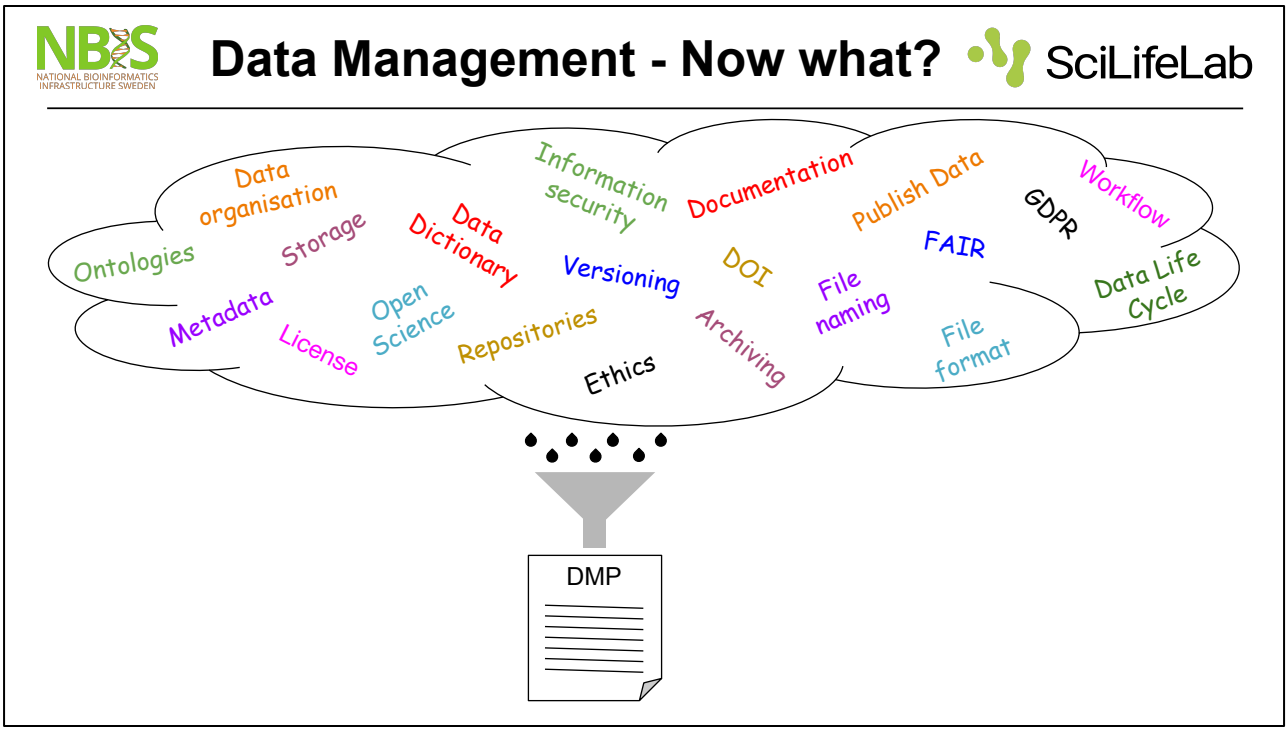
Data Management Plans (DMPs)

Introduction to Data Management Practices course

NBIS DM Team
data-management@scilifelab.se

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



During this course you have come across many terms regarding data management and learned how to implement best practices, but there is one more cornerstone to cover: How do we gather all the decisions made and how do we know that we have covered it all, that our data will be well managed throughout its life cycle? The answer is to write a data management plan.

What is a DMP?

- A document **addressing requirements and practices** for managing the project's data, code and documentation throughout the data life cycle
- Outlines the data management strategies in a project, how the data is:
  - collected
  - documented
  - organized
  - preserved

A data management plan (DMP) is a document addressing requirements and practices for managing the project's data, code and documentation, throughout the data life cycle, i.e from the initial planning until the project ends and beyond.

It outlines the data management strategies in a project.

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

Why write a DMP?

- Identify **gaps** in current data management strategies (reduce time spent later on)
- Facilitate collaboration by setting project wide **standards** (data organisation, naming conventions, metadata standards, vocabulary, …)
- **Cost** estimation
- Clarify **responsibilities**
- Obtain **well-managed** research data (which enables reproducibility and reusability)
- First step towards **FAIR**ness
- **Requirement** from funders and other stakeholders for reasons of **transparency**, **openness** and **return on investment**

There are several reasons why writing a data management plan is a very good idea:

- Think of the DMP as a checklist, comparable to a pilot's checklist before take-off, and going through the checklist allows you to identify gaps in current data management strategies. Identifying the gaps early on saves a lot of headache and time spent later. Going through the process of planning is more important than the actual plan itself.
- In a project with several members, it is important to decide on standards that all collaborators should adhere to, e.g. regarding how to organise the data, how to name it, which metadata standards to use, what vocabularies to use, etc.
- Writing a DMP also enables you to estimate costs with regards to data production, storage, data management, etc.
- It is also a good way to clarify responsibilities regarding the data and the data management, e.g. who is responsible for the execution of the DMP.
- By planning how the data will be managed, there's greater chance that the research data will be well-managed (no guarantee, since you still need to have good strategies and actually implement them for this to happen). Of course there are many benefits with well-managed data but the main ones are:
    a. reproducibility, so that the results can be verified
    b. reusability, so that this data can be used for answering other scientific questions, thus reducing redundancy
- A DMP is the first step towards being FAIR in your project.

If the reasons above don't persuade you, the last argument is that it is more and more a

requirement by funders and other stakeholders:

- **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** that will develop throughout the project:

- **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).

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?

Reuse and production of research 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.?

Documentation procedures and standards

3. **Storage and backup**
   - How is data security, storage and backup of data and metadata safeguarded during the research process?

Storage, security and backup

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?

Legal & ethical and assurance procedures

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?

Future reuse and long-term access aspects

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?

Responsibilities and resources required

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 MS Word?

DMP tools:

- DMPOnline - Common for universities; Good guidance but not Life Science specific; Free text answer type;
- Data Stewardship Wizard - Provided by SciLifeLab; Life Science specific guidance; Many questions with answer options;

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
  a. The tool most universities have chosen to offer (check with your institute)
  b. Good guidance but typically generic and not Life Science specific
  c. Most often free text answers
- Data Stewardship wizard
  a. Provided by ScilifeLab
  b. Gives Life Science specific guidance
  c. Less free text answers, instead many questions with answer options

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.