

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** 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 3 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 campus⁴:

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](#).

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).

5. References

- [1] National Science Foundation. NSF Award and Administration Guide, Chapter VI.D.4. http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4 [cited 4 May 2012]
- [2] National Science Foundation. Discipline specific advice on data management plans from NSF directorates. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp> [cited 4 May 2012]
- [3] NanoParticle Ontology. <http://www.nano-ontology.org/> [cited 27 April 2012]
- [4] Ecological Metadata Language (EML). <http://knb.ecoinformatics.org/software/eml/> [cited 27 April 2012]

- [5] Indiana University. 2009. Empowering People: Indiana University's Strategic Plan for Information Technology. <http://ep.iu.edu/> [cited 6 Jan 2011]
- [6] Inter-University Consortium for Political and Social Research. Data management plan examples. <http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/other-plans.jsp> [cited 4 May 2012]
- [7] MIT Libraries. Data Planning Checklist. <http://libraries.mit.edu/guides/subjects/data-management/> [cited 4 May 2012]
- [8] Indiana University. IUScholarWorks. <http://scholarworks.iu.edu/> [cited 4 May 2012]
- [9] Indiana University Purdue University-Indianapolis. IUPUIDataWorks. <http://dataworks.iupui.edu/> [cited 4 May 2012]