Data Management Planning

NSF funding applicants

Version 1.5 June 2021

University of Bristol

Research Data Service

Image: *Auliscus oamaruensis* by Anatoly Mikhaltsov,

University of BRISTOL

SUMMARY

The National Science Foundation (NSF) expects grant-holders to share data generated by NSF-funded projects with other researchers. The following points are intended as a summary of NSF expectations on data sharing:

- Investigators must share data generated by NSF-funded work; this includes samples, physical collections, software, and models
- Where data cannot be openly released due to privacy or other confidentiality issues, you should consider whether altering the form of your data (for example, anonymisation or aggregation) or applying re-use conditions (for example, a licence or non-disclosure agreement) would allow your data to be safely used by others
- A two-page data management plan is required with all NSF grant applications to describe how your proposal meets the NSF data sharing policy
- Data management costs may be included in your budget and budget justification

INTRODUCTION

The NSF policy on dissemination and sharing of research results states that grant holders are expected to share "primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants."[1] This should be done within a 'reasonable time', which is typically at the time of publication of the main findings, but may be later if this is the standard in your research community. An exception is made for privileged or otherwise confidential information, which should be released "only in a form that protects the privacy of individuals and subjects involved."[2] Applicants are required to submit a Data Management Plan (DMP) outlining how their proposal complies with this policy.

Applicants to the following directorates should follow the directorate- or division-specific guidance below; otherwise, the general guidance[3] outlined in this document applies.

- Biological Sciences Directorate: http://www.nsf.gov/bio/biodmp.jsp
- Computer & Information Sciences & Engineering (CISE): http://www.nsf.gov/cise/cise_dmp.jsp
- Education & Human Resources Directorate: http://www.nsf.gov/bfa/dias/policy/dmpdocs/ehr.pdf
- Engineering Directorate: http://nsf.gov/eng/general/ENG_DMP_Policy.pdf
- Geosciences Directorate: http://www.nsf.gov/geo/geo-data-policies/index.jsp
- Mathematical and Physical Sciences Directorate
  - Division of Astronomical Sciences: http://www.nsf.gov/bfa/dias/policy/dmpdocs/ast.pdf

[1] NSF policy on Dissemination and Sharing of Research Results, http://www.nsf.gov/bfa/dias/policy/dmp.jsp

[2] NSF Award and Administration Guide, chapter VI, http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp

[3] NSF Grant Proposal Guide, chapter II, http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp

- Division of Chemistry: http://www.nsf.gov/bfa/dias/policy/dmpdocs/che.pdf
- Division of Materials Research: http://www.nsf.gov/bfa/dias/policy/dmpdocs/dmr.pdf
- Division of Mathematical Sciences: http://www.nsf.gov/bfa/dias/policy/dmpdocs/dms.pdf
- Division of Physics: http://www.nsf.gov/bfa/dias/policy/dmpdocs/phy.pdf

- Social, Behavioral and Economic (SBE) Sciences Directorate: https://www.nsf.gov/sbe/DMP/SBE_DataMgmtPlanPolicy_RevisedApril2018.pdf

What counts as data?

The NSF recognises that the nature of research data will vary from discipline to discipline, but may include samples, physical collections, software and models as well as outputs such as text documents, spreadsheets/databases, images, video and audio recordings. There may be Directorate-specific guidance outlining the scope of material to be covered by a data management plan so you should consult with the relevant program officer.

Data Management Plan format

A data management plan is an integral part of the grant application and explains how the research proposal conforms to the NSF policy on dissemination and sharing of research results. For NSF grant applications, your DMP should be no more than two pages. It is considered under either or both of the application review criteria (Intellectual Merit and Broader Impacts), and should cover the following areas:

- Types of data
- Data and metadata standards
- Access policies
- Re-use policies
- Archiving

If you do not anticipate generating any data, you should still complete a DMP. However, in this case, it can simply consist of a statement that there will be no data to manage or share, along with a clear justification. Conversely, if your data management plan is complex and you are unable to fit it within the two-page limit, you may use part of the 15-page Project Description to capture additional details.

Types of data

Here you should note the types of data generated in your study – for example, the format and likely volume of any digital files, the nature and quantity of any samples collected, and any original software or models developed. You should also think about how and where your data will be stored and backed up. Where possible, you should use open file formats; if this is not possible whilst you are actively working with your data, consider converting proprietary formats into open standards at the point of archiving.

Data and metadata standards

In order for your data to be usable for your own project and re-usable by others, it will need to be documented and organised. There are many data and metadata standards available to aid data

interoperability, both general and subject-specific; these can help you standardise terminology, define minimum information required for data capture, and select formats which allow for effective data exchange. In this section, you should note existing standards that you plan to use; if there are none, state this, and describe what you plan to do instead. Also, note here any policies and procedures related to data capture and cleaning.

Access policies

In this section, you should consider access to your data both during the lifetime of your project and any subsequent sharing. If your project involves collaboration between two or more institutions, think about how intellectual property and other rights will be distributed. The RED contracts team[4] can help you draw up a formal collaboration agreement to address this. If you are collecting sensitive data (e.g. personal information about project participants, commercially sensitive information, data with national security implications or data concerning rare or endangered species), how will you ensure security and confidentiality? Do you need to anonymise data before sharing, and if so, how will you achieve this?

Re-use policies

If your study uses sensitive data as outlined above, or data from a third party, you may not be able to openly share your raw data. However, by applying appropriate re-use policies (for example a licence or non-disclosure agreement) sharing under more restrictive conditions may be possible. In this section, you should outline the terms and conditions that will apply to others wishing to use your data, including any restrictions on the production and dissemination of derivatives. In particular, if your data has been anonymised, consider whether re-identification of participants is possible via linkage with other datasets, and if so, how this will be managed. The RED commercialisation team[5] can provide advice on applying licences to your data.

Archiving

The length of time that you are required to preserve and maintain access to your data for will vary depending on the standards within your research community; however, 10 years is typical. In this section, you should outline how you plan to archive your research data – for example, in a repository or biobank (for physical samples), or on a project website. If you are planning to use a project website or otherwise administer data sharing yourself, you should explain how you plan to ensure access is maintained over the long term. Please note that this can be a significant administrative burden over the long term, so the use of a repository for data preservation is strongly recommended. If you are unable to find a subject-specific repository to host your data, you can use generic repositories such as Figshare[6] or Zenodo[7], or the University's own Research Data Repository, data.bris[8]. If you choose to use data.bris, there are a number of different access levels which can be applied

[4] RED contracts team, http://www.bristol.ac.uk/red/contracts/
[5] RED commercialisation team, http://www.bristol.ac.uk/business/research-commercialisation/
[6] https://figshare.com/
[7] https://zenodo.org/
[8] https://data.bris.ac.uk/data/

to your data, and a standard data access agreement is available governing access to controlled datasets.[9]

Costs

Applicants should include data management costs in the budget and budget justification. When thinking about data management costs, you should consider whether additional equipment for data storage will be required as well as staff time for tasks such as processing (for example data cleaning prior to analysis) and archival preparation tasks such as anonymisation and format conversion.

Help and advice

Please contact the Research Data Service (data-bris@bristol.ac.uk) for advice on creating your data management plan.

[9] data.bris policies on dealing with sensitive research data, http://www.bristol.ac.uk/staff/researchers/data/dealing-with-sensitive-data/