Data Management Plans (DMPs)

Introduction to Data Management Practices course

NBIS DM Team
data@nbis.se

https://nbisweden.github.io/module-open-science-dm-practices/index.html

What is a DMP?

- Document
- Outlines the data management strategies in a project
- How the data is
  - collected
  - documented
  - organized
  - preserved

A data management plan (DMP) is a document that describes the data produced during a research project.

It outlines the data management strategies that will be implemented both during the active phase of the research project and after the project ends.

Making plans for how you will collect, document, organize, and preserve your data are all part of the data management strategy.

The main parts of a DMP

1. **Description of data**
   - What types of data will be created and/or collected, in terms of data format and amount/volume of data?
2. **Documentation**
   - How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, etc.?
3. **Storage and backup**
   - How is data security, storage and backup of data and metadata safeguarded during the research process?

cont. The main parts of a DMP

4. **Legal and ethical aspects**
    - How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?
5. **Accessibility and long-term storage**
    - How, when and where will research data or information about data (i.e. metadata) be made accessible?
    - In what way is long-term storage safeguarded, and by whom?
6. **Responsibility and resources**
    - Who are the responsible persons for data management?
    - What resources (costs, labour input or other) will be required for data management?

Why write a DMP?

Well-managed research data allows for:

- verification of published research
- reduce the potential for scientific fraud
- enable re-use of existing data
- discourage unintentional redundancy in research
- serve as training resource for new researchers

Funding agencies requires a DMP:

- For **transparency** and **openness**: publicly funded research data must be discoverable, accessible, and reusable to the public
- **Return on investment**: well planned data maximizes the research potential of the data and provides greater returns on public investments and research

A DMP is a good way to obtain **well-managed research data**, and this in turn allows for:

- verification or refinement of published research results
- reduces the potential for scientific fraud
- promotes new research through the use of existing data
- provides resources for training new researchers
- discourages unintentional redundancy in research

Also, **funding agencies** often requires a DMP:

- **For transparency and openness**: publicly funded research data must be discoverable, accessible, and reusable to the public
- **Return on investment**: well planned data maximizes the research potential of the data and provides greater returns on public investments and research

When write a DMP?

A DMP is a living document:

- **Project planning**: Outline the strategies to be able to estimate the resources needed, so this can be included in the proposal for funding.
- **Project start**: Complete with details e.g. about documentation, data quality measures, file and folder strategies, etc.
- **Project end**: Update with e.g. links to published data and details about archiving (what data and where).

A DMP is a **living document**, the initial version is written the same time as a new project idea is emerging, and then successively updated as the project continues and new decisions are made:

Project planning: The DMP should outline the strategies for data management in sufficient detail to be able to estimate the resources needed to implement the DMP, so that this can be included in the proposal for funding (e.g. data production, data analysis, storage during and after project, costs related to publishing of data).

Project start: The DMP is completed with more details e.g. about documentation, data quality measures, file and folder strategies, etc.

Project end: The DMP is updated a final time with e.g. links to published data and details about archiving (what data and where), so that this document enables future re-use of the project (by yourself or others).

How write a DMP?

DMP templates:

- Provided by funding agencies, e.g. Swedish Research Council and Science Europe
- High-level questions, with no guidance on how to answer
- Use Word?

DMP tools:

- DMPOnline - The tool most universities have chosen to offer; Good guidance but typically generic and not Life Science specific
- Data Stewardship Wizard - Provided by SciLifeLab; Gives Life Science specific guidance

Standard DMP templates can typically be found at funder agencies, e.g. Swedish Research Council and Science Europe, and it is of course possible to write in your favorite text editor.

However, the questions in these templates are quite high-level, with little or no guidance on how to answer them.

Luckily, there are tools to assist you:

- DMPOnline - The tool most universities have chosen to offer; Good guidance but typically generic and not Life Science specific
- Data Stewardship wizard - Provided by SciLifeLab; Gives Life Science specific guidance

Key Points

- A data management plan (DMP) is a document that describes the data produced in the course of a research project.
- A DMP allows for well-managed data, and funding agencies often requires a DMP for transparency and return on investment.
- A DMP is a living document, the first version is written during project planning, and is then updated as the project proceeds.
- There are standard templates available e.g. at funder agencies, and tools to assist when writing.