Data Management Plan Tools

Indiana University Guidance on NSF Data Management Plans

September 28, 2012

Effective for proposals submitted on or after January 18, 2011, the National Science Foundation (NSF) requires the inclusion of a supplementary document of no more than two pages entitled "Data Management Plan" (DMP). The plan should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (see AAG Chapter VI.D.4 [1]).

Following is a guide to writing your DMP, consisting of the following sections:

- **Section 1** gives a template and consideration points for completing a data management plan.
- **Section 2** is a short set of boilerplate language for your use when composing your DMP.
- Because no single template works for every discipline and community, **Section 3** lists additional resources that could be helpful in figuring out what works for your needs.
- **Section 4** is explains the Indiana University (IU) resources that are available for your use. It is useful if you want to use UITS storage and one of the institutional repositories (IUScholarWorks or IUPUIDataWorks) as your data preservation solution.

This document is prepared with fonts and margins consistent with the NSF Grant Program Guide, so researchers may cut and paste directly from this document when preparing data management plans.

This document is derived, in part, from a report by a blue-ribbon panel of IU experts led by Professor Beth Plale, of the School of Informatics and Computing (SOIC). As such, this guidance reflects the combined effort and consensus thinking of IU's top experts in data management and reflects IU guidance for NSF Data Management plans endorsed by the Office of the Vice President for Research and offered in a manner that is consistent with the Indiana University Information Technology Strategic Plan [6].

1. Data Management Plan Template

A data management plan meeting the general NSF requirement can be organized by the below template, though not all items will be relevant for all disciplines, Directorates, or solicitations. See [2] for discipline specific advice. It may be helpful to begin your DMP with a few sentences describing the research project in general, to provide general context for the detailed information in each section. In each section, you should describe your reasoning, particularly if you are deviating from common practice or standards used in your discipline or community of practice. Identifying a particular person or role to carry out these activities is also vital.

1. Describe the **types of data and products** that will be generated in the research such as samples, physical collections, software, and curriculum materials. Characterize the data with details such as the types of data (text, numeric, images, audio, video, etc.) and an approximate number and size of files to be generated or used. Provide a brief description of the data collection process, including instruments or tools, sites, and process for getting data into a secure location. In addition, briefly describe the storage/backup plan and the IU cyberinfrastructure to be used.

2. Describe the **format in which the data, metadata, and other products are stored**. Describe the formats in which the data will be stored, preferably using a common or open file format standard. Include a description of the metadata that will make the actual data products useful to the general researcher. Some examples of discipline-based metadata standards include the NanoParticle Ontology [3] and Ecology Metadata Language (EML) [4].

3. Describe the **policies for general access** to data including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. "Access to data" refers to data made accessible without explicit request from the interested party. Policies for access and reuse should clearly when, how, and to whom the data will be made available. Describe the policies and mechanisms for access to the data and other products, including specific provisions (described in the next section) for appropriate protection of privacy, confidentiality, security,

intellectual property, or other rights. Mechanisms should provide for access beyond the life span of the project, preferably via institutional or community infrastructure (i.e., institutional or subject repository).

Reminder: NSF allows grantees to retain principal legal right to intellectual property developed under NSF grants.

4. Describe **policies and provisions for re-use, re-distribution, and the production of derivatives**. Clearly describe the proposed policies and rationale for limitations on others' ability to re-use, re-distribute, and produce derivatives of the data and other products. These policies may be developed in response to the ethical and legal issues identified in the previous question (#3).

5. Describe **plans for archiving data**, samples, and other research products, and for preservation of access to them. Identify the data that is appropriate for preservation and the means through which preservation of digital and physical materials will be ensured. If the data will be preserved by a third party, refer to their preservation plans. If the data will be preserved at your institution, describe the cyberinfrastructure that will be used.

Depositing data into an institutional or subject repository ensures access to the data beyond the life span of the project. If you are interested in using an Indiana University repository (IUScholarWorks, IUPUIDataWorks) for your data, see Section 4 below.

2. Boilerplate Language

Introduction

This plan describes the management, dissemination, retention, and archiving of the research data produced during the proposed project. The staff of [INSERT YOUR DEPARTMENT OR LAB NAME HERE], with the assistance of the [IU Libraries-Bloomington/IUPUI University Library] and University Information Technology Services (UITS), will provide for sustainable discovery, access to, and preservation of these data for use by other researchers, instructors, and interested members of the public for the length of this project and at least three years beyond. This will be facilitated through data and publication deposits in existing open-access disciplinary and/or institutional repositories.

Data Formats and Description

We will utilize the Dublin Core metadata scheme to capture information about the data collected during the course of our research. We will work with a metadata expert from the [IU Libraries/IUPUI University Library] to create a working template that captures each dataset's metadata throughout the research process. Upon completion, we will export this data to Dublin Core format, which conforms to the data submission requirements of the IUScholarWorks and many other relevant museums/repositories.

Access, Re-Use, Re-Distribution, and Derivative Works Policies

[If no sensitive or personally-identifiable information is used, include this:]

All data produced during this research will be available freely to the public; we anticipate no sensitive or confidential data. Under the terms of the Creative Commons Zero Universal 1.0 Public Domain Dedication (CC0 1.0; http://creativecommons.org/publicdomain/zero/1.0/), users may share, create, and/or adapt these data/databases.[1]

---

[1] If you wish to retain attribution rights so that anyone who uses your data must credit you as the creator, IU recommends you apply the Open Data Commons Attribution License (ODC-BY; http://opendatacommons.org/licenses/by/) to your data instead of CC0. In your plan, replace the noted sentence with the following: "Under the terms of the Open Data Commons Attribution License (ODC-BY; http://opendatacommons.org/licenses/by/), users may share, create, and/or adapt these data/databases with proper attribution."

[If sensitive or personally-identifiable information is used, include this instead:]

Results, data, and collections will be made available to other researchers in a timely basis with [EXAMPLE] limitations. Sensitive and confidential data collected will be treated following [HIPAA/IRB] regulations, and an added layer of security will be implemented using [STRATEGIES SUCH AS DATA ENCRYPTION, RESTRICTED ACCESS, OR THE SEPARATION OF IDENTIFIABLE DATA]. Under the terms of the Creative Commons Zero Universal 1.0 Public Domain Dedication (CC0 1.0; http://creativecommons.org/publicdomain/zero/1.0/), users may share, create, and/or adapt data/databases made freely available.[2]

Data Archiving and Preservation

To increase access to the published research that has been funded, the research collaborators will deposit peer-reviewed or pre-print manuscripts (with linked supporting data where possible) in the [IUScholarWorks/IUPUIScholarWorks][3] institutional repository. Other works, including presentations and white papers, will also be made accessible via the [IUScholarWorks/IUPUIScholarWorks] institutional repository.

Digital data will be stored using the Indiana University Scholarly Data Archive (SDA; https://pti.iu.edu/storage/sda), a distributed storage service that is centrally supported across mirrored tape silos in Bloomington and Indianapolis. Data stored on the SDA that will be made freely available will be archived in the [IUScholarWorks/IUPUIDataWorks] repository, which will provide a user-friendly interface for the organization, context, and discoverability of data. This combination of [IUScholarWorks/IUPUIDataWorks] and the SDA provides mirroring, redundancy, media migration, access control, file integrity validation, embargoes, and other security-based services that ensure the data are appropriately archived for the life of the project and beyond the project if necessary.

3. Resources

NSF funds a wide range of research. Some directorates and programs have provided specific guidance, which can be found at Dissemination and Sharing of Research Results [2]. In the absence of specific guidance, the Award & Administration Guide (AAG) Chapter VI.D.4 [1] applies.

Data management plan examples spanning a range of disciplines are available from the Inter-University Consortium for Political and Social Research [6]. Additionally, a Data Planning Checklist [7] can be helpful in preparation.

For more help: Skilled Librarians and grant writers are available to assist you in developing a data management plan, identifying appropriate data and metadata standards, finding resources on developing policies for sharing and reuse of data, locating community- or discipline-based data repositories, and finding resources on data management and preservation. To arrange a consultation to meet your needs, contact the data services program for your campus[4]:

---

[2] If you wish to retain attribution rights so that anyone who uses your data must credit you as the creator, IU recommends you apply the Open Data Commons Attribution License (ODC-BY; http://opendatacommons.org/licenses/by/) to your data instead of CC0. In your plan, replace the noted sentence with the following: "Under the terms of the Open Data Commons Attribution License (ODC-BY; http://opendatacommons.org/licenses/by/), users may share, create, and/or adapt these data/databases with proper attribution."

[3] IUPUIDataWorks is the data repository for the IUPUI campus. All other IU campuses should use IUScholarWorks as their institutional repository of choice.

[4] A list of regional campus research data specialists can be found on the IUScholarWorks Data Management Service website.

Indiana University Guidance on NSF Data Management Plans
http://www.libraries.iub.edu/secure/defiles/NSF_DMP_Boilerplate_IUB-IUPUI_Fall_2012.doc

IUB
Stacy Konkiel, E-Science Librarian

Email: iuswdata@indiana.edu
Phone: 812-856-5295

Meryl W. Bertenthal, Proposal Development Specialist
Email: mbertent@indiana.edu
Phone: 812-856-5245

IUPUI
Heather Coates, Digital Scholarship & Data Management Librarian
Email: dataserv@iupui.edu
Phone: 317-278-7125

Ann Kratz, Proposal Development Manager
Email: akratz@iupui.edu
Phone: 317-274-6732

4. IU Storage Systems and Institutional Repository

University Information Technology Services (UITS) maintains a large suite of storage systems. These are described in an extensive document that can be used (in whole or in part) in the Facilities section of an NSF proposal. This document is available online in a link accessible from: http://kb.iu.edu/data/anwu.html. It describes the storage systems operated by UITS and the backup facilities and plans for those storage systems. It also describes data security.

Indiana University has institutional repositories for archiving scholarly and scientific works called IUScholarWorks [8] at Bloomington and IUPUIScholarWorks [9] at Indianapolis. These repositories will accept digital data generated by IU researchers and from National Science Foundation funded efforts with PIs outside IU when there is a formal collaboration with an IU researcher (documented via a Memorandum of Understanding or via a Statement of Work associated with funding to an IU researcher as part of said project). IU, through IUScholarWorks and IUPUIDataWorks, will provide replicated storage of all data sets (as described in detail in the cyberinfrastructure facilities statement).

Both IUScholarWorks and IUPUIDataWorks accept data in all formats. Classified or confidential data requiring formal, contractual, or legal restrictions to access, such as HIPAA-designated Protected Health Information, will not be accepted for deposit, but may be stored on the SDA. In this case, the searching and metadata management facilities that help make these repositories so valuable in disseminating data are not available for use. However, de-identified datasets are eligible for deposit into the repositories. The PI is responsible for ensuring that datasets are appropriately and fully de-identified.

If you intend to use one of the institutional repositories, you should consider budgeting funds for data management:

1. Funding for a person to manage data and metadata. This may be funded within your own research group, or you may consider a consulting arrangement with the IU Bloomington Libraries (contact iuswdata@indiana.edu) or the IUPUI University Library (contact dataserv@iupui.edu).

2. Funding for storage services for exceptionally large data storage needs (more than 50 TB). In this case, please contact researchtechnologies@iu.edu for more information. UITS may be able to offer storage above the default 50 TB limit as part of matching support for grant proposals.

NSF allows for adding data management costs to your proposal (typically Line G2).

Questionnaire to Help with the Creation of a Data Management Plan

JHU Data Management Services of the Sheridan Libraries; datamanagement@jhu.edu

How to use this document

This questionnaire distills NSF's guidelines for what to address in your data management plan. You can use the section headings in your own document. The questions can help you structure the content of each section of the plan. The table in section 1 facilitates listing the different data types for your study. Some researchers are including the table in their plan. Please note that you may not need to address all questions under a numbered category, and in some cases, you may not need to address each category, though any omissions should be justified in your plan. See endnotes for more tips, (view by placing mouse over the blue numbers in text). If you have any questions on the content of this questionnaire, please contact a JHU data management consultant at datamanagement@jhu.edu.

1. **Data Products and Standards**

    Research Outputs

	Data Source	**Data Product**	**Format(s)**	**Estimated Size or Amount**
1				
2				
3				

    You may include this table and use numbers to reference corresponding data types in your plan, or use numbered text paragraphs if needed to fit the 2-page format. 1

    Data and Metadata Standards 2

    1. Do the listed data products use standards for formats or metadata, and why are you using them? If not, will your project develop and maintain standardized formatting and metadata?

    2. What details (metadata) are necessary for others to use your data?

    3. How will metadata be generated (automatically, manually, or both)?

    4. What naming conventions/schema will be used for your data, if any?

    5. What data dictionaries/taxonomies/ontologies will you use for your data, if used within your field?

    6. How will lineage/provenance of some or all of your data be documented (e.g., processing steps executed on raw source data)?

    7. What tools will be required to read the data (e.g., software, instruments)?

2. **Data Storing and Long-Term Preservation** 3

    Storage during project

    8. What digital and non-digital data will be retained during the project?

    9. How (i.e., media) and where (i.e., location(s)) will the data be stored and who is responsible for it?

    10. How and where will the data be backed up and who is responsible for it?

    11. If data need to be secured through access controls (e.g. password-protected network space), how will they be applied? (e.g. local passwords, institutional LDAP or Shibboleth)

    12. If data are stored in an unusual or not generally accessible format, will they be converted to a more common format for storage or sharing?

Revised 2/16/12

Preserving data after project

13. Which digital and non-digital data will be stored or archived after the project? Why will you preserve these data?

14. Will "raw data" (not processed, analyzed or associated with publications) be relevant to store for reuse in your or others' future projects? If so, describe.

15. Where and for how long will data be stored or archived after the project? 4

16. Who will manage and administer the stored or archived data?

17. Will security and access codes be retained on archived data after the project? How?

18. If using a service other than your project group to archive research data, please describe the services that the archive provides in preserving and disseminating research data. Will there be a formal archiving agreement? 5

3. **Data Sharing**

Research to be shared 6

19. Of the data products generated during the project, which data will be shared? 7

20. Which data will be publicly-accessible, if at all? 8

21. When will you share those data? 9

22. How will the data be shared with other stakeholders? (e.g., made available for general access through a public website or database, or released only upon specific request from an interested party. Specify any 3rd party resources or services used.)

23. Who is expected to use the shared data?

Policies for access and sharing 10

24. Identify who will be allowed to use your data, and how data are to be used and disseminated. Explain any restrictions on re-use, production of derivatives and how you will communicate these restrictions, (e.g., requiring citation, or Creative Commons licensing.)

25. Are there any data with confidentiality issues (e.g., embargo period)? If so, what are the conditions of use, sharing, and dissemination?

26. Are there any data with specific security or regulatory concerns with sharing (e.g. classified information or FDA handling requirements), and how will they be addressed?

27. Are there any data with intellectual property (e.g., patent, copyright) concerns with sharing? If so, what are the conditions of use, sharing, and dissemination?

28. Are there any data with privacy concerns to sharing (e.g., human subjects)? If so, what policies need to be adhered to and how will policies be enforced? 11

29. Is any of the data owned by someone else? If so, what are the conditions of use, sharing, and dissemination?

Additional Tips and Instructions (See corresponding endnote number in the text)

1   **Data source** can include instruments, people, and data centers. **Data product** examples: transcripts, tables, 3D models, digital audio, geospatial data. Format examples: RTF text, MS Excel converted to CSV, MATLAB, WAV audio, shapefile. (Specify any instrument-specific formats or software packages). Estimated amount can include rate produced, e.g. 1 TB/year, 50GB/experiment. Include any sources and data products created by others that you are using. It may help to think through the steps of your research workflow to identify data types and sources requiring management.

2   Metadata is the information that captures the who, what, when, where, why and how of your data, providing the details necessary for another researcher to use your data sets. Some scientific communities have established metadata standards, such as Content Standard for Digital Geospatial Metadata (CSDGM), Data Documentation Initiative (DDI), Climate and Forecast (CF) metadata convention, and Dublin Core. Metadata may take the form of "readme files" that explain variables and file structures; however, it is preferable if metadata files are machine readable for better re-usability and processing.

3   Storing data is defined differently than archiving data. Storage is a necessary step towards archiving your data; however, storing data (e.g., on an external drive) does not safeguard against media degradation (e.g., CD file corruption), obsolescence of data formats (e.g., VisiCalc spreadsheets) or providing easy access in the future. Archiving encompasses both active preservation of the digital object and increased discoverability and access to those data. Your plan should discuss how you will store your research data during the project and your preservation strategy for after the project, particularly of research data that will be reused and shared. The next two sections help frame these different topics.

4   JHU requires retention of research data for a minimum period of 5 years after the date of any publication upon which it is based (http://jhuresearch.jhu.edu/Data_Management_Policy.pdf). The NSF Engineering Directorate requires retention for 3 years after conclusion of the award or 3 years after public release, whichever is later.

5   Different data archives provide different kinds of services, such as the creation of persistent, unique identifiers for citation, format migration, disaster recovery plans, and free, publicly-accessible downloading of data files. If you plan to use a data repository, we strongly recommend that you contact the repository to ensure that their archive can handle your data, and determine their archiving fees to include in your budget. Johns Hopkins University has built a research data archive. Please contact datamanagement@jhu.edu to learn more about it.

6   Briefly address the following questions for each data product in Table 1. (You might refer to each by number).

7   NSF expects data sharing to follow the norms of your research community, but encourages efforts to broaden the range of data shared and of potential users beyond your field. Data can often be of unanticipated interest in the future if it can be located, understood, and cited.

8   "Accessible" generally means unmediated public access to your data distributed through a "cyber resource," unless you specify conditions, such as embargo periods. "Sharing" can include direct release to interested parties upon request.

9   Specify a time period, e.g., "Data will be made available for sharing, in principle, two years after acquisition."

10   This section will detail any reasons for sharing delays (e.g., embargo, publisher, patent, or political reasons) or restrictions (e.g. ecological endangerment concerns, IRB restrictions of sensitive data). You should also address granular methods for control and access (e.g., maintaining formal consent agreements, anonymous data, and restricted access to secured networks.)

11   State if there are IRB restrictions on data and steps to prepare accessible datasets, such as deidentifying transcripts. NSF requires fewer details than IRB forms, and respects when IRB restrictions put sharing beyond a reasonable effort, but they do sometimes ask for some attempt to create sharable datasets.

Note: Your DMP for NSF grants should not exceed two pages. Contact Research Services in the Libraries for consultation (ljohnsto@umn.edu).

Data Management Plan

V1 last updated **MM-DD-YYYY**

Name of student/researcher(s)	**Your Name**
Name of group/project	**Project Name or Research Lab (for group plan)**
Funding body(ies)	
Partner organisations	
Project Duration	Start: **MM-DD-YYYY** End: **MM-DD-YYYY**
Date Written	**MM-DD-YYYY**

Table of Contents

1. Introduction

The research project described in this data management plan (DMP) ….

2. Data Types

This types of data generated and/or used in this project include …

Note: Your DMP for NSF grants should not exceed two pages. Contact Research Services in the Libraries for consultation (ljohnsto@umn.edu).

Section 2 Checklist

o What type of data will be produced?
o How will data be collected? In what formats?
o How to document data collection?
o Will it be reproducible? What would happen if it got lost or became unusable later?
o How much data will it be, and at what growth rate? How often will it change?
o Are there tools or software needed to create/process/visualize the data?
o Will you use pre-existing data? From where?
o Storage and backup strategy?

3. Data Organization, Documentation and Metadata

The plan for organizing, documenting, and using descriptive metadata to assure quality control and reproducibility of these data include …

Section 3 Checklist

o What standards will be used for documentation and metadata?
o Is there good project and data documentation format/standard?
o What directory and file naming convention will be used?
o What project and data identifiers will be assigned?
o Is there a community standard for metadata sharing/integration?

4. Data Access and Intellectual Property

The data have the following access and ownership concerns …

Section 4 Checklist

o What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights?
o Does your data have any access concerns? Describe the process someone would take to access your data.
o Who controls it (e.g., PI, student, lab, University, funder) ?
o Any special privacy or security requirements (e.g., personal data, high-security data) ?
o Any embargo periods to uphold?

UNIVERSITY OF MINNESOTA

Data Management Plan

https://docs.google.com/a/umn.edu/document/d/1MxQP-BqDv_fMF12F2ANQ0jYDw__
1xPTTXW14xdLhIyE/edit

Note: Your DMP for NSF grants should not exceed two pages. Contact Research Services in the Libraries for consultation (ljohnsto@umn.edu).

5. Data Sharing and Reuse

The data will be released for sharing in the following way …

Section 5 Checklist

o If you allow others to reuse your data, how will the data be discovered
and shared?
o Any sharing requirements (e.g., funder data sharing policy) ?
o Audience for reuse? Who will use it now? Who will use it later?
o When will I publish it and where?
o Tools/software needed to work with data?

6. Data Preservation and Archiving

The data will be preserved and archived in the following ways …

Section 6 Checklist

o How will the data be archived for preservation and long-term access?
o How long should it be retained (e.g., 3-5 years, 10-20 years, permanently) ?
o What file formats? Are they long-lived?
o Are there data archives that my data is appropriate for (subject-based? Or institutional)?
o Who will maintain my data for the long-term?

UNC UNIVERSITY LIBRARIES

Research Data Toolkit

Home | Agency Rqs/Guidance | Templates/Examples | Services | Training | Assess Data Needs | Tools | Contact Us | About

Templates

Please Note: These examples are not officially sanctioned by any UNC office. They are only intended to serve as examples for what you *might* do. Likewise, the sample plans linked below are very context-specific and are intended only to give a general idea of what others have done.

DMPTool - service of the University of California Curation Center (UC3) and the California Digital Library but customized for UNC at Chapel Hill. Select UNC from the list of institutions and login with your Onyen to see resources specific to our campus. Walks you through requirements for specific funding agencies. Allows you to work in sequence or jump around, save drafts, and export text files.

Sample Plans
Odum Institute's sample plans

ICPSR's sample plan (for deposit with ICPSR)

Natural Science examples, from a wide range of projects and agencies (links collected on the ICPSR website)

Guides for Formulating Data Management Plans
Guidelines for Effective Data Management Plans (ICPSR)

Managing and Sharing Data: Best Practice for Researchers (UK Data Archive)

IRB application with sections that relate to data management

Applicable sections (on pages 9 and 10) include:

- A.4.11 Confidentiality of Data;
- A.4.12 Data sharing;
- A.4.13 Data security for storage and dissemination; and
- A.4.14 Post-study disposition of identifiable data or human biological materials

Odum Institute's data management plan checklist

Data Management & Frequently Asked Questions (FAQs) (NSF)

Other Resources
IRB-required consent form templates for research on human subjects (scroll down to "Consent")

Example Language

For each of the five clauses presented in NSF's Grant Proposal Guide, Chapter II - Proposal Preparation Instructions, Section j. Special Information and Supplementary Documentation (second bulleted point), we have outlined below various points to consider in writing your plan. Where possible, we have adapted text from actual data management plans (although not necessarily plans for NSF) under the heading *Sample Text*. Such examples are not available for every section.

Please Note: These examples are not officially sanctioned by any UNC office at this time. They are only intended to serve as examples for what you *might* do.

If you are willing to share text from your own plan, please **contact us**.

Here is a **Word version** in which to begin drafting your own plan.

From the NSF's Grant Proposal Guide: "Plans for data management and sharing of the products of research. Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan." This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (see AAG Chapter VI.D.4), and may include:

Show All

1. **Types of Data**
   - **Data Description**
   - **Existing Data**
2. **Standards**
   - **Format**
   - **Metadata**
   - **Data Organization**
   - **Quality Assurance**
   - **Responsibility**
3. **Access and Sharing (Including Protected Data)**
   - **Storage and Backup**
   - **Data Access**
   - **Ethics and Privacy**
   - **Proprietary Data**
   - **Intellectual Property**
   - **Legal Requirements**
4. **Re-use**
   - **Access and Sharing**
   - **Re-Use**
5. **Archiving Data**
   - **Archiving and Preservation**
   - **Disaster Preparedness**
   - **Budget**
   - **Selection and Retention**

Admin Sign In