

Guidance for Penn State Researchers on the NSF Data Management Plan Requirement

Introduction

In the interest of ensuring and broadening access to federally funded research data, the National Science Foundation now requires researchers to submit a **data management plan (DMP)** with their grant proposal application. This guidance is intended to help Penn State faculty develop a data management plan that explains how research outcomes will be described, shared, and preserved for future access. (For further information about this requirement, please see the GPG, Chapter II.C.2.j (http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp).

Please note: The guidance in this document should be considered only as advice based on experience working with researchers on data management planning, and should not be construed as legal or compliance advice on specific matters. If you have specific questions or concerns on legal or compliance matters, you should consult with appropriate legal counsel, the Office of the Vice President for Research or the appropriate program officer for your research proposal.

A Quick Glance: The Five Components of a Data Management Plan

In no more than two pages, a DMP generally addresses the following five areas:

1. **Types of data** – what kinds of data will your research produce?
2. **Data and metadata format and content** – how will you document and describe your data?
3. **Access and sharing policies** – how will you provide access to your research data, and what privacy considerations, if any, are at play?
4. **Policies and procedures for reuse, redistribution, creating derivatives** – what restrictions will there be on the data?
5. **Archiving and preserving the data** – how will the data be preserved, to continue to make it accessible?

Use the guidelines below to develop your data management plan.

Section 1: Types of Data

“Data” are defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings. This includes original, derived and processed data, but also “metadata” (e.g. experimental protocols, code written for statistical analyses, etc.). In this section please describe the types of data that will be produced in the project. These may include samples, physical collections, software, curriculum materials, or other materials to be produced in the course of the project.

Points to Consider:

- Types of data that will be generated by your research

- For example, human subjects related surveys, interviews, field data, model output data
- Data format(s) and file types (e.g., .txt, .pdf, .xls, .csv, .jpeg, etc.)
- How the data will be collected or accessed (if using existing data)
- Which data will be handled in special ways (e.g., human subjects, special agents proprietary data) and show appropriate compliance committee approval paths

Section 2: Data and Metadata Format and Content

Documentation of data – how they are described - is an important factor in the development of a data management plan. How well data are documented determines how easily it can be discovered and re-discovered by others who may need to use these files. Creation of metadata is a key way to keep your data well organized. You may be familiar with some metadata standards for your specific discipline; if not, then please contact the Research Data Management Services Team at the University Libraries (UL-RDMST@lists.psu.edu), and we can offer further guidance on this point.

Points to consider:

- Information about your data you will need to save (i.e., experimental design, environmental conditions, global positioning information, etc.)
- What metadata standard you will use to document your data (i.e., some research domains have widely accepted formats, others may not and you may target how that decision may be made in the project)
- How you plan to record your metadata

Section 3: Access and Sharing Policies

This section is a key reason why the NSF now requires a data management plan as part of the proposal. Here, you should provide a statement about access to the data and any policies related to sharing the data, particularly if the data your project will produce will require restricted access (e.g., because of human subject research), or a period of embargo (e.g., for publisher, patent, or commercial reasons). That is, any issues regarding "privacy, confidentiality, security, intellectual property, or other rights or requirements" need to be stated here, as suggested in the NSF Grant Proposal Guide.

PSU policies that apply to this section and that should be reviewed:

- AD20: Computer and Network Security: <http://guru.psu.edu/policies/AD20.html>
- AD23: Use of Institutional Data: <http://guru.psu.edu/policies/AD23.html>
- AD35: University Archives and Records Management: <http://guru.psu.edu/policies/AD35.html>
- IP01: Ownership and Management of Intellectual Property: <https://guru.psu.edu/policies/IP01.html>
- IPG01 - Faculty Guidance on Student Intellectual Property Rights: <https://guru.psu.edu/policies/IPG01.html>
- IPG02: Special Student Intellectual Property Agreement Forms: <https://guru.psu.edu/policies/IPG02.html>
- RA12: Technology Transfer and Entrepreneurial Activity: <http://guru.psu.edu/policies/RA12.html>

- RA14: The Use of Human Participants in Research: <http://guru.psu.edu/policies/RA14.html>
- RA22 - HIPAA and Research at Penn State University: <http://guru.psu.edu/policies/RA22.html>
- RA23 - HIPAA and Research at The Milton S. Hershey Medical Center and Penn State College of Medicine: <http://guru.psu.edu/policies/RA23.html>
- RAG16 - The Responsible Conduct of Research: <http://guru.psu.edu/policies/RAG16.html>

Points to consider:

- Expected availability of the data during the project period
- The way(s) in which you will make the data available
 - Modes of access (website? data repository?)
 - Levels of access (everything accessible? restricted access? embargo period?)
- List/Explain any ethical or privacy issues incurred by the data
- Address any intellectual property rights issues (e.g., who holds these rights to the data?)
 - It's appropriate to aggregate the experience of PIs or Co-PIs in terms of the number of non-disclosure/confidentiality agreements executed to reflect experience, if applicable to your project
 - Who will administer any non-disclosure, confidentiality or IP agreements?
- Compliant with IRB requirements? (explain clearly whether the IRB data will be handled differently, and briefly discuss the processes performed on those IRB that will be shared – e.g., anonymous, aggregated)

Section 4: Policies and Procedures for Reuse, Redistribution, Creating Derivatives

Providing a statement about policies and procedures for repurposing and redistributing the data from your project is a natural next step after mapping out policies for access and sharing. For example, if in the above section you have asserted access restrictions for your data, then such a policy will influence whether (and how) the data may be reused, redistributed, or allowed to be a source for derivations.

Points to consider:

- What you will permit in terms of reuse and redistribution of the data, based on policies for access and sharing
- Think about what other researchers (whether in your subject domain or others) may find your data useful
- Identify the lead person or committee on the project who will make the decisions on redistribution on a case-by-case basis
- If you have already identified a national repository (or national repositories) or intend to explore those possibilities, then please state them

Section 5: Archiving and Preservation

In this section, describe where and how the data (or samples, or other research products) will be archived and preserved over time. In addition, an archiving and preservation plan should account for any materials that should be stored with the data in order for them to be read, understood, worked with, etc.

Points to consider:

- Will all of the data produced on your project be preserved, or only some?
- Context for your data (e.g., tools, project documentation, metadata etc.) required to make it accessible and understandable
- Anticipated transformations of the data in order to deposit it and make it available
- The length of time the repository will be available to the public and/or maintained (some directorates have a suggested minimum for the time after a project ends or after publication of certain data)
 - If beyond life of project, state your plan for funding long-term storage (if significant storage capacity is involved, this could be included in a commitment letter if one is allowed in the solicitation) – for guidance, consult the Libraries’ Data Management List: l-data-mgmt@lists.psu.edu
- Review external data repositories to see if your data would be appropriate for deposit in one of them
 - DataBib is a growing registry or bibliography of research data repositories around the world: <http://databib.org/>
 - This is a reliable working list of data repositories: <http://www.datacite.org/repolist>

Examples of data entry/submission requirements:

- Inter-University Consortium for Political and Social Research (ICPSR)
 - Deposit Guidelines: <http://www.icpsr.umich.edu/icpsrweb/deposit/>
- Australian Social Science Data Archive
 - Deposit Guidelines: <http://www.assda.edu.au/deposit.html>
 - Deposit Form: http://www.assda.edu.au/forms/ASSDA_Deposit_Form.pdf
- The Netherlands’ Data Archiving and Networked Services (DANS)
 - Deposit Guidelines: <http://www.dans.knaw.nl/en/content/data-archive/depositing-data> (home page)
- Dryad Deposit Recommendations:
 - Data Archiving Policy: <http://datadryad.org/jdap>
 - Data Submission Guidelines: <http://datadryad.org/pages/faq#depositing>
 - Dryad FAQs: <http://datadryad.org/pages/faq>