

This is a reformatted version of Springer Nature's response to the NIH Request for Information on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research, which was submitted online at <https://osp.od.nih.gov/provisions-data-managment-sharing/> on 10th December 2018

NIH Office of Science Policy
NIH Office of the Director
SciencePolicy@od.nih.gov

10th December 2018

Springer Nature response to Request for Information on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research

Prepared by Iain Hrynaszkiewicz, Head of Data Publishing and Rebecca Grant, Research Data Manager

To Whom it May Concern,

Introductory comments

Springer Nature is committed to open research and offers researchers, institutions and their funders open access (OA) options for journals, books and sharing research data. We also offer expertise and services that support research data management and sharing.

Springer Nature has an established interest in policies on data sharing and data management and welcomes this opportunity to provide comments to the NIH on the development of its policies. To our knowledge, more than 80 funding agencies globally now have data sharing policies with around half of those with a policy requiring data management plans (DMPs). However, how policies are implemented, and their compliance monitored, and how these activities are resourced are vitally important^{1,2}.

Many Springer Nature journals publishing in the life sciences, including Nature and the BMC journals, have operated data sharing policies for many years. More recently, in 2016, Springer Nature launched an initiative to standardise, harmonise, and implement consistent data sharing policies to all its journals. Several other large publishers, including Elsevier, Wiley, Taylor & Francis, BMJ journals, have since introduced similar initiatives and globally publishers are collaborating via the Research Data Alliance (RDA) to standardise and implement journal policies.

Specific comments requested by NIH

Definition of Scientific Data

Our first comment is regarding the term “scientific data”. Given research is increasingly interdisciplinary and international, and to ensure an inclusive policy to all NIH grantees, we suggest

using “research data” instead of “scientific data”. Springer Nature - which works with researchers from the humanities through to the life and physical sciences - consciously uses “research data” in the context of data policy development and implementation across our organisation, and product and service portfolios (<https://www.springernature.com/gp/authors/research-data>).

In our experience of research data policy and service development and implementation, defining the scope of the policy is critical. As a research publisher, this generally means focusing on research data that specifically support scholarly publications and we welcome this explicit definition (“including, but not limited to, data used to support scholarly publications”) in the policy draft. In addition, it may be helpful to reference or support established community norms for data sharing (such as DNA/RNA sequence deposition), in cases where they are not covered by the first definition. We additionally recommend NIH define whether or not their definition of “scientific data” includes:

- All data generated by a grant
- All data analysed (including data from third parties) by grant holders in their research

Material from (electronic) laboratory notebooks, and preliminary analyses, can be used to support scholarly publications (for example, in³) so we suggest these not be excluded from the policy and definition. Also, we recommend being explicit about whether research software and code are included in the policy.

In addition, we would caution referencing “Data sharing...in a manner that is consistent with the FAIR principles”. While the FAIR principles are important framework for data sharing and increasingly used by policy makers, institutions and those already active in the research data management (see Jisc report for more information⁴), a recent survey by Digital Science found that more than 60% of researchers were not familiar with FAIR⁵. This might suggest that, where researchers are the audience, a more general or accessible definition for data sharing may be helpful.

Requirements for Data Management Plans

As part of our Nature Research Academies training on research data sharing, we also provide training to researchers in preparing data management plans and interpreting the guidance provided by their funder or institution. We have found that it is beneficial to focus on the practical elements of data management planning, to ensure that a realistic and actionable plan is developed. For this reason, it might be beneficial to include additional resources for researchers to assist them in preparing their plans, e.g. examples of well-written plans and a list of sources where they can find information on good practice in data storage, metadata creation, and repository selection.

More detailed guidance could be provided on the following:

- Required retention periods for the data, including a defined minimum, for example 10 years from the date the data were last accessed.
- Data types which do not have a discipline-specific repository available.
- Issues of copyright and data ownership.

It seems somewhat arbitrary to have a two-page limit on these plans - a one- to two-page minimum might work better in order to encourage detailed planning.

In relation to compliance and enforcement, consideration should be given to the likelihood of a plan changing from the time that a grant proposal is drafted, to when the data collection begins, through data analysis and archiving. It would be more beneficial for the researcher to consider the plan as a living document which should continuously be referred to and continuously be updated. For this reason, annual reviews could aim to ensure that the plan is being updated on a regular basis, rather than being complied with rigidly.

To provide additional opportunities for researchers to consider their plan in a practical way, it might also be helpful to add the following topics:

- Short term storage (while the research is being conducted).
- Resourcing (both technical infrastructure required and the skills and/or training needs of the researchers to carry out the plan).

Timing, implementation, resources

The focus of the policy as worded appears to be on the requirements for data sharing and data management plans, with fewer details on compliance and enforcement, and on how researchers, support staff (including programme officers) will be supported to implement and monitor compliance with the policy. An analysis of research data management policy implementation from several case studies by Cameron Neylon (2017), highlighted, amongst other relevant findings, that “for DMP requirements to be supportive of culture change they need to be well supported with expertise, systems and guidance in place,” and that “Requirements imposed on data sharing...must be auditable and audited”².

Where and how researchers can best receive support - such as through formal training, through supervision or dedicated data management staff and help desks - will vary. Support for creating and implementing DMPs can be available from institutions, funding agencies and third parties such as the Digital Curation Centre and scholarly publishers’ researcher services

(<https://www.springernature.com/gp/open-research/institutions/research-data-services>).

Leveraging a wider “market place” for resources to support open science has become a key part of the European Commission’s Open Science Cloud (<https://www.eosc-portal.eu/for-providers>). We would welcome more detail on how NIH will support researchers in preparing and implementing their DMPs. As well as support in terms of tools and training, we recommend explicit clarity about use of NIH-funding for costs associated with data management and curation, storage and publication fees of articles that describe datasets (such as data papers, data notes, data descriptors).

We recommend more consideration be given to the mechanisms by which compliance with the policy can be demonstrated by researchers. Monitoring of compliance with data sharing policies and data management plans is, understandably, challenging because the practices, expectations and available infrastructure (including repositories) for sharing research data vary by discipline. Other funding agencies, and research publishers have implemented requirements for transparency in reporting of information about data availability, as a prerequisite to monitoring compliance. This is

achieved by requiring researchers provide, in their publications, a 'Data availability statement' (sometimes called 'Data accessibility statement' or 'Data sharing statement').

Provision of Data availability statements is a requirement of hundreds - if not thousands - of journals and publishers' research data policies. These statements are a common feature of publishers' data sharing policies. Data availability statements in researchers' publications are also part of the policies a number of funding agencies including the seven UK Research Councils (<https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/>). There is, also, evidence that consistent statements of data availability, combined with a mandatory deposition policy, is the most effective approach for journals to ensure data availability supporting publications long term⁶.

Journals and publishers that require these statements in publications include all the Nature and BMC (BioMed Central) journals, as well as PLOS, BMJ, and others. More recently, publishers and the Belmont Forum (which includes the US National Science Foundation) group of agencies are collaborating to introduce consistent requirements for Data availability statements for the member agencies, their grantees, and publishers⁷. This more consistent reporting of data availability also supports the utility of literature search and evaluation tools and services designed to determine availability of data supporting scholarly publications (for example, <http://blog.europepmc.org/2018/11/mapping-out-path-to-data.html>).

For the above reasons, we recommend NIH consider including requirements for the provision of data availability (accessibility) statements in its Data Sharing and Data Management policy requirements.

Other common features of research data policies, to publishers and funding agencies, include support for formal citation (referencing) of research data in reference lists (bibliographies), provision of researcher support (helpdesk) functions, and collaboration with trusted data repositories for implementation.

In addition to reviewing DMPs for compliance, consideration should be given to how researchers can be credited and rewarded for good practice. This can be facilitated, for example, by promoting the sharing of datasets in repositories that assign persistent identifiers (such as DOIs) to datasets so that datasets can be formally cited, and tracked, in scholarly publications. Data citation was a topic of a previous NIH RFI, to which Springer Nature responded, and more information on our proposals for data citation can be found in our response⁸. We note that the National Science Foundation encourages researchers to list their "research products", including datasets and software, on their bibliographic sketches⁹. It is also possible to share DMPs publicly, in repositories and furthermore some journals, such as *BMC Research Notes*, will consider publishing them as peer-reviewed articles.

Other remarks

Regarding the statement: "NIH encourages the sharing of data for as long as it is useful to the scientific community." It would be helpful to clarify if this refers to the NIH's overall position on data sharing, and if so if this represents any change in stringency of the policy compared to the 2003 policy. That is, does the NIH *encourage* data sharing by all grant holders or *require* it?

Regarding the length of time that data archiving, the UK Research and Innovation (UKRI)'s guidelines reference a minimum of 10 years from date of last access. Similarly, when assessing data repositories for Springer Nature's recommended repository list¹⁰, we typically look for sustainability of infrastructure for a minimum of 10 years (<https://www.nature.com/sdata/policies/data-policies#repo-suggest>).

Acknowledgements

For comments on the first draft of this proposal thanks to: Sowmya Swaminathan, Head of Editorial Policy, Nature Research Group, Springer Nature, and Grace Baynes, VP Research Data and New Product Development, Springer Nature.

Springer Nature is happy to be contacted for more information about our response (researchdata@springernature.com or iain.hrynaszkiewicz@nature.com) and encourages NIH's further participation in international fora such as the RDA data policy standardisation and implementation Interest Group (<https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation-ig>).

Yours faithfully,

Iain Hrynaszkiewicz

Bibliography

1. Grant, R. & Hrynaszkiewicz, I. The impact on authors and editors of introducing Data Availability Statements at Nature journals. *BioRxiv* (2018). doi:10.1101/264929
2. Neylon, C. Building a culture of data sharing: policy design and implementation for research data management in development research. *RIO* **3**, e21773 (2017).
3. Panetta, J. L. *et al.* Reptile-associated *Borrelia* species in the goanna tick (*Bothriocroton undatum*) from Sydney, Australia. *Parasit. Vectors* **10**, 616 (2017).
4. Allen, R. & Hartland, D. FAIR in practice - Jisc report on the Findable Accessible Interoperable and Reuseable Data Principles. (2018).
5. Science, D.. The State of Open Data Report 2018. (2018). doi:10.6084/m9.figshare.7195058.v2
6. Vines, T. H. *et al.* Mandated data archiving greatly improves access to research data. *FASEB Journal* fj.12-218164- (2013).

7. Murphy, F. & Samors, R. J. Belmont Forum Data Accessibility Statement Policy and Template - Endorsed 18 October 2018. (2018).
8. Springer Nature response to NIH RFI on Strategies for NIH Data Management, Sharing, and Citation. (2017). doi:10.6084/m9.figshare.4616293.v1.
9. Piwowar, H. Altmetrics: Value all research products. *Nature* **493**, 159 (2013).
10. Data, S. Scientific Data recommended repositories. (2018). doi:10.6084/m9.figshare.1434640.v12