Fleerackers, Alice et al.

Preprint

Open data journalism: A narrative synthesis of how, when, and why data journalists use open data sources

SocArXiv

Suggested Citation: Fleerackers, Alice et al. (2023) : Open data journalism: A narrative synthesis of how, when, and why data journalists use open data sources, SocArXiv, Cornell University, Ithaka, https://doi.org/10.31235/osf.io/wh8jx

This Version is available at:
http://hdl.handle.net/11108/578

Open data journalism: A narrative synthesis of how, when, and why data journalists use open data sources

Alice Fleerackers[1], Natascha Chtena[2], Monique Oliveira[3], Isabelle Dorsch[4], Juan Pablo Alperin[2], Stephen Pinfield[5]

[1]Interdisciplinary Studies, Simon Fraser University, Vancouver, Canada, [2]School of Publishing, Simon Fraser University, Vancouver, Canada, [3]Laboratory of Advanced Studies in Journalism, State University of Campinas, Campinas, Brazil, [4]ZBW – Leibniz Information Centre for Economics, Kiel, Germany, [5]Information School, University of Sheffield, Sheffield, UK

Correspondence concerning this article should be addressed to Alice Fleerackers, Interdisciplinary Studies, Simon Fraser University, Vancouver, BC, V6B 5K3. Email: afleerac@sfu.ca

Alice Fleerackers https://orcid.org/0000-0002-7182-4061

Natascha Chtena https://orcid.org/0009-0006-3586-5372

Monique Oliveira https://orcid.org/0000-0002-7642-0971

Isabelle Dorsch https://orcid.org/0000-0001-7391-5189

Juan Pablo Alperin https://orcid.org/0000-0002-9344-7439

Stephen Pinfield https://orcid.org/0000-0003-4696-764X

Acknowledgements

We would like to thank Esteban Morales and Kathleen Gregory for their thoughtful suggestions for improving the manuscript.

Funding statement

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.

Disclosure statement

The authors report there are no competing interests to declare.

Open data journalism: A narrative synthesis of how, when, and why data journalists use open data sources

In this exploratory descriptive review, we synthesize the limited body of literature that has investigated whether, how, and why data journalists engage with open data. Fuelled by the open science movement, research and government data are increasingly shared online. While data journalists are ideally positioned to turn these datasets into compelling, accessible, and socially relevant stories, little is known about their use of these information sources and how this might affect their audiences. Through a critical reading of the literature, we provide an overview of key findings and gaps, as well as directions for future research. Given the rise of open data during the COVID-19 pandemic and the role data journalists have played in reporting on the evolving crisis, we also explore differences in literature published before and during the COVID-19 pandemic. We find that theoretically grounded, mixed methods 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 journalistic cultures, media market structures, and political contexts. Such research could enable scholars to better support data journalists and their audiences in taking advantage of open data.

Keywords: data journalism, open data, open science, sources, COVID-19

Introduction

Increasingly, scholars are starting to explore the potential benefits of open science (OS)—i.e., the movement to make scientific research and its dissemination accessible to all levels of an inquiring society—for individuals and communities outside the academic field. Haim and Puschmann (2023), for instance, argued that openly sharing data, code, and other research products isn't just helpful for increasing research transparency and combating the 'replication crisis' in science, but also for those outside academia who have a stake in scientific discovery and progress, including journalists and the public. Lauerer (2022) similarly proposed that sharing research data openly provides "non-members of the scientific community" (p. 395) important opportunities to examine their own research questions and hold those in power to account.

Data journalists are an obvious example of the type of 'non-members' who would benefit from access to open data (OD), notably open research data (ORD) and open government data (OGD), 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 (Cheruiyot & Ferrer-Conill, 2018; Pereira & Mastrella, 2022). The potential benefits of such access became even more relevant during the COVID-19 pandemic, as organizations around the world committed to sharing data "as rapidly and widely as possible" to curb the spread of the virus (Carr, 2020) and data journalists became key brokers of COVID-19-related knowledge (Pentzold et al., 2021).

Yet, despite this clear *potential* for OD to benefit data journalists—especially during a public health emergency like the COVID-19 pandemic—we know very little about whether, how, and why they engage with these information sources. This lack of knowledge is surprising, given increasing rates of data sharing in academia (Kaiser & Brainard, 2023; Khan et al., 2023; Tenopir et al., 2020) and government (Zuiderwijk & Reuver, 2021); the data-driven nature of data journalism stories, particularly those related to COVID-19 (Wu, 2021); and the continued challenges data journalists face in accessing high quality data, especially those working with greater resource constraints around the world (Mutsvairo, 2019; Wright et al., 2019). Moreover, examining data journalists' engagement with OD has the potential to provide insight into the epistemologies of this evolving profession (Cheruiyot & Ferrer-Conill, 2018; Westlund & Hermida, 2021).

In light of these gaps, we conducted an exploratory descriptive review of the limited body of scholarship that has investigated data journalists' use of ORD and OGD. With a goal of guiding future research in this area, we focus on both synthesising key findings from this emerging body of work and identifying the most pressing gaps. To do so, we searched Google Scholar for literature published since 2018 using keywords such as "data journalism," "open data," "access," "transparency," and "data sources." We read through these articles, retained those that were relevant, and identified additional

literature by checking their reference, including seminal articles from outside of our search dates. This first phase of searching allowed us to first establish a 'baseline' of data journalists' OD usage pre-pandemic (Lee, 2021), which we then compared to their pandemic usage by gathering additional literature using COVID-19-related keywords (e.g., "pandemic," "coronavirus"). Throughout, we actively worked to seek out relevant literature from regions outside of the Global North. However, given known biases in Google Scholar's search algorithm, our review was largely restricted to English-language literature (Rovira et al., 2021). We then applied an adapted qualitative metasummary approach that involved summarizing and grouping concepts and results (Sandelowski & Barroso, 2007) to synthesize key findings and arguments represented within the literature. As we found few, if any studies, that explicitly examined the connection between OD and data journalism, we drew heavily on literature that addresses this topic tangentially, rather than as a focus.

We found that scholars have only begun to examine whether, how, and why data journalists engage with OD, with a particular lack of research that investigating their use of ORD. In addition, although the pandemic has served to raise interest in discussions about the value of OS, data sharing, and journalism, it does not seem to have encouraged scholars to turn their attention to the intersection of these interconnected domains. As is the case with data journalism scholarship in general (Ausserhofer et al., 2020; Wright et al., 2019) we noted a lack of geographic diversity within the literature, with most studies focusing on the US and Europe. We conclude this paper with recommendations for future research, including rigorous, cross-cultural investigations into journalists' preferred sources and uses of OD, the barriers they face in accessing or reporting on OD, their adherence to OS and open-source values, and the impact of open data journalism on audiences—especially in the wake of the COVID-19 pandemic.

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, and philosophies (Coddington, 2018).

The COVID-19 pandemic brought about an increased interest in OD (Gutierrez & Li, 2020; OECD, 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, the movement 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 OS (UNESCO, 2021), which seeks to make scientific processes and outputs—including research data—available and accessible to the broadest possible audience (Bartling & Friesike, 2014). 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). However, studies suggest that (re)use of ORD is hampered by cultural, legal, organizational, and infrastructural factors (Pasquetto et al., 2019; Wallis, 2014). More broadly, much of the potential of OD remains untapped (Collins & Alexander, 2022; Temiz et al., 2022).

Data journalism and OD have the potential to benefit one another in several ways. First, because data journalism is not possible without high-quality, accessible data, it is likely that journalists working in this area could benefit from ORD (Wormer, 2018). Second, scholars have argued that data journalism is premised on values that align with open-source culture, and by extension, OS culture, such as transparency and openness (S. C. Lewis & Usher, 2013; Loosen et al., 2020), some have argued that the defining elements of data journalism are its "use of open-source tools and open data" (Zhang & Feng, 2019, p. 1283). Finally, because data journalists have the skills to make large data sets understandable, share data publicly, and actively engage citizens, they are well-equipped to be the *empowering data intermediaries* (Baack, 2015) that OD advocates have argued are essential for democratizing knowledge (Cheruiyot et al., 2019; Ekaputra et al., 2017). This connection to democracy also aligns with the function of journalism within the US and many other societies (Kovach & Rosenstiel, 2021; Pereira & Mastrella, 2022).

Data journalists' use of open data before the pandemic

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). Much of the existing research suggests that data journalism stories draw most heavily on publicly available data provided by government institutions and political actors (i.e., OGD), while data from universities or academic repositories (i.e., ORD) are used less frequently (or not at all). 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 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% relied on data from other non-commercial organizations, including universities and research institutes and non-governmental organizations (stories could cite multiple data sources) (Loosen et al., 2020).

Journalists' reliance on government datasets is surprising given the difficulties they face in accessing these data (Borges-Rey, 2016; Porlezza & Splendore, 2019; Zhang & Feng, 2019). Government data are also often shared in processed forms (e.g., as PDFs rather than spreadsheets), making it difficult for journalists to process, verify, and use—particularly in the Global South (N. P. Lewis & Nashmi, 2019; Muthmainnah et al., 2022). Even researchers have had to resort to alternative tools for gathering data due to the lack of publicly available official data in Venezuela, such as using Google Trends to estimate the number of Zika virus cases (Strauss et al., 2020). Some strategies journalists have developed for overcoming data access barriers include supplementing openly available datasets with data they have collected themselves (e.g., via web scraping), gathered in partnership with a third party (e.g., an NGO), or received via a leak by a source (Muthmainnah et al., 2022; Porlezza & Splendore, 2019).

Still, poor-quality, incomplete, and out-of-date OGD sources have resulted in a situation where Freedom of Information (FOI) requests remain the method of choice for gathering government data for many journalists (Stoneman, 2015), In countries where FOIs are not possible, lack of access to reliable, usable data remains a common barrier for data journalists (N. P. Lewis & Nashmi, 2019; Mutsvairo, 2019; Wu, 2021). Geographic and cultural contexts may play a role here; journalists in Lewis and Nashmi's (2019) study reported that they preferred to rely on data from the Western NGOs, because they deemed this data more trustworthy than locally available data. Media access to OGD also appears to be particularly compromised in countries with little or no transparency infrastructure supporting OGD, 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., 2022, p. 16).

While government data is typically data journalists' preferred data source, one study of Indonesian journalists found that research data sometimes played a supplementary role; when government data about a newsworthy topic was not available, journalists explored other options, such as data from scientific journals (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, 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% of the visualizations used academic or university data. 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; Peters et al., 2016).

It is important to note that some of the reported low levels of OD use may be because researchers have not used terms such as "open research data," "open government data," or "open repository" 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.

Journalists' apparent lack of engagement with ORD specifically may also 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). 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 not often 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.

Beyond the use of ORD, there is limited demonstrated engagement within data journalism with the values underpinning OS and OD. While some data journalists appear to have "embrace[d] open-source ideals" by working collaboratively with individuals internal and external to their organizations (Borges-Rey, 2016, p. 6), in other studies, large collaborations were found to be the exception rather than the rule (Young et al., 2018). Although many data journalists actively use Github as part of their working routines, this use does not appear to generate many interdisciplinary collaborations (Haim & Zamith, 2019). Other OS values, such as transparency and openness, appear to be much less relevant in practice than scholars have proposed. Some studies have documented a strong commitment to transparency and openness, both in the meta-journalistic discourse about data journalism and in data journalism stories themselves (Pereira & Mastrella, 2022; Stalph & Heravi, 2021). However, others have found that between 40% and 90% of data journalism stories do not cite their data sources or provide access to the underlying data (Loosen et al., 2020; Young et al., 2018), with one analysis of 290 stories

failing to find a single hyperlink to a data source (Zhang & Feng, 2019). Similarly, more than 80% of stories in Zamith's (2019) study did not offer any additional information about the nature of the data or the methods that were used to collect and analyze it, again suggesting a lack of transparency. Boyles and Meyer (2016) found that, while many data journalists believed in the value of making datasets publicly available, only a handful did so in practice—likely due to a lack of skills or knowledge.

Some evidence suggests that embodiment of OS 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-US-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). Another study found that journalists employed at smaller, local media outlets tended to rely more on collaborations with external developers or third-party organizations than those at larger, national outlets, presumably because they had fewer in-house resources and skills (Borges-Rey, 2016). 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 OS values into practice (Cheruiyot et al., 2019; Cheruiyot & Ferrer-Conill, 2018).

Assessing whether data journalists have lived up to their potential by acting as "empowering data intermediaries" who make data understandable, publicly available, and engaging (Baack, 2015) is more challenging, as relatively few studies have analyzed audience responses to data journalism (Ausserhofer et al., 2020). Research has found that data journalists see it as part of their professional role to render otherwise "chunky" datasets more understandable and to "promote discussions around the data so that audiences can fully comprehend the final product" (Boyles & Meyer, 2016, p. 949). However, in

practice, journalists' limited computational skills and access to software can prevent them from achieving these goals (Boyles & Meyer, 2016). In this respect, data "journalistic output appears to be shaped less by what could be considered the best way of representing/exploring the data and more by what can be done and is available for free" (Young et al., 2018, p. 127). Some newsrooms have found that data journalism stories are among the only journalistic products that audiences are willing to pay for, meaning that these stories—which are ideally positioned to make publicly available data more "accessible"—may paradoxically be the stories most likely to end up behind paywalls (Boyles & Meyer, 2016).

Data journalists' use of open data during the pandemic

The COVID-19 pandemic underscored the importance of access to high quality data for supporting informed public decision making (Cancela-Kieffer, 2021) and the role that data journalism can play in facilitating that access (Desai et al., 2021). Journalists were among the first to collect, analyze, and communicate 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). That is, data journalists not only helped to describe the early impacts of COVID-19 but also forecast potential outcomes; they provided a "first rough draft of history" (Shafer, 2010), as well as sketches of a post-pandemic future.

While COVID-19 highlighted the value of data journalism, it also brought new attention to barriers facing data journalists, such as lack of access to high quality data (Stollorz, 2021; Wu, 2021; Zhang & Wang, 2022). For example, in Brazil, journalism companies formed a consortium for data collection during the pandemic due to unexplained delays and blackouts on official data platforms ("Ação Colaborativa Une Veículos Para Evitar Apagão de Dados Sobre Pandemia," 2020). Beyond a

lack of physical access, a lack of training and skills for critically vetting and contextualizing data also impeded journalists' reporting on the pandemic (Cancela-Kieffer, 2021). This sense—that COVID-19 revealed both strengths and weaknesses of journalism—is not specific to data journalism but is rather an attribute of the wider meta-journalistic discourse that emerged during the pandemic (Perreault & Perreault, 2020). Journalists have noted that access to data is both a question of availability and of ability, in that "basic data skills" are needed to access OD portals (Cancela-Kieffer, 2021). Skill, Cancela-Kieffer (2021) argued, is also important for being able to verify, interrogate, critique, and put data into context. Such critical data journalism is, in turn, needed to support audiences in making informed decisions about their health and safety (Nguyen, 2018). While some data journalists were able to provide these kinds of in-depth investigations into the nature of COVID-19 data (Desai et al., 2021), others simply amplified information provided by official sources: they "took the numbers for granted, without raising questions…they did not contextualize and forgot to mention how the raw data was gathered" (Cancela-Kieffer, 2021, p. 161). Different political contexts and the nature of relationships between governments and media outlets may explain differences in how (un)critical journalists in different countries were of the official data provided to them during the pandemic (Wu, 2021).

Much like pre-pandemic data journalism, COVID-19-related data stories tended to lean heavily on a narrow set of institutional sources: namely, national public health agencies, academic medical institutions, and the World Health Organization (WHO) (Pentzold et al., 2021; Wu, 2021). In the US, many journalists drew extensively on one institutional source, in particular: the COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University (Dong et al., 2020). This interactive online dashboard was appreciated by journalists because "it was well sourced and verifiable" (anonymous Qatar-based editor, quoted in Blanco & Sauras, 2021), but perhaps also because—as data published under a CC BY 4.0 license—it was free to use and adapt, pending appropriate attribution. Even as other primary data sources that met these criteria became available,

journalists reported that they lacked the time, resources, and skills needed to seek them out and use them (Blanco & Sauras, 2021). Other scholars have suggested that the narrow reliance on the same few institutional data sources may also have been an indication of "a data bottleneck with only few powerful public agencies providing that kind of information" (Pentzold et al., 2021, p. 1385). ORD published in institutional repositories were not identified as data sources by either Pentzold et al. (2021) or Blanco & Sauras (2021). In Germany, China, and Singapore, the lack of publicly available data was a barrier for data journalists covering the pandemic (Stollorz, 2021; Wu, 2021).

Data journalism seen during the pandemic has also been critiqued for being ad-hoc and unsystematic, potentially contributing to high levels of misinformation (Westlund & Hermida, 2021). Cabreros (2021) argued that data journalists were increasingly "doing science" while "side-stepping safeguards" afforded by the academic system, resulting in a situation where "the quickest and least-vetted science also wields the megaphone." To address this risk, Cabreros (2021) argued that data journalists should become more open and transparent about their data sources and analytical methods, again suggesting that data journalism has continued to fall short of the open source values that are often ascribed to it (Baack, 2015; S. C. Lewis & Usher, 2013). There is some anecdotal evidence that certain journalists and outlets may have adopted these values during the pandemic. For example, the Chinese outlet *The Paper* launched its own "open-data initiative" to provide access to raw COVID-19-related datasets, which journalists and audience members could use to perform their own analyses (Zhang & Feng, 2019). However, it is difficult to assess how common such initiatives have been during the pandemic, given a lack of research on this topic. While one recent (non-pandemic-related) study suggests that publishing data archives or repositories may be "an important trend in journalism" (Abhishek & Graves, 2023, p. 18), more research is needed to assess whether this "trend" is likely to persist.

Concluding remarks and recommendations for future research

Data journalism is still a developing research area, with many known gaps. Specifically, very little is known about equity-related aspects that may be relevant to data journalists' use of OD, nor about data journalism practices outside of Europe and North America. Research on data journalism in the Global South is limited in general (Ausserhofer et al., 2020), and has often relied primarily on exploratory, qualitative, single-nation case study approaches (Wright et al., 2019). We thus join scholars in calling for more geographically diverse, theoretically-grounded, mixed methods research (Ausserhofer et al., 2020; Wright et al., 2019) to yield a more globally relevant understanding of data journalism and its connection to OD. This is especially important given the varied forms and practices of data journalism that can emerge in different journalistic cultures, media market structures, and political contexts (Appelgren et al., 2019).

Additionally, our review of the literature revealed that more research is needed to understand how journalists across geographies use OD, including the sources who are creating and sharing these datasets (e.g., governments, non-governmental organizations, academics), the degree to which data journalists and their organizations support open source values and practices, the barriers journalists face in using OD, and the impacts of this use on audiences. We also lack insight into whether journalists' use of OD is limited to communicating and reorganizing information, or whether more critical or generative roles are possible.

Finally, our review revealed that we know very little about the influence of the COVID-19 pandemic on data journalists' use of OD. The few studies we identified suggest that the widespread support for openness seen during the pandemic has done little to reduce the barriers data journalists' face in accessing and using data, at least in nations such as the US and China. Why these barriers persist, whether they are universal across cultures or media outlets, and what, if anything, could help move the needle remain open questions. Addressing these known gaps could help build an understanding of the

degree to which data journalists act as data intermediaries in times of crisis and whether this has the intended effect of "empowering" audiences to engage with, understand, and use data (Baack, 2015). More broadly, critical, empirically grounded research at the intersection of OD and data journalism is essential if we truly hope to "make the debate around Open Science center-stage in journalism research" and "add a new perspective on Open Science that takes the specifics of digital journalism research seriously" (Haim & Puschmann, 2023, p. 2).

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

Appelgren, E., Lindén, C.-G., & van Dalen, A. (2019). Data journalism research: Studying a maturing field across journalistic cultures, media markets and political environments. *Digital Journalism*, *7*(9), 1191–1199. https://doi.org/10.1080/21670811.2019.1685899

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

Blanco, A., & Sauras, J. (2021, June 1). Drawing the curve: Data visualization and COVID-19: A pandemic year in graphics. *Tow Center Weekly Newsletter*. https://mail.google.com/mail/u/0?ui=2&ik=a70fc9b791&view=lg&permmsgid=msg-f%3A1701395960013595419&ser=1

Borges-Rey, E. (2016). Unravelling data journalism. *Journalism Practice*, *10*(7), 833–843. https://doi.org/10.1080/17512786.2016.1159921

Boyles, J. L., & Meyer, E. (2016). Letting the data speak: Role perceptions of data journalists in fostering democratic conversation. *Digital Journalism*, *4*(7), 944–954. https://doi.org/10.1080/21670811.2016.1166063

Cabreros, I. (2021, April 26). Side-stepping safeguards – Data journalists are doing science now. *Impact of Social Sciences*. https://blogs.lse.ac.uk/impactofsocialsciences/2021/04/26/side-stepping-safeguards-data-journalists-are-doing-science-now/

Camaj, L., Martin, J., & Lanosga, G. (2022). 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

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

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

Coddington, M. (2018). Defining and Mapping Data Journalism and Computational Journalism: A review of typologies and themes. In *The Routledge Handbook of Developments in Digital Journalism Studies*. Routledge.

Collins, A., & Alexander, R. (2022). Reproducibility of COVID-19 pre-prints. *Scientometrics*, *127*(8), 4655–4673. https://doi.org/10.1007/s11192-022-04418-2

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

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Ekaputra, F. J., Novak, N. M., Kiesling, E., Aryan, P., Do, B., Trinh, T., & Tjoa, A. (2017). Towards open data mashups for data journalism. *"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), 075002. https://doi.org/10.1289/EHP4808

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*.

Haim, M., & Puschmann, C. (2023). Opening up data, tools, and practices: Collaborating with the future: Introduction to the special issue analytical advances through Open Science: employing a reference dataset to foster best-practice data validation, analysis, and reporting. *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

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

Kaiser, J., & Brainard, J. (2023). Ready, set, share: Researchers brace for new data-sharing rules [Data set]. In *Science Magazine* (Vol. 379, Issue 6630). https://doi.org/10.1126/science.adg8470

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

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*(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

Lee, E.-J. (2021). Making sense of pandemic-induced changes in journalism and beyond. *Digital Journalism*, *9*(9), 1431–1437. https://doi.org/10.1080/21670811.2021.1997149

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

Lewis, S. C., & Usher, N. (2013). Open source and journalism: Toward new frameworks for imagining news innovation. *Media, Culture & Society*, *35*(5), 602–619. https://doi.org/10.1177/0163443713485494

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

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

Nguyen, A. (Ed.). (2018). *News, numbers and public opinion in a data-driven world*. Bloomsbury Publishing.

OECD. (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

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and reuses of scientific data: The data creators' advantage. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.fc14bf2d

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

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. (2016). Research data explored: An extended analysis of citations and altmetrics. *Scientometrics*, *107*(2), 723–744. https://doi.org/10.1007/s11192-016-1887-4

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. 263–287prie). MIT Press. https://doi.org/10.7551/mitpress/9445.003.0019

Shafer, J. (2010, August 31). Who said it first? *Slate*. https://slate.com/news-and-politics/2010/08/on-the-trail-of-the-question-who-first-said-or-wrote-that-journalism-is-the-first-rough-draft-of-history.html

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

Strauss, R., Lorenz, E., Kristensen, K., Eibach, D., Torres, J., May, J., & Castro, J. (2020). Investigating the utility of Google trends for Zika and Chikungunya surveillance in Venezuela. *BMC Public Health*, *20*(1), 947. https://doi.org/10.1186/s12889-020-09059-9

Temiz, S., Holgersson, M., Björkdahl, J., & Wallin, M. W. (2022). Open data: Lost opportunity or unrealized potential? *Technovation*, *114*, 102535. https://doi.org/10.1016/j.technovation.2022.102535

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

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

Wallis, J. (2014). Data producers courting data reusers: Two cases from modeling communities. *International Journal of Digital Curation*, *9*(1), Article 1. https://doi.org/10.2218/ijdc.v9i1.304

Westlund, O., & Hermida, A. (2021). Data journalism and misinformation. In *The Routledge Companion to Media Disinformation and Populism* (pp. 142–150). Routledge.

Wormer, H. (2018). Mind the statistics gap: Science journalism as a bridge between data and journalism. In A. Nguyen (Ed.), *News, numbers and public opinion in a data-driven world* (pp. 226–241). Bloomsbury Academic.

Wright, K., Zamith, R., & Bebawi, S. (2019). Data journalism beyond majority world countries: Challenges and opportunities. *Digital Journalism*, *7*(9), 1295–1302. https://doi.org/10.1080/21670811.2019.1702472

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

Young, M. L., Hermida, A., & Fulda, J. (2018). What makes for great data journalism? A content analysis of data journalism awards finalists 2012–2015. *Journalism Practice*, *12*(1), 115–135. https://doi.org/10.1080/17512786.2016.1270171

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

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