General Finnish DMP guidance

Version: 9 Jun 2017

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

- Because it is good research practice!
- You will maintain and ensure data integrity and replication.
- You will reduce the risk of your data becoming obsolete or even lost.
- You will be able to tackle complex ownership and user rights issues in advance.
- It helps you support open access in order to promote new discoveries and productive future collaborations.
- You will meet the requirements of external funding agencies.
- At the end of the day, it will save you time and money.

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

Your DMP should describe how you 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 dictate how your data will be Findable, Accessible, Interoperable, and Re-usable.

Good luck with your DMP!

1. General description of data	
1.1 What kinds of data are collected or reused?	Briefly describe your research data. Explain what kind of data you are collecting or producing. Outline how the data will be collected: e.g. via surveys, interviews, laboratory experiments, or observations. Moreover, explain what kind of existing data you will reuse. Describe in short what types of data will be used and are expected to be produced: e.g. texts, images, photographs, statistics, physical samples, or codes. Tips for best practices · Describe your data in such a way that you can refer to it later in the plan. Your answer to this question forms the basis of the whole plan. · Explain your methods in more detail in the research plan. · By reusing data produced by you or others, you will avoid duplicating work already done.
1.2 What file formats will the data be in?	File format is a primary factor in the accessibility and reusability of your data in the future. List the file formats the data will be stored in. Note that a file format used during the project might not be the one most suitable for long-term preservation and reuse. Tips for best practices · List the file formats that data will be in: e.g. .csv, .txt, .docx, .xslx, .tif. · When listing the data formats you will be using, make sure to include any software necessary to view the data.

This guidance for writing a data management plan is used in Finnish research organisations. The guidance has been composed in the national Tuuli-project https://wiki.helsinki.fi/x/g5w6Cg

	· Favour software and formats based on open standards to enable data reuse, interoperability and sharing.

2. Documentation and Quality	
2.1 How will the data be documented?	Data documentation enables data sets and files to be discovered, used, and properly cited. Metadata is essentially information regarding the data: e.g. where, when, why, and how were the data collected, processed and interpreted. Metadata may also contain details about experiments, analytical methods, and research context. Metadata elements can include descriptive metadata which enables indexing, discovery and retrieval (e.g. keywords); technical metadata which describes how data sets were produced, structured and how they should be used (e.g. file naming); as well as rights to metadata which define who owns and who can access the data, and who has the right to manage it. Tips for best practices · Consider how the data will be organized during the project. Describe e.g. your file naming conventions, version control and folder structure. · Identify the types of information that should be captured to enable a researcher like you to discover, access, interpret, use, and cite your data. · Repositories for long-term preservation 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. · While utilizing research instruments which create metadata automatically, use standard metadata formats, when available. Then data can be moved from one manufacturer tool to another. · The national service for publishing metadata is the Etsin research data finder which contains metadata of data sets.
2.2 How will the consistency and quality of data be controlled and documented?	Data quality control ensures that no data will be lost or accidentally changed during the research process. Quality control of data is an integral part of all research and takes place during data collection, data entry or digitization, and data checking. Tips for best practices · Explain how the data collection methods used will affect the quality of data. You can provide evidence of data quality by documenting in detail how the data is collected. · Quality control measures can include e.g. using standardized methods and protocols for capturing observations, alongside recording forms with clear instructions, taking multiple measurements, observations or samples and calibration of instruments.

3. Storage and Backup	
3.1 How will the data be stored and 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 details 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 organization's IT support is preferable. · If you use commercial cloud services (e.g. Google Drive), make sure not to store or share unencrypted personal or sensitive data with them.
3.2 How will you control access to keep the data secure?	It is vital to consider data security issues, especially if your data is sensitive e.g., personal data, politically sensitive information or trade secrets! Describe **who has access to your data and what they are authorized to do with it**. Who will be responsible for access control? Tips for best practices · Access controls should always be proportionate to the kind of data and level of confidentiality involved. · Please note that there may be institutional data security policies which you are required to adhere to.

4. Ethics and Legal Compliance

4.1 How will ethical issues be managed?	Describe how you will maintain high ethical standards and comply with relevant legislation. Ethical issues must be considered throughout the whole research life cycle, from planning to publication as well as in paving the way for future reuse. 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 anonymization. Tips for best practices · Check your institutional Ethical Guidelines and Security Policy and prepare to follow instructions that are given in these guidelines. · Check whether an ethical review is required for your research project · If your research is to be reviewed by an ethical committee, outline in your DMP how you will comply with the protocol (i.e. how to remove personal or sensitive information from your data before sharing it to ensure privacy protection; or, how you will use restricted access procedures). · See e.g. 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 · See e.g. General Data Protection Regulation
4.2 How will ownership, copyright and Intellectual Property Right (IPR) issues be managed?	Describe who will own the data and who can issue permissions to reuse it. If you use research material or data collected or produced by a third party, consider the copyright issues and potential licenses which may affect its distribution. These issues should be solved already at the planning stage of the research project. If ownership issues have not been considered early enough in the research life cycle, sharing and reusing the data may become impossible. Tips for best practices · Check your organizational data policy for ownership guidelines · 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. The recommended CC license according to open science principles is the CC-BY.
5. Data Sharing and Long-Term Preservation	
5.1 How, when, where and to whom will the data be made available?	Describe whether you will share all your data or only parts of it, and for how long will it be made available. If your data or parts of it cannot be shared, explain why. Valid explanations might include confidentiality, trade secrets or ownership issues (license, copyright). Sometimes data cannot be shared due to the unreasonable effort required for its sharing (e.g. legacy data or large volumes of analog data). Tips for best practices · Consider data sharing both during and after research. · The openness and sharing of research data promotes its reuse. · When sharing your data, it is recommended that it be made available for reuse e.g. under Creative Commons or another relevant license. The recommended CC license for open science is the CC-By license. · Use persistent identifiers (PID) to enable access to the data via a persistent link (e.g. DOI, URN).
5.2 How and where will data with long-term value be made available?	The aim of long-term preservation is to store and keep data usable and comprehensible for dozens or even hundreds of years. Data selected for long-term preservation will be submitted to a data repository or data archive. Long-term preservation will ensure your data can be found, understood, accessed and used in the future, even for generations. Tips for best practices · Briefly describe what data to preserve and for how long – as well as what data to dispose of after the project. · Remember to check funder, disciplinary or national recommendations for data repositories, data archives or data banks. · Use persistent identifiers (PID) to enable access to the data via a persistent link (e.g. DOI, URN, Handle).
5.3 Have you estimated costs in time and effort for preparing the data for preservation and sharing?	Tips for best practices · Remember to remark that you will specify your data management costs in the budget. · Will you need to hire expert help to manage, preserve and share the data? · Do you have sufficient storage space, or will you need to include charges for additional services? Consider the additional computational facilities and resources that need to be accessed, and what the costs associated will amount to. · How will responsibilities for data management and costs be split across partner sites in collaborative research projects?