

Indiana University

Guidance on Preparing NSF Data Management Plan

The National Science Foundation released a new requirement for proposal submissions regarding the management of data generated using NSF support. Starting January 2011, all proposals must include a data management plan. The plan should be short, no more than two pages. 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]). Section I gives a template and suggestions for completing a Data Management Plan (DMP). Since no single template works for every discipline and community, Section II list additional resources that could be helpful in figuring out what works for your needs. Section III is useful if you want to use IU's institutional repository as your data management solution.

I. Template

A general data management plan can be organized by the below template though not all items will be relevant for all disciplines. See [2] for discipline specific advice.

1. Describe the types of data and products that will be generated in the research such as samples, physical collections, software, and curriculum materials.
2. Describe the format in which the data or products are stored. Include a description of the metadata that will make the actual data products useful to the general researcher.
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. Could include a website maintained by your research group, or direct contribution to public database. If maintenance of a web site or database is direct responsibility of your group, provide information about the period of time the web site or database is expected to be maintained. IU recommends that your DMP address a time period longer than the duration of your proposed project.

NSF allows grantees to retain principal legal right to intellectual property developed under NSF grants. You should consider contributing your digital data and data collections to the public domain such as can be achieved through adoption of the Open Data Commons Public Domain Dedication and License (PDDL) [3].

4. Describe policies and provisions for re-use, re-distribution, and the production of derivatives.

5. Describe plans for archiving data, samples, and other research products, and for preservation of access to them. Archiving data with a public database or institutional repository ensures access to the data for periods of time far longer than project duration. If you are interested in using Indiana University's institutional repository for your data, see Section III below. The PDDL license described in #3 above provides the archiving entity with the rights it needs to archive your digital data on your behalf.

II. Resources

NSF funds a wide range of research. Some directorates and programs have provided specific guidance, and that can be found at Dissemination and Sharing of Research Results [2], <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>. In the absence of specific guidance, the [Award & Administration Guide \(AAG\) Chapter VI.D.4](#) [1] applies.

This URL links to data management plan examples that span a range of natural science disciplines

<http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/other-plans.jsp>

And a Data Planning Checklist from [4] can be helpful in preparation:

1. What type of data will be produced? Will it be reproducible? What would happen if it got lost or became unusable later?
2. How much data will it be, and at what growth rate? How often will it change?
3. Who will use it now, and later?
4. Who controls it (PI, student, lab, university, funder)?
5. How long should it be retained? *e.g.* 3-5 years, 10-20 years, permanently
6. Are there tools or software needed to create/process/visualize the data?
7. Any special privacy or security requirements? *e.g.*, personal data, high-security data
8. Any sharing requirements? *e.g.*, funder data sharing policy
9. Any other funder requirements? *e.g.*, data management plan in proposal
10. Is there good project and data documentation?
11. What directory and file naming convention will be used?
12. What project and data identifiers will be assigned?
13. What file formats? Are they long-lived?
14. Storage and backup strategy?
15. When will I publish it and where?
16. Is there an ontology or other community standard for data sharing/integration?
17. Who in the research group will be responsible for data management?

III. IU Institutional Repository

Indiana University has an institutional repository for archiving scholarly and scientific works called IUScholarworks [5]. IU Bloomington will accept digital data generated from a National Science Foundation funded effort on which the principle investigator is formally affiliated with IUB. IUB through ScholarWorks will provide replicated storage of all data sets. The current formats accepted are listed in [6].

If you intend to use the IUB Scholarworks, you will need to budget into your grant proposal funds for providing the access and safe storage of your data. Except in rare cases, this cost is substantially lower than the cost of providing access yourself and is much safer. NSF allows for adding data management costs to your proposal (typically Line G2). The cost depends on a number of factors, particularly: How much data will it be, and at what growth rate? How often will it change? Will data need to be available in multiple formats simultaneously? How many formats? Are the files long-lived? How long should the collection be retained? *e.g.* 3-5 years, 10-20 years, permanently?

One can expect to budget \$2000 in one's grant proposal to store 1 TB for 4 years, or \$500/yr for a terabyte of storage.¹

References:

- [1] NSF Award and Administration Guide, Chapter VI.D.4
http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4
- [2] Discipline specific advice on data management plans from NSF directorates
<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
- [3] ODC Public Domain Dedication and Licence (PDDL)
<http://www.opendatacommons.org/licenses/pddl/1-0/>
- [4] Data Planning Checklist <http://libraries.mit.edu/guides/subjects/data-management/>
- [5] Data management plan examples
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/other-plans.jsp>
- [6] IUScholarWorks <http://scholarworks.iu.edu/>.
- [7] IUScholarWorks current accepted data formats <http://libraries.mit.edu/dspace-mit/build/policies/format.html - formats>

¹ While \$500/yr for 1 TB of storage may seem high compared to using a cloud provider, keep in mind that the storage cost with a cloud provider such as Amazon S3 does not include the cost incurred every time someone accesses your data. Nor does it factor in the automatic replication of your data that you get with IU, nor the human labor of managing access to the data over a long period. Finally, IU is a public research university that will still be around long after cloud providers have come and gone.