General Finnish DMP guidance 2019

Why should you manage your research data and write a data management plan (DMP)?

- Because it is good research practice!
- You will reduce the risk of losing your data.
- You will be able to anticipate complex ownership and user rights issues in advance.
- It helps you support open access to create productive future collaborations.
- You will meet funder requirements.
- It helps you save time and money.
- The DMP reflects your managerial skills as a project leader.

Data is understood as a broad term that includes ”all information that is needed to replicate a study ---, and everything that is potentially useful for others.” – Sarah Jones /DCC

Your DMP should describe how you will manage data **during the whole research life cycle**. The DMP is **a living document** which should be updated as the research project progresses.

Your research data management practices should follow the FAIR principles, which require that your data will be Findable, Accessible, Interoperable and Re-usable.

You can use this general set of questions to structure your DMP. Answer the questions where applicable.

Good luck with your DMP!

1. General description of data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Also give a rough estimate of the size of the data produced/collected?

Consider your DMP as a part of your research plan. The standalone readability of a DMP is not necessary. DMP complements your research plan with a description of the technical management of your data. To avoid redundancy, refer to your research plan in your DMP and vice versa.

Briefly describe what types of data you are collecting or producing. Also explain what kinds of already existing data you will use. For example, the types of texts, images, photographs, measurements, statistics, physical samples or codes.

Categorise your data in such a way that you can refer to it later in the plan. That is, your answer to this question can form a general structure for the rest of the plan, for example, A) data collected for this project, B) data produced as an outcome of the process, C) previously collected existing data which is reused in this project, D) managerial documents and project deliverables, and so on.

List the file formats for each data set. In some cases, the file formats used during the research project may differ from those used in archiving the data. List both. The file format is a primary factor in the accessibility and reusability of your data in the future.

CC 0 ZERO

Tips for best practices

- Data analysis and methodological issues related to data and materials should be described in your research plan.
- Examples of file formats: .csv, .txt, .docx, .xslx, .tif
- When listing the file formats, you will be using, make sure to include any special or uncommon software necessary to view or use the data, especially if the software is coded in your project.
- Use a table or bullet points for a concise way of presenting data types, file formats, the software used and so on.
- You can also estimate the increase in data production or collection during the project, e.g. per week: "The project is producing/collecting data approximately 100 GB per week."

1.2 How will the consistency and quality of data be controlled?

Explain how the data collection, analysis and processing methods used may affect the quality of data and how you will minimise the risks related to data accuracy.

Data quality control ensures that no data is accidentally changed and that the accuracy of data is maintained over its entire life cycle. Quality problems can emerge due to the technical handling, converting or transferring of data, or during its contextual processing and analysis.

Tips for best practices

- Transcriptions of audio or video interviews should be checked by someone other than the transcriber.
- Analog material should be digitised in as high resolution as possible for accuracy.
- In all conversions, maintaining the original information content should be ensured.
- Software-producing checksums should be used.

2. Ethical and Legal Compliance

2.1 What ethical issues are related to your data management, for example, in handling sensitive data, protecting the identity of participants, or gaining consent for data sharing?

Describe how you will maintain high ethical standards and comply with relevant legislation when managing your research data. Ethical issues must be considered throughout the whole research data life cycle.

For example, following the guidelines regarding informing research participants is considered an ethical requirement for most research. Moreover, if you are handling personal or sensitive information, describe how you will ensure privacy protection and data anonymisation or pseudonymisation.

Tips for best practices

CC 0 ZERO

- Check your institutional Ethical Guidelines and Data Security Policy, and prepare to follow the instructions that are given in these guidelines.
- If your research is to be reviewed by an ethical committee, outline in your DMP how you will comply with the protocol (e.g., how to remove personal or sensitive information from your data before sharing it to ensure privacy protection).
- See, e.g., the Finnish Advisory Board on Research Integrity for more information about the responsible conduct of research.
- See, e.g., the European Code of Conduct for Research Integrity.

2.2 How will data ownership, copyright and Intellectual Property Right (IPR) issues be managed? Are there any copyrights, licenses or other restrictions which prevent you from using or sharing the data?

Describe who will own the data and how the ownership issues have been agreed upon. Describe who can issue permissions to (re)use it.

Tips for best practices

- Check your organisational data policy for ownership, right of use and right to distribute.
- Ownership agreements should be made as early as possible in the project life cycle.
- Also consider the funder's policy on copyrights or IPR.
- It is recommended to make all research data, code and software created within a research project available for reuse, e.g., under Creative Commons, GNU, MIT or another relevant license.

3. Documentation & metadata

3.1 How will you document your data in order to make it findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?

Data documentation enables data sets and files to be discovered, used and properly cited by other users (human or computer). Metadata is essential information regarding the data, for example, where, when, why and how the data were collected, processed and interpreted. Metadata may also contain details about experiments, analytical methods and the research context.

Tips for best practices

- Describe all the types of documentation (README files, metadata, etc.) you will provide to help secondary users find, understand and reuse your data.
- Following the FAIR principles will help you ensure the Findability, Accessibility, Interoperability and Re-usability of your data.
- Use research instruments which create standardised metadata formats automatically. Then your data can be moved from one manufacturer tool to another.
- Consider how the data will be organised during the project. Describe, for example, your file-naming conventions, version control and folder structure.

CC 0 ZERO

- Identify the types of information that should be captured to enable other researchers to discover, access, interpret, use and cite your data.
- Repositories often require the use of a specific metadata standard. Check whether a discipline/community- or repository-based metadata schema or standard (i.e., preferred sets of metadata elements) exists that can be adopted.

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will it be backed up?

Describe where you will store and back up your data during your research project. Methods for preserving and sharing your data after your research project has ended are explained in more detail in Section 5.

Consider who will be responsible for backup and recovery. If there are several researchers involved, create a plan with your collaborators and ensure safe transfer between participants.

Tips for best practices

- The use of a safe and secure storage provided and maintained by your organisation's IT support is preferable.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

It is essential to consider data security issues, especially if your data is sensitive, for example, personal data, politically sensitive information or trade secrets. Describe who has access to your data, what they are authorised to do with it, or how you will ensure the safe transfer of data to your collaborators.

Tips for best practices

- Access controls should always be in line with the level of confidentiality involved

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?

Describe whether you will publish or otherwise make all your data or only parts of it openly available. If your data or parts of it cannot be opened, explain why. The openness of research data promotes its reuse.

Tips for best practices

CC 0 ZERO

- You can publish a description (i.e., the metadata) of your data without making the data itself openly available, which enables you to restrict access to the data.
- Publish your data in a data repository or peer-reviewed data journal.
- Check re3data.org to find a repository for your data.
- Remember to check funder, disciplinary or national recommendations for data repositories.
- It is recommended to make all research data, code and software created within a research project available for reuse, e.g., under Creative Commons, GNU, MIT or another relevant license.
- Consider using repositories or publishers which provide persistent identifiers (PID) to enable access to the data via a persistent link (e.g. DOI, URN).

5.2 Where will data with long-term value be archived, and for how long?

Briefly describe what data to archive and for how long – as well as what data to dispose of after the project. Describe the access policy to the archived data.

Tips for best practices

- Remember to check funder, disciplinary or national recommendations for data archives.

6. Data management responsibilities and resources

6.1 Who will be responsible for specific tasks of data management during the life cycle of the research project? Also estimate the resources (e.g. funding, time, and effort) required for data management.

Some of the responsibilities have already been described in the previous answers. You can refer to them or summarize all responsibilities with an estimation of the resources needed.

Estimate the need to hire expert help to manage, preserve and share the data. Consider the additional computational facilities and resources that need to be accessed, and what the associated costs will amount to.

Tips for best practices

- Remember to specify your data management costs in the budget, according to funder requirements.

CC 0 ZERO