General Finnish DMP guidance 2020

How do I write a DMP? - READ THIS FIRST!
- Read all of the questions first!
- Use a DMP to complement your research plan – avoid overlaps with the research plan!
- The research plan describes the scientific, analytical and methodological processing of data.
- The data management plan describes the technical and administrative management of data.
- To avoid redundancy, refer to your research plan in your DMP and vice versa.
- Use the DMP as a risk evaluation document – it shows that you can recognise, anticipate and handle the risks related to your data management workflow.
- The DMP should be drawn from your own research project – do not copy/paste examples from somewhere else.
- Write only sentences you yourself understand.
- Answer the questions where applicable – if a certain question is not applicable in your case, justify why not.
- Answer at least the main categories – each sub-question does not need to be answered separately.
- Include background information such as the name of the applicant and the project, the project number, the funding programme and the version of the DMP.
- Demonstrate your data management and version control skills, for example, when considering the name of the DMP file.
- Follow the organisation's or funder's requirements.

Why should you manage your research data and write a data management plan (DMP)?
- 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 your funder's requirements.
- It helps you save time and money.
- Your DMP reflects your managerial skills as a project leader.

In the DMP context, 'data' is understood as a broad term. Data covers all of the information and material your research results are based on. You can concentrate on the data, which is your responsibility.

Your DMP should describe how you will manage the data throughout the life cycle of your research. The DMP is a living document, which should be updated as the research project progresses.

Your research data management practices should aim to produce reusable data, which follows FAIR principles, that is, your data will be Findable, Accessible, Interoperable and Re-usable.

Good luck with your DMP!

1. General description of the 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? Additionally, give a rough estimate of the size of the data produced/collected.

Briefly describe what types of data you are collecting or producing. In addition, explain what kinds of already existing data you will (re)use. List, for example, the types of texts, images, photographs, measurements, statistics, physical samples or codes.

Categorise your data in a table or with a clear list, for example:
A) data collected for this project,
B) data produced as an outcome of the process,
C) previously collected existing data which is being reused in this project,
D) managerial documents and project deliverables, and so on.

The categorisation follows the license policy of your data sets. For example, briefly describe the license according to which you are entitled to (re)use the data. The categorisation can form a general structure for the rest of the DMP.

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 after the project. List both. The file format is a primary factor in the accessibility and reusability of your data in the future.

In the DMP, what is important is to describe the required disk space, not how many informants participated in the project. A rough estimation of the size of the data is sufficient, for example, less than 100 GB, approx. 1 TB or several petabytes.

Tips for best practices

- Use a table or bullet points for a concise way to present data types, file formats, the software used and the size of the data.
- Examples of file formats are .csv, .txt, .docx, .xslx and .tif.
- Make sure to describe any special or uncommon software necessary to view or use the data, especially if the software is coded in your project.
- You can also estimate the increase in data production or collection during the project for a specific time period: "The project is producing/collecting approximately 100 GB of data per week."
- AVOID OVERLAPS WITH THE RESEARCH PLAN! Data analysis and methodological issues related to data and materials should be described in your research plan.

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 the 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 the 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 the highest resolution possible for accuracy.
- In all conversions, maintaining the original information content should be ensured.
- Software-producing checksums should be used.
- Organise training sessions and set guidelines to ensure that everyone in your research group can implement quality control and anticipate the risks related to the quality of the data.
- AVOID OVERLAPS WITH THE RESEARCH PLAN! Issues related to data analysis, methods and tools should be described in your research plan, that is, do not include, for example, instrument calibration descriptions here.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

All types of research data involve questions of rights and legal and ethical issues. Demonstrate that you are aware of the relevant legislation related to your data processing. If you are handling personal or sensitive information, describe how you will ensure privacy protection and data anonymisation or pseudonymisation.

Tips for best practices

- Check your institutional ethical guidelines, data privacy 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 you will remove personal or sensitive information from your data before sharing data to ensure privacy protection).
- Will you process personal data? If you intend to do so, please detail what type of personal data you will collect.
- All data related to an identified or identifiable person is personal data. Information such as names, telephone numbers, location data and information on the congenital diseases of the individual's grandparents is personal data.

- Office of the Data Protection Ombudsman (https://tietosuoja.fi/en/processing-of-personal-data)
- AVOID OVERLAPS WITH THE RESEARCH PLAN! Details of the ethical issues, the ethical committee statements and the use of laboratory animals should be described in the research plan.

2.2 How will you manage the rights of the data you use, produce and share?

Describe how you will agree upon the rights of use related to your research data – including the collected, produced and (re)used data of your project. Here, you can employ your categorisation in the first question. Each of these categories involves different rights and licenses. Describe the transfer of rights procedures relevant to your project. Describe confidentiality issues if applicable in your project.

Tips for best practices

- Check your organisational data policy for ownership, the right of use and the right to distribute.
- Have you gained consent for data preservation and sharing?
- Agreements on ownership and rights of use should be made as early as possible in the project life cycle.
- Consider the funder's policy.
- It is recommended to make all of the research data, code and software created within a research project available for reuse, e.g., under a Creative Commons (https://creativecommons.org/choose/), GNU (https://www.gnu.org/licenses/gpl-3.0.en.html) or MIT license (https://opensource.org/licenses/MIT), or under another relevant license.

3. Documentation and metadata

3.1 How will you document your data in order to make the data 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). Documentation includes essential information regarding the data, for example, where, when, why and how the data were collected, processed and interpreted. Without the proper documentation, your data is useless. Describe the tool, such as Qvain, that you will use to describe your data sets. Do not mention metadata standards if you do not use them. You can anticipate the open accessibility of your data and its description already here. However, a detailed description of which part of your data can be set openly available will be included in Section 5 below.

AVOID OVERLAPS WITH THE RESEARCH PLAN! The data-level documentation (https://www.ukdataservice.ac.uk/manage-data/document/data-level.aspx) and details about

experiments, analytical methods and the research context belong to the research plan. In the DMP you should concentrate on the study-level documentation (https://www.ukdataservice.ac.uk/manage-data/document/study-level.aspx).

Tips for best practices

- Describe all the types of documentation (README files, metadata, etc.) you will provide to help secondary users to find, understand and reuse your data.
- Following the FAIR (https://www.force11.org/group/fairgroup/fairprinciples) principles will help you ensure the Findability, Accessibility, Interoperability and Re-usability of your data.
- Know the minimum requirements for data documentation; see, for example, Qvain Light (https://www.fairdata.fi/en/qvain/qvain-light-user-guide/).
- Use research instruments, which create standardised metadata formats automatically.
- Identify the types of information that should be captured to enable other researchers to discover, access, interpret, use and cite your data.

4. Storage and backup during the research project

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

Describe where you will store and back up your data during your research project. Explain the methods for preserving and sharing your data after your research project has ended 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.

Show that you are aware of the storing solutions provided by your organisation. Do not merely refer to IT services. In the end, you are responsible for your data, not the IT department or the organisation.

Tips for best practices

- The use of a safe and secure storage provided and maintained by your organisation's IT support is preferable.
- Do NOT USE external hard drives as the main storing option.

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 include sensitive data, personal data, politically sensitive information or trade secrets. Describe who has access to your data, what they are authorised to do with the data, 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 make openly available or publish all your data or only parts of the data. If your data or parts of the data cannot be opened, explain why.

In the case of sensitive data, which cannot be opened, describe the opening of its metadata. Describe the secured preservation procedure of sensitive data in Section 5.2.

The openness of research data promotes its reuse.

Tips for best practices

- 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 a data journal.
- Check re3data.org (https://www.re3data.org/) to find a repository for your data.
- Remember to check the funder, disciplinary or national recommendations for data repositories.
- It is recommended to make all of the research data, code and software created within a research project available for reuse, for example, under a Creative Commons (https://creativecommons.org/choose/), GNU (https://www.gnu.org/licenses/gpl-3.0.en.html) or MIT license (https://opensource.org/licenses/MIT), or under 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).
- AVOID OVERLAPS WITH THE PUBLICATION PLAN! The research article publication does not equal data publication. The data journal is a publication forum specialised in publishing research data.

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

Briefly describe what part of your data you will preserve and for how long. Categorise your data sets according to the anticipated preservation period:

A) Data to be destroyed upon the ending of the project
B) Data to be archived for a verification period, which varies across disciplines, e.g., 5–15 years
C) Data to be archived for potential re-use, e.g., for 25 years

D) Data with long-term value to be archived by a curated facility for future generations for tens or hundreds of years

Describe which part of the data you will dispose of after the project and how you will destroy the data. Describe the access policy to the archived data. Consider using archives with a curation policy.

Tips for best practices

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

6. Data management responsibilities and resources

6.1 Who (for example role, position, and institution) will be responsible for data management (i.e., the data steward)?

Summarise here all the roles and responsibilities described in the previous answers.

Tips for best practices

- Outline the roles and responsibilities for data management/stewardship activities, for example, data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name the responsible individual(s) where possible.
- For collaborative projects, explain the co-ordination of data management responsibilities across partners.
- Indicate who is responsible for implementing the DMP and for ensuring that it is reviewed and, if necessary, revised.
- Consider scheduling regular updates of the DMP.
- Finally, consider who will be responsible for the data resulting from your project after your project has ended.

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

Estimate the resources needed (for example, financial and time) 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.

Account for the costs of the necessary resources (for example, time) to prepare the data for sharing/preservation (data curation). Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, the costs of preparing data for deposit and repository charges.