

Writing Data Management Plans

Many funding agencies require that grant proposals address how data will be managed and shared with other researchers. For example, the National Science Foundation now requires that all proposals for funding include a data management plan “...to stimulate new advances as quickly as possible and to allow prompt evaluation of the results.” NSF expects investigators to share with other researchers the primary data created or gathered in the course of work under NSF grants. The data management plan must show how data will be made available within a reasonable amount of time and at no more than incremental cost. This guide is designed to assist PIs by describing what information is required in a data management plan and providing links to other resources that may be helpful when preparing a plan.

Contents:

1. [What is a data management plan?](#)
2. [What constitutes “data”?](#)
3. [Where do I find the guidelines that apply to my proposal?](#)
4. [Guidance from NSF](#)
5. [Guidance from the EHR Directorate](#)
6. [MSU Resources](#)
7. [Elements of a Data Management Plan](#)
 - a. [Types of data produced](#)
 - b. [Data and metadata standards](#)
 - c. [Policies for access and sharing](#)
 - d. [Policies for re-use, redistribution](#)
 - e. [Plans for archiving & preservation](#)
8. [Additional Resources](#)

1. What is a data management plan?

A data management plan is a supplementary document, no longer than two pages, that describes the kinds of data to be produced, how they will be managed, and how research results will be made accessible to other researchers. The plan is a supplementary document and is not part of the 15-page limit for the Project Description. NSF will reject any proposal that lacks a data management plan for competitions that have a due date or target date on or after January 18, 2011. A proposal must include a DMP even if no data are produced. Fastlane will not permit you to upload a proposal that lacks a data management plan.

Data management plans must address the following issues:

1. What data are produced?
2. What standards will be used for data and metadata format?
3. What policies exist for access to and sharing data?
4. What are the plans for archiving data, samples and other research products and preserving access to them?

Collaborative proposals and proposals with subawards are treated as a single unified project and should include only one combined Data Management Plan, regardless of how many non-lead collaborative proposals or subawards are included.

2. What constitutes “data”?

According to the federal government, data is defined as:

“...the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleges. This recorded material excludes physical objects (e.g. laboratory samples). Research data also do not include:

- (A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
- (B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.”

[Source: OMB Circular A-110](#)

3. Where do I find the guidelines that apply to my proposal?

Guidelines for data management plans are laid out in the NSF's [Proposal and Award Policies and Procedures Guide \(PAPPG\)](#). In addition, the Education and Human Resources Directorate has issued its own [guidelines and recommendations](#). These guidelines are not intended to replace the guidance given in the PAPPG; if there is a conflict, the PAPPG takes precedence. Finally, some NSF solicitations may impose their own data management requirements (see [NSF 09-514](#) for one example).

4. Guidance from the National Science Foundation

The NSF Policy on Dissemination and Sharing of Research Results ([Award Administration Guide, Section VI.D.4](#)) states:

- a. Investigators are expected to promptly prepare and submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work conducted under NSF grants. Grantees are expected to permit and encourage such publication by those actually performing that work, unless a grantee intends to publish or disseminate such findings itself.
- b. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF

grants. Grantees are expected to encourage and facilitate such sharing. Privileged or confidential information should be released only in a form that protects the privacy of individuals and subjects involved. General adjustments and, where essential, exceptions to this sharing expectation may be specified by the funding NSF Program or Division/Office for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interest of investigators. A grantee or investigator also may request a particular adjustment or exception from the cognizant NSF Program Officer.

c. Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.

d. NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.

e. NSF program management will implement these policies for dissemination and sharing of research results, in ways appropriate to field and circumstances, through the proposal review process; through award negotiations and conditions; and through appropriate support and incentives for data cleanup, documentation, dissemination, storage and the like.

5. Guidance from the Education and Human Resources Directorate

NSF's Education and Human Resources Directorate (EHR) published [guidelines for data management plans](#) in March 2011. In it, the EHR states that:

"The DMP should reflect best practices in the PI's area of research and should be appropriate to the data generated. For proposals submitted to EHR, the plan should address two main questions:

- What data are generated by your project?

- What is your plan for managing the data?”

EHR encourages PIs to address the following issues in a DMP:

- The types of data that their project might generate and eventually share with others, and under what conditions
- How data are to be managed and maintained until they are shared with others
- Factors that might impinge on their ability to manage data, e.g. legal and ethical restrictions on access to non-aggregated data
- The lowest level of aggregated data that PIs might share with others in the scientific community, given that community’s norms on data
- The mechanism for sharing data and/or making them accessible to others
- Other types of information that should be maintained and shared regarding data, e.g. the way it was generated, analytical and procedural information, and the metadata

6. MSU Resources

To respond to the emerging concern surrounding research data management, the MSU Libraries and University Archives have joined forces to create the [Research Data Management Guidance service](#). The service is available to provide guidance to faculty in the development and execution of research data management plans.

Data management consultants are available to assist faculty with their data management plans. Consultants meet with researchers to learn about the specifics of their research projects and tailor data management best practices to meet their specific proposal requirements. This service is offered free of charge. Advising sessions may cover funding requirements and data management best practices for:

- Organizing Data (best practices, file plans, software)
- Storing and Preserving Data (MSU storage options, external repositories, file formats)
- Documenting Data (barebones documentation, metadata standards)
- Sharing Data (publishing data, ensuring access, licensing data, citing data)

Contact the Research Data Management Guidance service at: Researchdata@mail.lib.msu.edu

7. Elements of a Data Management Plan

While there is not an official template for a data management plan, the following is a description of the topics that need to be addressed in a plan.

a. Types of data produced

In this section, you give a short description of the data, including the amount of data to be produced (if known) and the content. If the project will be collecting data of a sensitive nature, you should note that here and reflect upon it in subsequent sections. Examples of data types could include text, curriculum materials, spreadsheets, images, 3D models, software, audio files, video files, field observation reports, surveys, assessment records, etc. If you will be using existing data, state that fact along with where and how you will obtain it.

b. Data and metadata standards

Here you will describe the format of your data and how it will be created or captured. You should consider what details (metadata) another researcher would need to be able to use the data. For example, a "readme file" may be needed to explain variables and how files are structured. If existing standards are absent or inadequate, you should document this along with any proposed remedies. If the data will be of a sensitive nature that precludes its release to the public in raw form, you should address here how access will be achieved (e.g. making the data anonymous, level of aggregation, formal consent agreements, restricting access to make it available only through a secure network.) Consider the steps that you have taken to comply with your obligations in your Institutional Review Board Protocol when addressing this issue. Discuss what form the metadata will take, what details are needed to make the data meaningful, and which metadata standard you will use and why you have chosen them (e.g. accepted domain-local standards, widespread usage).

c. Policies for access and sharing

In this section you should explain how and when the data will become available. Issues that you should address here include:

- How will you make the data available? (Address resources that are needed to make the data available: equipment, systems, expertise, etc.)
- When will you make the data available?
- What is the process for gaining access to the data?
- Will access be chargeable?
- How long will the principal investigator retain the right to use the data before making them available for wider distribution?
- Are there any embargo periods for political/commercial/patent reasons? (If so, give details.)
- Who will hold the intellectual property rights to the data and how might this affect data access?

d. Policies for re-use, redistribution

In addressing this issue, give some thought to who is likely to be interested in the data. Identify who will be allowed to use your data, how they will be allowed to use your data and whether or not they will be allowed to disseminate your data. If you are planning on restricting access, use or dissemination of the data, you must explain how you will codify and communicate these restrictions. If permission restrictions need to be placed on the data, discuss that here.

e. Plans for archiving & preservation

In this section, discuss the long-term strategy for maintaining, curating and archiving the data. Address where your data will be deposited (name the archive, repository, or database) and discuss the procedures that your intended long-term storage site has to preserve and backup data. If transformations are needed to prepare data for preservation and data sharing, be sure to mention that along with metadata/documentation that will be submitted with the data or created on deposit/transformation in order to make the data reusable. State how long the data will be kept beyond the life of the project.

8. Additional Resources

Below are links to some resources that may be useful when preparing a data management plan.

Inter-university Consortium for Political and Social Research
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/index.jsp>

ICPSR is an international consortium of about 700 academic institutions and research organizations that provides leadership and training in data access, curation, and methods of analysis for the social science research community. ICPSR maintains a data archive of more than 500,000 files of research in the social sciences. It hosts 16 specialized collections of data in

education, aging, criminal justice, substance abuse, terrorism, and other fields.

MIT Data Management Website

<http://libraries.mit.edu/guides/subjects/data-management/>

California Digital Library (CDL)

<http://www.cdlib.org/services/uc3/datamanagement/>

University of Virginia Library, Scientific Data Consulting Group

http://www2.lib.virginia.edu/brown/data/NSFDMP_EHR_Template_with_Guidance.doc

DataONE DMPTool

<https://dmp.cdlib.org/>.

DataONE, supported by NSF, has created a web-based tool called "DMPTool" that creates a ready-to-use DMP tailored for specific funding agencies. The tool gives you step-by-step instructions and guidance for your data management plan as you build it. DataONE recently added support for a DMP submitted to the EHR directorate.