of what it means to "own" competitive resources. The extent to which data are subject to traditional ownership and property rights remains a point of contention in Europe. Regardless, control over resources such as data, skills, and knowledge are currently not part of traditional media concentration measurements.

Indeed, existing constraints fail to recognise how changing business models and sources of power have challenged power dynamics, and thus fail to effectively address Threats 3, 4, and 5. According to Ferrer-Conill and Tandoc (2018, p. 448), data analytics and algorithmic tools have reshaped the relationship with the audience, allowing for far more fine-grained control over the flow of audience attention by new actors, particularly platforms. Journalists' and editors' reliance on audience metrics, often defined and controlled by external platforms, is another example of power shifting within newsrooms (Dodds et al., 2023). As a result, newsrooms are becoming increasingly dependent on platforms that collect and analyse data to develop and build new tools (Simon, 2022), leading editors in today's newsrooms to be constrained by and reliant on the technological affordances of the tools they use (Ferrer-Conill & Tandoc, 2018, p. 448). Latest advancements in generative AI have highlighted this issue even further. Although the consequences of generative AI such as ChatGPT on news organisations are far from foreseeable, it is apparent that few individuals and corporations control the resources and talents that underpin these AI capabilities (Murgia, 2023), with potentially severe long-term institutional consequences. And although there may be more competition in the "AI race," smaller actors will still depend on large companies for cloud access, computing infrastructure, and data, as well as financially. Hence, in the long-term, these developments may lead to further concentrating on the digital media environment and potentially exacerbating dependence.

In addition to being gateways to news and providing technologies and data, platforms remain relevant for funding skills, expertise, and research projects (Diakopolous, 2019, p. 179; Simon, 2022, p. 4). Although previous research has not shown any *direct* interferences with journalists' editorial independence through funding (Fanta & Dachwitz, 2020), limitations on foreign ownership do not capture the money flowing from foreign private companies into (European) news organisations, which might pose potential for *indirect* influence. More concretely, ownership transparency rules that focus on media ownership disclosure and reporting and restrictions on foreign ownership of media companies (Threat 4) are relevant for traditional media markets but outdated for the digital age. In most EU member states, a foreign shareholder's maximum stake in a country may not exceed 49% (CMPF et al., 2022, p. 219). The goal here is to limit the influence of a foreign owner (natural or legal person) on the European media market and ensure that the majority owner will never be a non-EU or EEA company, to protect from undue influence. It is evident that traditional ownership limits generally disregard the power dimension of platforms, which act not only as gatekeepers but also as political players (Helberger, 2020). Therefore, transparency obligations regarding the amounts and beneficiaries of funding (especially from Google and Meta) may need to be included in the current toolbox (CMPF et al., 20222, pp. 401–402; Papaevangelou, 2023). This is significant as platforms always pursue their own political and commercial agendas, and even Meta withdrawing from funding news could affect news organisations.

As a result, power and control inside newsrooms are increasingly characterised by *control over things that cannot be owned* in the traditional sense. This is problematic because those in control (often) fall outside the scope of ownership limitations despite wielding significant technological, commercial, and political power.

3.2.3. Policy Goals and Potential Remedies

The media serve a public and democratic purpose by upholding professional and journalistic ideals such as independence and autonomy (McCombs & Shaw, 1972). And although states have positive obligations towards media freedom, any regulatory remedies must be proportionate in protecting editorial independence and promoting pluralism, while ensuring that states refrain from interfering too heavily with journalistic freedoms.

3.2.3.1. Independence (Editorial and Organisational)

Media concentration rules set objectives to safeguard media from political interference (Threat 5) and protect independence. In automated newsrooms, the specific rules may need to be rethought, as they have started to erode borders between editorial teams, business departments, and third-party technology companies that often either fund and support R&D or externally develop and build technology is the norm (van Drunen & Fechner, 2022, p. 6). In addition to regulatory measures, van Drunen and Fechner (2022, p. 22) argue that "internal organisational matters have increasing relevance in ensuring editorial independence in the context of automation," also to prevent external influences. Despite some (limited) internal strategies and (risk) assessments of whether to use a certain tool, platforms are often the providers of new technology, skills, and knowledge

and offer these based on their own terms and conditions (Simon, 2022). Thus, alongside non-regulatory internal organisational measures, balanced contractual relationships between platforms and news organisations through contractual obligations could be established while respecting contractual freedoms (Helberger, 2020). This could enable more direct channels of communication, more balanced negotiations, and fairer relationships between news organisations and platforms. Another approach could focus on procurement rules and creating better conditions for news organisations to develop and build their own in-house technology or in cooperation with academia or public interest research institutions.

3.2.3.2. Media Sustainability and Resilience

The role of platforms calls for remedies that address the principle of control better than ownership limitations and overcome the lack of media ownership transparency and opacity of funding sources (Threat 4). To address the deepening dependencies on platforms for technology, data, tools, staff, knowledge, services, and networks, a digital media concentration law needs to empower the media and enable sustainability and resilience. One way of doing so could be to look at media privileges (e.g., funding, distribution, tax breaks, and subsidies) and (long-term) financial support in the form of public funding and potentially increased state aid. Tambini (2021, p. 48) states that "fiscal treatment of the press is…one of the key means through which states can create the conditions for the sustainability of journalism." Accordingly, Pickard (2020) and others have called for public funds and digital or "public media" taxes on platforms' earnings to fund public interest journalism (CMPF et al., 2022, p. 400). This requires a delicate balance to be struck between regulatory action and public support while avoiding public interference with journalistic freedom.

3.2.3.3. Transparency of "Ownership"

Lastly, a redefinition of "ownership" limitations and information transparency obligations to measure concentration is needed. Ownership concerns two kinds of control over the media: allocation control (company policy and strategy and controlling mergers, acquisitions, or cutbacks) and operational control (internal distribution of resources, setting editorial strategies, delegating editorial control; Sjøvaag & Ohlsson, 2019). The European Media Freedom Act proposes to mitigate "the risk of undue public and private interference in editorial freedom" (European Commission, 2022a) To do this, both public and private sources of financing for technology, skills (such as fellowships, research projects, and development), and services need to be made transparent. As the resources of control change, so must the approach towards "ownership." That means that, to limit "control," not only those who "own" media but also those who "control" the sources of power need to be accounted for.

3.3. Media Ecosystem Level

3.3.1. Normative Conceptual Framework: Systemic Power of and Structural Dependencies on Platforms

Finally, platforms wield systemic opinion power which creates structural dependencies and influence over other democratic players (Helberger, 2020, p. 846). Platforms have the (economic) monopolistic and political power to influence policymaking (Helberger, 2020; Seipp et al., 2023), while existing transparency and oversight mechanisms, merger controls, and competition laws seem to fall short. Hence, the overarching normative goals at the media ecosystem level focus on structural media pluralism and on a democratic and balanced media ecosystem to enable equality and diversity. Since journalistic power stems from the economic power of media companies (Knoche, 2021), measures to promote fair competition in the media market are indispensable in enabling media pluralism and power dispersal.

3.3.2. Gaps

To address Threats 1 and 2, media ownership limits and merger controls focus primarily on horizontal concentration and traditional media and "rarely take into consideration the take-up of new services and platforms, resulting in a scarcity of set limits for digital news media" (CMPF et al., 2022, p. 210). Rules at the media ecosystem level focus specifically on limiting concentrations of power and anti-competitive behaviour, addressing Threats 1 and 2. They are sector-specific rules and are contained, for instance, in national telecommunications laws and include duties to interconnect, provide number portability, or transparency obligations (Just, 2022). Those rules have been criticised as ineffective, causing general competition rules to be applied to the media. Only Germany and Austria explicitly recognise the media's democratic and opinion power for general competition and merger controls (CMPF et al., 2022, p. 214; Just, 2022). Most ownership limitations constitute purely competition-focused rules and are evaluated by national competition authorities (NCAs; e.g., Slovakia, Estonia, and Luxemburg; CMPF et al., 2022, pp. 209, 246). Although traditional media merger assessments are conducted by both, national media regulatory authorities (NRAs) and NCAs, the former acts merely in a non-binding advisory role while the latter makes the final decision (CMPF et al., 2022, pp. 214–218). NCAs can evaluate mergers' economic effects but not political power, which derives from opinion power (CMPF et al., 2022, p. 39). Media merger controls also typically focus on horizontal mergers and often lack clear restrictions on vertical mergers, where an individual or a company controls key elements of production, distribution,

and related activities like advertising (CMPF et al., 2022). This is a major limitation in effectively preventing digital media concentration. Broadly speaking, besides ineffective merger controls, no concrete rules can be found that address the imbalanced and asymmetric negotiation and market relations between platforms and news organisations effectively. This triggers the need to control power allocations of vertical integration and convergence in the media market better. Policymakers also need to regulate for enabling fairer conditions in negotiations, such as putting procedural safeguards in place to enforce "good faith" negotiations.

3.3.3. Policy Goals and Potential Remedies

To address the concerns described above in light of platforms' vast systemic power, the following policy goals should guide potential remedies: break structural dependencies, create counterpowers, enable fairer competition, combat negotiating power imbalances, and enhance collaboration mechanisms amongst regulatory bodies.

3.3.3.1. Breaking Dependencies and Empowering Journalism

To build an environment conducive to media pluralism and resilience, one priority is to ensure the survival of journalism, particularly independent and local forms (Pickard, 2020). Empowering local media has two main purposes: to reduce the dependence on platforms that provide resources (e.g., technical, financial, talent) and to counterbalance large media outlets (e.g., Springer), which is pivotal to avoid the emergence of media concentration. Journalism is at a competitive disadvantage for attention and advertising because of platform power, leading to an asymmetric power dynamic and increased reliance on platforms for services, data, and revenue (Nielsen & Ganter, 2022), sparking increased concentration trends. Hence, a pluralistic media environment requires special protections for independent, high-quality local journalism.

Collective bargaining agreements and other methods to strengthen news organisations' rights against platforms are an increasingly popular remedy. Regulatory initiatives in Australia, Canada, and the UK address this competition and market power imbalance. However, Australia's News Media Bargaining Code, for instance, has been criticised for not boosting small outlets' negotiation capacity sufficiently (Bossio et al., 2022). Hence, small outlets do still depend, to a degree, on collective bargaining agreements facilitated by external parties, like public interest foundations, to pursue negotiating objectives (Minderoo Foundation, 2022). Creating the conditions for more collective action could be important to further boost local news media's power in an asymmetric relationship. That may require an antitrust exception to allow publishers to negotiate with platforms collectively and share information about the progress of negotiations (Bossio et al., 2022, p. 8).

Other remedies could concentrate on defining a balanced contractual relationship between platforms and news media while respecting contractual freedoms (Helberger, 2020). Not only users but also the news media find themselves in situations where platforms have contractual control over their infrastructure and services (Simon, 2022, p. 12). Besides regulatory options, internal rules on procurement and instructions on implementation processes of new technologies based on public values and standard-setting guidelines (e.g., Council of Europe Expert Committee on Resilience in Journalism) could offer contractual protections (Council of Europe, n.d.).

3.3.3.2. Fair Competition

Current rules are limited mainly to horizontal merger controls and traditional media, thereby neglecting to account for platforms' increasing power over production, distribution, and infrastructure. Therefore, new rules on fair competition must better capture threats of vertical integrations to prevent concentration (Threats 1 and 2; CMPF et al., 2022, p. 38). Given the increasing vertical convergence of media sectors and businesses and the role of platforms as multi-sided markets (as gatekeepers, technology service providers, business partners, investors, and political players), it is vital to address media attention markets. That means explicitly considering power over data, technology, and infrastructure.

Especially data-driven advantages make newsrooms more dependent on platforms and the data they collect (CMPF et al., 2022, p. 39). At this moment, it seems unrealistic for a (large) news organisation's R&D team (nor researchers) to develop and build their own AI models to become more competitive since data and computing power remain heavily concentrated under the control of a select few platform companies. Rules on fair data access for all actors competing in the media ecosystem are needed to enable fair competition. The Digital Markets Act and the Data Act aim to set new laws on who can use and access data in the EU across all economic sectors and limit strategic advantages from data power and lock-ins (European Commission, 2022c; European Commission, 2022d). It remains to be seen whether such provisions will make news organisations more competitive and less dependent on the data market. In addition and complementary to the Data Governance Act (European Commission, 2020), the EU (European Commission, 2022e) has also announced a "European Data Space" to support media companies in sharing data and developing innovative solutions, which would better equip the media to scale up and become more competitive. It aims to support EU media stakeholders in handling data-driven business models and pool together sets of content, data, and metadata to produce new products and formats targeting expanded

audiences. Such initiatives are vital elements for fairness, reduced (structural) dependencies, and power asymmetries from data monopolisation.

3.3.3.3. Accessibility and Non-Discrimination of Communication Infrastructures

Because platforms and media are all part of the larger digital communication infrastructure, it may be useful to take an infrastructural approach, seeking to learn lessons from (tele)communications law for avoiding "infrastructural capture" of the media, where a "scrutinising body is incapable of operating sustainably without the physical or digital resources and service" of a business (Nechushtai, 2017, p. 1043). Instagram and YouTube provide important platforms for civic engagement, social participation, and public opinion formation. Power is also increasingly concentrated in cloud infrastructures and data centres (e.g., Google Cloud, Amazon Web Services), smartphones, digital assistants, and wearables (iOS/Apple, Android/Google, Alexa/Amazon; Busch, 2021). For digital news media markets, such dependence poses severe risks since media are supposed to scrutinise the power of these platforms, which control the infrastructures needed to connect with audiences and gather, produce, and disseminate news (Nechushtai, 2017).

In line with the EU's electronic communications policy, regulating platforms as infrastructures for general-interest services offers a source of inspiration for improving competition, driving innovation, and boosting consumer rights. Indeed, a digital media concentration law could incorporate several new legal areas, as the focus of current discussions on the market and opinion power of digital platforms may be too narrow to fully reflect the deeper sources of platform power. Busch (2021, p. 5) suggests a "platform infrastructure law" to address platforms as societal infrastructures and key actors in the sphere of services of general interest. Hence, a digital media concentration law could include stricter rules to guarantee infrastructural accessibility and non-discrimination, as well as a right to access "data-collection-free" and "non-personalised" digital services.

3.3.3.4. Transparency and Oversight Guarantees

NRAs, NCAs, and data protection authorities should increase efforts to cooperate and share data and expertise in assessing and preventing concentrations. Therefore, rules may be needed to improve cooperation between authorities and enable joint decisions, such as in media mergers. In spite of the fact that "the NCA should always have the power to block a merger on the basis of its competition concerns (including consumer choice)," the NRA should have the authority to do so based on its commitment to media pluralism (CMPF et al., 2022, p. 390). Collaborations could improve the exchange of data among authorities and create joint committees for institutional cooperation. Because economic and media plurality aims are linked, media mergers need to be "subject to the double, sometimes coordinated, but ultimately independent filter of two authorities" (CMPF et al., 2022, p. 390).

Furthermore, NRAs need enhanced access to data and information for monitoring and transparency purposes to address Threat 4. Measuring concentration and opinion power in digital media markets is extremely difficult. To do so effectively, other parameters than (TV) audience share measures need to be assessed. EU member states lack effective tools to measure cross-media concentration (CMPF et al., 2022, p. 244), despite some previous proposals (e.g., KEK, 2018). We need platform-based, dynamic media metrics, including metrics for algorithmic exposure. As proposed, centralised data-gathering frameworks to measure pluralism could be a starting point to assess exposure diversity in online news consumption (CMPF et al., 2022, p. 376).

Finally, law and policymakers must be wary of platforms' political and lobby power and their influences over democracy. The regulatory process of the News Media Bargaining Code in Australia and the emerging pressures from Meta on the Canadian government demonstrate platforms' political power, which in turn presents ineffective democratic protections (Roth, 2022). In Europe, where ambitious regulations are underway to curb platform power and ensure fair competition, the largest platforms aggressively push their own agendas. Big tech's lobbying clout in Brussels is not new, but how they aim to conceal their political influence is (Goujard, 2022). This calls for enhanced transparency reporting obligations. Media concentration controls should acknowledge the broader political economy in which platforms operate and push agendas. Enhanced lobbying controls and transparency reporting obligations are significant first remedies.

4. Conclusion

In this article, I demonstrated that to account for growing opinion power and concentration in digital media markets, it is not enough to simply update existing media concentration laws. Seeing the complex dynamics between platforms, media organisations, and users, a new, far more holistic approach towards dealing with media concentration in the digital age is needed. This is an approach that considers the effects of the power of platforms to influence and control opinion at the individual citizen, the institutional newsroom, and the ecosystem levels.

More specifically, I propose a greater focus on the role of users and enabling autonomous choices. This means that a new media concentration law will also have to incorporate elements of data protection, privacy, and consumer law. Finally, avoidance of lock-ins and network effects must be addressed by developing fairer data access rules that consider the characteristics of "attention markets" and the respective effects for individuals.

At the institutional newsroom level, I particularly call for measures to safeguard editorial independence and empower a resilient media. To do so, policymakers need to focus on the new sources of control, rather than solely on ownership limitations, to fulfil the normative goals of media concentration laws for the digital age. The sources of control in the digital, notably over data, technologies, funding, expertise, and knowledge, play a decisive role but are not "owned" in the traditional sense. Because of the necessity to strike a careful balance between public regulatory interference and journalistic freedom and independence, non-regulatory approaches, in particular, are significant in this context. Internal organisational measures, (procurement) rules, guidelines for implementing new technologies, and fair data access and sharing conditions could all be valuable avenues.

At the media ecosystem level, remedies should focus on platforms' systemic opinion power and structural dependencies. More specifically, collective bargaining agreements, protecting local journalism, and ensuring contractual fairness between news media and platforms are ways to balance and challenge dominant power. Additionally, NRAs, NCAs, and data protection authorities need to build better cooperation mechanisms, such as for media mergers. Further, due to platforms' indispensable infrastructural power, platform infrastructures can be seen as "utilities for democracy," which need to be considered in any potential remedies. Lastly, the political power of platforms and the challenges they pose to democracy must be understood better and addressed through enhanced transparency reporting obligations.

States have positive obligations to create a media system capable of sustaining democracy. Thus, I argue that the normative goals at each level should guide choices in each policy field; only then can a digital media ecosystem based on public values be created. I have shown that elements of a digital media concentration law can be found in different policy fields (see Figure 1) as well as scattered across the new emerging regulatory framework from Brussels. Despite the complexity of the topic and the need for additional research, what this article contributes is a unifying theoretical framework that anchors these individual elements as part of a more comprehensive reform of the rules on addressing media concentration with the goals to promote pluralism, equality, and democracy in digital media markets.

Acknowledgments

Thank you to Prof Natali Helberger, Prof Claes de Vreese, and Dr Jef Ausloos for their helpful feedback and comments. Also, thank you to Prof Dwayne Winseck for his comments, and to Dr Paddy Leerssen for the native language proofread. Last, I acknowledge that the article draws from the mapping of national rules as part of the EU-wide study on "Media Pluralism and Diversity Online." Hence, I want to also acknowledge the valuable work done by the (co-)authors of the study.

Conflict of Interests

The author declares no conflict of interests.

References

Armitage, C., Botton, N., Dejeu-Castang, L., & Lemoine, L. (2023). *Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers*. European Commission. https://op.europa.eu/en/publication-detail/-/publication/8b950a43-a141-11ed-b508-01aa75ed71a1

Baker, C. E. (2002). *Media concentration: Giving up on democracy*. SSRN. http://dx.doi.org/10.2139/ssrn.347342

Baker, C. E. (2007). *Media concentration and democracy: Why ownership matters*. Cambridge University Press.

Bossio, D., Flew, T., Meese, J., Leaver, T., & Barnet, B. (2022). Australia's news media bargaining code and the global turn towards platform regulation. *Policy & Internet*, *14*(1), 136–150. https://doi.org/10.1002/poi3.284

Busch, C. (2021). *Regulation of digital platforms as infrastructures for services of general interest* (WISO Diskurs Report No. 09/2021). Friedrich-Eber-Stiftung. https://library.fes.de/pdf-files/wiso/17836.pdf

Centre on Media Pluralism and Media Freedom, Centre for Information Technology and Intellectual Property of KU Leuven, European Commission's Directorate-General for Communications Networks, Content and Technology, Institute for Information Law of the University of Amsterdam, & Vrije Universiteit Brussels. (2022). *Study on media plurality and diversity online*. Publications Office of the European Union. https://data.europa.eu/doi/10.2759/529019

Charter of Fundamental Rights of the European Union. (2012). *Official Journal of the European Union*, OJ C 326. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT

Cohen, J. E. (2013). What privacy is for. *Harvard Law Review*, *126*(7), 1904–1933. https://www.jstor.org/stable/23415061

Cornils, M. (2020). *Designing platform governance: A normative perspective on needs, strategies, and tools to regulate intermediaries*. Algorithm Watch. https://algorithmwatch.org/de/wp-content/uploads/2020/05/Governing-Platforms-legal-study-Cornils-May-2020-AlgorithmWatch.pdf

Council of Europe. (n.d.). *MSI-RES Committee of Experts on Increasing Resilience of Media*. https://www.coe.int/en/web/freedom-expression/msi-res

Council of Europe. (1950). European convention on human rights as amended by protocols Nos. 11, 14 and 15. https://www.refworld.org/docid/3ae6b3b04.html

Deuze, M., & Beckett, C. (2022). Imagination, algorithms and news: Developing AI literacy for journalism.

Digital Journalism, *10*(10), 1913–1918. https://doi.org/10.1080/21670811.2022.2119152

Diakopolous, N. (2019). *Automating the news: How algorithms are rewriting the news*. Harvard University Press.

Die Medienanstalten. (2020). *Interstate media treaty*. https://www.die-medienanstalten.de/fileadmin/user_upload/Rechtsgrundlagen/Gesetze_Staatsvertraege/Interstate_Media_Treaty_en.pdf

Dodds, T., de Vreese, C., Helberger, N., Resendez, V., & Seipp, T. (2023). Popularity-driven metrics: Audience analytics and shifting opinion power to digital platforms. *Journalism Studies*, *24*(3), 403–421. https://doi.org/10.1080/1461670x.2023.2167104

European Commission. (2020). *Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act)* (COM/2020/767 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52020PC0767

European Commission. (2021). *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (COM/2021/206 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52021PC0206

European Commission. (2022a). *Proposal for a regulation of the European Parliament and of the Council establishing a common framework for media services in the internal market (European Media Freedom Act) and amending Directive 2010/13/EU (COM/2022/457 final)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0457

European Commission. (2022b). 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) (Text with EEA relevance). *Official Journal of the European Union*, L 277. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32022R2065

European Commission. (2022c). Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance). *Official Journal of the European Union*, L 265. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN

European Commission. (2022d). *Proposal for a regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)* (COM/2022/68 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2022:68:FIN

European Commission. (2022e). *A European strategy for data*. https://digital-strategy.ec.europa.eu/en/policies/strategy-data

Fanta, A., & Dachwitz, I. (2020). *Google, the media patron*. Otto Brenner Foundation. https://www.otto-brenner-stiftung.de/fileadmin/user_data/stiftung/02_Wissenschaftsportal/03_Publikationen/AH103_Google_EN.pdf

Ferrer-Conill, R., & Tandoc, E. C., Jr. (2018). The audience-oriented editor. *Digital Journalism*, *6*(4), 436–453. https://doi.org/10.1080/21670811.2018.1440972

Ganter, S. A. (2022). Governance of news aggregators' practices across five emblematic cases: Policy regimes between normative acceptance and resistance. *The Information Society*, *38*(4), 290–306. https://doi.org/10.1080/01972243.2022.2076180

Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, *22*(6), 854–871. https://doi.org/10.1080/1369118x.2019.1573914

Goujard, C. (2022, October 14). Big tech accused of shady lobbying in EU Parliament. *Politico*. https://www.politico.eu/article/big-tech-companies-face-potential-eu-lobbying-ban

Graef, I. (2021, September 2). Why end-user consent cannot keep markets contestable: A suggestion for strengthening the limits on personal data combination in the proposed Digital Markets Act. *VerfBlog*. https://verfassungsblog.de/power-dsa-dma-08

Harambam, J., Bountouridis, D., Makhortykh, M., & van Hoboken, J. (2019). Designing for the better by taking users into account: A qualitative evaluation of user control mechanisms in (news) recommender systems. In T. Bogers & A. Said (Eds.), *RecSys '19: Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 69–77). Association for Computing Machinery. https://doi.org/10.1145/3298689.3347014

Helberger, N. (2020). The political power of platforms: How current attempts to regulate misinformation amplify opinion power. *Digital Journalism*, *8*(6), 842–854. https://doi.org/10.1080/21670811.2020.1773888

Helberger, N., Micklitz, H. W., Sax, M., & Strycharz, J. (2021). Surveillance, consent, and the vulnerable consumer. Regaining citizen agency in the information economy. In N. Helberger, O. Lynskey, H.-W. Micklitz, P. Rott, M. Sax, & J. Strycharz (Eds.), *EU consumer protection 2.0: Structural asymmetries in digital consumer markets* (pp. 1–91). BEUC.

Helberger, N., Pierson, J., & Poell, T. (2017). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, *34*(1), 1–14. https://doi.org/10.1080/01972243.2017.1391913

Helberger, N., van Drunen, M., Vrijenhoek, S., & Möller, J. (2021). Regulation of news recommenders in the Digital Services Act: Empowering David against the very large online Goliath. *Internet Policy Review*, *26*. https://policyreview.info/articles/news/regulation-news-recommenders-digital-services-act-empowering-david-against-very-large

High-Level Group on Media Freedom and Pluralism. (2013). *A free and pluralistic media to sustain democracy*. https://ec.europa.eu/information_society/media_taskforce/doc/pluralism/hlg/hlg_final_report.pdf

Jarren, O. (2018). Normbildende Macht. Intermediäre als gesellschaftliche Herausforderung [Norm-building power. Intermediaries as a social challenge]. *Epd Medien*, *24*, 35–39.

Just, N. (2022). Media concentration: Problem and regulation. In J. Krone & T. Pellegrini (Eds.), *Handbook of media and communication economics: A European perspective* (pp. 1–14). Springer. https://doi.org/10.1007/978-3-658-34048-3_70-2

Karppinen, K. (2013). *Rethinking media pluralism*. Fordham University Press.

KEK. (2018). *Sicherung der Meinungsvielfalt im digitalen Zeitalter: Bericht der Kommission zur Ermittlung der Konzentration im Medienbereich (KEK) über die Entwicklung der Konzentration und ueber Maßnahmen zur Sicherung der Meinungsvielfalt im privaten Rundfunk* [Securing opinion diversity in the digital age: Report of the Commission on Concentration in the Media (KEK) on the development of concentration and on measures to ensure diversity of opinion in private broadcasting]. https://www.kek-online.de/publikationen/medienkonzentrationsberichte?tx_news_pi1%5Bnews%5D=4688&cHash=86fca16a250b6250b0018e9cd791c262

Kerber, W., & Specht-Riemenschneider, L. (2021). *Synergies between data protection law and competition law*. Federation of German Consumer Organisations. https://www.vzbv.de/sites/default/files/2021-11/21-11-10_Kerber_Specht-Riemenschneider_Study_Synergies_Betwen_Data%20protection_and_Competition_Law.pdf

Knoche, M. (2021). Media concentration: A critical political economy perspective. *tripleC*, *19*(2), 371–391. https://doi.org/10.31269/triplec.v19i2.1298

Lin, B., & Lewis, S. C. (2022). The one thing journalistic AI just might do for democracy. *Digital Journalism*, *10*(10), 1627–1649. https://doi.org/10.1080/21670811.2022.2084131

Lobigs, F., & Neuberger, C. (2018). Meinungsmacht im Internet und die Digitalstrategie von Medienunternehmen [Opinion power on the internet and the digital strategy of media companies]. *die Medienanstalten. VISTAS*, *51*, 12–271. https://www.kek-online.de/fileadmin/user_upload/KEK/Publikationen/Gutachten/Meinungsmacht_im_Internet_ALM51_web_2018.pdf

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *The Public Opinion Quarterly*, *36*(2), 176–187. https://www.jstor.org/stable/2747787

Mendel, T., García Castillejo, A., & Gómez, G. (2017). *Concentration of media ownership and freedom of expression: Global standards and implications for the Americas*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000248091

Minderoo Foundation. (2022, May 31). *Minderoo Foundation and Google sign agreement for Google to support 24 digital publishers* [Press Release]. https://www.minderoo.org/frontier-technology/news/minderoo-foundation-and-google-sign-agreement-for-google-to-support-24-digital-publishers

Murgia, M. (2023, March 22). Risk of industrial capture looms over AI revolution. *Financial Times*. https://www-ft-com.ezp.lib.cam.ac.uk/content/e9ebfb8d-428d-4802-8b27-a69314c421ce

Napoli, P. M. (1999). Deconstructing the diversity principle. *Journal of Communication*, *49*(4), 7–34. https://doi.org/10.1111/j.1460-2466.1999.tb02815.x

Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy*, *39*(9), 751–760. https://doi.org/10.1016/j.telpol.2014.12.003

Nechushtai, E. (2017). Could digital platforms capture the media through infrastructure? *Journalism*, *19*(8), 1043–1058. https://doi.org/10.1177/1464884917725163

Neuberger, C. (2018). Meinungsmacht im Internet aus Kommunikationswissenschaftlicher Perspektive [Opinion power on the internet from a communication studies perspective]. *UFITA*, *82*(1), 53–68.

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2022). *Digital news report 2022*. Reuters Institute for the Study of Journalism.

Nielsen, R., & Ganter, S. A. (2022). *The power of platforms*. Oxford University Press.

Ofcom. (2021). *The future of media plurality in the UK: Ofcom's report to the Secretary of State on the media ownership rules and our next steps on media plurality*. https://www.ofcom.org.uk/__data/assets/pdf_file/0019/228124/statement-future-of-media-plurality.pdf

Papaevangelou, C. (2023). Funding intermediaries: Google and Facebook's strategy to capture journalism. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2155206

Pickard, V. (2020). *Democracy without journalism? Confronting the misinformation society*. Oxford University Press.

Popiel, P. (2022). Regulating datafication and platformization: Policy silos and tradeoffs in international platform inquiries. *Policy & Internet*, *14*(1), 28–46. https://doi.org/10.1002/poi3.283

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). (2016). *Official Journal of the European Union*, L 119. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Roth, E. (2022, October 23). Facebook warns it could block news in Canada over proposed legislation. *The Verge*. https://www.theverge.com/2022/10/23/23418928/facebook-warns-block-news-canada-legislation-meta

Schulz, W. (1998). *Gewährleistung Kommunikativer Chancengleichheit Als Freiheitsverwirklichung* [Guaranteeing equal opportuntiy to communicate as the realisation of freedom]. Nomos.

Seipp, T. J., Helberger, N., de Vreese, C., & Ausloos, J. (2023). Dealing with opinion power in the platform world: Why we really have to rethink media concentration law. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2161924

Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, *10*(10), 1832–1854. https://doi.org/10.1080/21670811.2022.2063150

Sjøvaag, H., & Ohlsson, J. (2019). Media ownership and journalism. In M. Powers (Ed.), *Oxford research encyclopedia of communication*. https://doi.org/10.1093/acrefore/9780190228613.013.839

Tambini, D. (2021). *Media freedom*. Polity Press.

van der Vlist, F. N., & Helmond, A. (2021). How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society*, *8*(1). https://doi.org/10.1177/20539517211025061

van Drunen, M. Z., & Fechner, D. (2022). Safeguarding editorial independence in an automated media system: The relationship between law and journalistic perspectives. *Digital Journalism*. Advance online publication. https://doi.org/10.1080/21670811.2022.2108868

van Drunen, M. Z., Helberger, N., & Bastian, M. (2019). Know your algorithm: What media organizations need to explain to their users about news personalization. *International Data Privacy Law*, *9*(4), 220–235. https://doi.org/10.1093/idpl/ipz011

van Drunen, M. Z., Zarouali, B., & Helberger, N. (2022). Recommenders you can rely on: A legal and empirical perspective on the transparency and control individuals require to trust news personalisation. *JIPITEC*, *13*(3). https://www.jipitec.eu/issues/jipitec-13-3-2022/5562

Viljoen, S., Goldenfein, J., & McGuigan, L. (2021). Design choices: Mechanism design and platform capitalism. *Big Data & Society*, *8*(2). https://doi.org/10.1177/20539517211034312

About the Author

Theresa Josephine Seipp is a PhD candidate at the Institute for Information Law at the University of Amsterdam and is part of the AI, Media, and Democracy Lab. Her research focuses on governing AI in the media, opinion power, and media (concentration) law. Theresa holds degrees from the University of Groningen (LLB, LLM cum laude) and Ghent University (LLM), and previously researched at the Hans Bredow Institute (Leibniz Institute for Media Research) on topics related to EU media law and policy.

cogitatio

Media and Communication (ISSN: 2183–2439)
2023, Volume 11, Issue 2, Pages 406–409
https://doi.org/10.17645/mac.v11i2.6610

Commentary

The Fact of Content Moderation; Or, Let's Not Solve the Platforms' Problems for Them

Tarleton Gillespie [1,2]

[1] Microsoft Research New England, USA; tarleton@microsoft.com
[2] Department of Communication, Cornell University, USA

Submitted: 15 December 2022 | Accepted: 19 January 2023 | Published: 28 June 2023

Abstract
Recent social science concerning the information technology industries has been driven by a sense of urgency around the problems social media platforms face. But it need not be our job to solve the problems these industries have created, at least not on the terms in which they offer them. When researchers are enlisted in solving the industry's problems, we tend to repeat some of the missteps common to the study of technology and society.

Keywords
content moderation; governance; industry research; platforms

Issue
This commentary is part of the issue "A Datafied Society: Data Power, Infrastructures, and Regulations" edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

1. Problem-Solving and the Fallacy of Urgency

Over the last two decades, public and scholarly attention to the information industries has expanded and deepened. Critical questions about data-centrism, privacy, inequity, labor exploitation, and new forms of monopoly power have gained traction. The area I've been most concerned with—content moderation by social media platforms—has seen an explosion of academic attention to match the explosion of concern among the public, journalists, and policymakers around the world. This is undoubtedly good news. Senior leadership at most of the major platforms seem to take trust and safety more seriously than ever before. Detection tools have become more sophisticated. More attention is being paid to the labor that moderation requires, though not enough. The individual and societal harms being perpetrated via these platforms are now understood to be urgent, in a way they were not less than a decade ago.

But this is also a public relations strategy on the part of the tech companies, a grasp for "perceived corporate authenticity" (Hanlon & Fleming, 2009). Whenever troubled industries begin to acknowledge the concerns of their customers, they step up their corporate social responsibility efforts, recommit to the health of the public, the environment, the labor force, etc., and downplay the tactical value of these gestures. Even if well-intentioned, these gestures help to stabilize the understanding of the problems at hand, valorize the role of those companies in addressing them, demarcate appropriate solutions, and normalize the relations between the company, public, market, and state on which they depend (Baker & Hanna, 2022; Busch & Shepherd, 2014). The professionalization of Trust & Safety inside the companies for example, which is a welcome change in general, has also affirmed specific approaches to content moderation—reifying who counts as users, what registers as legitimate harm, and what reads as a reasonable intervention. And, as has been most painfully apparent in Elon Musk's takeover of Twitter in 2022, the content moderation debate can itself be deployed as a political cudgel, turning the problem into a game of "should he or shouldn't he."

These efforts also reshape what kinds of research get done, and what kinds enjoy the greatest visibility. As public and regulatory scrutiny has intensified, it is

the administrative, problem-solving research that has increasingly taken center stage: legal and economic analysis displace the sociological; American cases displace comparative ones; data-centric efforts to measure problems displace interpretive efforts to investigate them. Industry-friendly analysis crowds out the critical, the feminist, and the postcolonial, anything that might challenge the industry itself.

The widely shared sense of urgency about these issues—an urgency academics feel too, as users and citizens—has also drawn some researchers into lockstep with the social media companies, privileging a problem-solving mentality that takes for granted the definition of the problem and the aims of the stakeholders. Junior scholars are being lured by funded research projects and cajoled into taking unpaid advisory roles by the platform teams facing these controversies. Funding organizations have poured money into what Anderson (2021, p. 44) calls "consequence-driven and interpretation free" research on digital media and its effects. Funders like to see direct "engagement" with the companies as evidence of impact. Industry-academic partnerships, journals, and conferences enjoy outsized prominence, focusing attention on measuring and reducing harm while overshadowing equally important research about labor and inequity, subcultural expression, and the alternative approaches to moderation being squeezed out of view.

So social scientists, take note: It need not be our job to solve the problems these industries have created, at least not on the terms in which they offer them. Given my own employment, I get how odd, maybe problematic, it is for me to say this (see, for example, Sætra et al., 2022). But research oriented to problem-solving, while it may solve problems, also accepts the questions posed by industry stakeholders themselves—questions that are by no means innocent.

This may seem to run counter to the widespread concern that social media companies have frustrated academics by withholding access, to the massive data and rarified computational systems needed to conduct their research (Couch, 2020; Social Science Research Council, 2018), and with NDA arrangements deployed so as to thwart qualitative inquiry (Starr, 2020). But I believe the concern is in fact the same. Access to data is just not a wall, but a gate. By doling out access churlishly, the tech industry often can "capture" those researchers they do interact with, drawing a select few academic institutions into a cozy orbit (Whittaker, 2021)—a coziness that can leave people suspicious of the research they subsequently generate (Matias, 2020). And access to the people and inner workings of these companies can also be granted and withheld in ways designed to protect them from unfavorable assessment by researchers.

So I don't begrudge the field's demands for access to social media data. And I share the growing concerns about research partnerships with Silicon Valley. Mine is a complementary concern: Even researchers who do not enjoy access can nevertheless be captured—by embracing a problem-solving orientation that accepts the way the tech industries define their problems. This is akin to what Gitlin (1978)—commenting on mass media research of the time—called an "administrative mentality": research that "poses questions from the vantage of the command-posts of institutions that seek to improve or rationalize their control over social sectors in social functions" (1978, p. 225). Silicon Valley companies need researchers now more than ever, as signals of their good faith efforts, as they face multiple crises that have stirred public discontent and regulatory scrutiny that threaten their very existence. Plus, their ability to enlist researchers is stronger than ever, as they play on our genuine concern for the public welfare—that these companies jeopardized (Benson & Kirsch, 2010).

2. The Fact that Content Moderation Exists

Solving the problems the industry created on the terms they offer can lead us to overlook the problems we are not being invited to solve, the communities the industry tends to ignore, the solutions that challenge the business models embraced by the industry, and those dilemmas that are in fact not solvable, but are actually meant to be perennially contested. We are kept from thinking about how else moderation might be, or how the very fact of content moderation configures public power.

This kind of work is being done, certainly. Over more than a decade, scholars have rightly focused on expanding our understanding of the practices and dynamics involved in content moderation. Researchers in information studies, communication and media, and sociology have considered the entire sociotechnical ensemble being fixed into place: technological, institutional, social, and legal.

But when we adopt a problem-solving approach on terms borrowed from social media companies, we risk accepting as a precept that content moderation exists, and must exist in the way that it does—to accept that social media exists in the way it does. Even as this field grows more impactful, it takes too much for granted. This is an enormous mistake. Because it may be that what will matter about social media platforms and other information industries—their lasting significance—will not be about the specific dilemmas the technology sparks, or how well or poorly the industry stakeholders address them. It may be the very fact of content moderation as a societal project, the very fact of these industries and the roles they have inhabited, to which we should attend. "The fact of" is borrowed from Cavell's 1982 essay "The Fact of Television," and even more so from Streeter's (1996) use of Cavell in the start of his book *Selling The Air*.

Content moderation is an illustrative example. We can debate the facts *about* content moderation, how content moderation is or should be done, and how harm is or should be addressed. A judge in Texas may want there to be less moderation, while a feminist activist

harassed by misogynist trolls may want more. But both positions require there to be moderation, of some sort, and that means some apparatus that can accomplish moderation. Do the kinds of decisions platforms make matter? Of course. Does the scope of the specific problems they face matter? Absolutely. But it also matters that Silicon Valley has assembled an enormous labor force to do the work of moderation that didn't exist before, fitted with specific labor dynamics. It matters that the imposition of content moderation is driving some users to alt-sites that assert different moderation policies, cleaving political discourse in a particular way. It matters that the idea of moderation has enhanced and altered the cultural power of Silicon Valley companies. It matters that long-standing theories of regulation are shifting when it comes to the technology industry. It matters that moderation is helping reassert a new form of American cultural imperialism, under the guise of care. It matters that debates about content moderation and the responsibility of platforms have forced a subtle redefinition of "media" itself. Only research scoped so as to take in that entire sociotechnical ensemble shifts our attention from how moderation is done, to the very fact *of* it.

Studying the fact of content moderation means paying attention to what would still matter even if all the social media platforms disappeared. New legal and regulatory regimes are not only imposing obligations on social media platforms, but they're also generating new government agencies, new policy techniques, and whole new categories into which information intermediaries must adjust to fit. Tech firms are generating new managerial positions and adjudicative practices and forms of expertise inside the companies. Some are even attempting to manufacture institutional forms where none existed: Facebook's Oversight Board, which is designed to appear like an institutional partnership, though it is more like Facebook extruding out a part of itself so as to partner with it, may end up a model for other companies.

Or consider everyone being enlisted by the platform companies into playing sustained roles in the project of content moderation: non-profits and advocacy groups, content creator coalitions, institutional partners like fact-checking teams from news organizations, law enforcement, and regulatory agencies (Ananny, 2018). What is the relationship between these organizations and the platforms? Who defines and funds their efforts? What financial and political pressure does this impose on them, and how do they bear that pressure? What legitimacies and expertise are being called for and brought to bear, and how is all that borrowed authority used to legitimate the entire undertaking? How does their partnership with platforms, such as it is, alter how they understand their own public mission?

What social media is remains unsettled. What will later seem true about it may have more to do with the shifting institutions and arrangements being pulled together to stabilize it. The consequences that we will later mistakenly attribute to the technology of social media platforms will depend on the growth of new institutions, and the adjustments of existing institutions, around platforms and their governance (Johns, 1998). These arrangements will almost surely outlast the platform companies themselves.

3. We Are Implicated

When we focus on solving platforms' problems for them, our critical attention is obscured: by the sense of urgency, by the tactical way platforms define the problems for us, and by the fetishization of data science approaches and scalable solutions. Our concerns are replaced by theirs, or by none at all: Too often, calls for social science to be more oriented toward problem-solving seem to demonstrate a troubling disregard as to where those problems come from (Watts, 2017). I am not saying that the critical study of a datafied society needs to be irrelevant or utopian. But as researchers, we can and should opt to stand aside from this rush to solve immediate problems, to instead ask questions about the underlying dynamics that manifest themselves in these problems, about what the problem assumes or overlooks—and about what arrangements might provide solutions far down the road, but have to wait until our current institutional commitments have shifted (Splichal, 2008). While today's problems are urgent, they've been urgent far longer than the social media companies that now face them have been around.

In fact, if we take seriously the idea that it is the arrangements of institutions that will matter in the long run, then being enlisted in solving platforms' problems is us becoming part of that arrangement. We are implicated because we are among those being ensnared in this institutional and sociotechnical ensemble. If "don't solve problems" sounds counterintuitive, it is a reminder of how much our fields have already been enlisted in this project, by Silicon Valley and by the public outcry. Whatever the future fortunes of Facebook, or Meta, or whatever, the proximity between researchers and social media companies itself will matter: the implicit agreements being established about who takes on what responsibilities, who bears what costs, and who defines which goals.

Now, if we do refuse to dutifully engage in agnostic problem-solving, we still have a duty to meet with our research:

> Critical theory is, of course, not unconcerned with the problems of the real world. Its aims are just as practical as those of problem-solving theory, but it approaches practice from a perspective which transcends that of the existing order, which problem-solving theory takes as its starting point. (Cox, 1981, p. 130)

We should aspire to offer insights into the deeper assumptions embedded in how the problems are defined, the very fact of these sociotechnical systems as a part of the world, and the institutional arrangements being fixed into place. Understanding these, in their historical, sociological, and political-economic contexts, can be put into the service of more profound changes.

Acknowledgments

My gratitude to Elizabeth Fetterolf for their superb research support, and to Robyn Caplan, Mary Gray, and Dylan Mulvin for their comments on drafts of this essay.

Conflict of Interests

It is worth reiterating that I am employed by Microsoft, in the Microsoft Research lab in New England. The opinions expressed here are my own.

References

Ananny, M. (2018). *The partnership press: Lessons for platform-publisher collaborations as Facebook and news outlets team to fight misinformation*. Tow Center for Digital Journalism, Columbia University. https://doi.org/10.7916/D85B1JG9

Anderson, C. W. (2021). Fake news is not a virus: On platforms and their effects. *Communication Theory*, *31*(1), 42–61. https://doi.org/10.1093/ct/qtaa008

Baker, D., & Hanna, A. (2022, June 7). AI ethics are in danger. Funding independent research could help. *Stanford Social Innovation Review*. https://doi.org/10.48558/VCAT-NN16

Benson, P., & Kirsch, S. (2010). Capitalism and the politics of resignation. *Current Anthropology*, *51*(4), 459–486. https://doi.org/10.1086/653091

Busch, T., & Shepherd, T. (2014). Doing well by doing good? Normative tensions underlying Twitter's corporate social responsibility ethos. *Convergence*, *20*(3), 293–315. https://doi.org/10.1177/1354856514531533

Cavell, S. (1982). The fact of television. *Daedalus*, *111*(4), 75–96.

Couch, C. (2020). *The data driving democracy: Understanding how the internet is transforming politics and civic engagement*. American Academy of Arts & Sciences.

Cox, R. W. (1981). Social forces, states and world orders: Beyond international relations theory. *Millennium: Journal of International Studies*, *10*(2), 126–155. https://doi.org/10.1177/03058298810100020501

Gitlin, T. (1978). Media sociology: The dominant paradigm. *Theory and Society*, *6*(2), 205–253.

Hanlon, G., & Fleming, P. P. (2009). Updating the critical perspective on corporate social responsibility. *Sociology Compass*, *3*(6), 937–948. https://doi.org/10.1111/j.1751-9020.2009.00250.x

Johns, A. (1998). *The nature of the book: Print and knowledge in the making*. University of Chicago Press.

Matias, J. N. (2020). *Why we need industry-independent research on tech & society*. Citizens and Technology Lab. https://citizensandtech.org/2020/01/industry-independent-research

Sætra, H. S., Coeckelbergh, M., & Danaher, J. (2022). The AI ethicist's dilemma: Fighting big tech by supporting big tech. *AI and Ethics*, *2*(1), 15–27. https://doi.org/10.1007/s43681-021-00123-7

Social Science Research Council. (2018). *To secure knowledge: Social science partnerships for the common good*. https://www.ssrc.org/to-secure-knowledge

Splichal, S. (2008). Why be critical? *Communication, Culture & Critique*, *1*(1), 20–30. https://doi.org/10.1111/j.1753-9137.2007.00003.x

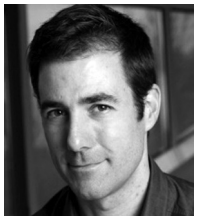
Starr, P. (2020, January 22). How money now tries to bury the truth. *The American Prospect*. https://prospect.org/power/how-money-now-tries-to-bury-the-truth

Streeter, T. (1996). *Selling the air: A critique of the policy of commercial broadcasting in the United States*. University of Chicago Press.

Watts, D. J. (2017). Should social science be more solution-oriented? *Nature Human Behaviour*, *1*(1), Article 0015. https://doi.org/10.1038/s41562-016-0015

Whittaker, M. (2021). The steep cost of capture. *Interactions*, *28*(6), 50–55. https://doi.org/10.1145/3488666

About the Author



Tarleton Gillespie is a senior principal researcher at Microsoft Research, and an affiliated associate professor in the Department of Communication and Department of Information Science at Cornell University. He is the author of *Wired Shut: Copyright and the Shape of Digital Culture* (MIT, 2007), co-editor of *Media Technologies: Essays on Communication, Materiality, and Society* (MIT, 2014), and author of *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media* (Yale, 2018).

MEDIA AND COMMUNICATION
ISSN: 2183-2439

Media and Communication is an international open access journal dedicated to a wide variety of basic and applied research in communication and its related fields. It aims at providing a research forum on the social and cultural relevance of media and communication processes.

The journal is concerned with the social development and contemporary transformation of media and communication and critically reflects on their interdependence with global, individual, media, digital, economic and visual processes of change and innovation.

cogitatio

www.cogitatiopress.com/mediaandcommunication