Opinion

It Takes a Village! Editorship, Advocacy, and Research in Running an Open Access Data Journal

Mandy Wigdorowitz 1,2, Marton Ribary 3,*, Andrea Farina 4, Eleonora Lima 5, Daniele Borkowski 6, Paola Marongiu 7, Amanda H. Sorensen 8, Christelle Timis 9 and Barbara McGillivray 4

1 Theoretical and Applied Linguistics, University of Cambridge, Cambridge CB2 1TN, UK; mwigdorowitz@ua.edu or mw738@cam.ac.uk
2 Department of Psychology, University of Johannesburg, Johannesburg 2092, South Africa
3 Department of Law and Criminology, Royal Holloway, University of London, London TW20 0EX, UK
4 Department of Digital Humanities, King’s College London, London WC2R 2LS, UK; andrea.farina@kcl.ac.uk (A.F.); barbara.mcgillivray@kcl.ac.uk (B.M.)
5 Trinity College Dublin, The University of Dublin, D02 PN40 Dublin, Ireland; limae@tcd.ie
6 Department of Greek & Latin, University College London, London WC1E 6BT, UK; daniele.borkowski.23@ucl.ac.uk
7 Institute of Language Sciences, University of Neuchâtel, 2000 Neuchâtel, Switzerland; paola.marongiu@unine.ch
8 College of Information (INFO), University of Maryland, College Park, MD 20742, USA; asorens1@umd.edu
9 Department of Contemporary and Digital History, University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg; christelle.timis.001@student.uni.lu
* Correspondence: marton.ribary@rhul.ac.uk

Abstract: Partaking in the editorial process of an academic journal is both a challenging and rewarding experience. It takes a village of dedicated individuals with a vested interest in the dissemination and sharing of high-quality research outputs. As members of the editorial team of an open access data journal, we reflect on the emergence of data-driven open research, a new journal genre (data paper), and a new journal type (data journal) in the Arts, Humanities, and Social Sciences (AHSS). Access to data—the currency of empirical research—is valuable to the research community, crucial to scientific integrity, and leads to cumulative advancements in knowledge. It therefore requires significant investment and appropriate venues for dissemination. We illustrate the necessity of raising awareness about data-driven open research and best practices in data-driven publishing. We discuss how it involves building a community of authors and readers, establishing a company of editors, reviewers, and support staff, and passing on the practice, which has been challenging the status quo in research and publishing. Potential future directions are considered, including data peer review and reward, recognition, and funding structures for data sharing.

Keywords: open research; open access; data paper; data journal; arts; humanities; social sciences; publishing

1. Introduction

Partaking in the editorial process of an academic journal is both a challenging and rewarding experience. It requires selfless commitment, energy, and time from many people working in unison and without financial incentive. It takes a village of dedicated individuals with a vested interest in the dissemination and sharing of high-quality research outputs. This is true for all types of journals, whether established, new, or on the brink of obsolescence. Without a team of people running a journal, it is unlikely to succeed in the ever-competitive publishing ecosystem. In this paper, we reflect on our experiences of being on the editorial team of the Journal of Open Humanities Data (JOHD), an open access (OA) data journal that broadly publishes in the Arts and Humanities. While our experience
is associated with this journal, this paper can be more generally applied to data-driven open research (OR) publishing in the Arts, Humanities and Social Sciences (AHSS).

The emergence of JOHD as a data journal may be considered part of a trend from the late 2010s, which has made quantitative research methods an important part of the Humanities, resembling methods used in the Sciences. This is mostly due to digitization, which favors an easier and more systematic collection of data for the Humanities [1–3], and consequent factors such as the growing importance of quantitative methods in Humanities research [4,5], and the increasing interdisciplinary collaboration between the Humanities and the Sciences [6,7]. This is also reflected in empirical and scientific methods that have been adopted by Humanities research [8].

Through collective effort led by an enthusiastic Editor-in-Chief, the editorial team of JOHD were able to revive this virtually inactive journal and transform it into a thriving publishing venue for data papers and discussions around open data sharing that touch on broad topics in the OR space. The growth trajectory of the journal was largely the result of a complete editorial change in 2019, where the journal was reinvigorated because of the effort and care of its editorial and community members (see Figure 1 showing the number of articles published in JOHD over time). While the journal’s article count is comparably lower than data papers in the Sciences, this number should be interpreted in the context of it being a relatively new journal with a recent editorial change, of a very particular kind (data journal), run entirely by unpaid volunteers, and one that publishes in Arts and Humanities, which are newcomers to data sharing and have consistently received less funding compared to the Sciences.

![Figure 1. Number of papers published by the Journal of Open Humanities Data from 2015 to 30 July 2024. Publications have increased steadily from 2019 when the editorship changed.](image)

The wider community of the data journal includes the editorial team and support staff as well as the authors, reviewers, and readers. Sustaining this journal and its associated community takes more than just having someone subsuming an editorial role; it requires that the community itself be nurtured and built upon so that this publishing tradition can be passed on to the next generation of scholarly beneficiaries. Similarly, editorial teams running data journals cannot simply wait for authors to discover them. They need to advocate for embracing the principles of OR and corresponding dissemination practices. Becoming involved with a data-driven journal means embracing new publishing methods and spreading the word about them in academic circles and beyond. As our experience shows, this is not perceived as a burdensome task, but a great opportunity to
raise awareness about OR and data sharing. It additionally allows scholars to grow their academic profile and contribute to a worthwhile publishing endeavor.

In what follows, we illustrate how data have garnered significant interest in the research landscape and have accordingly led to a new genre of output (data paper) and a new journal type where they are housed (data journal). We then present examples of data sharing and associated initiatives and activities that the editorial team have partaken in as a consequence of their involvement with the journal. We end with a note on emerging opportunities in this area, the importance of passing the tradition on, and how we see that once revived, an OA data-driven journal and its associated OR culture can be sustained.

2. A Place for Data Sharing and Data Journals

Typically, journal publishing models rely on voluntary efforts from researchers who have expertise in a journal’s subject matter to oversee the publication cycle of manuscripts. Journals are heavily reliant on the labor and fervor of the academic community to drive research forward, stay relevant, and ensure the journal’s long-term success. Traditionally, data-driven research articles follow a standard structure, where research questions/hypotheses are posed, a methodology is described, the data are analyzed, results are presented, and the findings are discussed with reference to the proposed questions/hypotheses. Enormous effort and care are involved in collecting and managing the data used to make scientific claims. As such, the research community has increasingly realized the value and impact of data in their own right (e.g., [9–11]).

In parallel, there are more datasets available than ever before [12,13]. Accordingly, interest in data production, curation, sharing, and reuse has risen and, in turn, created a need for publication outlets with a data-focused goal. This is especially the case given that data availability is relevant for research transparency and reproducibility and sharing data aligns with the ethos of the OR movement [14,15]. Grant requirements are also motivating the practice of data sharing (e.g., [16,17]). So too is the fact that data sharing is increasingly guided by conscientious principles, ensuring that what is shared is ‘as open as possible, as closed as necessary’ [18]. Most notably are the FAIR (Findable, Accessible, Interoperable, Reusable) principles, which were proposed for scientific data management and stewardship [19]. More recently, the CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles were recommended in relation to Indigenous knowledges [20], the CORE (Collected, Organized, Recontextualized, Explained) principles for Humanities data [21], and the DEAR (Dialogic, Emergent, Abundant, Relational) principles for qualitative data [22].

Even so, gaps in data sharing and discoverability still persist [23], which can largely be attributed to the lack of professional incentives to share data [24] or a fear of the data being ‘scooped’ by other researchers [25]. However, the tide is slowly turning with researchers expressing more favorable sentiments about data sharing and actually changing their behaviors by depositing their datasets into repositories [16]. There are also external forces and incentives contributing to these changes such as the European Union’s Plan S, and the ever-growing OA requirements mandated by the UK’s periodic Research Excellence Framework. In addition, grassroot initiatives such as the Research Data Alliance and Make Data Count champion data-sharing principles and advocate for suitable ways to identify, evaluate, and reward these practices.

One result of this landscape is the emergence of a newer type of journal, the data journal. Data journals play a pivotal role in challenging the status quo about data retention and ownership. They publish articles (also called data papers, data articles, data notes, dataset descriptors; see Appendix A in [26]) that describe the compilation, storage, metadata, and reuse potential of proficiently curated datasets that are openly available, free to access, and have a persistent identifier such as a DOI [27,28]. They provide a platform for researchers to describe datasets in detail and note the potential for the ways in which the data can be used in further investigation and analysis. It is not necessary to report on or interpret the data themselves in a data paper, but reference can be made to research where the data have
been used. Importantly, data papers are peer-reviewed, allowing researchers to receive appropriate and recognized credit for their work.

Data papers are catalysts for data sharing and foster transparency and OR that add to the discourse on data-driven practices and methodologies. They can have a transformative impact on the academic community and on research in general—they challenge traditional concepts of the ‘academic article’ by stressing the multifaceted nature involved in data creation. This is particularly relevant for interdisciplinary research in AHSS, wherein datasets may be reused by multiple disciplines to address different questions.

The number of data journals has significantly grown in recent years as more and more datasets have been generated and shared [26,29,30]. There is an asymmetrical proportion of data papers published in scientific (STEMM-related) journals compared to AHSS journals, and as far as we are aware, only two multidisciplinary journals are specifically dedicated to publishing data papers in AHSS disciplines: JOHD, launched in 20157, and the Research Data Journal for the Humanities and Social Sciences, launched in 20168. In addition, data journals dedicated to specific HSS disciplines include the Journal of Open Archaeology Data, the Journal of Open Psychology Data, and the Journal of Cultural Analytics. Since data journals, specifically those in AHSS, are a relatively new and niche form of academic currency, establishing credibility and fostering widespread recognition within the research community demand substantial effort, largely driven by a FUBU (For Us By Us) approach.

Standard performance metrics biased towards STEMM subjects are ill-suited to measuring the value and quality of such efforts. Instead, we suggest that a more helpful way to look at the quality and impact of journals is in line with the San Francisco Declaration of Research Assessment (DORA)9 and the Coalition for Advancing Research Assessment (CoARA)10. For example, listing a journal in the Directory of Open Access Journals (DOAJ)11 and admission to a reputable indexing service (such as Web of Science) are better signs of quality and added value than crude performance metrics.

AHSS data journals have a rightful place in the publishing ecosystem, but one which needs to be created. Our experience has shown that we cannot simply wait for manuscript submissions. The editorial team needs to initiate a dialogue with authors about data privacy, consent, and responsible data sharing so that the establishment of best practices in data-driven OR in AHSS align with ethical concerns. Editors need to be proactive in creating a market for AHSS data papers and be at the forefront of wider discussions around open data publishing. In the next section, we discuss some of the strategies and initiatives we have leveraged to amass interest in both the authorship and readership of the journal.

3. Intertwining Editorship, Advocacy, and Research

The roles in the editorial team of a data-driven journal are varied, but all members need to work together for the journal to succeed. In addition to the external editorial board, the journal’s team includes the Editor-in-Chief, senior and associate editors, editorial assistants, copy editors, and social media and newsletter editors. These roles may resemble those of journal editors in the contemporary research landscape, but may differ from journal to journal depending on each journal’s needs and goals (e.g., [31–33]). In terms of JOHD’s focus, we have tailored roles to ensure oversight of the best practice in data sharing, quality control via our in-house copy editing offering, and exposure and outreach through various channels and activities. We aim to stay abreast of the editorial landscape and do so by evolving our own editorial team (e.g., recruiting members, creating new roles when needed) in response to best practice needs in (data) publishing. As a team, we drafted a code of conduct outlining the responsibilities and benefits of volunteering for the journal, which each team member has to adhere to.

We are also keenly expanding our efforts to systematically address data peer review, which has been a largely neglected area of editorial focus in general. Data peer review extends the rigor of peer review applied to (data) papers to the datasets themselves by involving professional data curators, librarians, and other experts in a process traditionally dominated by researchers [34]. Given the limited investigation into data
peer review [35–37], a recent pilot project, in which members of our editorial team were involved, looked at different review models including “editorial”, “scientific”, and “technical” review and proposed that data peer review may need to adopt a hybrid circular model (“publish–review–curate”) involving all stakeholders (publishers, research peers, and data curators) [38]. This is a starting point in developing this area further, but more needs to be done to ensure data peer review meets a high enough and uniform standard. Currently, editors carry out preliminary data peer review on submissions such as checking that the data are openly available in a reputable repository, and evaluating the structure, metadata, and documentation of the deposited data against a manuscript template before sending them out for review. At the review stage, reviewers are asked (but are not required) to access and comment on the deposited data in addition to the submitted data paper. In the future, the journal seeks to formalize the data peer-review process and is considering having an in-house data curator and potential external data reviewers.

When the current Editor-in-Chief started her collaboration with the journal in 2019, she actively recruited a team of early-career academics to join different aspects of the editorial work. Given the voluntary nature of editorship, it soon became evident that maintaining motivation is a crucial part of members’ long-term commitment. This has been facilitated by providing opportunities for members to grow professionally within the journal in a constant dialogue with the Editor-in-Chief, and to ensure that everyone’s role adapts to their career aspirations and experience level. Moreover, having an open communication channel between more and less experienced members of the team creates an atmosphere that fosters mentorship and collaboration. Some of the significant opportunities that have arisen from our connection to the journal include outreach and social media activities, organizing special collections, conducting scientometric research, and gaining experience in grant writing, as detailed below.

3.1. Outreach and Social Media Presence

Outreach activities are coordinated through a consistent strategy to firstly create and then sustain a market for data-driven research and data papers, which help the journal’s reach and growth. Our authors are required to follow highly structured templates and adhere to strict open access and quality guidelines. We regularly provide updates about new publications, participation in conferences, and a variety of activities that members of the journal are involved in. For this purpose, a social media team was established to ensure that sufficient energy and ideas are devoted to the journal’s online presence, first via its X (formerly Twitter) channel, and then also via its YouTube and Instagram. We also have a newsletter that members of the journal’s community receive.

Through tailored social media initiatives, the journal has been expanding its role in the OR community and facilitating dialogue on emerging trends, challenges, and opportunities offered by data sharing and reuse. In particular, the journal’s “show-me-your-data” campaign on X invites authors to share an image of their dataset to demystify the concept of “data” in the Humanities, and to provide additional visibility to their datasets. To evaluate whether our social media activities increase the visibility of data papers, we measured the impact of the “show-me-your-data” campaign on the views and downloads of the JOHD papers featured in the campaign. We then compared these results with the metrics of JOHD papers not featured in the campaign. Our analysis showed that data papers that appeared in the campaign had a significantly higher number of views and downloads compared to those that were not a part of the campaign [28]. Managing our social media presence is an important tool for outreach and advocacy and something all members are encouraged to contribute to, regardless of their editorial role.

3.2. Special Collections

Special collections (SCs) that center on a topic of interest and bring together editorship, advocacy, and research have been instrumental to the journal’s growth. Organizing a SC involves reaching out to specific research communities and raising awareness about
the significance of sharing data and writing a corresponding data paper. SCs give guest editors the opportunity to partake in the editorial process of data papers by pooling their complementary subject knowledge, skills, and experience together to create a relevant collection of data papers that address a specific data need. SCs are led by guest editors who connect with communities already familiar to them, opening up further avenues for research collaboration. External guest editors are assisted by a member of the editorial team to provide authors with guidance in structuring and documenting their data and preparing it for publication.

SCs are generally a great success with research communities who have a sufficient level of data and OA maturity. Our most successful special collection to date is ‘Representing the Ancient World through Data’\textsuperscript{13}, devoted to datasets related to the Ancient World in different subfields (e.g., linguistics, archaeology, history). Other SC themes cover computational humanities, language documentation, COVID-19, and urban archives\textsuperscript{14}. Future SCs will give a platform to diverse datasets such as the galleries, libraries, archives, and museums (GLAM) sector, and those with a focus on the origins, development, and spread of ideas across time and space. The goal of SCs is to improve the journal’s inclusivity and diversity, and raise awareness about data-driven research and dissemination in different sectors with varying levels of OR engagement.

3.3. Scientometric Research

We are particularly interested in whether data papers actually make an uptake and usage difference. With prior experience in scientometric research, some of the editorial team have explored the impact that data papers have on the overall performance of data-driven research. We curated a dataset of what we call ‘golden triangles’, linking (1) datasets, (2) data papers, and (3) research papers that form part of the same research effort [28]. We compared the performance of our ‘golden triangles’ to that of a baseline consisting of only datasets (views and downloads) and only research papers (citations and altmetric scores). We were encouraged to discover a statistically relevant uptick in visibility and impact (views, downloads, citations, altmetric scores) when a data paper is one of the outputs of a research project, confirming our hypothesis that publishing a data paper is worth the effort. Our paper was highlighted in \textit{Publications} as a “feature article” and was awarded the 2022 Best Paper Award for \textit{Publications’} 10th Anniversary. The award comes as a timely boost, inspiring us to continue our research about the performance of data papers and how to establish links between the nodes of the ‘golden triangle’ in an automated way.

3.4. Grant Activities

The editorial team is constantly looking at funding opportunities for scientometric and data-driven research as well as for core publishing activities. We were part of a successful grant application for a pilot project on data peer review initiated by a research team from the University of Science and Technology in Seoul, South Korea. Our research partnership produced internal reports and an informative roundtable discussion on open data publishing in the Digital Humanities at SciDataCon 2023\textsuperscript{15} with input from researchers, publishers, and data infrastructure managers.

Article processing charges (APCs) are often a considerable financial burden to many authors wishing to publish their research openly. The burden often presents itself in communities where the research environment still operates according to the standard subscription-based publishing model, where publishing costs are covered by those consuming research outputs rather than those producing them [39]. One possible solution to expand our scope and coverage is to apply for funding as a collaborative effort between, for example, journal editors, publishers, and learned societies, specifically those representing niche academic disciplines and varied geographical areas. The publisher’s tangible and practical support in the grant-writing process is pivotal to avoid the impression that the efforts of unpaid volunteer editors are going towards securing money for for-profit commercial publishing enterprises. Making an APC-based open data journal accessible to
new communities by utilizing external grants is unchartered territory, but one we are eager to put to the test.

4. Sharing for the Future

It is clear that data sharing is pivotal to transparent and open research practices, not least because the adoption of such practices promises to generate unanticipated opportunities, alternative avenues of scientific exploration, and unique challenges. Data sharing is likely to become an important indicator in the evaluation of research contributions and an incentive for academic promotion and other award structures [40,41]. It therefore requires thoughtful attention and monitoring to ensure it is assessed and rewarded responsibly.

Growing acceptance and implementation of data sharing and data citation have already led to various offshoots of investigation around data practices. For example, recent discussions have focused on responsible data stewardship and data metrics [42,43]. By responsibly tracking the usage and citation of publicly available data over time, it is possible to provide novel insights about the groups who are sharing and reusing data and the rationale (academic, educational, financial, and so on) behind such practices. Another important OR practice arising from data sharing is data peer review, which, as mentioned, we are in the process of formalizing as part of our editorial services. Given the evolving data-driven landscape, we look forward to witnessing future questions, directions, and research that emerge as a result.

In this paper, we have discussed the emergence of data-driven OR in AHSS, which brought a new genre (data paper) and a new journal type (data journal) to life. We have reflected on the importance of building the community and its leadership and raising awareness about data-driven OR. While data journals publishing data papers are a direct response to a dynamic research and publishing context, their status has not yet been fully consolidated. For this reason, it is critical to consider the sustainability of data journals and data papers and how a culture of data sharing can be passed on and embraced by future generations of scholars. Even though the publishing landscape is dominated by an “oligopoly” of a small number of large publishing conglomerates [44], especially in STEMM subjects, data journals in the AHSS associated with smaller publishing houses or even independent ones have the opportunity to find alternative routes for growth and expansion. For example, the Journal of Cultural Analytics, which is published independently by faculty at McGill University and run on the Scholastica publishing platform, is one of the most impactful journals in its field.

Continuous advocacy needs to position data papers as not simply something good to have, but as an invaluable academic asset that is integrated seamlessly into the research workflow. By highlighting their utility, we will continue to encourage AHSS researchers to share their data and submit data papers as the standard way of working. Advocating for data sharing in AHSS contexts involves thoughtful consideration about how to reach more people and how we can work towards removing any stigma about employing so-called ‘scientific methods’ in non-Science disciplines. Like other data-focused journals in AHSS, we are headed towards a future where OR practices, open-access publishing, and transparent data sharing are becoming the norm rather than the exception. It is up to communities like us to ensure their presence is nurtured and preserved.

Author Contributions: Conceptualization, M.W. and M.R.; writing—original draft preparation, M.W. and M.R.; writing—review and editing, M.W., M.R., A.F., E.L., D.B., P.M., A.H.S., C.T. and B.M.; supervision, M.W. and M.R. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the 2022 Best Paper Award for Publications’ 10th Anniversary.

Acknowledgments: We would like to thank Samuel A. Moore for reading an earlier draft of this manuscript and providing valuable comments.

Conflicts of Interest: The authors are volunteer members of the editorial team at the Journal of Open Humanities Data.
Notes

1. For instance, *Scientific Data* published 49 articles in its first year (2014) and has since increased significantly in its number of publications, with 879 articles being published in 2023 (https://www.nature.com/sdata/; accessed on 29 July 2024).

2. For example, according to the H2020 factsheet of the European Research Council (ERC), 2950 projects of a total value of EUR 5.58 billion were funded in Physical Sciences and Engineering, along with 2064 projects of a total value of EUR 3.93 billion in the Life Sciences, compared to 1594 projects of a total value of EUR 2.80 billion in the Social Sciences and Humanities (https://erc.europa.eu/sites/default/files/document/file/FACTSHEET-Overview-2022.pdf, accessed on 29 July 2024).

3. The website of cOAlition S describes the European Commission and the European Research Council’s Plan S as “an initiative for Open Access publishing that was launched in September 2018 [which] requires that, from 2021, scientific publications that result from research funded by public grants must be published in compliant Open Access journals or platforms” (https://www.coalition-s.org/, accessed on 29 July 2024).

4. The upcoming REF 2029 in the UK seeks to broaden the exercise’s open access policy in an open consultation (https://www.ref.ac.uk/guidance/ref-2029-open-access-policy-consultation/, accessed on 29 July 2024).

References


23. Sixto-Costoya, A.; Robinson-García, N.; van Leeuwen, T.; Costas, R. Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: A methodological discussion. *Scientometrics* 2021, 126, 7149–7165. [CrossRef]


30. Walters, W.H. Data journals: Incentivizing data access and documentation within the scholarly communication system. *Insights* 2020, 33, 18. [CrossRef]


32. Nemeth, B.; Cobb, J. Editorial team roles and responsibilities. *Interpretation 2020*, 8, 1N-T1095. [CrossRef]


35. Lawrence, B.; Jones, C.; Matthews, B.; Pepler, S.; Callaghan, S. Citation and peer review of data: Moving towards formal data publication. *Int. J. Digit. Curation* 2011, 6, 4–37. [CrossRef]


40. Araujo, P.; Bornatici, C.; Heers, M. Recognising open research data in research assessment: Overview of practices and challenges (1.0). *Zenodo* 2024. [CrossRef]


42. Lowenberg, D.; Chodacki, J.; Fenner, M.; Kemp, J.; Jones, M.B. Open data metrics: Lighting the fire (Version 1) [Computer software]. *Zenodo* 2019. [CrossRef]
43. Puebla, I.; Lowenberg, D. Building trust: Data metrics as a focal point for responsible data stewardship. *Harv. Data Sci. Rev.* 2024. [CrossRef]

44. van Bellen, S.; Alperin, J.P.; Larivière, V. The oligopoly of academic publishers persists in exclusive database. *arXiv* 2024. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.