

EDITORIAL

Increasing best practice data sharing at *PLOS Pathogens*Lauren Cadwallader^{1*}, Kasturi Haldar², Michael H. Malim³

1 Public Library of Science, San Francisco, California, United States of America and Cambridge, United Kingdom, **2** Boler-Parseghian Center for Rare and Neglected Diseases, University of Notre Dame, Notre Dame, Indiana, United States of America, **3** Department of Infectious Diseases, King's College London, London, United Kingdom

* lcadwallader@plos.org

Introduction

PLOS Pathogens is a trusted venue for impactful research that readers everywhere rely on to view, download, and share. The journal empowers authors to decide how they want to share their work by removing barriers, taking action on the feedback from our community, and providing transparency every step of the way. As with all PLOS journals, *PLOS Pathogens* strives to be a conduit for good open science practices. However, we know that there are barriers to the adoption of open science practices [1], and our approach to increasing adoption should be driven by evidence and community need.

Current data sharing at *PLOS Pathogens*

Since 2014, all PLOS journals have mandated the sharing of research data, and *PLOS Pathogens* has seen a steady increase in the use of repositories from 16% of research articles with data depositing in a repository in 2014 to 29% in 2020. *PLOS Pathogens* authors share their data in a number of different places—discipline-specific repositories (e.g., Sequence Read Archive and Gene Expression Omnibus), generalist repositories (e.g., Dryad and Figshare), code repositories (e.g., Github), and within the Supporting information accompanying an article. In fact, in 2020, 75% of research articles published in *PLOS Pathogens* shared at least some of the data underlying their research in the Supporting information.

Sharing data via a repository—either discipline specific or generalist—has benefits for both authors and readers. Authors are able to demonstrate their compliance with the increasing number of funder and institutional policies that require data sharing [3,4]. The dataset can be cited independently of the article, which can be tracked using the assigned DOI, demonstrating the impact and value of the research to funders. And sharing data in a repository has been linked to a 25% increase in citations to the article [5]. Readers benefit as they are able to find data more easily as it is indexed in search engines and aggregators. Licensing information is clear so the potential for reuse is immediately obvious. Finally, repositories offer the possibility of machine-readable datasets, allowing for computational methods to access and reuse data.

Depositing in a repository is the ideal way to meet the requirements of the findable, accessible, interoperable, reusable (FAIR) principles for research data [2]. The data become a research output itself. Depositing it in a repository makes it findable, citable, and available for the long term (in accordance with the repository's preservation guarantees). It is clear that sharing data via a repository has tangible benefits for the community whether they are authors or readers and that *PLOS Pathogens* could, and should, do more to promote sharing of data in this way.



OPEN ACCESS

Citation: Cadwallader L, Haldar K, Malim MH (2021) Increasing best practice data sharing at *PLOS Pathogens*. *PLoS Pathog* 17(11): e1010021. <https://doi.org/10.1371/journal.ppat.1010021>

Published: November 11, 2021

Copyright: © 2021 Cadwallader et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is being supported by the Wellcome Trust (grant number 223821/Z/21/Z).

Competing interests: The authors have declared that no competing interests exist.

Integration

To enable more *PLOS Pathogens* researchers to share their data via a repository, we are experimenting with solutions that will make data deposition easier and more efficient. We have integrated the journal's submission system with the Dryad repository (<https://datadryad.org>), which will be available for 1 year initially. Dryad is a generalist repository that will accept any data, with the exception of identifiable human data, as part of the manuscript submission process. Dryad will curate the dataset to ensure that it meets the minimum requirement, and *PLOS Pathogens* will cover the costs of this service for an unlimited number of authors for the initial 1-year period. These actions aim to further remove barriers to submission and therefore promote robust participation.

This integration is part of a Wellcome Trust-funded PLOS open science initiative to explore how PLOS can improve data sharing and engagement with the data that have been shared [6].

Evidence-based solutions

This approach was chosen because research conducted by PLOS, involving mostly PLOS authors, suggests that busy researchers are unlikely to significantly alter their behavior when it comes to data sharing [7] and are more likely to participate in public data sharing if the barriers to entry (including time, multiple processes, and cost) are reduced or removed. Our view is that increased data sharing is more likely to occur if good practices are embedded into existing author workflows, such as the journal manuscript submission. By introducing a small, simple, and optional extra step to the submission process, we aim to promote good open science practice in a way that respects author choice and does not increase workload.

PLOS Pathogens is well placed to trial Dryad as the research data authors commonly put into Supporting information are primarily not suited to a discipline-specific repository but could be of value to other researchers who may wish to reproduce or utilize their results. Not all the outputs shared via Supporting information are suitable for a repository (e.g., tables supporting statistical analysis presented in the article), and so Supporting information will still be an option for authors wanting to share other materials.

Success measures

The Dryad integration with *PLOS Pathogens* will initially be available for 1 year from October 2021 to all authors starting a new submission. During this time, we will be monitoring the uptake of the integration and conducting research with the community to further understand their attitudes toward data repositories and their experience of the integration. What we do after the first year will depend on what the community tell us during our consultations and if there has been a measurable increase in data sharing via a repository in the journal. We hope that this integration will empower *PLOS Pathogens* authors to practice data sharing in a way that benefits them and their community.

Acknowledgments

The authors would like to thank Eileen Clancy, Iain Hrynaszkiewicz, and Lindsay Morton for comments on an earlier draft.

References

1. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE*. 2015; 10(8). <https://doi.org/10.1371/journal.pone.0134826> PMID: 26308551

2. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016 Mar 15; 3(1):160018. <https://doi.org/10.1038/sdata.2016.18> PMID: 26978244
3. Gaba JF, Siebert M, Dupuy A, Moher D, Naudet F. Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PLoS ONE*. 2020 Aug 20; 15(8): e0237464. <https://doi.org/10.1371/journal.pone.0237464> PMID: 32817724
4. Wellcome Trust. Data, software and materials management and sharing policy—Grant Funding [Internet]. Wellcome. [cited 2021 Sep 10]. Available from: <https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy>
5. Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B. The citation advantage of linking publications to research data. Wicherts JM, editor. *PLoS ONE*. 2020 Apr 22; 15(4):e0230416. <https://doi.org/10.1371/journal.pone.0230416> PMID: 32320428
6. Hrynaszkiewicz I. Wellcome Trust awards PLOS a grant to test ways to increase sharing and discovery of research data [Internet]. The Official PLOS Blog. 2021 [cited 2021 Sep 14]. Available from: <https://theplosblog.plos.org/2021/09/wellcome-trust-awards-plos-a-grant-to-test-ways-to-increase-sharing-and-discovery-of-research-data/>
7. Hrynaszkiewicz I, Harney J, Cadwallader L. A Survey of Researchers' Needs and Priorities for Data Sharing. *Data Sci J*. 2021; 20(1): p.31. <http://doi.org/10.5334/dsj-2021-031>.