

Paving the Way to Open Data: Postprint

Authors: Yan Wu, Elizabeth Moylan, Hope Inman, Chris Graf, Yan Wu

Date: 2022-11-29T00:00:00+00:00

Abstract

It is easy to argue that open data is critical to enabling faster and more effective research discovery. In this article, we describe the approach we have taken at Wiley to support open data and to start enabling more data to be FAIR data (Findable, Accessible, Interoperable and Reusable) with the implementation of four data policies: “Encourages”, “Expects”, “Mandates” and “Mandates and Peer Reviews Data”. We describe the rationale for these policies and levels of adoption so far. In the coming months we plan to measure and monitor the implementation of these policies via the publication of data availability statements and data citations. With this information, we’ll be able to celebrate adoption of data-sharing practices by the research communities we work with and serve, and we hope to showcase researchers from those communities leading in open research.

Full Text

Preamble

IMPLEMENTATION PAPER

Paving the Way to Open Data

Yan Wu^{1†}, Elizabeth Moylan², Hope Inman³ & Chris Graf²

¹Wiley, 805-808 Sun Palace, No. 12A Taiyanggong Middle Road, Beijing 100028, China

²Wiley, 9600 Garsington Road, Oxford, OX4 2DQ, UK

³Wiley, 111 River Street, Hoboken, NJ 07030, USA

Keywords: Open data; Data sharing; Data citation; Open research

Citation: Y. Wu, E. Moylan, H. Inman & C. Graf. Paving the way to open data. *Data Intelligence* 1(2019), 380-392. doi: 10.1162/dint_a_{00021}

Received: March 29, 2019; **Revised:** June 20, 2019; **Accepted:** July 5, 2019

ABSTRACT

It is easy to argue that open data is critical to enabling faster and more effective research discovery. In this article, we describe the approach we have taken at Wiley to support open data and to start enabling more data to be FAIR data (Findable, Accessible, Interoperable and Reusable) with the implementation of four data policies: “Encourages”, “Expects”, “Mandates” and “Mandates and Peer Reviews Data”. We describe the rationale for these policies and levels of adoption so far. In the coming months we plan to measure and monitor the implementation of these policies via the publication of data availability statements and data citations. With this information, we’ll be able to celebrate adoption of data-sharing practices by the research communities we work with and serve, and we hope to showcase researchers from those communities leading in open research.

1. BACKGROUND AND MOTIVATION

“Open research” and “open science” are two interchangeable terms that encompass a number of practices that are becoming widely adopted [1,2]. While definitions of open research and open science come in many flavors (see Table 1), their core elements include open accessibility and dissemination of research outputs including more than traditional journal articles.

At Wiley, the researcher is our “North Star” as explained by Judy Verses (Executive Vice President, Wiley) in her keynote talk at the APE2019 conference in Berlin, Germany [7]. This means that we put researchers at the heart of our research publishing and educational services. We listen to the research communities we serve and –by tailoring open research initiatives to the needs of researchers in particular disciplines –we support their open research aspirations. Adopting open practices, but phasing their implementation to suit different communities, is our focus. We organize our work in five key areas: open access, open practices, open collaboration, open recognition and reward, and of course, open data [8].

“Open data” is an often-used term for sharing data, and is perhaps made more meaningful by the term FAIR (Findable, Accessible, Interoperable and Reusable) [9]. After open access, “open data” (or better: FAIR data) is probably one of the most important elements of open research [10]. FAIR data have the potential to revolutionize the way research is done and communicated and we are seeing benefits in research discoveries as a result [11]. Open research initiatives, like open data, bring many benefits including increased transparency as well as, potentially, enhanced reproducibility and amplified impact [12]. Funders and institutions recognize this and are increasingly requiring researchers to share data [13].

However, given the scale and variety of data, the complexity of how best to share data, the need for new practices and habits by research communities, and the need for technology and infrastructure to support data sharing, it is clear

that collaboration across all stakeholders is key. This is a challenge we all must embrace, if we are going to make progress.

To reflect our commitment to open research and to support researchers in sharing their data, Wiley recently updated its data sharing and citation policies [14]. In the rest of this article, we will share the approach we took with our data policies, how this fits with approaches taken by other publishers, and how this helps Wiley begin to achieve the goals of FAIR data.

† Corresponding author: Yan Wu (E-mail: ywu2@wiley.com; ORCID: 0000-0001-7610-9465).

2. RESEARCH DATA SHARING POLICY AT WILEY

At Wiley, we are making open research not just the future of research and research communication, but the here and now. We have four policy-level requirements for data sharing, adopted across our portfolio of journals [15]:

1. **“Encourages data sharing”** is our entry-level policy to encourage data sharing. It enables journals serving researchers in communities where data sharing is not common to start their journey toward data sharing. There are no enforced requirements.
2. **“Expects data sharing”** is a policy for journals that require from every author a data availability statement to confirm presence or absence of shared data, and a data citation. It is equivalent to the Transparency and Openness Promotion (TOP) level 1 guidelines [16].
3. **“Mandates data sharing”** is a policy for journals that require a data availability statement, a data citation, and sharing of data (It is equivalent to TOP level 2 [16]).
4. **“Mandates data sharing and peer reviews data”** is a policy for journals that take the additional step of peer reviewing data (It is the equivalent to TOP level 3 [16]).

Of course, we recognize that the process of adopting open research practices can be challenging and requires cultural change as emphasized by Henriikka Mustajoki (Head of Development, Federation of Finnish Learned Societies) [17]. Our four policy levels give flexibility so that journals can adopt policies that are right for their research communities.

Tiered policies like these adopted by major publishers enable journals to adapt to the communities they serve [18]. The Wiley data sharing policies are shown in Table 2, which maps each against the Transparency Openness Promotion (TOP) guidelines [16] that are used by publishers and funders to increase transparency.

3. THE RESEARCH DATA SHARING LANDSCAPE

Many publishers are adopting data sharing policies either encouraging or requiring researchers to share their underlying data [18]. These developments are going hand-in-hand with requirements from institutions and funders [13]. The characteristic features of data policies from major publishers can be compared with how they map to TOP guidelines [17]. However, while publisher data policies have common elements there is a recognized need for further standardization [19].

With the adoption of data sharing policies comes the possibility of evaluating the impact of these policies. Are researchers compliant? How are data shared? Findings from a recent analysis suggest that the majority of researchers share data within a published article (rather than via a repository) [20] but more research is needed to understand the issues researchers face in sharing their data.

4. UNDERSTANDING RESEARCHER NEEDS

The 2016 Wiley Open Science survey, built on earlier work by Wiley in 2014, gathered opinions on data-sharing from over 4,600 researchers worldwide [21]. It identified researchers' motivations to share data (Figure 1 [Figure 1: see original paper]), as well as what they find most challenging about data sharing. Wiley is continuing to collect data on how researchers across all disciplines approach open access, open data, peer review and collaboration, and will report the new data in 2019.

5. PROMOTING DATA SHARING

In November 2018, during International Data Week [23] we began a campaign to implement the Wiley “Expects Data” data sharing policy more broadly. Our goal was to step up the support we offered to researchers who want or need to share their data, by transitioning journals from our “Encourages Data” data sharing policy to “Expects Data” [14].

Our first step was to create a toolkit that would brief publishing colleagues, so they could effectively liaise with editors of journals, and then –together –to implement the requirements of the “Expects Data” policy, namely by including data availability statements and data citations in every article. The data sharing team provided everything that journals would need, including support for authors in the form of template data availability statements, instructions for how to cite the data they are sharing, and advice on finding appropriate repositories at which to share their data [15]. We began our implementation plan by selecting journals serving disciplines that were most ready for data sharing, and introduced our new Expects Data policy to those journals first.

At that time, November 2018, more than 1,500 journals had the entry-level “Encourages Data” data sharing policy, and had no specific requirements for data

sharing by researchers. We also published a much smaller number of journals (more than 20) that had adopted an earlier version of our “Expects Data” policy, which emphasized the benefits of sharing data to researchers, but that still had no specific requirements for data sharing. Alongside this, we published a similarly small number of journals with “Mandates Data” policy (about 20), among which are the leading journals from the Wiley evolutionary biology portfolio.

Since 2018, we have made significant progress at Wiley. By April 2019, 90 journals have adopted and implemented our “Expects Data” policy and 70 journals have adopted “Mandates Data” policy. Examples of these journals are shown in Table 3 below. Each of these journals now requires data availability statements in every article it publishes, as well as data citations. To make the whole process straightforward for research authors, we created a series of standard templates to complete their data availability statements, shared in Table 4 .

We also publish several journals -including *EMBO Reports*, *The EMBO Journal*, and *EMBO Molecular Medicine* -that have adopted our highest data policy of “Mandates and Peer Reviews Data” , setting the standard for data transparency (and also data citation, discussed in the section that follows). Beyond our data sharing policy, we partner with repositories like Figshare and Dryad to make it easier for authors to share data in approved repositories. We develop standards and guidance that enable researchers to share and cite their research data more readily [24, 25]. We adopt and encourage the use of Center for Open Science badges, and over 30 journals use these to recognize and celebrate authors who share data. We are launching an Open Science Ambassador Program in China, and Open Data contribution and sharing will be important components.

6. CITING DATA

Wiley also endorses the FORCE11 Joint Declaration of Data Citation Principles [26], a set of guiding principles for data within scholarly literature, another data set, or any other research object. We recommend the format for data citation proposed in this Joint Declaration, and that data held within institutional, subject-focused, or more general data repositories should be cited. At the same time, we do not intend to replace community standards such as in-line citation of GenBank accession codes, instead we hope to supplement those with formal data citations. This is one way to begin to enable researchers who share data to be recognized in the same way that researchers are recognized when they collect citations to their research articles. Data citation like this is not new to Wiley policies. But the emphasis on data citation within the new Wiley data sharing policies is new and is in-line with industry standards and initiatives to recognize data as a primary research object.

7. WORKING TOWARD FAIR DATA

At Wiley, we believe that the introduction of data sharing policies is the first step toward supporting and embracing the FAIR guiding principles [9]. While our policies actively support data sharing (“Expects Data Sharing”) or require data sharing (“Mandates Data Sharing”), the task of making shared data actually FAIR remains with researchers. For many this will be a new responsibility, and it can present some challenges.

We have begun work to help overcome those challenges. For example, research authors who select the first of our template data availability statements (“Data openly available in a public repository that issues data sets with DOIs”) [15] are indicating that their data is “F” (Findable; with a unique and persistent identifier, the DOI) and “A” (Accessible; retrievable by that identifier). Journals that adopt our level 4 policy “Mandates Data Sharing and Peer Reviews Data” conduct peer review on data submitted alongside journal articles, and by doing that, help research authors make their data ready to be “R” (Reusable).

Each of these steps moves us closer to the goal of turning open data into FAIR data, although often the “I” of FAIR (Interoperable) remains a challenge. The following section shares examples of work we are leading or contributing to at Wiley that take us even closer to that goal. Collaboration from all parties – researchers, funders, institutions, policy-makers, infrastructure providers (like repositories), and publishers – is vital to make FAIR data a reality.

8. EXAMPLES OF PROGRESS TOWARD FAIR DATA

American Geophysical Union and Enabling FAIR Data. The American Geophysical Union (AGU) together with Wiley and other partners (including repositories and supporting organizations) have an on-going project to enable FAIR data across the earth and space sciences, sensibly called Enabling FAIR Data [27]. This builds on the work of the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) [28]. Large and complex data sets are common in the earth and space sciences, which makes this initiative particularly welcome.

Remarkable practice at GeoScience Data Journal and American Journal of Political Science. Journals that “Mandate Data Sharing and Peer Review Data,” for example, the *GeoScience Data Journal* published by Wiley [29] and the *American Journal of Political Science* published by Wiley for the Midwest Political Science Association [30], have already adopted remarkable practices toward sharing FAIR-compliant data, and are to be applauded.

SourceData at EMBO Press. The team at EMBO Press, for which Wiley provides publishing services, introduced SourceData in 2018 [31]. SourceData provides a significant step on the road to FAIR data: It makes data findable, accessible, interconnected and downloadable.

“Next” journals at Wiley: Genetics & Genomics Next and Neuro-

science Next. Wiley's new Next journals, *Genetics & Genomics Next* [32] and *Neuroscience Next* [33] support FAIRsharing.org [2] and endorse the FAIR data principles in their own research disciplines.

Research Data Alliance and standard data sharing policies. Wiley is a member of the community that comes together as the Research Data Alliance (RDA). RDA is creating the social and technical infrastructure that researchers need to share data successfully. For example, the RDA's Data Policy Standardization Interest Group is creating a unified approach to setting data policies, by providing identifying standard requirements for data sharing, as well as how these can be put together into a robust data policy [34].

9. CONCLUSIONS: NEXT STEPS

At Wiley, we believe that open research is not just the future of research communications; it is the here and now [8]. Publishers are fundamentally service providers for researchers, whether those researchers are acting as authors, peer reviewers, editors or readers. Our careful implementation of open research practices, including data-sharing policies and open data badges, is intended to help researchers adopt new practices and to benefit from extra impact. We are excited about seeing the results of this work in terms of published data availability statements, and data citations in future. Looking further ahead we intend to measure the success of our "Expects Data" policy implementation, and to measure publication of data availability statements and data citations. With this information, we will be able to celebrate adoption of new practices by the research communities we work with and serve, and showcase researchers from those communities leading in open research.

AUTHOR CONTRIBUTIONS

All authors made substantial contributions to the design of this paper. Y. Wu (yw2@wiley.com) wrote the first draft; E. Moylan (emoylan@wiley.com), C. Graf (cgraf@wiley.com) and H. Inman (hinman@wiley.com) revised the first draft. All authors approved the version to be published and are accountable for the paper.

ACKNOWLEDGEMENTS

We are grateful to Wiley's data sharing team and the people who helped design and implement the data sharing policies including: Erin Arndt (Associate Director, Editorial System), Hope Inman (Manager, Editorial Operations & Communications), Kate Perry (Product Manager), Kathryn Sharples (Director, Editorial Development), Terri Teleen (Director, Editorial Operations & Communications), Natasha White (Director, Open Access Product Marketing). We would like to thank Elisha Morris (Editorial Assistant), and Sarah Pedder (Intern) for support as we made revisions to this manuscript following peer review.

REFERENCES

- [1] Realizing the potential: Final report of the Open Research Data Task Force. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/681112/the-potential-ORDTF-July-2018.pdf.
- [2] FAIR sharing: Standards, databases, policies. Available at: <https://fairsharing.org>.
- [3] Open Science definition. Available at: <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>.
- [4] FET through the keyhole. (2015). Available at: http://ec.europa.eu/information_society/newsroom/cf/dspecific-archive-issue.cfm?newsletter_service_id=129&newsletter_issue_id=398&pdf=true&fullDate=20150701.
- [5] M. Nielsen. Open Science now! Available at: https://www.ted.com/talks/michael_nielsen_open_science.
- [6] Center for Open Science. Available at: <https://cos.io/>.
- [7] Adapting to customer needs: Judy Verses at APE 2019. Available at: <https://www.youtube.com/watch?v=SeaErs9h6Bc>.
- [8] Open research. Available at: <https://authorservices.wiley.com/open-research/index.html>.
- [9] The FAIR data principles. Available at: <https://www.force11.org/group/fairgroup/fairprinciples>.
- [10] R. Harington. Openness: An interview with Daniel Hook, CEO of Digital Science. Available at: <https://scholarlykitchen.sspnet.org/2019/03/11/openness-an-interview-with-daniel-hook-ceo-of-digital-science/>.
- [11] S. Scheer, S. Ackloo, T.S. Medina, M. Schapira, F. Li, J. A. Ward, A. M. Lewis, ... & C. H. Arrowsmith. A chemical biology toolbox to study protein methyltransferases and epigenetic signaling. *Nature Communications* 10(2019), Article No. 19. doi: 10.1038/s41467-018-07905-4.
- [12] What are the benefits of Open Science. Available at: <https://www.fosteropenscience.eu/content/what-are-benefits-open-science>.
- [13] M. Hahnel. Global funders who require data archiving as a condition of grants. doi: 10.6084/m9.figshare.1281141.v1.
- [14] C. Graf. How and why we're making research data more open. Available at: <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-we-re-making-research-data-more-open>.
- [15] Wiley's data sharing policies. Available at: <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>.
- [16] Open center for science: TOP guidelines. Available at: <https://cos.io/our-services/top-guidelines/>.

- [17] C. Graf. What does Open Science require? Cultural change. Available at: <https://www.wiley.com/network/researchers/licensing-and-open-access/what-does-open-science-require-cultural-change>.
- [18] D. Mellor. The landscape of open data policies. Available at: <https://cos.io/blog/landscape-open-data-policies/>.
- [19] I. Hrynaskiewicz, N. Simons, S. Goudie, & A. Hussain. Data policy standardization and implementation IG. Available at: <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation-ig>.
- [20] L.M. Federer, C.W. Belter, D.J. Joubert, A. Livinski, Y.L. Lu, L.N. Snyders, & H. Thompson. Data sharing in PLoS ONE: An analysis of data availability statements. *PLoS ONE* 13(5) (2018), e0194768. doi:10.1371/journal.pone.0194768.
- [21] B. Vocile. Open Science trends you need to know about. Available at: <https://www.wiley.com/network/researchers/licensing-and-open-access/open-science-trends-you-need-to-know-about>.
- [22] Wiley Open Science researcher survey 2016 infographic. Available at: https://figshare.com/articles/Wiley_{Open}_{Science}_{Researcher}_{Survey}_{2016}_{Infographic}/4910714.
- [23] International Data Week. Available at: <https://www.internationaldataweek.org>.
- [24] H. Cousijn, & A. Kenall. A data citation roadmap for scientific publishers. *Scientific Data* 5(2018), Article No. 180259. doi: 10.1038/sdata.2018.259.
- [25] C. Graf. Why share and cite my research data? A guide to making open research easier. Available at: <https://www.wiley.com/network/researchers/latest-content/why-share-and-cite-my-research-data-a-guide-to-making-open-research-easier>.
- [26] M. Martone. Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. doi: 10.25490/a97f-egy.
- [27] S. Stall. Enabling findable, accessible, interoperable, and reusable data. Available at: <https://eos.org/editors-vox/enabling-findable-accessible-interoperable-and-reusable-data>.
- [28] COPDESS: Coalition for publishing data in the earth and space sciences. Available at: <https://copdess.org/>.
- [29] Geoscience Data Journal guidelines for reviewers. Available at: <https://rmets.onlinelibrary.wiley.com/hub/journal/20496060/features/guidelines-for-reviewers>.
- [30] American Journal of Political Science verification policy. Available at: <https://ajps.org/ajps-replication-policy/>.
- [31] SourceData and BioStudies bring open data sharing to EMBO Press. Available at: <http://www.embo.org/news/press-releases/2018/sourcedata-and-biostudies-bring-open-data-sharing-to-embo-press>.

[32] Genetics & Genomics Next: Editorial process overview. Available at: <https://onlinelibrary.wiley.com/page/journal/26416573/homepage/author-guidelines>.

[33] Neuroscience Next: Editorial process overview. Available at: <https://onlinelibrary.wiley.com/page/journal/26416573/homepage/author-guidelines>.

[34] C. Graf. Powering up: How the Research Data Alliance is making open research a reality. Available at: <https://www.wiley.com/network/researchers/licensing-and-open-access/powering-up-how-the-research-data-alliance-is-making-open-research-a-reality>.

AUTHOR BIOGRAPHY

Yan Wu is currently Associate Director for Open Research and Editorial Development at Wiley, responsible for developing Open Research strategy in journal editorial development and China, also shaping strategy collaborations. She has been working in academic publishing over 8 years with various roles.

Elizabeth Moylan is Publisher for Research Integrity and Publishing Ethics at Wiley where she works with colleagues to develop and implement policies which support open publishing practices, with emphasis on peer review and open research. She is an Editorial Board Member for *Research Integrity & Peer Review* and is an Advisory Board Member for two EU-funded projects “EnTIRE” and “Virt2ue” that aim to support researchers in upholding and practicing the principles of research integrity. Elizabeth completed her PhD and post-doctoral training at the University of Oxford, UK.

Hope Inman is Senior Manager, Publishing Operations and Communications at Wiley. Hope has been with Wiley for over 14 years in roles of peer review management and journal management before her current role where she leads initiatives supporting operational excellence and adoption of open research policies throughout Wiley’s proprietary and society-owned title list.

Chris Graf is Director, Research Integrity and Publishing Ethics at Wiley, and is Trustee and Past Co-Chair of COPE, Committee on Publication Ethics. Academic and researcher behavior and publishing are changing fast. The integrity of what we peer review and publish has never been more important. Academic and research publishers need editorial processes, technology, and a focus on quality that fits the research environment we work in, and that fits what our communities and customers expect of us. How we make that focus real is important: What we do is always a collaboration with editors, publishing teams, with researchers and academic communities, with societies and associations, and it is always shaped by funders and institutions. Chris leads initiatives at Wiley that focus on transparency, research integrity and publishing ethics.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.