University of Massachusetts Amherst

From the SelectedWorks of Thea P Atwood

Winter December 12, 2014

NSF Data Management Plans

Thea P Atwood, University of Massachusetts Amherst

Available at: https://works.bepress.com/tpatwood/5/
NSF Data Management Plans

Tips & Guidance on applying to your SBE DDRIG

UMass Amherst Libraries
Today...

1. Framing the Data Management Plan
2. Components of a DMP & sample responses
3. Post-award monitoring
4. Resources for more help
5. Q & A
Framing the DMP

DMPs address two main topics:

*What data are generated by your research?*
*What is your plan for managing the data?*

No more than two pages in length.

Costs are explained in the Budget Justification pages.
Framing the DMP

Essentially, your funding agency wants you to *share the data you generate* in a way that others can easily *access* and *re-use* it*.

The DMP is meant to help you plan ahead and “clearly articulate” how you will manage, take care of, and disseminate the data generated under the auspices of the proposal.

*Access and use parameters will differ based on what kind of data you’re working with. e.g., working with human subjects has different privacy concerns than working with cell lines.*
Framing the DMP

A note on data ownership & custody (part 1 of 2):

“The Amherst campus of the University of Massachusetts is the owner or joint owner of all data that is created or collected by its employees or contractors...When a creator of data ceases to be an employee or contractor of the University, the creator must leave the data in the physical possession of the owner(s), but will continue to have access rights to the data. The creator may take a copy of the data, at creator’s expense.”

"The researcher(s) who created the data typically serve as the custodian of the University’s data. Such researchers act on behalf of the University, without limiting the University’s ownership rights...The custodian of the data shall take all reasonable steps to protect the data from damage or loss, including damage or loss due to catastrophic events. The owner of the data shall provide storage space and financial support as necessary to maintain the data.”
What if I don’t produce any data?

The NSF will not accept any full proposal submitted, or due, to NSF on or after January 18, 2011, that is lacking a DMP… Even if no data are to be produced, e.g. the research is purely theoretical or is in support of a workshop, a DMP is required. In this case, the DMP can simply state that no data will be produced.
What if I don’t produce any data?

The work for this project is purely theoretical. No data will be produced.
Components

SBE directorate specifically asks for you to include:

1. Roles and responsibilities of all parties.
2. Expected data.
3. Period of data retention.
4. Data formats and dissemination.
5. Data storage and preservation of access.
6. Additional possible data management requirements.
Roles & Responsibilities

Outline all rights and obligations of all parties, including changes in these roles and responsibilities should individuals (e.g., yourself & your PI) leave the institution.

Include anyone who has a role to play in managing your data.

Cover how you will delegate any data management responsibilities.
Roles & Responsibilities: Example

The data management plan will be primarily executed by the PI. The lab manager will be in charge of coordinating and ensuring data storage and access, and will help ensure that the data management plan (DMP) is being followed. The lab manager will also help document any changes to the DMP. Furthermore, both the lab manager and the PI are responsible for executing the data management plan in alignment with the internal data management best practices guide, which outlines how files should be named, where they should be saved, instructions for what to do when creating a new file and how to disambiguate any acronyms, processes for data backup and wrapping up of any project, and other pertinent data management issues. Should the PI leave the institution, the lab manager will continue to be responsible for maintaining the storage and accessibility of the data, and will be responsible for transferring responsibility for the data to other lab personnel.
The Federal government defines ‘data’ in OMB Circular A-110 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 colleagues. 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.”
Expected Data

Describe the types of data, samples, physical collections, software, curriculum materials, or other materials to be produced in the course of the project.

Explain what data means in the context of your research project.

Explain the types of data you plan to generate, including size, file formats, and number of files.
This project will obtain fMRI data from a Siemens MAGNETOM Trio, and response time data for twenty subjects. Raw data will be saved as DICOM files and response times will be recorded in Excel spreadsheets from a custom-built response paddle located in the fMRI environment. The total data collected for the project is estimated at 90 GB.
Address how you will meet the SBE’s commitment to timely and rapid dissemination of research data. Outline how long you will retain the exclusive right to use the data before opening it up to wider use.

SBE recognizes that there are discipline-specific requirements and norms for data release. If you will embargo your data, explain why and for how long. Include information on political/commercial/patent or publisher requirements.
Data will be kept by the lab until it is cleaned and de-identified in its entirety. Data collection is estimated to take one year. Upon acceptance of publication of our first paper, we will release the dataset underlying this publication.
Data Formats & Dissemination

Describe the data formats, media, and disseminations approaches that will be used to make \textit{data} and \textit{metadata} available to others.

Here, describe the formats your data and metadata will take. Describe any policies for public access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.
Ideally, use file formats that are open, freely available, non-proprietary in nature, and well documented.

Whatever file formats you select, provide your reasoning for using this particular format (e.g., standard in my discipline).
Data Formats & Dissemination

As much as you are able, get granular with this section. Describe how you will share your data, and its context.

**File formats**
- Original formats
- Transformations
- Metadata
  - Project specific info
  - Standards
  - Creation & capture methods
- Reasoning

**Dissemination**
- Sharing/access strategy
  - E.g., repository, website, contact PI
  - Licensing: CC-BY, etc.
- Managing privacy issues
  - Anonymization
- Compliance with IRB
Data Formats & Dissemination

Policy from UMass:

“Researchers shall endeavor to make their data publicly available as soon as possible, and to the extent possible. Access may be delayed while the correctness of the data is being verified, until an initial publication based on the data appears, for the minimum period needed to file a patent application, or for any other reasonable need. Data should be released early if benefit to the public is likely.”

Note that confidential info is to be protected!

Participant data will be shared as de-identified DICOM files. DICOM files are the standard data type for medical imaging. Response rates will be saved as comma-separated values (csv). Both DICOM and csv files are importable by various programs, and benefit from a high degree of interoperability.

Our instrumentation automatically generates metadata for the parameters in which the data were collected (e.g., slice angle, pulse frequency, etc.). Metadata regarding the project itself will also be included and collected according to the Dublin Core Metadata Standard.

Before distribution of our de-identified data, we will seek approval from the IRB. Data will be shared in a discipline specific repository.
Data Storage & Preservation of Access

Describe the physical and cyber resources and facilities that will be used for the effective preservation and storage of research data.

Describe your long-term strategy for storing, archiving, and preserving the data you will generate or use.
Data Storage & Preservation of Access

Note that UMass has a policy on storing data:

“Data shall be retained for at least three years after its creation. If the data were created as part of a sponsored research project, then the data shall be retained at least three years after the final report to the sponsor has been submitted, or the ending date of the project, whichever is later...If the data led to the granting of a patent, then the data shall be retained for the life of the patent and its extensions...The data shall otherwise be retained for as long as anyone expresses, in writing, an interest in its retention. In no case will the data be discarded or destroyed when it is known to be in use.”

Data Storage & Preservation of Access

Include:

- The long-term plan for maintaining, curating, and archiving your data.
  - What data you will select for long-term preservation
  - How long data will be kept after the project ends
- Your selected archive or data repository.
- Any procedures your repository has in place to ensure preservation of access.
  - E.g., redundancies, mirrors, etc.
- Metadata or other documentation included in your deposit.
Data will be stored on a dedicated server for three years after the completion of the project, as per the University’s Policy on Data Ownership, Retention, and Access. After acceptance of our first publication, when our data will be made publicly available, and will be deposited in the OpenfMRI repository with a Public Domain Dedication and License v1.0. Due to the size of our generated data, only the raw data, response times, and metadata will be stored in the repository.
Additional Possible Data Management Requirements

This section is included in case a particular institution or the NSF solicitation itself has more stringent data management requirements. Additional requirements will be specified in the program solicitation and award conditions. PIs to be supported by such programs must discuss how they will meet these additional requirements in their DMPs.

Describe here any additional program-specific data management requirements. If none exist you may leave this section blank.
Post-Award Monitoring

1. Annual Reports
   “...must provide information on the progress on data management and sharing of the research products.”

Example info:
- Citations of relevant publications (can include data publications as well)
- Conference proceedings
- Descriptions of other types of data sharing and dissemination of results (blog posts, tweets & retweets)
2. Final Project Reports

“...must discuss execution and any updating of the original DMP.”

Describe:

- Data produced during the award;
- Data to be retained after the award expires;
- Verification that data will be available for sharing;
- Discussion of community standards for data format;
- How data will be disseminated;
- The format that will be used to make data available to others, including any metadata; and
- The archival location of the data.
3. Subsequent Proposals

“Data management must be reported in subsequent proposals by the PI and Co-PIs under “Results of prior NSF support.”

You might feel overwhelmed today...


By the time you need to report your data management practices in a subsequent proposal, you’ll be a pro!

More resources for you!

Librarians!
The Data Working Group
Your advisor
Others in your field

Data Management Best Practices

- **DataONE** best practices: [https://www.dataone.org/best-practices](https://www.dataone.org/best-practices)
- **MANTRA** - [http://datalib.edina.ac.uk/mantra/](http://datalib.edina.ac.uk/mantra/)
Even MORE resources for you!

General DMP guidance:
**DMPTool**: [https://dmptool.org/](https://dmptool.org/)

Locate a data repository:
**DataBib**: [http://databib.org/](http://databib.org/)

Track your impact in one place:
**ImpactStory**: [https://impactstory.org/](https://impactstory.org/)
  - Non-profit, but requires yearly subscription
Gosh, just one more...

Data Management & Sharing FAQ from the NSF:
Thank you
&
Let’s chat!

More questions? Get in touch: datamgt@library.umass.edu
This presentation available under a Creative Commons Attribution (CC-BY) license.