We need more carrots: give academic researchers the support and incentives to share data

*Making data available for other researchers to find, use, reuse, and reproduce is fundamental to open science, and ultimately makes research more efficient and effective. Yet despite funder policies that encourage and require data sharing, researchers in both the UK and the US report lower percentages of data sharing than the global average. In addition to progressive policies, **Grace Baynes** suggests researchers should be given incentives, expert support, training, and the infrastructure to make it seamless and easy to share data, and worth their while. Governments, funders, research institutions, libraries, and publishers all have a role to play to unlock the huge potential of research data.*

To achieve real "open science" we need to open up all areas of research, including research data. Making data available for other researchers to find, use, reuse, and reproduce will make research more efficient and effective. Members of the newly formed UK Research and Innovation, an independent organisation that brings together the seven Research Councils, Innovate UK and Research England, the Wellcome Trust, and other UK funders have moved early to encourage and require data sharing. Yet researchers in the UK report lower percentages of data sharing than the global average. Policy must be coupled with greater support and education for researchers, and faster, easier routes to sharing data optimally. Incentives and credit for data sharing are also needed.

As a publisher I firmly believe research articles and scholarly books and monographs are important summaries and conclusions of years of work for researchers. However, the real building blocks of discovery are the data they produce.

Data sharing brings many benefits to society. According to the Open Data Institute, the value of public sector open data is between 0.4% and 1.5% of an economy's GDP. An independent report found that the European Bioinformatics Institute returns £1 billion in annual efficiency savings to researchers worldwide. Data archiving can double the publication output of research projects, according to a study of 7,000 National Science Foundation and National Institutes of Health-funded research projects in the social sciences. Citation impact of research papers has also been shown to increase by as much as 50% when data are made available. It can help reduce duplication of effort and is a foundation for reproducibility research. Despite all these benefits, in 2017 only about half of research data were shared and a much smaller proportion were shared openly or in ways that maximise discoverability and reuse.

Last year, Springer Nature asked over 7,000 researchers about data sharing at the point of publishing a research article. We wanted to understand how much data sharing is actually happening, how and where researchers are sharing, the challenges they face, and where they need help. Our findings, *Practical Challenges for Researchers in Data Sharing*, are openly accessible in Figshare along with the survey data.

When submitting to a journal, 63% of respondents shared data files either as supplementary information, in a repository, or both. A slightly lower proportion share data in a repository (41%) than as supplementary information files (42%). Yet the willingness is there, with 80% of researchers surveyed in the *State of Open Data 2017 report* willing to share their data and the same proportion either already or amenable to using others' data.

To their credit, UK and US funders have moved early to encourage and require data sharing through policies, pilots, and infrastructure; yet in our survey, researchers in the UK and US report lower percentages of data sharing than the global average of 63%:

Country	Percentage of respondents
Poland	76%
Germany	75%
Switzerland	69%
Greece	69%
Italy	68%
Spain	66%

Country	Percentage of respondents
France	65%
Netherlands	64%
Norway	64%
Sweden	61%
Denmark	60%
Belgium	59%
United Kingdom	58%
Portugal	56%
Australia	55%
United States	55%
Canada	50%

Table 1: Percentages of respondents sharing data through a repository, as supplementary information files, or both, in countries with >100 respondents. Source: *Practical Challenges for Researchers in Data Sharing*.

So while funder mandates continue to be essential, policy must be coupled with greater support and education for researchers, and faster, easier routes to sharing data optimally. The challenges facing researchers include a lack of time and expertise. In our survey, "*Organising data in a presentable and useful way*" was the most stated reason for not sharing data (46% of respondents). Other common challenges were: "*Unsure about copyright and licensing*" – 37%; "*Not knowing which repository to use*" – 33%; and "*Lack of time to deposit data*" – 26%.
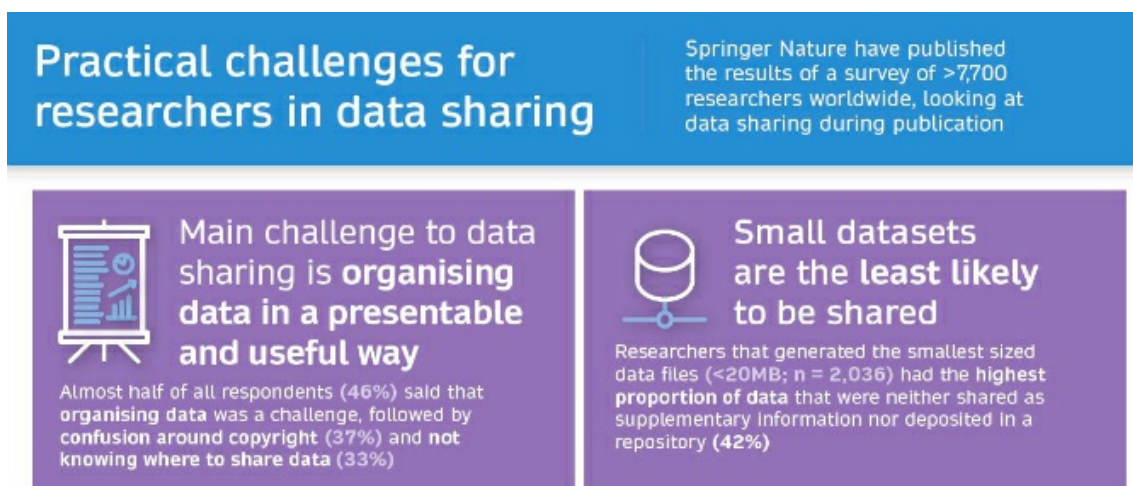
From my conversations with scholarly communications officers working in UK research institutions, I think that time may be a more important issue than is reported in surveys such as ours. The issue for researchers may not be purely "lack of time" but "is it worth my time?" Published, citable datasets need to be viewed as research outputs on a par with a research article in terms of career advancement and assessment. We need to measure the usage and citations of datasets, and communicate the impact and benefits of data sharing. In the meantime, data publishing, and better data citation and linking, are part of the solution.

While sharing data as supplementary information is better than not sharing data at all, it is a sub-optimal solution. Data deposited in a repository is more findable and accessible. A number of publishers, including Springer Nature, are now depositing supplementary information into publicly accessible repositories. Through the Research Data Alliance, a group of publishers, funders, and research institutions are collaborating to agree a framework for journal data policies, to reduce complexity for authors and encourage good practice. Initiatives such as DataCite and the Joint Declaration of Data Citation Principles help make research data more citable and discoverable.

Scholarly communications offices and libraries have a key role to play in supporting researchers. In many research institutions, libraries and research data management teams are now offering expert advice, support, and infrastructure. Researchers in these institutions are fortunate to have such support. Governments, funders, institutions, libraries, and service providers like publishers all have a role to play to unlock the huge potential of research data. For example, at Springer Nature we offer a free Research Data Support Helpdesk and recommended repositories list, as well as an optional Research Data Support service to help researchers and institutions deposit their data in repositories and make it easier to find and use.

I don't underestimate the size of the challenge. We are talking about shifting expected norms, skills, and behaviour so that data sharing and good practice becomes standard research practice. As well as policy, researchers need incentives, expert support, training, and infrastructure to make it seamless and easy to share data, and worth their while. That support needs to come when they need it, in ways that are accessible, easy to use, and that work with their research workflow. Achieving this is too complex, and the potential benefits too great, for a fragmented approach.

This blog post is based on the white paper, "Practical Challenges for Researchers in Data Sharing", which is available on Figshare (DOI: 10.6084/m9.figshare.5975011.v1).

Featured image credit: The Bait by nist6dh (licensed under a CC BY-SA 2.0 license).

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

Grace Baynes *is Vice President, Data & New Product Development for Open Research at Springer Nature. She is responsible for promoting open data and good research data practice; data publishing, including the journal Scientific Data; data services; and new product development across open science and open research.*