2018

Op Ed — Open Data, the New Frontier for Open Research

Tim Britton

Springer Nature, researchdata@springernature.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Britton, Tim (2018) "Op Ed — Open Data, the New Frontier for Open Research," *Against the Grain*: Vol. 30: Iss. 1, Article 14.

DOI: <https://doi.org/10.7771/2380-176X.8003>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.
Please contact epubs@purdue.edu for additional information.

Op Ed — Open Data, the New Frontier for Open Research

by **Tim Britton** (Managing Director of Open Research at Springer Nature) <researchdata@springernature.com>

Open science should be about opening up all areas of research, including the data underlying it. The evidence clearly demonstrates that open data and good data management makes research more productive, more likely to be cited and unlocks innovation for the good of society including unexpected new discoveries and economic benefit.

Yet while open access to research articles and books is now increasingly the expected norm, and while up to half of the world's research data are shared (which is still not enough), much, much less data are made *openly* available (let alone FAIR as in Findable, Accessible, Interoperable or Reusable).

But why is this? Why is it that open data is not yet the norm given that it is fifteen years since the **National Institutes of Health (NIH)** released its trail-blazing 2003 Statement on Sharing Research Data?

The **NIH** statement reads:

"We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The NIH endorses the sharing of final research data to serve these and other important scientific goals. The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers. Starting with the October 1, 2003 receipt date, investigators submitting an NIH application seeking \$500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible."

And this statement was *before* **NIH's** 2004 Public Access Policy which, together with the **Wellcome Trust**, led the way in funders encouraging, and then requiring, open access to articles, with the consequent growth of openly available article content.

Why is it that since the WorldWideWeb was invented in 1989 to help researchers connect and collaborate, nearly thirty years on the majority of the world's research data still lays dusty, unloved and undiscoverable in the proverbial or literal desk drawer?

I believe macro trends are at play here which help explain why open article accessibility has outstripped the underlying data accessibility: first and

foremost is that it is easier to achieve the first. The tools emerging from our digital revolution are often designed to deliver pre-formed, consumable content to users. Whether it is YouTube footage, financial news or academic articles, digital dissemination and consumption tools are widely available and easy to use for this purpose.

The consumer culture is one of expecting consumable content to be freely available: from music and film to news and analysis, pay walls are few and far between. This is the way the world has changed in the last 15 years and it is no surprise that academic publishing is following suit with open access.

Sharing data in a meaningful, usable way requires more complex tools and a degree of user expertise. There is a cultural point here too, in that the underlying building blocks of content (in our case data) are not necessarily expected to be as freely available. But, as we are witnessing in other sectors, these structural and attitudinal barriers are starting to come down.

Opening up data to make it useful to others is a challenge not unique to the research community. I came to scholarly communication and publishing from a long career in market research and using data to draw new insights. Before joining **Springer Nature**, I was EMEA Chief Operating Officer of **YouGov**, the international Internet-based market research and data analytics firm and later head of strategy and transformation for **PwC's** global data research and insight centre, **r2i**. In both these roles I saw a move away from focussing on data production and collection, or more accurately repetitive data creation, to a focus on the reuse of data already amassed with the application of new experimental efforts on where data had not previously been collected.

The benefits from this are clear: time and money is not needlessly spent repeating the work of others; researchers can focus on the analysis and interpretation of, rather than production of, data and the unintended benefits of a particular dataset are much more likely to be uncovered. Indeed, since open government data has become the norm (from social statistics through weather

data to procurement records) there is clear evidence of both economic and societal benefit. A 2013 **McKinsey** study found that "seven sectors alone could generate more than \$3 trillion a year in additional value as a result of open data, which is already giving rise to hundreds of entrepreneurial businesses and helping established companies to segment markets, define new products and services, and improve the efficiency and effectiveness of operations." According to the Open Data Institute, "Those studies focused on the value of public sector open data alone found that it is worth between 0.4% and 1.5% of an economy's GDP."

In the same way that the value of data reuse and sharing has been recognised in other sectors, there is great potential for it in academic research and publishing. The Human Genome Project is a success we can build on. From an investment of \$14.5 billion by the U.S. government, the Wellcome Trust and others, the whole human genome was sequenced and the data is open to anyone. The estimated contribution to the U.S. economy alone was \$1 trillion in the first decade after this data became available, according to a report by the **Battelle Memorial Institute**.

As a publisher I am keen to recognise and acknowledge the importance of research articles and books. These are important summaries and conclusions of years of work for researchers. However, the real building blocks and arguably the real value of discovery lie in the data — the excel tables sitting on a PhD student's desktop, the thousands of microbiology images taken to produce an exquisite figure in a research paper, the coded responses to qualitative surveys.

The world's research data is a rich but untapped seam of new insights and unexpected connections. Today, we have the ability through algorithms to explore discarded datasets from experiments that produced negative or "null" results and potentially dig out a few precious gems from these forgotten mines. Sharing this data ensures that the hours spent pipetting in a laboratory is never a waste, and public funding is not spent needlessly repeating the same tests over and over.

continued on page 33

Many funders are now driving data sharing through policy including U.S. federal agencies, the **Bill & Melinda Gates Foundation**, **Howard Hughes Medical Institute**, the **UK Research Councils**, **Wellcome Trust** and **European Commission**. The **NIH** continues to drive change through its trans-**NIH** initiative The Big Data to Knowledge (BD2K) program.

Yet survey after survey shows that researchers' efforts to archive, publish and share data continue to be hampered by time constraints and a lack of knowledge around data standards, metadata and curation expertise, repository options, and funder requirements. There are still few incentives for researchers to share their data, and many challenges, despite their motivations to do so.

According to the recent State of Open Data 2017 (a **Digital Science** and **Figshare** report, produced in collaboration with **Springer Nature** and **Wiley**), the proportion of researchers sharing their data has grown, with an ever greater willingness to reuse open datasets. Eighty per cent of survey respondents are willing to share their data, with the same number already or open to using others' data. Yet only 60% of respondents make their data openly available "frequently" or "sometimes." The most common ways of sharing data are still supplementary information in a journal article or peer-to-peer, which do often not make data openly available or discoverable. We have to make it easier for data to be truly and usefully open.

Over the past twelve months the Open Research team at **Springer Nature** has talked with librarians and funders, whilst continuing to explore the attitudes and behaviours of researchers. The concerns are clear: a lack of researcher expertise in metadata and data curation to describe data so that it can be found easily by others, the lack of established data standards, along with more fundamental questions around identifying which data to archive and preserve and where to deposit it. Funders and institutions also want to support researchers to follow best practice, ensure compliance with data sharing policies, and help bring about a change in culture — where data sharing and data management is the norm.

Institutions and libraries have a key role to play in supporting researchers: helping them understand and comply with funder requirements, and establishing local research data management support, training and solutions where needed. In many research institutions, libraries and research data management teams are now offering expert advice and support, often reskilling teams as experts in data management. Partnering with data initiatives, repositories and other useful parties, including publishers, will help reduce potential duplication of effort and ensure sustainability.

Springer Nature is committed to supporting good research data management and data sharing. We want to help researchers who seek to adopt open approaches to their data wherever possible, and to partner with funders, institutions and community initiatives such as the Research Data Alliance (RDA) and CODATA. Our starting point is to recommend and partner with community and general repositories, rather than keeping data in a proprietary publisher ecosystem. Our list of more than eighty recommended repositories is publicly available under a CC-BY license and regularly updated.

We have long advocated for data availability through editorial and journal policies, and in 2016 we launched our standardised journal data policies to further encourage good practice. As a publisher, you would expect us to offer publishing options to improve data sharing and we do. Our flagship research data journal, *Scientific Data* and our new data note article type in *BMC Research Notes* both offer authors publication credit, while making data easier to find, understand and reuse.

The insights we gained from researchers and institutions led us to consider how we could help them further, drawing on the experience and expertise we have built in data policy, standards and curation. As well as a free helpdesk to advise researchers and editors on journal data policies and choice of repositories, we have been working on new solutions for researchers and institutions. In 2017, we introduced a new Research Data Support service to enable authors to share their data. Research Data Editors curate and enhance metadata to improve discoverability and encourage reuse in accordance with the FAIR data principles, and help authors draft data summaries and data availability statements to improve human readability and data linking and citation.

This January we extended Research Data Support to make it available to all researchers. We also extended the service to institutions who want to support best practice data de-

positing by their researchers, or complement in-house research data management teams. Our Data Curation Editors, Research Data Editors and in-house data policy experts are now available for training including on policy and best practice in data management, and hands-on data curation workshops.

At the Research Data Access and Preservation (RDAP) Summit in Chicago this March, we will consider the "Intersection of Publishing and Data" and share some of our learning from the implementation of our Research Data Support service. This will include how data curation standards have been developed and how the data services integrate with the established publishing workflow. Also in March, we will be revealing the results of a recent survey which explores how researchers are sharing data associated with published research.

Open science offers huge opportunities through open data. Good data practice should speed the pace of discovery by increasing reproducibility, making research more productive, and delivering return on investment to the global economy. As with open government data, we now need the infrastructure, funding and skillset embedded into the research community to make data sharing and open data the new normal in academic research. Many of the building blocks exist today, including positive researcher attitudes, community repositories, funder policies, data publishing options, and the growing research data management expertise in institutions, driven by libraries and librarians.

Yet the open data challenge is too big, and the potential benefits too great, for a fragmented approach. We need to join up and collaborate to find shared solutions, and look to and learn from other industries. By working together, governments, funders, institutions, publishers and libraries can unlock the huge potential of open data to improve our knowledge, the global economy, our health and environment. 🌱

Tim Britton, Managing Director of Open Research Group, Springer Nature

Tim Britton joined **Springer Nature** in October 2016 as Managing Director of the Open Research Group. With decades of experience working in the marketing and research information industries, **Tim** is responsible for the open research portfolio across **Springer Nature**, including **BioMed Central**, **SpringerOpen**, the open access journals from **Nature Research** and open access monographs from **Springer** and **Palgrave Macmillan**.

Prior to joining **Springer Nature**, **Tim** was Head of Strategy & Transformation at **r2i** (PwC's global data research and insight centre of excellence). Before that, he spent eight years at international Internet-based market research firm **YouGov** in a number of leadership roles, most recently as Chief Executive UK and Head of Global Reports.

To find out more about the new **Springer Nature** research data services available to institutions, please email <researchdata@springernature.com>. 🌱