

Journal data sharing policies are moving the scientific community towards greater openness but clearly more work remains



*Data sharing is a key part of the drive towards greater openness in scientific research, allowing readers to reproduce and confirm an article's findings, or even reuse its data as part of a new study. Many journals have policies requiring researchers to share their data in full, with PLOS being a forerunner in this area. But how effective has the PLOS policy been in increasing the availability of data associated with articles? **Lisa Federer** reports on an analysis of the Data Availability Statements of more than 45,000 PLOS articles, finding that the ideal of open data is far from fully realised, with researchers' use of repositories clearly an area for improvement.*

If you're involved in scientific research of any kind, you've probably heard about the open science movement – the practice of making the products of research, including research data, openly available. Proponents suggest that openness helps increase the transparency and reproducibility of research, creates a greater return on the investment of research dollars, and helps democratise science. However, not all researchers embrace the open science movement, especially when it comes to sharing their data. Some fear they will be “scooped” if they share their data and someone beats them to publication, while others see researchers who reuse data as “[research parasites](#)”.

Regardless of your personal stance on data sharing, if you're getting research funding or publishing in scientific journals, chances are you'll be required to share your data at some point. Major funders around the world have created policies requiring researchers to share data resulting from their grants. Likewise, the International Committee of Medical Journal Editors, an organisation of large science publishers, has [announced](#) that manuscripts submitted to ICMJE journals after 1 July 2018 must contain a data sharing statement.

Though such policies are becoming more widespread, some journals have been requiring researchers to share their data for several years. One of the forerunners of journal data sharing policies is PLOS, which publishes several subject-specific journals as well as the interdisciplinary journal *PLoS ONE*. Since March 2014, PLOS journals have required all submissions to include a [Data Availability Statement](#) describing how to access “all data and related metadata underlying the findings reported in a submitted manuscript”. The policy gives authors some options about mechanisms for sharing their data (and of course makes exceptions for sensitive data that can't be shared for privacy or security reasons), but strongly encourages authors to deposit their data in one of the thousands of research data repositories that preserve, curate, and make accessible scientific data. Compared to self-archiving or making data available only upon request, using a repository helps ensure that data remain available over time, and can help researchers discover datasets more easily.

Researchers have been reporting their findings in the scientific literature for centuries, but the data that appear in an article are typically summary data, distilling many data points down to just what is needed to tell the story. The PLOS policy – and others like it – asks researchers to share *all* the data underlying their articles. A reader should be able to use the data to reproduce, and thereby confirm, the findings of an article, or even reuse the data for a novel study of his or her own. Neither of these uses is possible with just the summary data typically found in an article. Although sharing data is common in some fields, [research has shown](#) that data sharing practices can differ widely across scientific disciplines, so sharing requirements may be a significant culture change to researchers from some disciplines.



Image credit: [Richard Balog](#), via Unsplash (licensed under a [CC0 1.0](#) license).

Given some researchers' reluctance to sharing data and the challenges that can come with making data reusable and accessible, we wondered how effective the PLOS policy had been in increasing the availability of data associated with articles. To gain a better understanding of the extent to which researchers had shared data and the ways they had done so, we analysed the Data Availability Statements of more than 45,000 research articles published in PLOS in the 28 months since the policy took effect.

While the scientific community is making progress, our findings suggest that the ideal of open data is far from fully realised, even in a journal with a strong data sharing policy. Despite PLOS encouraging the use of repositories (and their inclusion of a list of suggested repositories on their data policy page), only 18% of Data Availability Statements indicated that the data were available in a repository. Instead, over 70% of Data Availability Statements noted that the data were in the paper or its supplements. This analysis did not investigate whether these papers did in fact contain a full, reproducible dataset or merely the type of summary data often found in papers, but these findings still suggest that most authors are not sharing their data in ways that conform with best practices. Even when authors did indicate they had shared in a repository, their Data Availability Statements didn't always provide all the necessary information; some gave only a repository name without a dataset name, accession number, or persistent unique identifier that would allow a reader to actually locate the dataset.

Policies like PLOS's are a good first step toward increasing openness and availability of research data, but clearly more work remains to ensure that articles can easily be connected with their supporting data. Our findings suggest that PLOS and other journals with similar policies may want to consider including Data Availability Statements in the peer review process to help increase compliance with the policy. Repositories could also play a role in making it easier for reader to find datasets (as well as easier for authors writing Data Availability Statements) by providing suggested template language that includes the relevant information. These findings also suggest there may be an opportunity for libraries and other research institutions to provide greater data management support to researchers who face new data sharing requirements.

As with most changes in policy and practice, the move toward open data won't happen overnight, but early evidence suggests that policies like PLOS's are helping the scientific community make progress on increasing openness.

This blog post is based on the author's co-written article, "[Data sharing in PLOS ONE: An analysis of Data Availability Statements](#)", published in PLoS ONE (DOI: [10.1371/journal.pone.0194768](#)).

Note: This article gives the views of the author, and not the position of the LSE Impact Blog, nor of the London School of Economics. Please review our [comments policy](#) if you have any concerns on posting a comment below.

About the author

Lisa Federer is a research data informationist at the National Institutes of Health Library, Office of Research Services, NIH, in Bethesda, MD. She is also a PhD candidate at the University of Maryland.