Implementing publisher policies that inform, support and encourage authors to share data: two case studies

Open research data is one of the key areas in the expanding open scholarship movement. Scholarly journals and publishers find themselves at the heart of the shift towards openness, with recent years seeing an increase in the number of scholarly journals with data-sharing policies aiming to increase transparency and reproducibility of research. In this article we present two case studies which examine the experiences that two leading academic publishers, Taylor & Francis and Springer Nature, have had in rolling out data-sharing policies. We illustrate some of the considerations involved in providing consistent policies across journals of many disciplines, reflecting on successes and challenges.

**Keywords**
Research data policies; publishers; open data; data availability; open scholarship

**Introduction**
Open science/open scholarship/open research – whichever term is preferred – refers to a set of perspectives, techniques and tools that seek to enhance the transparency, reproducibility and overall robustness of research. While open access (OA) focuses on unrestricted access to the research article, conversations around open scholarship reach beyond this, considering how the whole research lifecycle can be opened up, with the ultimate aim of improving the quality and integrity of research.

Open research practices are gaining in traction, with increasing examples and discussion about the opportunities that openness brings for individual researchers, and for other players in this ecosystem, such as research councils, funders and international bodies.

One of the key strands in the expanding discourse around open scholarship centres on open research data. Recent research from figshare indicates that ‘open data is becoming more embedded in the research community’ with 64% of respondents to their State of Open Data survey claiming that they made their data openly available in 2018 (a 4% increase on 2017).
This rise in a positive attitude towards sharing data corresponds with a general industry development, with more institutions and funders recognizing the value of data sets.

In recent years research data policies have been used by funding agencies, publishers and research institutions to drive research data sharing. Requirements across these stakeholders vary, but tend to include common aspects, such as the preparation of data management plans, data archiving, citation of data sets and inclusion of data availability statements in published research papers; and in some cases, the peer review of research data. In October 2018 the Belmont Forum, which represents 26 funding agencies internationally, released a draft position paper outlining requirements for data availability statements to be made available outside any paywalls applied to articles, and the minimum features they should include, such as persistent identifiers for the data set and licensing and access information. These developments suggest a growing consensus among funding agencies and publishers on key features of data-sharing policies.

In journal publishing, which is our focus in this article, adherence to a data-sharing policy supports authors in making the data associated with a publication available as an open data set so that the conclusions reached in the publication can be checked and verified. As scholarly journals and publishers find themselves at the heart of the shift towards openness, it is not surprising that 2017–2018 saw an increase in data-sharing policies from major publishers, resulting in a significant number of scholarly journals with policies aiming to increase transparency.

In this article we present two case studies which examine the experiences that two different publishers have had in rolling out data-sharing policies. Both are leading academic publishers, publishing scholarly journals, books and educational reference material. Taylor & Francis Group is a publisher of scholarly journals, books, e-books, text books and reference works in all areas of the humanities, social sciences, behavioural sciences, science, technology and medicine sectors. Springer Nature is a research, educational and professional publisher and is home to brands including Springer, Nature Research, BioMed Central (BMC), Palgrave Macmillan and Scientific American. In the first case study, we reflect on Taylor & Francis’ experience as we approach the first anniversary of the launch of the publisher’s data-sharing policies. We focus on how the policies have been devised with consideration of very different research subject areas, on the response to the policies from the communities Taylor & Francis work with, and how this is shaping future work. In the second case study, the more ‘mature’ Springer Nature policies, which have been available since 2016, are examined, describing initiatives that are being undertaken to enhance compliance with these policies, and to raise the profile of data availability statements which the policies either recommend or require.

The intention in presenting these two case studies is to illustrate some of the considerations involved in providing consistent policies across journals of many disciplines, and how publishers inform, support and encourage authors to share the data underpinning their research. Following the two case studies, the work of other stakeholders, including the Research Data Alliance and Center for Open Science, is also outlined and future plans for aligning data policies and supporting data sharing are described.

Case study 1: Taylor & Francis
Data-sharing policies at Taylor & Francis

At Taylor & Francis, we launched our data-sharing policies at the start of 2018 to signal our support for increased transparency in academic research, and to communicate that we are ready to work with individual journals on necessary policy and workflow changes to facilitate this. We recognize that publishers are well positioned to catalyse conversations about data sharing within the research community via the societies, editors and authors we work with.

Taylor and Francis data-sharing policies are tiered, with the five standard policies offering increasing levels of expectations around how and when data should be shared. The basic, or
entry-level policy, encourages authors to share and cite data. At the more progressive end, the open and fully FAIR (findable, accessible, interoperable and reusable) policy mandates making data open under a CC BY, CC0 or equivalent licence in full compliance with FAIR principles. Full details of our suite of data-sharing policies can be viewed on our Author Services site. The five policies are:

- **basic**: journal authors are encouraged to share and make data open where this does not violate protection of human subjects or other valid subject privacy concerns. Authors are further encouraged to cite data and provide a data availability statement.

- **share upon reasonable request**: authors agree to make their data available upon reasonable request. It’s up to the author to determine whether a request is reasonable. Data availability statements are mandatory.

- **publicly available**: authors make their data freely available to the public under a licence of their choice. Data availability statements are mandatory.

- **open data**: authors must make their data freely available to the public, under a licence allowing re-use by any third party for any lawful purpose. Data availability statements and data citation are mandatory. Data shall be findable and fully accessible.

- **open and fully FAIR**: authors must make their data freely available to the public under a licence allowing reuse by any third party for any lawful purpose. Data availability statements and data citation are mandatory. Additionally, data must meet with FAIR standards as established in the relevant subject area.

For each policy, authors are encouraged or mandated to take specific action in relation to providing a data availability statement, citing data in line with the FORCE11 Joint Declaration of Data Citation Principles, providing a persistent identifier and selecting an appropriate licence for any data sets.

For all the policies there is, of course, the understanding that exceptions will need to be granted, for example where the sharing of data conflicts with a need to protect personal identities, where authors do not have ownership of the data in question, where release of the data poses a security risk, and other reasonable situations.

Recognizing that the launch of these data-sharing policies represents a significant business and cultural change, we have taken a phased approach to the rollout, with a focus in the first year (2018) on securing adoption of the basic policy across the majority of our portfolio. As of the end of 2018, over 1,600 journals published by Taylor & Francis (which includes over 300 journals published in conjunction with a learned society) have the basic data-sharing policy in place. (See Figure 1, showing the percentage of these journals by subject area.) Additional journals are agreeing on a regular basis.

Figure 1. Taylor & Francis journals which have adopted the basic data-sharing policy by subject area (at the end of 2018)
The decision to roll out the basic policy across many journals in a relatively short period of time has required the internal co-ordination of multiple departments across our journals business, as well as consultation with editors and society partners.

To support authors with the rollout of the policies, we have information available on our Author Services site, for example guidance on Data Availability Statements and data citation best practice, information on the selection of data repositories, and policy FAQs (Frequently Asked Questions). We are adding to this bank of material regularly and as required. As well as this, we have set up a dedicated mailbox – datasharing@tandf.co.uk – for questions on internal and external data sharing. We respond to a steady flow of queries via this channel, and we have recently refreshed the FAQs on our Author Services site in line with the queries and cases received.

Recognizing readiness to share

The Taylor & Francis data-sharing policies were devised with consideration of the range of disciplines we publish – from AHSS (arts, humanities and social sciences) through to STEM (science, technology, engineering and medicine) – and to allow for the varying types of underlying research that will be emerging from these communities. We recognize that some subject areas and some regions are more ‘ready’ and positioned to share data than others.

The so-called ‘replication crisis’ in psychology, where scientists have found it difficult or impossible to reproduce the results of previous studies, has led to engagement in practices that facilitate and encourage transparency. Many journals in this area were early signatories of the Center for Open Science TOP guidelines,12 eight modular standards which support the drive for reproducible research. More recently the earth, space and environmental sciences have demonstrated growing support for best practice around open data, with a strong Commitment Statement emerging from the Enabling FAIR Data Project which aims to accelerate scientific discovery and enhance the integrity, transparency and reproducibility of this data. This will see many journals in this discipline making publication ‘conditional upon the concurrent availability of the data underpinning the research finding’.13

Our discussions about open data with editors, societies and authors in the humanities indicate that there are still important questions to consider around the type, amount and description of ‘data’ that these scholars work with. Humanities scholars collect complex resources during the course of their research, many of which are not traditionally structured so don’t easily fit the quantitative definition of ‘data sets’, e.g. ephemera in archives, photographs, oral testimonies, etc. Add to this the challenge of data ownership or control, particularly, for example, in ethnographic projects,14 which impacts on the ability to share openly. We understand that while many humanities scholars embrace the open principles of sharing, reuse and collaboration, the relationship that humanists have with their underlying research resources differs from that in other disciplines,15 and this needs to be reflected in data management and sharing policies.

With all this in mind, it was important for us to take an approach with our policies that would allow exploration of data sharing at levels that our editors and societies were comfortable with and that would feel appropriate for them, their disciplines and the authors they work with. At the same time, we wanted to present an ultimate end goal (i.e. open and fully FAIR) that could be achieved with our support and guidance.

Response to policy rollout … and future plans

It is fair to say that the response to the launch of our policies has been polarized, with some partners describing the change as ‘important and valuable’ while other editors or society officers expressed feeling ‘alarmed’ and ‘not enthusiastic’ about the prospect of open data for their subject area or journal(s). Key concerns raised include fears of
data misuse or unethical sharing: will data sharing mean more ‘scooping’ (when another author group publishes similar work before you do)? Some of the editors we work with are, understandably, worried about the impact of a new journal data-sharing policy on their already heavy workloads, wondering whether the change will mean more checks and challenges to contend with during the peer-review process. There is also the knotty question of credit and how researchers will/should be recognized for more transparent practices. This suggests that there is still work to do around answering the ‘What’s in it for me?’ question. We appreciate all this valuable feedback and we are conscious that this sort of policy needs to work for everyone involved.

While digesting the feedback received on our policies to date, we have also reflected on the rollout approach we have taken. Our decision to focus on driving adoption of our basic policy (where the key word is ‘encourage’) in the first year naturally has pros and cons. We still have a lot of work to do to move journals on to more progressive policies, but we look forward to working with our partners on that priority in 2019 and beyond. We appreciate that institutions are picking up on the opportunities for better management of underlying research data that has been conducted by researchers in their institution, and we will be looking to further gather the library’s point of view as our progress in this area continues.

Challenges aside, we have succeeded in rolling out a data-sharing policy for a significant number of journals in a short time. In doing so, we are exposing a large number of authors to the notion of a data-sharing policy. We are facilitating conversations about the benefits of open practices and encouraging changes to existing behaviours around data. We think that is a valuable starting point.

**Case study 2: Springer Nature**

**Data Policies at Springer Nature**

Aiming to address the issues of complexity and lack of clarity for authors across the data policy landscape, Springer Nature began rolling out standard research data policies in 2016. The policies offered are based on a framework developed to address the differing requirements of approximately 2,500 journals across multiple disciplines. The exact specifications and requirements of each policy are available on the Springer Nature policy website, where the policy text is licensed openly as CC BY and can be reused by any stakeholder requiring a similar standard policy.

Four standard policy types are available, with increasingly stringent requirements for data sharing by authors as they progress from Type 1 through to the Type 4 policy.

- **Type 1 journals**: authors are encouraged to share their data, preferably in repositories, and to cite publicly available data sets in their reference lists.

- **Type 2 journals**: authors are strongly encouraged to share their data, preferably in repositories, and to cite publicly available data in their reference lists. Authors are also encouraged to include a statement of data availability with their manuscript.

- **Type 3 journals**: authors are strongly encouraged to share their data, preferably in repositories, and to cite publicly available data in their reference lists. Authors are required to include a statement of data availability with their manuscript.

- **Type 4 journals**: authors are required to share their data in a repository, to include a statement of data availability with their manuscript, and to make their data available for peer review.

Slightly modified versions of the Type 2 and Type 3 policies are available for life sciences journals, which include more specific information regarding the obligations of authors publishing in these disciplines to share certain types of data – such as genetic sequences and protein structures – in recommended community repositories.
The communication of the recommendations (or requirements) of each policy type to authors is key to their successful implementation, as is making sure there are procedures and resources in place at the journals to ensure that authors comply with mandatory policy requirements. Updates to journal websites and manuscript submission systems, which communicate the policy requirements, need to be co-ordinated and rolled out in a way that is appropriate for the technical infrastructure each journal uses. As of November 2018, more than 1,500 journals Springer Nature have a standard policy, with additional journals being implemented on a weekly basis. At time of writing in early 2019, 39% of the journals with a policy have Type 1, 34% have Type 2, 26% Type 3, and less than one per cent (six journals) have Type 4 (Figure 2). The last two policy types, which mandate actions on the part of the authors, have a lower proportion of uptake from journals. In due course, it is anticipated that more journals will adopt the higher-level policies, and the publisher has an important role in supporting and enabling this transition.

![Figure 2: Overview of Springer Nature data policy implementations by policy type](image)

**Promoting data availability statements**

A key aspect of the Type 2, 3 and 4 data policies at Springer Nature is the recommendation (or requirement) that an author includes a statement of data availability with their manuscript. This statement indicates to readers where the data supporting the results reported in their article can be found, or, in the case of some authors, that no data were generated by the study. The use of these statements provides a standardized way for authors to describe how their data are shared. Data availability statements do not necessarily mean data are readily available, as authors may still choose to share data on request, or state that their study did not generate or analyse data. In spite of this, they are a visible and consistent means for demonstrating compliance with journal policies. These statements are also required by a growing number of research funding agencies as part of their data-sharing policies, including the seven funding agencies that form UK Research and Innovation (UKRI).

The implementation of data policies can be a driver for these statements becoming a standard feature across published research, encouraging transparency and enabling better monitoring of data-sharing practices and policy compliance. Data availability statements were first introduced in 2016 by Nature Research journals, bringing their policies in line with Springer Nature’s Type 3 data policy, and have since been made into a distinct section of each article, similar to the methods section. This gives information on data availability similar prominence to Methods and Results, and as part of the Nature Research journals they are freely and universally accessible in subscription and OA titles.
The Research Data team at Springer Nature have also sought to analyse the impact of introducing new data policies in terms of the impact on authors and the ways in which they describe the availability of their data, and the impact on publishers in terms of the time and cost of introducing new statements to published articles. The analysis indicated that when required to describe how their data were shared, life scientists were the most likely to make their data available in a repository, while those in the physical sciences were least likely to. The analysis also found that introducing mandatory data availability statements increased manuscript processing time by around ten minutes overall, giving an indication of the increased costs that stronger data policies may lead to. This is something that publishers and journals can consider alongside the anticipated benefits of a new data policy.

**Communicating policies and compliance to authors**

The data policy of a particular journal is communicated to its authors in two ways: through its instructions for authors and via its manuscript submission system. Some journals also introduce policy requirements in e-mail correspondence templates. The exact wording and presentation of the policies may appear differently across Springer Nature imprints, for example Springer, BioMed Central (BMC) or Nature journals, but the information provided is the same, and the purpose (to ensure that authors are aware of their obligations) is identical. It is essential that these authors are aware of any obligations they may have, and that a failure to comply may have consequences, for example preventing their manuscript from being sent out to peer review.

For Type 3 policies, another critical part of implementing the policy is ensuring those responsible for checking that articles contain the necessary mandatory sections, such as data availability statements, are trained and resourced to do so. Additionally for life science journals, it is also essential to ensure procedures are in place to check discipline-specific requirements (such as DNA/RNA sequence deposition) are followed. How these are implemented will depend on the editorial model of the journal. Springer Nature introduced additional administrative support for journals edited by academics, rather than professional editors, to help them ensure data availability statements were consistently provided along with functional links to data sets, where applicable.

Further to supporting compliance with editorial procedures at specific journals, Springer Nature launched a research data helpdesk at the same time the policy initiative was introduced. The research data helpdesk is a free service which is available to authors or editors at any Springer Nature journal, or anyone who has a query regarding research data. The helpdesk provides assistance for any query relating to research data, but the majority of correspondence is from authors who would like advice on complying with their journal’s data policy, including the selection of appropriate repositories and drafting data availability statements. In the first two years, however, the helpdesk answered a large proportion of queries from editors who had questions about which policy they should adopt and how they should implement it. To further support the rollout of the policies, guidance and FAQs on key policy requirements were provided on the publisher’s website, such as a curated list of recommended (trusted) data repositories arranged by research discipline and advice on writing data availability statements. For journals adopting the Type 4 policy, research data must be available to peer reviewers, and all supporting data sets must be publicly available when the articles are published (with some exceptions for sensitive research data permitted) and cited in the article’s reference list. The data journal *Scientific Data* exemplifies the implementation of the Type 4 policy, and has shared its experience of peer reviewing and enabling access to data sets supporting its publications, including for clinical research data.

**Future plans**

Springer Nature will continue to work towards more complete policy coverage for all of its journals in 2019. The process of implementation involves one-to-one correspondence with editors and their societies, where applicable, rather than policies being enforced.
by the publisher. Additionally, support and guidance will be given to journal editors who would like to move their journals to higher policy levels, with a focus on moving towards Type 3 policies.

Springer Nature will also continue to develop its list of recommended repositories in response to developments in different research communities. In January 2019, for example, more guidance was added for the earth sciences, which have begun to increase their expectations for data sharing.23 Along with other publishers, Springer Nature is also implementing data citation into its content production workflows, enabling data citations in reference lists to be specifically identified—supporting more sophisticated linking of articles and data in repositories and promoting data as a first-class scholarly output.23

Additionally, Springer Nature is developing other ways of supporting data sharing in response to the needs of researchers and is conducting research to determine what these needs are. In 2018 Springer Nature published a large survey—conducted in 2017—with more than 7,000 responses that found researchers face common practical challenges in data sharing, with a lack of skills, awareness and capacity to organize data for sharing being a key issue.24 In response, in 2018, we introduced an optional Research Data Support service, providing more active support to researchers who are interested in data sharing but need professional assistance to do so. Also in 2018, a training programme for researchers, available to institutions, was launched to provide researchers with the skills they need to organize and share data themselves. Finally, compliance with policies, and data sharing, is enabled with appropriate infrastructure—in particular, data repositories. As well as recommending discipline-specific repositories, Springer Nature is continuing to extend the availability of integrated figshare repositories to its journals and conference proceedings.

Conclusion

The two case studies described in this paper demonstrate the common concerns of publishers in providing standard research data policies to their authors. Although the introduction of standard data policies is intended to clarify data-sharing requirements for researchers, several factors must be considered by publishers when developing and introducing them. While encouraging good practice, policies must also be developed to reflect the culture of the discipline they apply to, and publishers must ensure that researchers understand what their obligations are. The communication of policies is crucial, as authors can only comply with policies when they are aware of what they need to do. Compliance is also key, and where policies mandate certain behaviours (such as the inclusion of data availability statements in published papers), editorial staff must be trained and resourced to make the required checks. For data-sharing policies to be successfully implemented and enforced, appropriate support must be available in terms of documentation, web-based resources and online helpdesks to provide feedback and guidance.

Research published by Jisc back in 2016 indicated that as data policies proliferated across journal publishers, the policy landscape for journals had become too complex.25 The development of standard policies by Springer Nature and Taylor & Francis has addressed this complexity through the provision of tiered policy types, aiming to accommodate journals across disciplines, including those with a less-developed culture of data sharing. Other scholarly publishers have taken similar approaches, including Elsevier which provides five data-sharing options for its journals,26 and Wiley which has developed four data-sharing policies.27 Although publishers’ data policies reflect similar concerns and requirements, the existing policy types and levels do not map exactly to one another, and authors’ obligations are not necessarily identical across every publisher. This could lead to confusion for authors, who still need to distinguish between differing requirements for data sharing depending on who they are publishing with.

*‘Springer Nature is … implementing data citation into its content production workflows’*

*‘where policies mandate certain behaviours … staff must be trained and resourced’*
To address the proliferation of multiple tiered policy types and provide a global forum for many stakeholders, a Research Data Alliance Interest Group was established in 2017 with the aim of defining common frameworks for research data policy, while allowing for different levels of commitment and requirements and disciplinary differences. This is a collaborative project which is chaired by representatives of the funder, publisher and data-sharing communities. The Interest Group has developed a Journal and Publisher Research Data Policy Master Framework which was shared with the community for review in 2018. The Framework is based on a review of requirements from existing scholarly publisher data frameworks and supporting documentation, as well as the Current Best Practice for Research Data Management Policies commissioned by CODATA (Committee on Data of the International Council for Science) and the Transparency and Openness Promotion Guidelines. Initiatives such as this framework will ensure that publishers continue to support and encourage data sharing across research disciplines while moving towards consistent policy requirements which can be easily understood by researchers.

Acknowledgements
The authors would like to thank Caroline Sutton, Vicky Farrimond and Katherine Burton at Taylor & Francis for their input.

Abbreviations and Acronyms
A list of the abbreviations and acronyms used in this and other Insights articles can be accessed here – click on the URL below and then select the ‘full list of industry A&As’ link: http://www.uksg.org/publications#aa

Competing interests
The authors have declared no competing interests.

References
6. Belmont Forum Data Accessibility Statement and Policy: https://docs.google.com/document/d/1AwrVzYI1yXsY5vXrPahIECGTt_Osw1EKCLc5c7-rGyHAX9mrhM/edit (accessed 8 March 2019).


23. Cousijn et al., “A data citation roadmap”.


25. Naughton and Kernohan, “Making sense of journal policies”.


