

**Open data in data journalism: Opportunities and future directions**

Alice Fleerackers, University of Amsterdam, Amsterdam, Netherlands

Natascha Chtena, Simon Fraser University, Vancouver, Canada

Monique Oliveira, Universidade Estadual de Campinas, Campinas, Brazil

Isabelle Dorsch, Leibniz Information Centre for Economics, Kiel, Germany

Juan Pablo Alperin, Simon Fraser University, Vancouver, Canada

Stephen Pinfield, University of Sheffield, Sheffield, UK

**ABSTRACT**

**Background:** In this commentary, we argue that it is time for communication scholars to turn their attention to how and why data journalists engage with the increasing amount of open research and government data available online.

**Analysis:** We review the limited scholarship that has investigated data journalists' engagement with open data and suggest directions for future research.

**Conclusions and implications:** Research that explicitly examines data journalists' use of open data is sorely needed, especially research that attends to the varied forms and practices that can emerge in different national and institutional contexts.

**Keywords:** data journalism, open data, open science, sources, science communication

## RÉSUMÉ

**Contexte :** Dans ce commentaire, nous soutenons qu'il est temps pour les spécialistes de la communication de s'intéresser à la manière dont les journalistes de données s'engagent dans la quantité croissante de données publiques et de recherche ouvertes disponibles en ligne, et aux raisons qui les poussent à le faire.

**Analyse :** Nous passons en revue les travaux limités qui ont étudié l'engagement des journalistes de données avec les données ouvertes et suggérons des orientations pour les recherches futures.

**Conclusions et implications :** Des recherches qui examinent explicitement l'utilisation des données ouvertes par les journalistes de données sont plus que nécessaires, en particulier des recherches qui tiennent compte des diverses formes et pratiques qui peuvent émerger dans différents contextes nationaux et institutionnels.

**Mots-clés:** journalisme de données, données ouvertes, science ouverte, sources, vulgarisation scientifique.

## Introduction

The COVID-19 pandemic is said to have revealed both strengths and weaknesses of journalism, demonstrating its public value but also its precarity (Perreault & Perreault, 2020). The same can be said of data journalism, as data journalists were among the first to collect, analyze, and communicate data about excess deaths due to COVID-19, drawing on a mix of techniques to rapidly gather and share insights (Desai et al., 2021; Zhang & Wang, 2022). Data journalists also took on tasks that have traditionally been ascribed to scientists, such as making predictions, comparing different computational models with one another, and exploring probable impacts of proposed measures for curbing the spread of the virus (Pentzold et al., 2021). As such, the pandemic not only underscored the importance of access to high quality data for informed public decision making (Cancela-Kieffer, 2021) but also the role data journalists can play in facilitating that access (Desai et al., 2021). At the same time, COVID-19 highlighted—and perhaps exacerbated—longstanding challenges data journalists face in accessing and using publicly relevant data (Cancela-Kieffer, 2021; Stollorz, 2021; Wu, 2021; Zhang & Wang, 2022). In countries such as Brazil, Germany, Singapore, and China, journalists attempted to cover the crisis amid data blackouts, unexplained delays on official data platforms, and inaccessible, low-quality, or incomplete data, all while operating with limited time and resources (“Ação Colaborativa Une Veículos Para Evitar Apagão de Dados Sobre Pandemia,” 2020; Stollorz, 2021; Wu, 2021). While the urgency of the pandemic may have waned, these barriers that prevent data journalists from bringing important insights to the public remain relevant, as other pressing social issues such as climate change or generative AI similarly warrant a data-driven reporting approach.

In this commentary, we propose that the open data (OD) movement—which seeks to make data freely accessible to all—has the potential to benefit data journalism by increasing the availability of high-

quality data sources, bolstering the profession's existing values, fostering democratic engagement, and supporting its mission to engage and inform citizens through data. To make our case, we review the limited body of scholarship that has investigated data journalists' engagement with OD sources and values, identify the most pressing gaps in the literature, and suggest directions for future research. As few, if any studies, have explicitly examined the connection between OD and data journalism (Papageorgiou et al., 2023), we draw heavily on literature that addresses this topic tangentially, rather than as a focus. In light of the potential gains of bridging these two domains and the limited scholarship that has investigated them, we argue that it is high time that scholars turn their attention to the intersection of OD and data journalism.

### **The argument for bridging data journalism and open data**

Data journalism has been broadly defined as “essentially any activity that deals with data in conjunction with journalistic reporting and editing or toward journalistic ends” (Coddington, 2015, p. 334). Journalism has traditionally been a largely qualitative profession, primarily concerned with text- and visual-based stories; however, some journalists have used computational approaches to information-gathering for decades (Coddington, 2015). Such computer-assisted reporting remained marginal to mainstream journalism until the early 2010s, when the profession took a “quantitative turn” (Petre, 2013). Enabled by new technology, reporters and editors increasingly used large datasets to craft stories, generating novel forms of data-driven storytelling that have become known as data journalism. While norms and practices vary widely across countries and media outlets, data journalism can be broadly understood as both a *process*—of using quantitative, computational methods to tell stories based on large datasets—and *product*—typically, one which involves some form of data visualization (Ausserhofer et al., 2020; Stalph & Heravi, 2021). Although data journalism remains a relatively small

part of journalism, it has captured an outsized share of scholarly attention, with a growing number of studies examining data journalists' practices, epistemologies, business models, and philosophies (de Lima-Santos, 2024; Erkmen, 2024).

The COVID-19 pandemic brought about an increased interest in OD (Carr, 2020; Gutierrez & Li, 2020; Organisation for Economic Co-operation and Development, 2021). Yet, the idea of OD—data available for anyone to access, use, and share—has been around since at least the 1990s (Kelty, 2008; Ostrom, 1990). With growing demands for organizational and governmental accountability and open access to research knowledge, OD has since gained momentum, mainly through open government initiatives and national mandates for making research data openly available (Attard et al., 2015; Goodey et al., 2022). This momentum has only grown further with the increased global push for open science (United Nations Educational, Scientific and Cultural Organization, 2021), which seeks to make scientific processes and outputs—including research data—available and accessible to the broadest possible audience (Bartling & Friesike, 2014). Today, rates of data sharing are increasing in academia (i.e., “Open Research Data” [ORD], Kaiser & Brainard, 2023; Khan et al., 2023; Tenopir et al., 2020) and government (i.e., “Open Government Data” [OGD], Zuiderwijk & Reuver, 2021). Democratic gains are expected from data openness and reuse (e.g., increased transparency, citizen participation, response to the crisis of trust in institutions), along with gains in scientific quality and efficiency (Fell, 2019; Jethani & Leorke, 2021). These democratic gains extend beyond those in academia or government, with scholars arguing that openly sharing data, code, and other research products provides “non-members of the scientific community” (Lauerer, 2022, p. 395) important opportunities to examine their own research questions and hold those in power to account (Haim & Puschmann, 2023).

Data journalists are an obvious example of the type of “non-members” who would benefit from access to OD, as they are heavy data users who are driven to investigate the kind of socially relevant questions and failings of the powerful that Lauerer (2022) described (Pereira & Mastrella, 2022). Moreover, although data journalism by its very nature “relies on the availability of data” (Tong, 2024, p. 3), studies have repeatedly demonstrated that data journalists often face challenges in obtaining access to high quality, useable datasets—either because data are not publicly available or have been shared in formats that are difficult to manipulate and analyze (e.g., aggregate formats, PDFs, printed Excel files) (deLima-Santos, 2024; Erkmen, 2024). In fact, lack of access to quality data was the top challenge (65%) reported by data journalists in The State of Data Journalism Survey 2023 conducted by the European Journalism Centre (2023). Open data thus has the potential to benefit data journalism by providing unfettered access to data that have been cleaned, curated, and shared with the express purpose of enabling reuse.

Second, because data journalists have the skills to make large datasets understandable, share data publicly, and actively engage citizens, they are well-equipped to be the “empowering intermediaries” (Baack, 2015) that OD advocates have argued are essential for enabling widespread, democratic engagement with data (Fell, 2019; Jethani & Leorke, 2021). That is, journalistic reports that use data that are accessible to everyone could help bridge the gap between institutions and citizens and encourage public involvement, providing audiences with evidence they need to develop empirically grounded attitudes and beliefs about the political and social world, as well as hold informed opinions on key public issues (Carpini, 2004). Such an intermediary role aligns with the values of transparency and participation that are commonly ascribed to both data journalism and OD (Jethani & Leorke, 2021; Loosen et al., 2020; Zhang & Feng, 2019), as well as with the democratic function journalism is

envisioned to hold within the United States and many other societies (Kovach & Rosenstiel, 2021; Pereira & Mastrella, 2022).

### **Data journalists' use of open government data**

Despite the potential for data journalists to benefit from using OD, empirical scholarship has yielded mixed results about the extent to which such use takes place in practice (Khan et al., 2021; Loosen et al., 2020; Zamith, 2019). As rates of OD use appear to differ depending on the type of data in question, we discuss journalists' engagement with OGD and ORD separately.

With respect to OGD, existing research suggests that data journalism stories draw heavily on publicly available data provided by government institutions and political actors, at least some of which would be considered open. One analysis of data journalism stories published in *The New York Times* and *The Washington Post* found that official government data was the most cited data source (whether it was OGD was not specified; Zamith, 2019). Similarly, more than two-thirds of award-nominated data journalism stories in another study relied on data from official institutions (e.g., statistical ministries or offices), while 42 percent relied on data from other non-commercial organizations, including universities and research institutes and non-governmental organizations (Loosen et al., 2020). More recently, Abhishek and Graves (2023) found that government data was among the most common types of data included in news organizations' GitHub repositories, although about a fifth of these government data were gathered through a Freedom of Information (FOI) request rather than downloaded from a freely available source.

Despite journalists' reliance on government datasets, they often face difficulties in accessing or using these data (Papageorgiou et al., 2023; Porlezza & Splendore, 2019; Zhang & Feng, 2019), suggesting

that OGD could play a bigger role in data journalism than research suggests it currently does. Even when government datasets are made available to journalists, they are often shared in forms that make it difficult to process, verify, and use (e.g., as PDFs rather than spreadsheets; Martin et al., 2022), particularly for journalists working in the Global South (Lewis & Nashmi, 2019; Muthmainnah et al., 2022). To overcome these access-related barriers, data journalists use strategies such as supplementing openly available datasets with data they have collected themselves (e.g., via web scraping), gathered in partnership with a third party (e.g., a non-governmental organization [NGO]), or received via a leak by a source (Martin et al., 2022; Muthmainnah et al., 2022). Some journalists even use physical resources, such as books, when OGD are not available (Wright & Nolan, 2021). Other journalists, especially those working in Latin America, use crowdsourced data as a complement or substitute for OGD (Amado & Tarullo, 2019; Borges-Rey, 2019; de-Lima-Santos & Mesquita, 2023; Palomo et al., 2019). Some of these efforts to gather data through alternative means ultimately result in the creation of OGD, as some media outlets make the datasets they develop publicly available to enable wider access and use (Palomo et al., 2019; see also Zhang & Feng, 2019).

Even more commonly, journalists often use FOI requests to circumvent poor-quality, incomplete, and out-of-date OGD sources—at least in countries where the necessary legislation is in place (de-Lima-Santos & Mesquita, 2023). In countries where FOIs are not possible, lack of access to reliable, usable data remains a common barrier (Borges-Rey, 2019; Lewis & Nashmi, 2019; Mutsvairo, 2019; Wiley, 2023; Wu, 2021). Media access to OGD also appears to be particularly compromised in countries with little or no transparency infrastructure or mandates supporting OGD (de-Lima-Santos & Mesquita, 2023; Kashyap et al., 2023; Martin et al., 2022), resulting in a situation in which “technical and economic inequalities that affect the implementation of the open data infrastructures can produce unequal data access and widen the gap in data journalism practices between information-rich and information-poor

countries” (Camaj et al., 2022b, p. 16). Similarly, lack of technical skills is repeatedly mentioned as a barrier to using OGD, meaning journalists with less access to training are at a disadvantage (Papageorgiou et al., 2023). Geographic and culture-specific factors may further exacerbate these inequalities; for instance, journalists in the Arab region report preferring data from the Western NGOs, which they deem more trustworthy than locally available data (Lewis & Nashmi, 2019). Differences in the type of OGD legislation and infrastructure in place also appear to influence the types of stories data journalists can produce, as journalists in countries where FOI requests are common produce more investigative, watchdog journalism while those in countries with solid OGD infrastructures tend to produce lighter content (i.e., “soft” news, Camaj et al., 2023).

### **Data journalists’ use of open research data**

While government data appears to be data journalists’ preferred data source, a few studies provide evidence that ORD may be used in data journalism stories at least occasionally. One study of Indonesian journalists found that research data, such as data from scientific journals, sometimes played a supplementary role when government data about a newsworthy topic was not available (Muthmainnah et al., 2022). Along these lines, Zhang and Feng (2019) found that Chinese data journalism stories often relied on publicly available data sourced from NGOs and research institutions—however, as these two data providers were treated as one category in their analysis, it is not possible to determine the proportion of stories that used academic sources specifically. Meanwhile, Stalph and Heravi (2021) found that, while almost two-thirds of award-winning data visualizations relied on open or publicly available data, only 4 percent of the visualizations used academic or university data. Camaj et al. (2023) found that about half of award-nominated data journalism projects included some type of “open data,” but that the use of data from universities and NGOs was decreasing over time (the two types of data

sources were treated as one category). Similarly, Abhishek and Graves (2023) found that data journalism stories use “a variety of private or academic databases” but did not specify the proportion that use academic versus private sources, nor whether those academic sources were openly available. Results from these content analyses are supported by studies measuring scholarly impact based on social media engagement with research (Priem, 2014), which have similarly found little or no use of ORD in blogs or news media (Khan et al., 2021).

Journalists’ apparent lack of engagement with ORD specifically may be connected to the nature of the data themselves. Academic research is often complex and highly specialized, with data collected for the purpose of providing insights that are relevant to scholars rather than citizens or journalists (Elliott & Resnik, 2019); in contrast, government data typically has more direct relevance to civic issues and can be used to hold governments to account (Lawson, 2022). Indeed, in countries where OD are more available, data journalists sometimes feel that there is “too much” data but not enough data that is relevant to the public (Martin et al., 2022). Relatedly, journalists working in diverse contexts have expressed that the stories they can tell are often limited by the nature of the data available to them, with simple, relatable stories based on government data typically winning out over more complex data narratives (Lawson, 2022; Muthmainnah et al., 2022). It does not help that journalists are seldom considered as stakeholders or users of ORD within the research community (Elliott, 2022), as evidenced by the lack of attention to journalism within manifestos and policy papers on ORD (Ekaputra et al., 2017; Stoneman, 2015). That is, ORD may be underutilized by journalists because, in their current forms and topics, they do not lend themselves easily to journalistic storytelling.

### **Open data values in data journalism**

On a practical level, OD could thus support data journalism by providing an accessible data source for journalists' stories. Yet, bridging data journalism and OD may also prove fruitful on a more philosophical level, due to synergies in the stated values underpinning these two domains. Some studies have documented a strong commitment within data journalism to OD values such as transparency, accessibility, and participation (Jethani & Leorke, 2021), both in the meta-journalistic discourse about data journalism and in data journalism stories themselves (Morini et al., 2023; Papageorgiou et al., 2023; Pereira & Mastrella, 2022; Stalph & Heravi, 2021). For instance, in the Arab region, transparency is seen as essential for establishing audience trust, particularly through the sharing of the data sources, metadata, and methodologies involved in data journalism stories (Jebril & Altef, 2023). A similar commitment to transparency has been voiced among journalists in other regions (Morini et al., 2023). More broadly, both scholars and practitioners have described data journalists as "activists" or "advocates" for openness and transparency who seek to empower the public to access and use data (de-Lima-Santos, 2024; Martin et al., 2022; Morini et al., 2023), in line with Baack's (2015) conception of empowering data intermediaries. This goal of advancing OD culture and practice appears to be particularly strong in Latin America, where data journalism has been actively supported by organizations that promote openness to knowledge, such as the Open Knowledge Foundation (de-Lima-Santos & Mesquita, 2021).

Yet other research suggests a weaker commitment to open data values, with one study of data journalism projects submitted to the Global Editors Network Data Journalism Awards finding that only 20 percent of submissions mentioned a desire to promote data transparency (Camaj et al., 2022a). Moreover, stated OD values are often inconsistently enacted in practice (Zamith, 2023). For instance, although many data

journalists actively use GitHub as part of their working routines and see participating in the open-source community as essential to their role (Boyles, 2020; Khalfan et al., 2024), this use does not often appear to lead to interdisciplinary or interorganizational collaborations or support public engagement with data (Dodds et al., 2024; Haim & Zamith, 2019; Zamith, 2023). Similarly, although many data journalists seek to “empower citizens to take ownership of public data,” their stories often tend to engage audiences in a passive or limited way (Martin et al., 2024, p. 11). Likewise, despite express commitments to transparency (Hannaford, 2023), sharing the code, software, methods, and data behind data journalism stories appear to be relatively rare (Chaparro-Domínguez & Díaz-Campo, 2023; Dodds et al., 2024; Loosen et al., 2020; Zamith, 2019), with one analysis of 290 stories failing to find a single hyperlink to a data source (Zhang & Feng, 2019). Indeed, some data journalists only share datasets with the public when requested, and sometimes only do so by (physical) mail (Martin et al., 2024). This reluctance to share data appears to be connected to data journalists’ beliefs that the public does not have the interest or literacy needed to use data themselves, as well as “trial and error” experiments in sharing data, which journalists report have yielded limited audience engagement (Martin et al., 2024).

As with use of OD sources, some evidence suggests that embodiment of OD values may vary across geographies and media outlets. For example, the value of openness appears to be more central to Italian data journalists than those in the mostly U.S.-focused studies discussed above; they report sharing datasets with other journalists and publishing detailed methodological notes so that “other actors such as NGOs and the wider public are able to access and use the data for further collaboration or investigation” (Porlezza & Splendore, 2019, p. 1241). Brazilian data journalists have expressed similar commitments, although support for these values was not universal among those interviewed (Pereira & Mastrella, 2022). Similarly, in China, transparency practices appear to be highly inconsistent across stories and media outlets (Wang et al., 2023). In addition, actors who have traditionally been peripheral to

mainstream journalism but contribute to producing data-driven stories, such as fact-checking organizations or civic technologists, may be more willing and able to put OD values into practice (Cheruiyot et al., 2019; Cheruiyot & Ferrer-Conill, 2018).

### **Concluding remarks and recommendations for future research**

Collectively, the research discussed in this commentary suggests that data journalists support values associated with the OD movement, such as transparency and participation, and see it as part of their role to act as empowering data intermediaries who make data understandable, publicly available, and engaging (Baack, 2015). They simultaneously struggle to access and use data, in part because publicly relevant datasets are not always publicly available and in part because those datasets that *are* made available are often shared in formats that are difficult to analyze. All of this suggests that data journalism and OD have the potential to strengthen and support one another, with data journalism enabling the kind of democratic engagement that has often been envisioned by OD advocates, but which has remained elusive in practice, and OD providing data journalists with readily accessible and easily useable fodder for their stories.

Yet, the literature discussed above also suggests that data journalists do not always perform their data intermediary role in practice, with inconsistent evidence that they use OD sources—especially ORD—in their reporting or embody the values they describe as being so central to their role. This lack of engagement with OD has important implications, as “low availability and inefficiency of accessing quality open data hinders the development of data journalism” (de-Lima-Santos, 2024), especially within countries and newsrooms where the resources to purchase data or skills needed to gather their own are scarce. We therefore join Papageorgiou et al. (2023) in calling for more research investigating “the barriers and benefits journalists encounter when using open data” (p. 165), which we see as

essential for guiding the development of education and resources that can better equip data journalists to take advantage of these rich data sources and for ensuring that governments and scholars share their data in ways that support rather than hinder journalists from using them. In addressing this call, we encourage scholars to attend particularly to the ways in which different geographic, institutional, and individual factors shape data journalists' willingness and ability to make use of OD, as existing research suggests that these patterns of data use are highly varied and context specific.

In addition, we urge scholars to more explicitly examine data journalists' engagement with OD, as existing studies that focus on this topic are extremely scarce, limiting the insights that can be gleaned from them. Indeed, the mixed evidence described in this commentary may in part be an artefact of inconsistent treatment of terminology and a lack of attention to OD within study designs, as terms such as "open data" and "public data" are often used interchangeably within the literature to refer to a broad range of data sources, including crowdsourced data, social media and website data, and data provided by government institutions in response to FOI requests (Martin et al., 2022; Papageorgiou et al., 2023).

While such types of data may be considered "open" or "public" in a colloquial sense, they do not technically qualify as OD, as they are not freely "available as a whole [. . .] in a convenient and modifiable form" under terms that allow reuse and redistribution by anyone, regardless of their profession or identity (*What Is Open Data?*, n.d.). Moreover, terms such as "open research data," "open government data," or "open repository" are seldom included as coding categories for content analyses, making it difficult to document how often these sources are used by journalists. This may soon change, as scholars have recently introduced an analytical framework for examining data visualization that includes an assessment of both the *provider* of the data (with "university/academia" and "government" included as options) and *method of access* (with "open data/available publicly on the web" as an option) (Stalph & Heravi, 2021). Wu (2021) similarly proposed a conceptual framework for future data

journalism research that includes *access* and *availability* of data as key variables. We urge scholars to take advantage of these frameworks and practice greater specificity in terminology in designing their studies. Doing so is crucial for building a more robust understanding how, why, and to what extent data journalists act as non-profit data intermediaries as societies grapple with urgent, data-related issues such as climate change and algorithmic bias, and whether this has the intended effect of “empowering” audiences to engage with, understand, and use data (Baack, 2015).

**Suggested citation:** Fleerackers, A., Chtena, N., Oliveira, M., Dorsch, I., Alperin, J. P., & Pinfield, S. Open data in data journalism: Opportunities and future directions. *Canadian Journal of Communication*, 50(1), 0–0.

**Acknowledgments:** We would like to thank Esteban Morales, Kathleen Gregory, and the anonymous reviewers for their thoughtful suggestions for improving the manuscript, as well as Manon Dessart for her help translating the abstract.

**Data accessibility:** All data will not be made publicly available. Researchers who require access to the study data can contact the corresponding author for further information.

**Funding:** The work was supported by the Trans-Atlantic Platform for Social Sciences and Humanities (T-AP) under grant 2021/07508-6, VOICES project, with contributions from: the Arts and Humanities Research Council (AHRC, UK), grant R/172830; the Social Science and Humanities Research Council (SSHRC, Canada), grant 2005-2021-0011; the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Brazil), grant 21/07577-8; and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant 495515545.

**Disclosures:** The authors have no conflicts of interest to disclose.

#### Author bios

**Alice Fleerackers** is Assistant Professor of Journalism and Civic Engagement at the Department of Media Studies, University of Amsterdam. Email: [a.l.fleerackers@uva.nl](mailto:a.l.fleerackers@uva.nl)

**Natascha Chtena** is a Research Associate at the School of Publishing, Simon Fraser University. Email: [achtena@sfu.ca](mailto:achtena@sfu.ca)

**Monique Oliveira** is a Postdoctoral Fellow at the Laboratory of Advanced Studies in Journalism, Universidade Estadual de Campinas. Email: [monique.oliveira@usp.br](mailto:monique.oliveira@usp.br)

**Isabelle Dorsch** is a Postdoctoral Researcher at ZBW – Leibniz Information Centre for Economics. Email: [I.Dorsch@zbw.eu](mailto:I.Dorsch@zbw.eu)

**Juan Pablo Alperin** is Associate Professor at the School of Publishing, Simon Fraser University, and Scientific Director of the Public Knowledge Project. Email: [juan@alperin.ca](mailto:juan@alperin.ca)

**Stephen Pinfield** is Professor of Information Services Management at the Information School, University of Sheffield and Senior Research Fellow at the Research on Research Institute (RoRI). Email: [s.pinfield@sheffield.ac.uk](mailto:s.pinfield@sheffield.ac.uk)

## References

- Abhishek, A., & Graves, L. (2023, May 29). *Analyzing code: What is the editorial role of computational analysis in data journalism stories?* The 73rd Annual International Communication Association Conference, Toronto, Ontario.
- Ação colaborativa une veículos para evitar apagão de dados sobre pandemia. (2020, June 8). *Abraji*.  
<https://www.abraji.org.br/noticias/acao-colaborativa-une-veiculos-para-evitar-apagao-de-dados-sobre-pandemia>
- Amado, A., & Tarullo, R. (2019). Journalism and civil society: Key to data journalism in Argentina. In B. Mutsvairo, S. Bebawi, & E. Borges-Rey (Eds.), *Data journalism in the Global South* (pp. 285–299). Springer International Publishing. [https://doi.org/10.1007/978-3-030-25177-2\\_15](https://doi.org/10.1007/978-3-030-25177-2_15)
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399–418.  
<https://doi.org/10.1016/j.giq.2015.07.006>
- Ausserhofer, J., Gutounig, R., Oppermann, M., Matiasek, S., & Goldgruber, E. (2020). The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism*, 21(7), 950–973.  
<https://doi.org/10.1177/1464884917700667>
- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2), 2053951715594634.  
<https://doi.org/10.1177/2053951715594634>

- Bartling, S., & Friesike, S. (Eds.). (2014). *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Springer Nature.  
<https://doi.org/10.1007/978-3-319-00026-8>
- Borges-Rey, E. (2019). Data journalism in Latin America: Community, development and contestation. In B. Mutsvairo, S. Bebawi, & E. Borges-Rey (Eds.), *Data journalism in the Global South* (pp. 257–283). Springer International Publishing. [https://doi.org/10.1007/978-3-030-25177-2\\_14](https://doi.org/10.1007/978-3-030-25177-2_14)
- Boyles, J. L. (2020). Deciphering code: How newsroom developers communicate journalistic labor. *Journalism Studies*, 21(3), 336–351. <https://doi.org/10.1080/1461670X.2019.1653218>
- Camaj, L., Lanosga, G., & Martin, J. (2023). A global perspective on data journalism materiality: Knowledge production across public transparency infrastructure environments. *Digital Journalism*, 0(0), 1–20. <https://doi.org/10.1080/21670811.2023.2278040>
- Camaj, L., Martin, J., & Lanosga, G. (2022a). Professional ideals of data journalists around the globe: Congruencies and divergences between role conceptions and narrated role performances. *Journalism Studies*, 23(12), 1450–1471. <https://doi.org/10.1080/1461670X.2022.2094822>
- Camaj, L., Martin, J., & Lanosga, G. (2022b). The impact of public transparency infrastructure on data journalism: A comparative analysis between information-rich and information-poor countries. *Digital Journalism*, Advance online publication.  
<https://doi.org/10.1080/21670811.2022.2077786>

- Cancela-Kieffer, M. (2021). Journalism, algorithms, and the people's right to know. In *News media innovation reconsidered* (pp. 155–173). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9781119706519.ch10>
- Carpini, M. X. D. (2004). Mediating democratic engagement: The impact of communications on citizens' involvement in political and civic life. In L. L. Kaid (Ed.), *Handbook of political communication research* (pp. 395–434). Lawrence Erlbaum Associates, Inc., Publishers.
- Carr, D. (2020, January 31). Coronavirus (COVID-19): Sharing research data. *Wellcome*.  
<https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak>
- Chaparro-Domínguez, M.-Á., & Díaz-Campo, J. (2023). Data journalism and ethics: Best practices in the winning projects (DJA, OJA and Sigma Awards). *Journalism Practice*, 17(6), 1321–1339.  
<https://doi.org/10.1080/17512786.2021.1981773>
- Cheruiyot, D., Baack, S., & Ferrer-Conill, R. (2019). Data journalism beyond legacy media: The case of African and European civic technology organizations. *Digital Journalism*, 7(9), 1215–1229.  
<https://doi.org/10.1080/21670811.2019.1591166>
- Cheruiyot, D., & Ferrer-Conill, R. (2018). “Fact-checking Africa”: Epistemologies, data and the expansion of journalistic discourse. *Digital Journalism*, 6(8), 964–975.  
<https://doi.org/10.1080/21670811.2018.1493940>
- Coddington, M. (2015). Clarifying journalism's quantitative turn. *Digital Journalism*, 3(3), 331–348.  
<https://doi.org/10.1080/21670811.2014.976400>

de-Lima-Santos, M.-F. (2024). The entanglements between data journalism, collaboration and business models: A systematic literature review. *Digital Journalism*, 12(2), 256–281.

<https://doi.org/10.1080/21670811.2023.2247449>

de-Lima-Santos, M.-F., & Mesquita, L. (2021). The strategic value of data journalism. In R. Salaverría & M.-F. de-Lima-Santos (Eds.), *Journalism, data and technology in Latin America* (pp. 97–136).

Springer International Publishing. [https://doi.org/10.1007/978-3-030-65860-1\\_4](https://doi.org/10.1007/978-3-030-65860-1_4)

de-Lima-Santos, M.-F., & Mesquita, L. (2023). Data Journalism in favela: Made by, for, and about forgotten and marginalized communities. *Journalism Practice*, 17(1), 108–126.

<https://doi.org/10.1080/17512786.2021.1922301>

Desai, A., Nouvellet, P., Bhatia, S., Cori, A., & Lassmann, B. (2021). Data journalism and the COVID-19 pandemic: Opportunities and challenges. *The Lancet Digital Health*, 3(10), e619–e621.

[https://doi.org/10.1016/S2589-7500\(21\)00178-3](https://doi.org/10.1016/S2589-7500(21)00178-3)

Dodds, T., Reséndez, V., von Nordheim, G., Araujo, T., & Moeller, J. (2024). Collaborative coding cultures: How journalists use GitHub as a trading zone. *Digital Journalism*, 1–22.

<https://doi.org/10.1080/21670811.2024.2342468>

Ekaputra, F. J., Novak, N. M., Kiesling, E., Aryan, P., Do, B., Trinh, T., & Tjoa, A. (2017). Towards open data mashups for data journalism. In *Proceedings of the Posters and Demos Track of the 13th International Conference on Semantic Systems - SEMANTiCS2017 Co-Located with the 13th International Conference on Semantic Systems (SEMANTiCS 2017)*, 2044, 6.

- Elliott, K. C. (2022). Open science for non-specialists: Making open science meaningful beyond the scientific community. *Philosophy of Science*, 1–20. <https://doi.org/10.1017/psa.2022.36>
- Elliott, K. C., & Resnik, D. B. (2019). Making open science work for science and society. *Environmental Health Perspectives*, 127(7). <https://doi.org/10.1289/EHP4808>
- Erkmen, Ö. (2024). Data journalism: A systematic literature review. *Journalism Studies*, 25(1), 58–79. <https://doi.org/10.1080/1461670X.2023.2289885>
- Fell, M. J. (2019). The economic impacts of open science: A rapid evidence assessment. *Publications*, 7(3), Article 3. <https://doi.org/10.3390/publications7030046>
- Goodey, G., Hahnel, M., Zhou, Y., Jiang, L., Chandramouliswaran, I., Hafez, A., Paine, T., Gregurick, S., Simango, S., Peña, J. M. P., Murray, H., Cannon, M., Grant, R., McKellar, K., & Science, L. D. D. (2022). *The state of open data 2022* [Report]. Digital Science, Springer Nature, Figshare. <https://apo.org.au/node/319974>
- Gutierrez, B., & Li, S. L. (2020, November 17). *The need for open data sharing in the era of global pandemics*. Impact of Social Sciences. <https://blogs.lse.ac.uk/impactofsocialsciences/2020/11/17/the-need-for-open-data-sharing-in-the-era-of-global-pandemics/>
- Haim, M., & Puschmann, C. (2023). Opening up data, tools, and practices: Collaborating with the future. *Digital Journalism*, 11(2), 247–254. <https://doi.org/10.1080/21670811.2023.2174894>

Haim, M., & Zamith, R. (2019). Open-source trading zones and boundary objects: Examining GitHub as a space for collaborating on “news.” *Media and Communication*, 7(4), 80–91.

<https://doi.org/10.17645/mac.v7i4.2249>

Hannaford, L. (2023). The discourses of data journalism. *Journalism*, 24(11), 2397–2417.

<https://doi.org/10.1177/14648849221107803>

Jebri, N., & Altef, S. (2023). Perceptions of data journalism and its impact on democratising Arab news media. *Journalism*, 14648849231221038. <https://doi.org/10.1177/14648849231221038>

Jethani, S., & Leorke, D. (2021). What is open data? Historical and instrumental perspectives. In S.

Jethani & D. Leorke (Eds.), *Openness in practice: Understanding attitudes to open government data* (pp. 15–32). Springer. [https://doi.org/10.1007/978-981-16-4251-7\\_2](https://doi.org/10.1007/978-981-16-4251-7_2)

Kaiser, J., & Brainard, J. (2023). Ready, set, share: Researchers brace for new data-sharing rules

[Dataset]. In *Science Magazine*, 379(6630). <https://doi.org/10.1126/science.adg8470>

Kashyap, G., Mishra, H., & Bhaskaran, H. (2023). Data journalists’ perception and practice of transparency and interactivity in Indian newsrooms. *Media Asia*, 50(1), 24–42.

<https://doi.org/10.1080/01296612.2022.2076324>

Kelty, C. M. (2008). *Two bits: The cultural significance of free software*. Duke University Press.

Khalfan, W., Lugo-Ocando, J., & Requejo Alemán, J. L. (2024). Coding and capability-building in nonprofit digital-native news organizations in the Middle East and North Africa (MENA) and

Latin America (LATAM). *Digital Journalism*, 0(0), 1–23.

<https://doi.org/10.1080/21670811.2024.2340589>

Khan, N., Thelwall, M., & Kousha, K. (2021). Measuring the impact of biodiversity datasets: Data reuse, citations and altmetrics. *Scientometrics*, 126(4), 3621–3639.

<https://doi.org/10.1007/s11192-021-03890-6>

Khan, N., Thelwall, M., & Kousha, K. (2023). Data sharing and reuse practices: Disciplinary differences and improvements needed. *Online Information Review* (ahead-of-print).

<https://doi.org/10.1108/OIR-08-2021-0423>

Kovach, B., & Rosenstiel, T. (2021). *The elements of journalism, revised and updated 4th edition: What newspeople should know and the public should expect* (4th edition). Crown.

Lauerer, C. (2022). Balancing efficiency and inclusivity: Open Science principles and practices in international collaborative research projects. *Digital Journalism*, 11(2), 390–398.

<https://doi.org/10.1080/21670811.2022.2123020>

Lawson, B. (2022). Realizing the benefits of open government data: Journalists' coverage of the NHS winter crisis, 2016–17. *The Information Society*, 38(1), 25–35.

<https://doi.org/10.1080/01972243.2021.1998274>

Lewis, N. P., & Nashmi, E. A. (2019). Data journalism in the Arab region: Role conflict exposed.

*Digital Journalism*, 7(9), 1200–1214. <https://doi.org/10.1080/21670811.2019.1617041>

- Loosen, W., Reimer, J., & De Silva-Schmidt, F. (2020). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, 21(9), 1246–1263. <https://doi.org/10.1177/1464884917735691>
- Martin, J. A., Camaj, L., & Lanosga, G. (2022). Scrape, request, collect, repeat: How data journalists around the world transcend obstacles to public data. *Journalism Practice*, 0(0), 1–18. <https://doi.org/10.1080/17512786.2022.2142837>
- Martin, J. A., Camaj, L., & Lanosga, G. (2024). Audience engagement in data-driven journalism: Patterns in participatory practices across 34 countries. *Journalism*, 14648849241230414. <https://doi.org/10.1177/14648849241230414>
- Morini, F., Dörk, M., & Appelgren, E. (2023). Sensing what’s new: Considering ethics when using sensor data in journalistic practices. *Digital Journalism*, 11(3), 465–483. <https://doi.org/10.1080/21670811.2022.2134161>
- Muthmainnah, A. N., Arief, M., & Fitriyani, N. (2022). Verification and data processing of open data for data-driven news stories in lokadata.id and katadata.co.id. *Jurnal Kajian Jurnalisme*, 6(1), 1. <https://doi.org/10.24198/jkj.v6i1.37167>
- Mutsvairo, B. (2019). Challenges facing development of data journalism in non-Western societies. *Digital Journalism*, 7(9), 1289–1294. <https://doi.org/10.1080/21670811.2019.1691927>
- Organisation for Economic Co-operation and Development. (2021). *OECD Science, Technology and Innovation Outlook 2020: Science and Innovation in Times of Crisis*. OECD. <https://doi.org/10.1787/75f79015-en>

- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511807763>
- Palomo, B., Teruel, L., & Blanco-Castilla, E. (2019). Data journalism projects based on user-generated content: How La Nacion data transforms active audience into staff. *Digital Journalism*, 7(9), 1270–1288. <https://doi.org/10.1080/21670811.2019.1626257>
- Papageorgiou, G., Euripides, L., Magnussen, R., & Charalabidis, Y. (2023). Open data journalism: A domain mapping review. *Proceedings of the 16th International Conference on Theory and Practice of Electronic Governance*, 159–166. <https://doi.org/10.1145/3614321.3614340>
- Pentzold, C., Fechner, D. J., & Zuber, C. (2021). “Flatten the curve”: Data-driven projections and the journalistic brokering of knowledge during the covid-19 crisis. *Digital Journalism*, 9(9), 1367–1390. <https://doi.org/10.1080/21670811.2021.1950018>
- Pereira, F. H., & Mastrella, B. (2022). The practice of data-driven journalism in Brazil: Between disruption and the reinstatement of professional values. *Digital Journalism*, 10(4), 599–625. <https://doi.org/10.1080/21670811.2022.2061550>
- Perreault, M. F., & Perreault, G. P. (2021). Journalists on COVID-19 journalism: Communication ecology of pandemic reporting. *American Behavioral Scientist*, 0002764221992813. <https://doi.org/10.1177/0002764221992813>
- Petre, C. (2013, October 30). *A quantitative turn in journalism?* Tow Center for Digital Journalism. <https://web.archive.org/web/20161024020758/http://towcenter.org/a-quantitative-turn-in-journalism/>

Porlezza, C., & Splendore, S. (2019). From open journalism to closed data: Data journalism in Italy.

*Digital Journalism*, 7(9), 1230–1252. <https://doi.org/10.1080/21670811.2019.1657778>

Priem, J. (2014). Altmetrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 263–287). MIT Press.

<https://doi.org/10.7551/mitpress/9445.003.0019>

Stalph, F., & Heravi, B. (2021). Exploring data visualisations: An analytical framework based on dimensional components of data artefacts in journalism. *Digital Journalism*, Advance online publication. <https://doi.org/10.1080/21670811.2021.1957965>

Stollorz, V. (2021). Challenges for science journalism in the Corona pandemic-some observations about a mediated world event. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, 64(1), 70–76. <https://doi.org/10.1007/s00103-020-03257-x>

Stoneman, J. (2015). *Does open data need journalism?* <https://ora.ox.ac.uk/objects/uuid:c22432ea-3ddc-40ad-a72b-ee9566d22b97>

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, 15(3), e0229003.

<https://doi.org/10.1371/journal.pone.0229003>

*The state of data journalism survey 2023*. (2023). European Journalism Centre.

Tong, J. (2024). Data journalism and audience engagement: Introduction to the special issue.

*Journalism*, 14648849241256457. <https://doi.org/10.1177/14648849241256457>

United Nations Educational, Scientific and Cultural Organization. (2021). *UNESCO Recommendation on Open Science* (SC-PCB-SPP/2021/OS/UROS; p. 34). United Nations Educational, Scientific and Cultural Organization. <https://en.unesco.org/science-sustainable-future/open-science/recommendation>

Wang, Q., Guan, T., & Yan, X. (2023). The “conditional unbiased” of data journalism in China’s News Industry. *Digital Journalism*, 1–20. <https://doi.org/10.1080/21670811.2023.2278045>

*What is open data?* (n.d.). <https://opendatahandbook.org/guide/en/what-is-open-data/>

Wiley, S. K. (2023). The grey area: How regulations impact autonomy in computational journalism.

*Digital Journalism*, 11(6), 889–905. <https://doi.org/10.1080/21670811.2021.1893199>

Wright, S., & Nolan, D. (2021). The development of data journalism in China: Influences, motivations and practice. *Digital Journalism*, 0(0), 1–18. <https://doi.org/10.1080/21670811.2021.1927779>

Wu, S. (2021). Data “objectivity” in a time of coronavirus: Uncovering the potential impact of state influence on the production of data-driven news. *Digital Journalism*, 9(9), 1303–1320.

<https://doi.org/10.1080/21670811.2021.1942111>

Zamith, R. (2019). Transparency, interactivity, diversity, and information provenance in everyday data journalism. *Digital Journalism*, 7(4), 470–489. <https://doi.org/10.1080/21670811.2018.1554409>

- Zamith, R. (2023). Open-source repositories as trust-building journalism infrastructure: Examining the use of GitHub by news outlets to promote transparency, innovation, and collaboration. *Digital Journalism*, 0(0), 1–22. <https://doi.org/10.1080/21670811.2023.2202873>
- Zhang, S., & Feng, J. (2019). A step forward? Exploring the diffusion of data journalism as journalistic innovations in China. *Journalism Studies*, 20(9), 1281–1300. <https://doi.org/10.1080/1461670X.2018.1513814>
- Zhang, S., & Wang, Q. (2022). Refracting the pandemic: A field theory approach to chinese journalists' sourcing options in the age of COVID-19. *Digital Journalism*, 10(6), 1115–1134. <https://doi.org/10.1080/21670811.2022.2029521>
- Zuiderwijk, A., & Reuver, M. de. (2021). Why open government data initiatives fail to achieve their objectives: Categorizing and prioritizing barriers through a global survey. *Transforming Government: People, Process and Policy*, 15(4), 377–395. <https://doi.org/10.1108/TG-09-2020-0271>