Getting Started with Data Management & DMPTool

Dartmouth College Libraries
Lora Leligdon, Physical Sciences Librarian
Pamela Bagley, Biomedical Research and Education Librarian
Getting Started with Data Management & DMPTool

1. Why is data management important?
2. What is a data management plan?
3. The DMPTool
Why is Data Management Important?

$191,200,000

FY16 sponsored project awards received by Dartmouth
Why is Data Management Important?

Data management allows for reproducibility, transparency, and data reuse in research.

It’s easier if data are managed well from the start of a project.
Data Management & Sharing Mandates

• Journals – PLOS, Nature, BioMedCentral, ...
• Funders – NSF, NIH, DOE,…
• Office of Science & Technology Policy mandate, February 2013
Getting Started with Data Management & DMPTool

1. Why is data management important?
2. What is a data management plan?
3. The DMPTool
What is a data management plan?

A document that describes what you will do with your data during your research and after you complete your project.
Why prepare a DMP?

• Saves time
• Increases research efficiency
• Satisfies funder requirements
• Makes reproducibility & sharing easier
A DMP is a Living Document

• Keep your plan current
• Incorporate changes
• Use as a guide for daily activities

DMP Tool
Where to Start with DMPs?

Small & Simple

• Document what you know **now**
• Simple & understandable
• Share the plan with your team
• Avoid procrastination and immobilization
Components of a Basic DMP

1. Types of data
2. Documentation and metadata
3. Storage and preservation
4. Policies
5. Budget
Ten Rules for Creating a Good DMP

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Define how the data will be organized
4. Explain how the data will be documented
5. Describe how the data quality will be assured

Ten Rules for Creating a Good DMP

6. Present a sound data storage and preservation strategy
7. Define the project’s data policies
8. Describe how the data will be disseminated
9. Assigned roles and responsibilities
10. Prepare a realistic budget

1. Sponsor requirements

NSF ENG Data Management Plan Requirements

Proposals submitted to NSF must include a supplementary document of no more than two pages labeled “Data Management Plan” (DMP). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. Proposals that do not include a DMP will not be able to be submitted. For more information about this requirement, please see the Grant Proposal Guide, Chapter II.C.2.i and NSF policies on Dissemination and Sharing of Research Results, including Frequently Asked Questions (FAQs).

Please note: The Engineering Directorate (ENG) has additional guidance for proposals submitted to ENG programs, http://nsf.gov/eng/general/ENG_DMP_Policy.pdf. Questions about data management plans for proposals to the Engineering Directorate may be addressed to Dr. Alexandra Medina-Borja at amedinab@nsf.gov.
2. Identify the data

**Types**
Collected or created data including physical items, code, models, spreadsheets, images, etc.

**Sources**
Complied, proprietary, human subjects, instrumentation, sensitive or private

**Volume**
Total number of files and quantity of data produced

**Formats**
Non-proprietary, open standards. Uncompressed, unencrypted, stored with standard encodings.
3. Data organization
4. Data documentation

• What information should be captured to enable others to discover, access, interpret, use, and cite your data?
• How will you create or capture these metadata?
• What tools can help you create this metadata?
5. Data quality

How quality control could save your science

It may not be sexy, but quality assurance is becoming a crucial part of lab life.

Monya Baker
6. Sound data storage and preservation

• How long will the data be accessible?
• How will the data be stored and protected during the project?
• How will the data be preserved and made available for future use?
7. Data policies for access, sharing, reuse

- Licensing or sharing arrangement of preexisting materials (data/software/code)
- Plans for retaining, licensing, sharing, and embargoing original materials
- Legal and ethical restrictions on access to sensitive or IP data
Policies

Content
Scope
All fields of science. All types of research artifacts. Content must not violate privacy or copyright, or breach confidentiality or non-disclosure agreements for data collected from human subjects.

Status of research data
Any status is accepted, from any stage of the research lifecycle.

Eligible depositors
Anyone may register as user of Zenodo. All users are allowed to deposit content for which they possess the appropriate rights.

Ownership
By uploading content, no change of ownership is implied and no property rights are transfer to CERN. All uploaded content remains the property of the parties prior to submission.

Versions
Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.

Data file formats
All formats are allowed - even preservation unfriendly. We are working on guidelines and features that will help people deposit in preservation friendly formats.

Volume and size limitations
Total files size limit per record is 50GB. Higher quotas can be requested on case-by-case basis. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.

Data quality
All information is provided "as-is", and the user shall hold Zenodo and information providers supplying data to Zenodo free and harmless in connection with the use of such information.

Withdrawal of data and revocation of DOIs
Revocation
Content not considered to fall under the scope of the repository will be removed and associated DOIs issued by Zenodo revoked. Please signal promptly, ideally no later than 24 hours from upload, any suspected policy violation. Alternatively, content found to already have an external DOI will have the Zenodo DOI invalidated and the record updated to indicate the original external DOI. User access may be revoked on violation of Terms of Use.

Withdrawal
In the event that the reason for withdrawal is not immediately clear, a withdrawal request will be served in its place. Withdrawal requests may be requested and fully justified withdrawal attempts will be made to preserve and facilitate access to the URL of the original object.

Metadata
Metadata access and reuse
Zenodo is provided free of charge for educational and informational use. Metadata is licensed under CC0, except for email addresses. All metadata is exported via DAI-PMH and can be harvested.

Metadata types and sources
All metadata is stored internally in MARC according to the schema defined in http://inveniosoftware.org/wiki/Project/OpenAIREplus/DevelopmentRecordMarkup. Metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema according to OpenAIRE Guidelines.

Language
For textual items, English is preferred but all languages are accepted.

Licenses
Users must specify a license for all publicly available files. Licenses for closed access files may be specified in the description field.

Access and reuse of data
Access to data objects
Files may be deposited under closed, open, or embargoed access. Files deposited under closed access are protected against unauthorized access at all levels. Access to metadata and data files is provided over standard protocols such as HTTP and DAI-PMH.

Use and re-use of data objects
Use and re-use is subject to the license under which the data objects were deposited.

Tracking users and statistics
Zenodo does not track, collect or retain personal information from users of Zenodo except as otherwise provided herein. In order to enhance Zenodo and monitor traffic, non-personal information such as IP addresses and cookies may be tracked and retained, and log files may be shared, in aggregation, with other community services (in particular OpenAIREplus partners). User provided information, like corrections of metadata or paper claims, will be integrated into the database without displaying its source and may be shared with other services. Zenodo will take all reasonable measures to protect the privacy of its users and to resist service interruptions, intentional attacks, or other events that may compromise the security of the Zenodo website.

Embargo status
Users may deposit content under an embargo status and provide and end date for the embargo. The repository will restrict access to the data until the end of the embargo period, at which time, the content will become publicly available automatically.

Restricted Access
Users may deposit restricted files with the ability to share access with others if certain requirements are met. These files will not be made publicly available and sharing will be made possible only by the approval of depositor of the original file.

Preservation of data
Retention period
Items will be retained for the lifetime of the repository. This is currently the lifetime of the
8. Data dissemination

• What/ When/ How data will be made available
• Passive approach
  • Project/personal website
  • “on request”
• Active approach
  • Open repository or archive
  • Publish data as SI in journals
  • Publish data/metadata/code as a “data paper”
9. Roles and responsibilities

NSF ENG Directorate

“[The DMP] should outline rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data. It must also consider changes to roles and responsibilities that will occur should a PI or co-PI leave the institution.”


DMPTool
10. Budget

• Time and Money costs of data preparation & documentation
  – Hardware, software
  – Personnel
  – Storage/ Archive fees
  – OA/ public access fees

• Request funding!
Getting Started with Data Management & DMPTool

1. Why is data management important?
2. What is a data management plan?
3. The DMPTool
Data Management Planning Tool
Create, review, and share data management plans that meet institutional and funder requirements.

PUBLIC DMPS
List of sample data management plans provided by DMPTool users.

DMPTOOL NEWS
Latest information about data management and the DMPTool.
- Getting our ducks in a row
- The 20:51 sprint (Roadmap team-building...)
- Mini release (and a mini maintenance wi...
- NIH Policy on Rigor and Reproducibility
- Roadmap team-building exercises: US ed...

DMPTOOL HELP
Overview of how to use the tool, plus resources and guidance on data management.
- Frequently Asked Questions
- Create a DMP
- Administer the DMPTool
- Data management guidance
- Community resources
Data Management Planning Tool
Create, review, and share data management plans that meet institutional and funder requirements.

Get Started

Library of public DMPs

Data management resources
# Data Management Planning Tool

Create, review, and share data management plans that meet institutional and funder requirements.

## Public DMPS

List of sample data management plans provided by DMPTool users.

- UNDERSTANDING THE ROLE OF PHYSICIAN INTEGRATION WITHIN NURSING HOMES IN POST-ACUTE CARE OUTCOMES
- A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy
- A unified approach to preserving cultural software objects and their development histories

## DMPTool News

Latest information about data management and the DMPTool.

- Getting our ducks in a row
- The 20:51 sprint (Roadmap team-building...)
- Mini release (and a mini maintenance wi...)
- NIH Policy on Rigor and Reproducibility
- Roadmap team-building exercises: US edi...

## DMPTool Help

Overview of how to use the tool, plus resources and guidance on data management.

- Frequently Asked Questions
- Create a DMP
- Administer the DMPTool
- Data management guidance
- Community resources

[Link to get started]
Dartmouth

NetID:

Continue

Lookup my NetID
Where do I enter my password?

Need more help? Email help@dartmouth.edu or call 603-646-2999. Alumni should call 603-646-3202 for help.
Either start with a template or copy an existing DMP

### Create New DMP

#### Start with a DMP Template

To create a new DMP, select a funder or institutional template.

- **Select Template >>**

### Copy an Existing DMP

The existing DMPs in this list are either publicly shared by any user, shared within your institution by other DMP creators, or are plans that you have previously created.

Select an existing DMP below and add text to the template.

<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>USGS CDR/ECY DMP</td>
<td>Cali Jenkerson</td>
</tr>
<tr>
<td>A unified approach to preserving cultural software objects and their development histories</td>
<td>DMP dmcurator</td>
</tr>
<tr>
<td>Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013</td>
<td>DMP dmcurator</td>
</tr>
<tr>
<td>Daymet Follow-On: Surface Weather Data with Uncertainty Quantification for Terrestrial Ecosystem Process Models</td>
<td>DMP dmcurator</td>
</tr>
<tr>
<td>Multimedia Text Annotation for Students</td>
<td>DMP dmcurator</td>
</tr>
</tbody>
</table>

1 2 3 4 5 ... Next » Last »
SELECT DMP TEMPLATE

Select one of the funder DMP Templates listed to proceed to the next step. The type of template chosen can affect what information you will need to provide in the following steps.

Select Funder
DMP OVERVIEW

DMP Template: National Institutes of Health

* DMP Title: Neonatal Neurobehavioral Impacts of Iodine Insufficiency and Pesticide Exposures

Proposal Solicitation Number:

Proposal Submission Deadline: 10/14/2015

Add Co-owners: Jonathan Wheeler <jwheel01@unm.edu>

Existing Co-owners:

Current Status: New

Owner Comments: View

Visibility: This DMP is Private

Note: when visibility is set to "Public", your DMP will appear on the Public DMPs page of this site and it will be downloadable and copy-able.

Available Actions: Preview DMP | Delete DMP

Fill out basic information about the plan

Co-owners can:
• Edit
• Preview
• Download

Save  Cancel  Save and Next >>
Visibility settings

Three options:
1. Private – owners & co-owners only
2. Within institution
3. Publicly
Investigator seeking $500,000 or more in direct costs in any year should include a description of how final research data will be shared, or explain why data sharing is not possible.

Guidance

The precise content of the data-sharing plan will vary, depending on the data being collected and how the investigator is planning to share the data. Consider the following:

- When will you make the data available?
- What file formats will you use for your data, and why?
- What transformations will be necessary to prepare data for preservation/data sharing?
- What metadata/documentation will be submitted alongside the data?
- Will a data-sharing agreement be required? What will the agreement state?
- What are your plans for providing access to your data?
- Which archive/repository/central database have you identified as a place to deposit data?
Plan was successfully created.

DMP DETAILS

National Institutes of Health
Click on a section below to edit it at any time.

= Complete
= Mandatory

Template Outline
- Data sharing plan
- Additional data sharing requirements

General Resources
- NIH data sharing guidance and resources
- NIH data sharing policy and implementation
- NIH Key Elements in Preparing a Data Sharing Plan (PDF)
- NIH Data Standards and Common Data Elements Resource Guide (doc)
- NIH Data Sharing Repositories
- NIH sharing policy statement

Dartmouth College Resources
- Dartmouth Office of Sponsored Projects: Data Management Plans

Links out to help

Instructions  Links

Save Response  Save and Next

Cancel Changes
1. Products of Research

The project will collect and analyze the following data: Conductivity and temperature from moorings and shipboard CTD surveys, Horizontal currents from Lowered ADCP and moorings, Horizontal currents from shipboard sonar, Fine and micro-scale velocity from the WHOI High Resolution, Fine and micro-scale temperature from fast-response thermistors.

2. Data Storage and Preservation

Data will be shared in matlab MAT format and/or as netCDF files. Data quality will be in accord with published uncertainty values for each instrument and within error bars for standard processing techniques. These PIs have experience with this mix of data types from previous collaborative reports. Data responsibilities include: PI Responsibility: A. Thurnherr. LADCP-CTD fixes: L. St. Laurent and E. Shroyer. HRP microstructure analysis: S. Jachec. Ongoing model output prediction: J. Moum, J. Microstructure data: M. Alford, J. Nash, J. MacKinnon.

3. Formats and Metadata

All field data collected under this program will be made available as per NSF guidelines within 2 years of collection via published manuscripts, publicly available final reports to NSF, and data archiving with NODC. Recognizing that any individual PI server may become unavailable over time, data will be made available by PI website locations and also by specific request to any colleague.

4. Data Dissemination & Policies for Data Sharing and Public Access

Aside from the LADCP-shipboard CTD there are currently no established standards for archiving data from many of the fine- and micro-scale sensors used in the proposed work. This is a concern of the Climate Process Team on Ocean Mixing, of which many PIs are members. We propose to work with the CPT to evolve formats for data and metadata suitable for archiving both sensor and (critically) model output from the experiment. Field data will be provided to NODC upon project completion. Ultimate archival formats will be determined in consultation with NODC and with the CPT. Adequate archiving is anticipated to be an expensive, time-consuming task. All PIs have included funds for this in their budgets.

5. Roles and Responsibilities

Models codes to be employed are all public domain. Published peer-reviewed manuscripts will document the simulations and forcing sufficiently.
<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
<th>Status</th>
<th>Visibility</th>
<th>Last Modification Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample NSF Physics</td>
<td>Lora Leligdon</td>
<td>New</td>
<td></td>
<td>07/26/2015 03:08PM</td>
</tr>
<tr>
<td>Neonatal Neurobehavioral Impacts of Iodine Insufficiency and Pesticide Exposures</td>
<td>Lora Leligdon</td>
<td>New</td>
<td></td>
<td>07/26/2015 03:07PM</td>
</tr>
<tr>
<td>Overview</td>
<td>Details</td>
<td>Share</td>
<td>Preview</td>
<td>Copy</td>
</tr>
<tr>
<td>Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013</td>
<td>Lora Leligdon</td>
<td>New</td>
<td></td>
<td>07/26/2015 02:26PM</td>
</tr>
<tr>
<td>Template</td>
<td>Funder</td>
<td>Funder Links</td>
<td>Sample Plans (if available)</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>Alfred P. Sloan Foundation</td>
<td>Alfred P. Sloan Foundation</td>
<td>Alfred P. Sloan Foundation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCO-DMO NSF OCE: Biological and Chemical Oceanography</td>
<td>National Science Foundation</td>
<td>NSF OCE Sample and Data Policy, May 2011 (PDF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Department of Energy: Office of Science</td>
<td>Department of Energy (DOE)</td>
<td>DOE Statement on digital data management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMP Template from DCC</td>
<td>University of California, Office of the President</td>
<td>Information on DCC website</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GoMRI Research Consortia DMP Template 2015</td>
<td>Gulf of Mexico Research Initiative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gordon and Betty Moore Foundation</td>
<td>Gordon and Betty Moore Foundation</td>
<td>Guidelines</td>
<td>GEMF: Sample Plan #1</td>
<td></td>
</tr>
<tr>
<td>IMLS (2014-): Digital Content</td>
<td>Institute of Museum and Library Services</td>
<td>Institute of Museum and Library Services (IMLS)</td>
<td>GEMF: Sample Plan #2</td>
<td></td>
</tr>
<tr>
<td>IMLS (2014-): New Software Tools or Applications</td>
<td>Institute of Museum and Library Services</td>
<td>Institute of Museum and Library Services (IMLS)</td>
<td>GEMF: Sample Plan #3</td>
<td></td>
</tr>
<tr>
<td>IMLS (2014-): Research Data</td>
<td>Institute of Museum and Library Services</td>
<td>Institute of Museum and Library Services (IMLS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institute of Education Sciences (US Dept of Education)</td>
<td>Institute of Education Sciences (US Dept of Education)</td>
<td>IES Data Sharing Implementation Guide</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

View All
NIH Policy on Rigor and Reproducibility

Posted on June 2, 2016 by stephaniesimms

You’ve all heard about the reproducibility crisis in science. But you may not be aware of a (relatively) new National Institutes of Health (NIH) policy designed to address the issue. The NIH Policy on Rigor and Reproducibility became effective for proposals received on or after January 25, 2016 and applies to most NIH and Agency for Healthcare Research and Quality (AHRQ) grant applications. We just learned about the policy ourselves thanks to the combined efforts of UCSD library and research staff to raise awareness on their campus (and here’s a noteworthy mention in a Nature review of 2015 science news). To aid researchers in meeting the new criteria, UCSD produced this handy guide that we (and they) would like to share with the wider community.

The new policy does not involve any changes to data sharing plans. It is related and important enough, however, that we inserted a statement and link in the “NIH-GEN: Generic” template (Please note the Rigor and Reproducibility requirements that involve updates to grant application instructions and review criteria [but not Data Sharing Plans]).

The policy does involve:

- Revisions to application guide instructions for preparing your research strategy attachment
- Use of a new “Authentication of Key Biological and/or Chemical Resources” attachment (example from UCSD library website)
- Additional rigor and transparency questions reviewers will be asked to consider when reviewing applications

These policies are all meant to achieve basically the same goals: to promote openness, transparency, reproducibility, access to, and reuse of the results of scientific research. We're grateful to the folks at UCSD—Dr. Anita Bandrowski, Ho Jung Yoo, and Reid Otsui—for helping to consolidate the message and for providing some new educational resources.