Data Management: A Guide for Union College Faculty

Overview

All National Science Foundation (NSF) proposals with a due date on or after January 18, 2011 must include a supplemental document of no more than two pages labeled "Data Management Plan" describing how the proposal will conform to NSF's policy on disseminating and sharing research results. Proposals for supplementary support to an existing award are not required to include a Data Management Plan (DMP).

The DMP will be reviewed as an integral part of the proposal, coming under Intellectual Merit or Broader Impacts or both, as appropriate for the scientific community of relevance. Specific data management requirements and plans exist for the following NSF Directorates and Programs:

Engineering Directorate (ENG)
- Directorate-wide guidance: http://nsf.gov/eng/general/ENG_DMP_Policy.pdf

Geological Sciences Directorate (GEO)
- Division of Earth Sciences: http://www.nsf.gov/geo/ear/2010EAR_data_policy_9_28_10.pdf
- Integrated Ocean Drilling Program: http://www.nsf.gov/bfa/dias/policy/dmpdocs/geo_iod.pdf
- Division of Ocean Studies: http://www.nsf.gov/pubs/2004/nsf04004/start.htm

Mathematical and Physical Sciences Directorate (MPS)
- Division of Astronomical Sciences: http://www.nsf.gov/bfa/dias/policy/dmpdocs/ast.pdf
- Division of Chemistry: http://www.nsf.gov/bfa/dias/policy/dmpdocs/che.pdf
- Division of Materials Research: http://www.nsf.gov/bfa/dias/policy/dmpdocs/dmr.pdf
- Division of Mathematical Sciences: http://www.nsf.gov/bfa/dias/policy/dmpdocs/dms.pdf
- Division of Physics: http://www.nsf.gov/bfa/dias/policy/dmpdocs/phy.pdf

Social, Behavioral and Economic Sciences Directorate (SBE)
- Directorate-wide guidance: http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf

For all other Directorates and Programs, or other areas for which specific guidance is not available, the requirements established in this section apply.

A valid DMP may include only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification. Proposers who feel that the plan cannot fit within the supplement limit of two pages may use part of the 15-page Project Description for additional data management information. Proposers are advised that the DMP may not be used to circumvent the 15-page Project Description limit.

To describe how the proposal will conform to NSF policy on disseminating and sharing research results, the DMP may include:

1. The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;

2. The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
3. Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. Policies and provisions for re-use, redistribution, and the production of derivatives; and
5. Plans for archiving data, samples, and other research products, and for preservation of access to them.

Considerations for Managing Research Data

Managing data before beginning research and throughout a research project's lifecycle is essential to ensure that it is both usable and accessible now and in the future as well. The following questions draw heavily from the Massachusetts Institute of Technology's data management site (http://libraries.mit.edu/guides/subjects/data-management/). For the NSF proposal, faculty PIs should answer the applicable questions presented in this checklist to develop the primary content of the supplemental DMP for NSF proposals. See the Additional Resources section at the end of this guide for reference material from the MIT site which gives examples and helpful tips to answering each question in this checklist.

Data Planning Checklist – Before you begin, answer these questions:

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, MIT, 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?

Developing a Written Data Management Plan

The MIT website referenced above includes a variety of examples and models of plans that may be useful to faculty preparing a DMP. These existing models may facilitate the process of taking the answers to the questions above that are specific to the proposed research and transforming this information into a written DMP for the NSF grant. The plan may cover the following topics or areas of focus:

- name of the persons responsible for data management, during your research project and over time; how you will ensure adherence to the plan over time
- description of data to be collected, method of collection, its nature and format
- explanation of how data will be documented throughout the research project
- discussion of data quality issues
- description of your short-term and long-term storage and preservation plans, including backup procedures
- explanation of how you will make the data available for public use and potential secondary uses
- explanation of your plans to preserve the data and ensure a long-lived format
- description of any arrangements that might be needed to protect participant confidentiality or intellectual property; any other legal or ethical considerations

Institutional Guidelines for Data Management
Union College recommends to its faculty researchers the following campus repositories for data management purposes: the Library archives, D-space, Blackboard

Union College has existing manuals, guidelines, and other academic resources available to faculty members. These include:
- Union College Faculty Manual:
  http://union.edu/Academics/AcademicAffairs/GovernanceCommittees/FacultyManual/index.php
- Union College Human Subjects Review Committee:
  http://www.union.edu/Resources/Academic/hsrc/index.php
- Union College Office of Foundation, Corporate & Government Relations (Grants):
  http://www.union.edu/Resources/Academic/Grants/Proposal_Prep/index.php
- Union College Schaffer Library: http://www.union.edu/library/index
- Union College Information Technology Services:
  http://www.union.edu/Resources/Technology/ITS/index

Once a DMP has been created and uploaded into FastLane for submission, the submitted proposal package (including the supplemental document labeled "Data Management Plan") shall be retained by the Sponsored Programs Office as record of the PIs management of its research results through sponsored agreements. If at any time the PI modifies, alters, or otherwise changes the methods, approach, or process for managing the research results of the sponsored agreement, the PI must update the DMP and provide an updated copy to the Sponsored Programs Office.

Additional Resources

Many resources are available from the MIT library to help faculty PIs develop a DMP:

Restricted Data, http://libraries.mit.edu/guides/subjects/data/access/restricted.html)

Documentation and Metadata, http://libraries.mit.edu/guides/subjects/data-management/metadata.html)

Organizing Your Files, http://libraries.mit.edu/guides/subjects/data-management/organizing.html

Citing Data, http://libraries.mit.edu/guides/subjects/data/access/citing.html)

File Formats for Long-Term Access, (http://libraries.mit.edu/guides/subjects/data-management/formats.html)

Backups and Security, http://libraries.mit.edu/guides/subjects/data-management/backups.html)

Sharing Your Data, http://libraries.mit.edu/guides/subjects/data-management/publishing.html)

Data Integration, http://libraries.mit.edu/guides/subjects/data-management/integration.html)