

Paving the Way to Open Data

Yan Wu¹, Elizabeth Moylan¹, Hope Inman¹, and Chris Graf¹

¹Wiley

April 28, 2020

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 FAIR (Findable, Accessible, Interoperable and Reusable) data 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.

Address for correspondence: Yan Wu, Senior Manager, OA and Editorial Development, John Wiley & Sons Commercial Service Co Ltd., Room 805-808, Floor 8, Sun Palace, No. 12A, Taiyanggong Middle Road, Chaoyang District, Beijing, P.R. China; ywu2@wiley.com

Draft for preprint (submitted to *Data Intelligence*).

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

Table 1. Definitions of open research and open science Attribution

	Definition
Foster Open Science	”Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.” [3]
European Commission	”A broad term, covering the many exciting developments in how science is becoming more open, accessible, efficient, democratic, and transparent. This Open Science revolution is being driven by new, digital tools for scientific collaboration, experiments and analysis and which make scientific knowledge more easily accessible by professionals and the general public, anywhere, at any time.” [4]
Michael Nielsen	”the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process” [5]
Center of Open Science	”Openness and reproducibility are core scientific values because science is a distributed, non-hierarchical culture for accumulating knowledge. No individual is the arbiter of truth. Knowledge accumulates by sharing information and independently reproducing results.” [6]

Table 1: Definitions of open research and open science.

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 has 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.

However, given the scale and variety of data, the complexity of how best to share data, the need for the adoption of 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 supporting researchers in sharing their data, Wiley recently updated its data sharing and citation policies [13]. In the rest of this article, we will share the approach we took and our findings so far.

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 [14].

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 towards 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 presences or absence of shared data, and a data citation. It is equivalent to the Transparency and Openness Promotion (TOP) level 1 guidelines [15].
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 [15]).
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 [15]).

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) [16]. 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 and journals enable journals to adapt to the communities they serve [17]. The Wiley data sharing policies are shown in Table 2, which maps each against the Transparency Openness Promotion (TOP) guidelines [15] that are used by publishers and funders to increase transparency.

Table 2. Four data-sharing policies adopted at Wiley and their features.

	Data availability statement is published ^a	Data has been shared ^b	Data has been peer reviewed ^c	Example Wiley journals	The TOP Guideline Level
Encourages Data Sharing	Optional	Optional	Optional		Not TOP compliant, i.e. "Level 0")
Expects Data Sharing	Required	Optional	Optional	https://onlinelibrary.wiley.com/doi/10.1002/for.120448309	TOP Level 1
Mandates Data Sharing	Required	Required	Optional	https://onlinelibrary.wiley.com/doi/10.1002/for.120448309	TOP Level 2
Mandates Data Sharing and Peer Reviews Data	Required	Required	Required	https://rmets.onlinelibrary.wiley.com/doi/10.1002/for.120448309 https://ajps.org/ajps-replication-policy/	TOP Level 3

a A data availability statement confirms the presence or absence of shared data.

b Links to data in data availability statements are checked to ensure they link to the data that the authors intended. If data has been stored in a data repository, the data availability statement includes a permanent link to the data. Shared data is also cited.

c Quality and/or replicability of linked data are peer reviewed. Depending on the journal, this may be to peer review the quality of the data by ensuring that the results in the paper and the data in the repository align (for example, sample sizes and variables match), or it may be to peer review the replicability of the data to ensure that the claims presented in the journal article are valid and can be reproduced.

Table 2: Four data-sharing policies adopted at Wiley and their features.

3. Understanding Author Needs

The 2016 Wiley Open Science survey gathered opinions on data-sharing from over 4,600 researchers worldwide [18] and identified researchers' motivations to share data (Figure 1), as well as what they find most challenging about data sharing.

Researchers sharing data by region



Data accessibility trends



Top 4 researcher motivations for sharing data



Top 4 reasons why researchers are hesitant to share their data

- 1 50% - Intellectual property or confidentiality issues
- 2 31% - Ethical concerns
- 3 23% - I am concerned about misinterpretation or misuse of my research
- 4 22% - I am concerned that my research will be scooped

Figure 1: Selected insights from Wiley Open Science survey, 2016. The whole infographic with many more insights is described in detail by Wiley [18] and is available on Figshare [19].

4. Promoting Data Sharing

In November 2018, during International Data Week [20] we began a campaign to implement the Wiley “Expects Data” data sharing policy 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” if they were ready [13].

First we created 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 [14]. 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, c. 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 (c. 20+) that had adopted an earlier version of our “Expects Data” policy, that 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 a “Mandates Data” policy (c. 20), among which are the leading journals from the Wiley evolutionary biology portfolio.

Since 2018 we have made significant progress at Wiley. By March 2019 over 160 journals have adopted and implemented our “Expects Data” policy, or our Mandates Data policy. Examples of these journals are shown in Table 3 below. Each now requires data availability statements in every article it publishes, as well as data citations. To make the whole process easy for research authors, we created a series of standard templates to complete their data availability statements, shared in Table 4.

Table 3. Ten journals as examples of those that have adopted the Wiley Expects Data policy.

Acta Neurologica Scandinavica
Applied Stochastic Models in Business and Industry
Brain and Behavior
Chemical Biology and Drug Design
Clinical Endocrinology
Clinical Genetics
Environmetrics
Immunity, Inflammation and Disease
Research Synthesis Methods
Pharmaceutical Statistics

Table 3: Ten journals as examples of those that have adopted the Wiley Expects Data policy.

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 enables researchers to share

Table 4. Template data availability statements from [14].	
Availability of data	Template for data availability statement
Data openly available in a public repository that issues datasets with DOIs	The data that support the findings of this study are openly available in [repository name e.g "figshare"] at http://doi.org/[doi] , reference number [reference number].
Data openly available in a public repository that does not issue DOIs	The data that support the findings of this study are openly available in [repository name] at [URL], reference number [reference number].
Data derived from public domain resources	The data that support the findings of this study are available in [repository name] at [URL/DOI], reference number [reference number]. These data were derived from the following resources available in the public domain: [list resources and URLs]
Embargo on data due to commercial restrictions	The data that support the findings will be available in [repository name] at [URL / DOI link] following an embargo from the date of publication to allow for commercialization of research findings.
Data available on request due to privacy/ethical restrictions	The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
Data subject to third party restrictions	The data that support the findings of this study are available from [third party]. Restrictions apply to the availability of these data, which were used under license for this study. Data are available [from the authors / at URL] with the permission of [third party].
Data available on request from the authors	The data that support the findings of this study are available from the corresponding author upon reasonable request.
Data sharing not applicable – no new data generated	Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Table 4: Template data availability statements from [14].

and cite their research data more readily [21, 22]. We adopt and encourage the use of Center for Open Science badges, and over 30 journals use these to recognise and celebrate authors who share data. In addition, we are launching an Open Science Ambassador Program in China, and Open Data contribution and sharing will be important components of that program.

5. Citing Data

Wiley endorses the FORCE11 Joint Declaration of Data Citation Principles [23], a set of guiding principles for data within scholarly literature, another dataset, 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.

6. 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. Looking further ahead we intend to measure the success of our

“Expects Data Sharing” policy implementation, and to measure publication of data availability statements and data citations. With this information, we’ll 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.

Acknowledgment

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) Elizabeth Moylan (Publisher), Kate Perry (Product Manager), Kathryn Sharples (Director, Editorial Development), Terri Teleen (Director, Editorial Operations & Communications), Natasha White (Director, Open Access Product Marketing), Chris Graf (Director, Research Integrity and Publishing Ethics).

References

1. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775006/Realising-the-potential-ORDTF-July-2018.pdf
2. <https://fairsharing.org>
3. <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>
4. http://ec.europa.eu/information_society/newsroom/cf/dae/newsletter-specific-archive-issue.cfm?newsletter_service_id=129&newsletter_issue_id=398&pdf=true&fullDate=Thu%2016%20Jul%202015&lang=default
5. https://www.ted.com/talks/michael_nielsen_open_science_now
6. <https://cos.io/>
7. <https://www.youtube.com/watch?v=SeaErs9h6Bc>
8. <https://authorservices.wiley.com/open-research/index.html>
9. <https://www.force11.org/group/fairgroup/fairprinciples>
10. <https://scholarlykitchen.sspnet.org/2019/03/11/openness-an-interview-with-daniel-hook-ceo-of-digital-science/>
11. <https://www.thesgc.org/news/chemical-biology-toolbox-study-protein-methyltransferases-and-epigenetic-signaling-nature>
12. <https://www.fosteropenscience.eu/content/what-are-benefits-open-science>
13. <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-we-re-making-research-data-more-open>
14. <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>
15. <https://cos.io/our-services/top-guidelines/>
16. <https://www.wiley.com/network/researchers/licensing-and-open-access/what-does-open-science-require-cultural-change>
17. <https://cos.io/blog/landscape-open-data-policies/>
18. <https://www.wiley.com/network/researchers/licensing-and-open-access/open-science-trends-you-need-to-know-about>
19. https://figshare.com/articles/Wiley_Open_Science_Researcher_Survey_2016_Infographic/4910714
20. <https://www.internationaldataweek.org>
21. <https://www.nature.com/articles/sdata2018259>.
22. <https://www.wiley.com/network/researchers/latest-content/why-share-and-cite-my-research-data-a-guide-to-making-open-research-easier>
23. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>

