Data Management Plan Template Guidelines

This sample plan is provided to assist grant applicants in creating a data management plan, if required by the agency receiving the proposal. A data management plan should contain the following sections:

- A description of the data to be collected or produced during the research project
- Standards for data and metadata format and content
- Policies for public access and sharing of the data
- Plans for storing and archiving digitized data, physical samples, software code, curricular materials, etc and plans for preservation of access to that data.

Every research project is different. Some of these sections may not apply to your research project so check the guidelines provided by the funding agency for specific information on which sections to include. A comprehensive list of these can be found on the RCG website under the Funding Agency Guidelines. If the data management plan includes using a hosted data management repository it is strongly encouraged that a letter of commitment confirming that the data repository will archive and share the research data should accompany the plan.

Note: The National Science Foundation states that a valid Data Management Plan may include: “only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification”.

We will collect submitted data management plans from MSU Principle Investigators and make them available to use as examples for writing your own plan. The template below is a suggested format for the plan and includes links to MSU specific policies that should be included.

How to use this template:
- Check the guidelines for your funding agency
- Create a document with the sections provided below, including sections as appropriate
- Each section includes instructions/suggestions for what should be included
- Some sections include MSU specific language for data stewardship, intellectual property and MSU specific policy language and links.
Data Management Plan Template

Description of the data generated by this project:

What to include in this section:
Provide a description of the data to be collected. Include descriptions of physical data samples, images, time series data, video and audio files, documents, curricular materials, software code etc. Include a brief description of the scope and scale of the data collection.

Data and metadata standards:

What to include in this section:
Provide information about the metadata that will accompany the produced data. Metadata provides context-dependent information about the data that makes it easier to use and share. For example, an image of a landscape should include at a minimum the geospatial coordinates, the date and time the image was taken, the file format (jpeg, tiff etc), and the experiment or study the image data was associated with. A microarray experiment should include the standard set of protocols for the experiment as defined by the MIAME data standards. Include data standards if they exist for your scientific domain. The best rule of thumb is to include all the information you would normally record in your lab notebook or study for your own purposes. Including a reference to a paper where the data was used may also be acceptable.

Data access and sharing:

What to include in this section:
Provide information about how the data produced in the project will be made available to others. Be sure to address both Public-use data and Restricted-use data and address the timeliness of data availability, for example, “data will be made available after publication”. Include a statement for how long the data will be preserved – NSF requires that data be stored for three years following the end of the grant. For example: “Data will be stored in a secure repository for three years following the completion of the project”.

A number of national archives are now available for the contribution of data:

- Dryad [http://datadryad.org/](http://datadryad.org/) (biological sciences, basic and applied); in partnership with bioscience publishers Dryad is a repository for datasets associated with published articles. The metadata describing the datasets are from [Dublin Core](http://dublincore.org/) and Darwin Core.
- ICPSR, [http://www.icpsr.umich.edu/icpsrweb/ICPSR/](http://www.icpsr.umich.edu/icpsrweb/ICPSR/) (University of Michigan hosted social sciences data)
- Protein Data Bank [http://www.pdb.org/pdb/home/home.do](http://www.pdb.org/pdb/home/home.do)
- National Virtual Observatory [http://www.us-vo.org/](http://www.us-vo.org/) (part of International Virtual Observatory, a portal to distributed datasets)
- The Neuroscience Information Framework [http://neuinfo.org/](http://neuinfo.org/) (a portal to a comprehensive set of data resources for neuroscience)
• Neuroimaging Informatics tools and Resources Clearing House: http://nitrc.org
• Open source software archives: sourceforge.net; github.org

Also, software exists for managing collaborative data sharing from workstation to workstation such as Dropbox, Box.net, or BitTorrent Sync. MSU Research Computing recommends using Globus for collaborative data sharing between researchers. Globus also allows researchers to move data to and from the above national repositories/archives.

Example of this section for data stored on servers within the Research Computing Group:
Collaborative data access is managed through the use of Globus [1] endpoints. Globus provides researchers with a convenient user interface and infrastructure for moving data between accessible Globus endpoints (including XSEDE resources and other research and academic infrastructure), as well as to and from their workstations using the Globus Connect Personal software. This will make it easy for researchers move large datasets asynchronously in a fault-tolerant way over unreliable networks.


Intellectual Property Rights:
What to include in this section:
Include a statement that addresses the intellectual property rights related to the data produced.

Example of this section written for protein structure research:
Principal investigators and their institutions own the research data they generate. By depositing with the Protein Data Bank, investigators do not transfer data ownership but instead grant permission for Protein Data Bank to transform and redistribute the data as necessary.

MSU specific language and policies:
Include a citation to the MSU policy on intellectual property (http://www.montana.edu/wwwvr/osp/900.00.html)
Include a citation to the MSU policy on Electronic Research Data http://www2.montana.edu/policy/research/Electronic%20Research%20Data%20Security%20Policy.htm
Include a citation to the MSU Data Stewardship Policy: http://www2.montana.edu/policy/itc/data_stewardship.htm

Ethics and Privacy:
What to include in this section:
Describe the process, procedures, and agreements will be put in place to ensure that participants in this project will maintain an acceptable level of confidentiality. Address issues of HIPPA, FERPA, FDA and other compliance standards.
Example of this section written for human subjects research:

**Informed consent:** For this project, informed consent statements, if applicable, will not include language that would prohibit the data from being shared with the research community.

**Disclosure risk management:** The research project will remove any direct identifiers in the data before deposit with [Repository]. Once deposited, the data will undergo procedures to protect the confidentiality of individuals whose personal information may be part of archived data. These include: (1) rigorous review to assess disclosure risk, (2) eliminating subject identifiers if necessary to protect confidentiality, (3) limiting access to datasets in which risk of disclosure remains high, and (4) consultation with data producers to manage disclosure risk.

**Data Formats:**

*What to include in this section:*

Describe the formats of the data for storage, distribution, and use by other applications. Examples for text based data include: PDF, XML, DOI (document identifiers), for image based data: jpeg, png, tiff, DICOM standard for MRI imaging data. Include information about common standards for data exchange where appropriate.

**Archiving and Preservation:**

In this section describe the long-term preservation plan for the data including migration to new storage media or formats.

Example of this section written for social sciences research:

[Provide initial information about the selected repository]. [Repository] will accept responsibility for long-term preservation of the research data upon receipt of a signed deposit form. This responsibility includes a commitment to manage successive iterations of the data if new versions are deposited. [Repository] will ensure that the research data are migrated to new formats, platforms, and storage media as required by good practice in the digital preservation community. Good practice for digital preservation requires that an organization address succession planning for digital assets. [Repository] has a commitment to designate a successor in the unlikely event that such a need arises.

**Data Storage and Backup:**

*What to include in this section:*

Describe how the data is managed locally from the point of collection to the point of archiving. Include data storage on servers, investigators computers, or other locations and what technology is used to ensure the data is backed up to avoid catastrophic data loss.

Example of this section for data stored on servers within the Research Computing Group:

Research data will be stored on a storage system with multi-disk and multi-server redundancy. Data will be stored only in keyed access-restricted, climate-controlled server rooms. Access to data will be controlled using share- and file-level permissions for remote access. Data will be backed up weekly as snapshots to a separate disk archive and at least five progressive snapshots will exist at any given time. Version control will be achieved through the use of a source control management system.
These data will be managed and stored according to the guidelines and institutional policies laid described in the MSU Data Stewardship Plan and conform to the security standards in the MSU Data Stewardship Guidelines.