Data Management Planning

What it is and how to do it:

Ulf Jakobsson, SND

Ulf.jakobsson@snd.gu.se

ARIADNE is funded by the European Commission's Seventh Framework Programme

Data Management Planning

Can you prove…

- how the work was carried out?
- that data weren't falsified?

What should have been done?

Data Management Planning

Data Management Plan (DMP)

Data Management Planning

Vast amounts of research data

- Can be re-used

A plan for what information to save and how to save it

Here a DMP will be useful!

Data Management Planning

A DMP:

- increases the research efficiency
- ensures that the data is
  - in a correct format
  - is well organized and well annotated

No need to:

- re-format
- re-organize
- try to remember details later

Easier to explain

Data Management Planning

In an early stage of the research process

- Aspects of data management
  - Metadata generation
  - Data preservation
  - Lifecycle documentation
- Ensures that the material is
  - Well managed
  - Prepared for preservation

Data Management Planning

Decide and contact an archive

Funding agencies

- Requires data management plans
- Part of the proposal and evaluation process

What is a Data Management Plan?

A DMP is:

- A formal document
- Developed at the start of a research process
- Outlines all aspects
  - During the research project
  - After the research project

Find or create a data planning checklist

DMP: Components?

Components in a DMP?

- Description of the project:
  - purpose of the research
  - organizations and staff involved

DMP: Collection/production of data

- Description of data:
  - Nature, scope, scale & format of the data
  - How the data will be collected
  - Where & when
- Overview of existing data
- Will the data be reproducible?

DMP: Organization/processing of data

- How will the data be organized?
  - Folder structures
  - File naming conventions
  - File versioning
- How will the data be processed?
  - File formats
  - Tools and software

DMP: Documentation

- How will the data be documented?
  - Metadata (data about data)
    - Communication between principal investigator and later researchers
    - A good description is essential
    - Important for the future
    - Comprehensive information
      - Temporal and spatial details
      - Parameters
      - Units
      - Etc.

DMP: Documentation

What metadata are needed?

- How will the metadata be created/captured?
  - Lab notebooks
  - GPS hand-held units
  - Auto-saved files on instruments, etc.
- Format for the metadata
- Any ontology or community standard

DMP: Local data management

- Short-term data management.
  - File formats
  - Version control
  - Project and data identifiers
  - Local storage and back-up procedures
  - Security and protection
  - Creation and maintenance
  - Compliance with the plan
  - Policies

DMP: Law and ethics

Legal and/or ethical issues

- Intellectual property or copyright issues
- Confidential information
  - Personal data
  - High security data
- Restrictions
- Privacy or ethical issues with data sharing
- Embargoes

Publication of data

Why publish researchdata:

- It simplifies the researcher's life
  - The archive
    - Houses the data
    - Disseminates the data
    - Provides user support
    - Prevailing standards
    - Long-term preservation

Publication of data

- Encourage re-publication and secondary analyses of the data
- Reduces cost of research
- Might lead to new discoveries
- Prevents duplication

DMP: Data sharing

How to share data

- When and how
  - Publication at an archive/repository
  - Deposition at a subject-specific database
  - Self-dissemination through a dedicated web site
  - Institutional repository at an academic institution
- Delayed, lost or selective publication of data
- Technical complications

DMP: Data sharing

- How to gain access to the data
  - Limitations
- Citation
- Persistent identifiers
  - Digital Object Identifier (DOI)
  - Archival Resource Keys (ARKs)
  - Persistent Uniform Resource Locators (PURLs)
  - Uniform Resource Names (URNs)
  - Extensible Resource Identifiers (XRIs)

DMP: Long-term preservation

Identify an archive early

- Format
- Transform
- Document

- Include a backup archive
- Primary contact person

DMP: Budget

Reuse of data ensured

- Preservation and data management costs
- Anticipate costs ahead of time
  - Personnel time for
  - Hardware and/or software needed for
  - Costs associated with submitting the data to an archive

Ulf Jakobsson

Swedish National Data Service

Ulf.jakobsson@snd.gu.se

ARIADNE is funded by the European Commission's Seventh Framework Programme