

Tutorials on Data Management

Lesson 3: Data Management Planning



CC image by Joe Hall on Flickr

Lesson Topics

- What is a data management plan (DMP)?
- Why prepare a DMP?
- Components of a DMP
- NSF requirements for DMPs
- Example of NSF DMP



CC image by Daria Hueske on Flickr

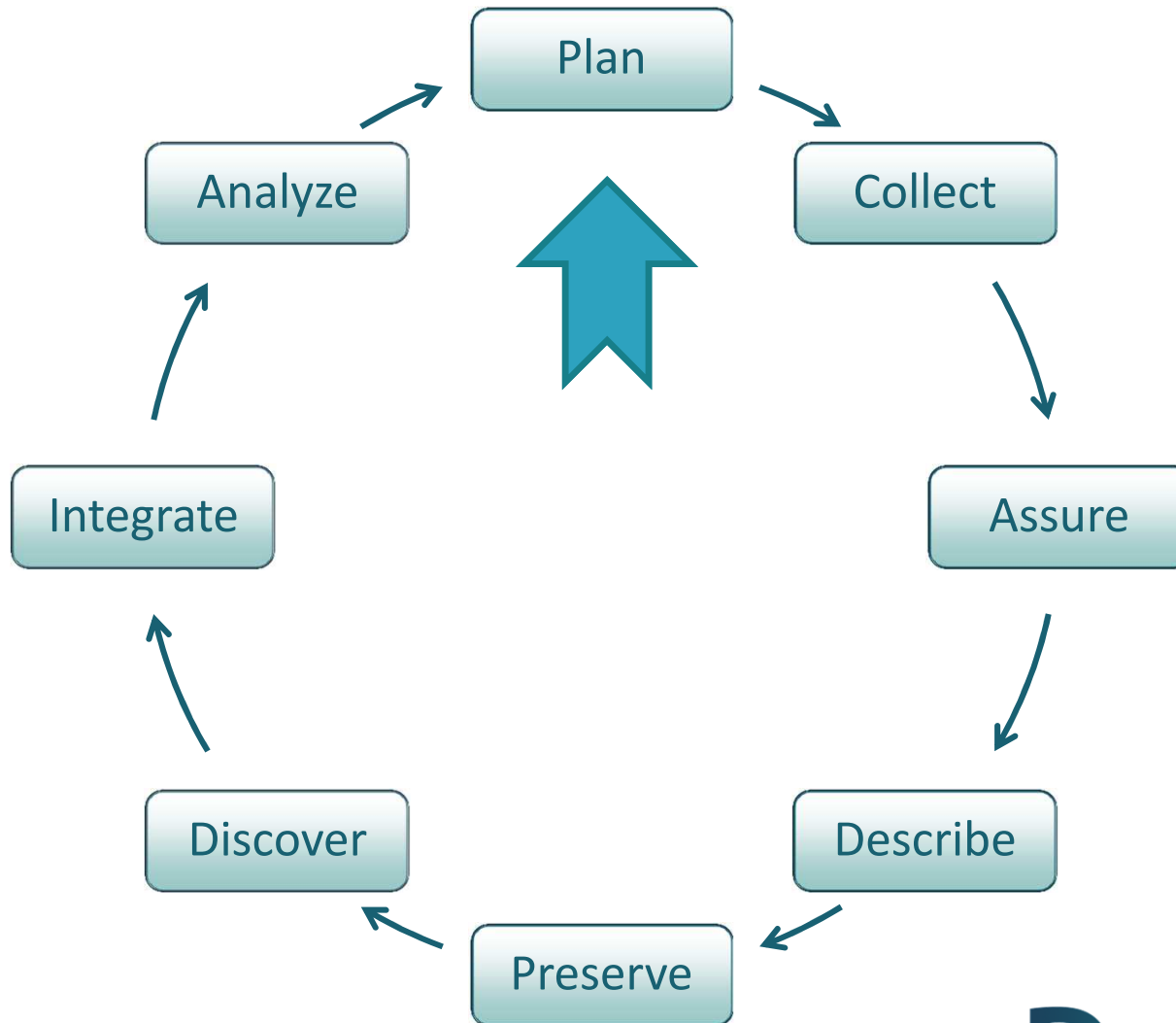
Learning Objectives

- After completing this lesson, the participant will be able to:
 - Define a DMP
 - Understand the importance of preparing a DMP
 - Identify the key components of a DMP
 - Recognize the DMP elements required for an NSF proposal



CC image by cybrarian77 on Flickr

The Data Life Cycle



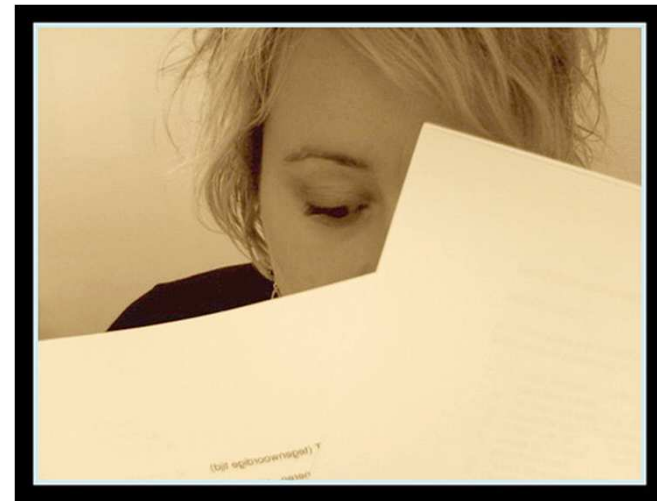
What is a Data Management Plan?

- Formal document
- Outlines what you will do with your data **during** and **after** you complete your research
- Ensures your data is safe for the **present** and the **future**

From University of Virginia Library

Why Prepare a DMP? (1)

- Save time
 - Less reorganization later
- Increase research efficiency
 - Ensures you and others will be able to understand and use data in future



CC image by Cathew on Flickr

Why Prepare a DMP? (2)

- Easier to preserve your data
- Prevents duplication of effort
- Can lead to new, unanticipated discoveries
- Increases visibility of research
- Makes research and data more relevant
- Funding agency requirement

Components of a General DMP

1. Information about data & data format
2. Metadata content and format
3. Policies for access, sharing and re-use
4. Long-term storage and data management
5. Budget

1. Information About Data & Data Format

1.1 Description of data to be produced

- Experimental
- Observational
- Raw or derived
- Physical collections
- Models and their outputs
- Simulation outputs
- Curriculum materials
- Software
- Images
- Etc...



CC image by Jeffery Beall on Flickr

1. Information About Data & Data Format

1.2 How data will be acquired

- When?
- Where?

1.3 How data will be processed

- Software used
- Algorithms
- Workflows



CC image by Ryan Sandridge on Flickr

1. Information About Data & Data Format

1.4 File formats

- Justification
- Naming conventions

1.5 Quality assurance & control during sample collection, analysis, and processing



CC image by Artform Canada on Flickr

1. Information About Data & Data Format

1.6 Existing data

- If existing data are used, what are their origins?
- Will your data be combined with existing data?
- What is the relationship between your data and existing data?

1.7 How data will be managed in short-term

- Version control
- Backing up
- Security & protection
- Who will be responsible

2. Metadata Content & Format

Metadata defined:

- Documentation and reporting of data
- Contextual details: Critical information about the dataset
- Information important for using the data
- Descriptions of temporal and spatial details, instruments, parameters, units, files, etc.

2. Metadata Content & Format

2.1 What metadata are needed

- Any details that make data meaningful

2.2 How metadata will be created and/or captured

- Lab notebooks? GPS units?
- Auto-saved on instrument?

2.3 What format will be used for the metadata

- Standards for community
- Justification for format chosen

3. Policies for Access, Sharing, Reuse

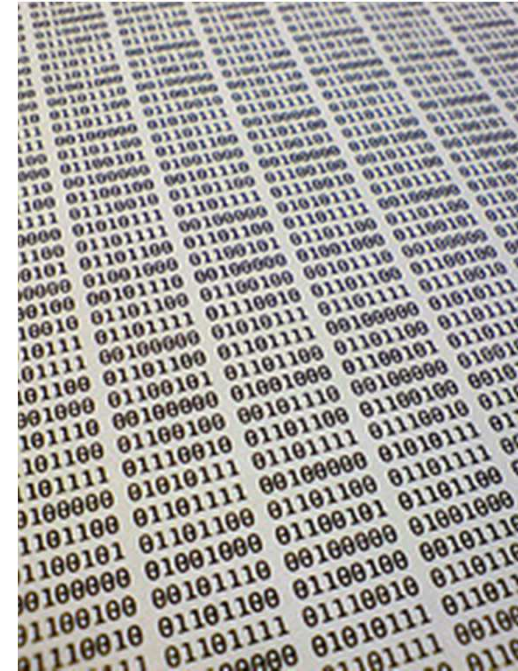
3.1 Obligations for sharing

- Funding agency
- Institution
- Other organization
- Legal

3.2 Details of data sharing

- How long?
- When?
- How access can be gained?
- Data collector rights

3.2 Ethical/privacy issues with data sharing



CC image by Jim Sher on Flickr

3. Policies for Access, Sharing, Reuse

3.4 Intellectual property & copyright issues

- Who owns the copyright?
- Institutional policies
- Funding agency policies
- Embargos for political/commercial reasons

3.5 Intended future uses/users for data

3.6 Citation

- How should data be cited when used?
- Persistent citation?



CC image by buddawiggi on Flickr

4. Long-term Storage & Data Management

4.1 What data will be preserved

4.2 Where will it be archived

- Most appropriate archive for data
- Community standards

3.6 Data transformations/formats needed

- Consider archive policies

4.4 Who will be responsible

- Contact person for archive



5. Budget

5.1 Anticipated costs

- Time for data preparation & documentation
- Hardware/software for data preparation & documentation
- Personnel
- Archive costs

5.2 How costs will be paid



CC image by Adria Richards on Flickr

Tools for Creating Data Management Plans

dmp.cdlib.org

DMPTool
Guidance and Resources for your Data Management Plan

Home About DMP Tool DMP News My Plans Funder Requirements Help

Meet funder requirements for data management plans.

Get Started!

Data Management Plan: Sample Plan Created at the DataONE Best Practices Workshop - Santa Fe NM 7/2011 Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013

1. Types of data produced

Recent DMP News

DataBib registry of repositories now linked

Survey on role of libraries in data management through April 9

Support for NSF Emerging Frontiers in Research and Innovation (EFRI) grants

More news >

The DMP Tool allows you to: 1 2 3 4

DMPTOOL is a service of the University of California Curation Center of the California Digital Library

DCC because good research needs good data

DMPonline

The DCC Data Management Planning Tool [Front Page](#) [About](#) [Instructions](#) [Future plans](#) [News](#) [Documents](#) [My Home](#)

Welcome to DMP Online

Funding bodies increasingly require their grant-holders to produce and maintain Data Management Plans (DMPs), both at the bid-preparation stage and after funding has been secured.

DMP Online has been developed by the Digital Curation Centre to enable you to build and edit DMPs according to the requirements stipulated by the major UK funders.

The tool also contains helpful guidance and links for researchers and other data professionals.

To get started, click on 'Create account' or 'Log in' above, or alternatively have a look at the [About DMP Online](#) page or the [Instructions for use](#), then get going by clicking 'Start a new plan', below.

If you already have plans stored in the system, click [My home](#) or [Return to a saved plan](#) to see them in a list.

- [Start a new plan](#)
- [Return to a saved plan](#)

News

- [1000 data management plans!](#)
posted by martin, about 1 month ago
- [Joint US-UK workshop at IDCC11](#)
posted by martin, 5 months ago
- [DMP Online at Future Perfect 2012, Wellington, New Zealand](#)
posted by sjones, 6 months ago

IISC

Data Management Planning

dmponline.dcc.ac.uk

DataONE

NSF DMP Requirements



From Grant Proposal Guidelines:

Plans for data management and sharing of the products of research. Proposals must include a supplementary document of no more than two pages labeled “Data Management Plan”. This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (in AAG), and 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**, re-distribution, and the production of derivatives
5. **plans for archiving** data, samples, and other research products, and for preservation of access to them

NSF DMP Requirements



Summarized from Award & Administration Guide:

4. Dissemination and Sharing of Research Results

- a) Promptly publish with appropriate authorship
- b) Share data, samples, physical collections, and supporting materials with others, within a reasonable timeframe
- c) Share software and inventions
- d) Investigators can keep their legal rights over their intellectual property, but they still have to make their results, data, and collections available to others
- e) Policies will be implemented via
 - Proposal review
 - Award negotiations and conditions
 - Support/incentives

Data in Real Life: A DMP Example

Project name: Effects of temperature and salinity on population growth of the estuarine copepod, *Eurytemora affinis*

Project participants and affiliations:

Carly Strasser (University of Alberta and Dalhousie University)

Mark Lewis (University of Alberta)

Claudio DiBacco (Dalhousie University and Bedford Institute of Oceanography)

Funding agency: CAISN (Canadian Aquatic Invasive Species Network)



Photo by C. Strasser; all rights reserved

Description of project aims and purpose:

We will rear populations of *E. affinis* in the laboratory at three temperatures and three salinities (9 treatments total). We will document the population from hatching to death, noting the proportion of individuals in each stage over time. The data collected will be used to parameterize population models of *E. affinis*. We will build a model of population growth as a function of temperature and salinity. This will be useful for studies of invasive copepod populations in the Northeast Pacific.

Video Source: Plankton Copepods. Video. Encyclopædia Britannica Online. Web. 13 Jun. 2011

Data in Real Life: A DMP Example

1. Information about data

Every two days, we will subsample *E. affinis* populations growing at our treatment conditions. We will use a microscope to identify the stage and sex of the subsampled individuals. We will document the information first in a laboratory notebook, then copy the data into an Excel spreadsheet. For quality control, values will be entered separately by two different people to ensure accuracy. The Excel spreadsheet will be saved as a comma-separated value (.csv) file daily and backed up to a server. After all data are collected, the Excel spreadsheet will be saved as a .csv file and imported into the program R for statistical analysis. Strasser will be responsible for all data management during and after data collection.

Our short-term data storage plan, which will be used during the experiment, will be to save copies of 1) the .txt metadata file and 2) the Excel spreadsheet as .csv files to an external drive, and to take the external drive off site nightly. We will use the Subversion version control system to update our data and metadata files daily on the University of Alberta Mathematics Department server. We will also have the laboratory notebook as a hard copy backup.

Data in Real Life: A DMP Example

2. Metadata format & content

We will first document our metadata by taking careful notes in the laboratory notebook that refer to specific data files and describe all columns, units, abbreviations, and missing value identifiers. These notes will be transcribed into a .txt document that will be stored with the data file. After all of the data are collected, we will then use EML (Ecological Metadata Language) to digitize our metadata. EML is one of the accepted formats used in Ecology, and works well for the type of data we will be producing. We will create these metadata using Morpho software, available through the Knowledge Network for Biocomplexity (KNB). The documentation and metadata will describe the data files and the context of the measurements.

Data in Real Life: A DMP Example

3. Policies for access, sharing & reuse

We are required to share our data with the CAISN network after all data have been collected and metadata have been generated. This should be no more than 6 months after the experiments are completed. In order to gain access to CAISN data, interested parties must contact the CAISN data manager (data@caisn.ca) or the authors and explain their intended use. Data requests will be approved by the authors after review of the proposed use.

The authors will retain rights to the data until the resulting publication is produced, within two years of data production. After publication (or after two years, whichever is first), the authors will open data to public use. After publication, we will submit our data to the KNB, allowing discovery and use by the wider scientific community. Interested parties will be able to download the data directly from KNB without contacting the authors, but will still be required to give credit to the authors for the data used by citing a KNB accession number either in the publication text or in the references list.

Data in Real Life: A DMP Example

4. Long-term storage and data management

The data set will be submitted to KNB for long-term preservation and storage. The authors will submit metadata in EML format along with the data to facilitate its reuse. Strasser will be responsible for updating metadata and data author contact information in the KNB.

5. Budget

A tablet computer will be used for data collection in the field, which will cost approximately \$500. Data documentation and preparation for reuse and storage will require approximately one month of salary for one technician. The technician will be responsible for data entry, quality control and assurance, and metadata generation. These costs are included in the budget in lines 12-16.

Summary

DMPs are an important part of the data life cycle. They save time and effort in the long run, and ensure that data are relevant and useful for others.

Funding agencies are beginning to require DMPs

Major components of a DMP:

1. Information about data & data format
2. Metadata content and format
3. Policies for access, sharing and re-use
4. Long-term storage and data management
5. Budget

Resources

1. University of Virginia Library
<http://www2.lib.virginia.edu/brown/data/plan.html>
2. Digital Curation Centre <http://www.dcc.ac.uk/resources/data-management-plans>
3. University of Michigan Library http://www.lib.umich.edu/research-data-management-and-publishing-support/nsf-data-management-plans#directorate_guide
4. NSF Grant Proposal Guidelines
http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp
5. Inter-University Consortium for Political and Social Research
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/index.jsp>
6. DataONE <http://www.dataone.org/plans>

The full slide deck may be downloaded from:
<http://www.dataone.org/education-modules>

Suggested citation:

DataONE Education Module: Data Management Planning.

DataONE. Retrieved Nov12, 2012. From

http://www.dataone.org/sites/all/documents/L03_DataManagementPlanning.pptx

Copyright license information:



No rights reserved; you may enhance and reuse for your own purposes. We do ask that you provide appropriate citation and attribution to DataONE.