

# To move towards a more open science, we must free the data



*Data sharing is a key principle of open science, and research funders are increasingly including this as a condition of grant awards. Despite this, **Jessica Couture** reports on research that found little more than a quarter of relevant research projects to be compliant. While there are valid reasons for certain data not to be made available – its sensitivity or the ease of its interpretation, for example – these findings indicate more needs to be done. A fundamental obstacle to data sharing is the absence of a professional reward structure, such as recognition that data citations are as valuable as article citations. Funders can also encourage data sharing compliance by creating dedicated data archives for funded projects and providing technological assistance to awardees.*

Open science can be incorporated into every step of the scientific process and emphasises data sharing. Making data publicly available facilitates its reuse by scientists, such as in synthesis research, and can thus have a much greater impact than data limited to the creator's initial analysis or intention.

With huge amounts of money dedicated each year to support scientific research, there is a growing push from funders to increase the impact and prestige of the money they award by requiring or encouraging data sharing. Particularly, when scientists receive public funds, research data is considered a public good and therefore carries an expectation of public accessibility. Additionally, new tools are emerging that make data annotation and sharing easier to incorporate into the research process.

However, while tools and protocols are changing to improve data sharing among researchers, colleagues and I found data was mostly not being made public in practice. In an [article published in PLoS ONE](#), our team of scientists tested compliance with funder-imposed data-sharing requirements among projects in the environmental sciences over a 20-year period. We were able to collect data from only 26 per cent of funded projects. As scientists, we believe everyone in the scientific community can play a role in increasing data publication and sharing, and it is our responsibility to do so to improve the efficiency of research.

In our analysis, data availability did differ based on the project's field of study, influenced by factors such as the time required to prepare data, whether a field has established data collection protocols and standardised methods, the sensitivity of data, and the ease of its interpretation. Nonetheless, we assert that a fundamental obstacle facing data sharing is the absence of a professional reward structure, such as the recognition that data citations are as valuable as paper citations. This discrepancy de-incentivises the time spent formatting, annotating, and preparing data to be shared.

While some publication platforms are starting to apply digital object identifiers (DOIs) to published data as a reliable way to enable attribution, similar to journal publications, it is ultimately up to the scientific community to recognise data citations as scientific currency that is equally valuable, and to encourage and practice the inclusion of data citations in their overall scientific output.



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

To move toward more open science, scientists must take on some of the responsibility of learning about the benefits of data sharing and incorporating open science methods into their daily work. Creating data in a way that others – and, in future, you – can access and easily interpret may require an extra initial step, but it will reduce additional work down the road.

Using data formats that are easy to share and read on multiple and open source platforms – for example, CSV files rather than MS Excel – and publishing data in open archives will also save time when other researchers or the funder request data. Refined data preparation protocols can also expedite the publication process, as many journals, similar to funders, now require proof of data publication.

Funders can also make changes that will incentivise data sharing. Many have long required their awardees to make data publicly available without following up on these requirements or providing any resources to help researchers preserve their data. Some funders, such as the National Science Foundation (NSF), are starting to ensure data sharing compliance by creating dedicated data archives for the projects they fund and provide technological assistance to awardees. For example, the [Arctic Data Center](#) houses all data about the Arctic collected under NSF grants and provides awardees with a team of technicians to assist with data attribution, metadata creation, formatting, and publication. NSF also requires funded Arctic researchers to publish their data in the archive, or prove their publication in a similar archive, before awarding further funding. This two-fold approach not only facilitates data publication but also provides funders with easy confirmation of data sharing compliance.

Data sharing is pivotal to ensuring open science and research efficiency. In the ways outlined above, scientists, funders, and publishers alike can play important roles in increasing data liberation. Thinking about data as a valuable scientific currency is an important step forward, and it requires support from the entire scientific community. It starts with how you think about and treat yours and other people's data.

*This blog post was originally published by the [National Center for Ecological Analysis and Synthesis](#) (NCEAS) and is reposted here with permission. It is based on the author's co-written article, "[A funder-imposed data publication requirement seldom inspired data sharing](#)", published in PLoS ONE (DOI: 10.1371/journal.pone.0199789).*

*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

**Jessica Couture** is a PhD student at the Bren School for Environmental Science and Management at UC Santa Barbara, studying the environmental impacts of aquaculture. Previously, she worked at NCEAS for four years on various data projects with DataONE and participated in three working groups.