Thesis data management plan

A thesis data management plan is an important part of a thesis plan.
The purpose of a data management plan is to ensure that good scientific practices are respected throughout the thesis process, and that the data will never be in danger during the process.
Data management plan includes a description on how the data is collected, used, stored, archived and how it could be used in future. A data management plan is an addition to your thesis plan, so try to avoid overlaps. Attach the data management plan to your thesis plan. You will need a data management plan also in case you apply for a research permit. This form will remind you at an early stage of your thesis process how ethical rules and data protection legislation affect writing and publishing, and what are the specific, separately agreed, rules for open access to the data. A good plan and thorough preparation will prevent unpleasant surprises and additional work at a later stage.
You also must specify and report the ethical issues and data collection details in the thesis.

Data Management Plan: one part of the thesis plan

Title of the thesis:

Authors:

1.What kind of data are you collecting or reusing in your thesis?

Briefly describe what type of data you will collect, or what existing data you plan to use. How will you collect the data? What type of data (for example a survey, an interview, an observation, picture/audio material, mechanical measuring results, modelling, samples)? If you reuse existing data, you avoid duplications and might gather a bigger data base. Follow the data access rules and conditions, mention the origin, authors, and sources according to good research methods and legislation.

Data collection

- collection dates, beginning and end
- data collection methods
- data collected in several phases, for instance an initial and final survey
- methods used to reach your data subjects
- author's details
- data storage methods
- lists of the existing sources, for example articles and register data

2. How do you document the data?

List the terminology, variables, codes, and abbreviations. How do you organize the data during the thesis process? For instance, what heading practices, version management and folder structure would you use?

A dataset could include one or more files. Enter:

- file name,
- location,
- file size,
- program used when creating the file,
- how to protect the file,
- user/access rights.

3. How do you ensure the quality of data?

How can you be sure that your data files are ethical, faultless, authentic, not falsified, internally consistent, complete, up-to-date, correct, and understandable? The following may impact on data quality: data input, technical details of audio and video recordings, naming and organizing the datasets, missing data, variables for quantitative data, transcriptions of qualitative data. Ensure the quality by documenting carefully how you collect and process the data.

4. Does your data contain personal details?

Is it necessary to process personal data to address the research problem?

Personal data is related to natural persons. When you describe someone's characteristics or living conditions, this information could be linked to a person, their families or other people living with them. Personal details mean such information that would disclosure the identity directly or indirectly.

The scope of processing personal data must always be based on the legislation. You can only process data that is essential for the scope. In principle, processing personal data is forbidden. If you process personal details, you are a holder of a personal data register.

Before you can process any personal data, you must plan how to collect, store, process, erase and destroy it and if you have the right to transfer the data to someone else. This information must be included in the register description. You can find a register format in Finnish at **https://web.samk.fi/public/viestinta/lomakkeet/ tutkimuksen_kehittamistyon_opinnaytetyon_tietosuojailmoitus.pdf.**

Informing research subjects about processing personal data is an essential part of your thesis and data management. This obligation is based on Personal Data Act. Find more information at **https://www.fsd.tuni.fi/en/services/data-management-guidelines/informing-research-participants/**

5. How do you store the data and how do you backup your data?

Make sure that your data is stored and backed up properly during and after processing. Remember that it is safer to use a protected cloud storage than a hard drive. Bear in mind that it is forbidden to store personal data or other confidential data in cloud services. Unnecessary personal data should not be collected. The data must be protected so that outsiders cannot reach it. Data backup is part of data security. Provide your application to use Satakunta UAS cloud storage to your thesis supervisor, who will contact the data management center. You will determine yourself who can access the data. If you are dealing with personal or confidential data, you must inform the research subjects about who will process or see the data before collecting the data from them.

6. How do you ensure that outsiders/external actors cannot access the data?
When you process data, you must ensure data security. Protect confidential and personal data, keep in mind data security and data protection rules.

7. Are there any research ethical issues in your thesis or datasets? How will you deal with them?
If you collect and process personal data, respect research ethical principles, data processing and data protection legislation. Be sure to protect privacy by: a) Considering each identifier or confidential or ethically questionable item separately in the light of whether it is really necessary for your research, b) use codes such as pseudonyms at the earliest possible stage, c) when the research is done, erase identifiers definitively (anonymization), d) destroy the data or save the anonymized data in a long term storage, e) make sure all results and publications are anonymized. More information at **https://www.fsd.tuni.fi/en/services/data-management-guidelines/**

Identification data forms a risk for research subjects in case some data is leaked. You are not allowed to transfer any data without ensuring data security. You must store the data carefully, for instance you should not copy the data, and if you do so, you must be able to destroy the copies according to the data security rules. Recordings are especially sensible and vulnerable material because the research subject could be directly identified. You must justify very well in your research paper if you need to store the recordings for long term and tell this to research subjects. Personal data is collected only if it is necessary for the thesis. You must present a planned, specific, scientific purpose for personal data collection.

8. Keep in mind intellectual property rights and copyright

Define the data ownership. If there are other partners in your project make sure that the data ownership and data access rights are clear. In case you are using data that belongs to a third party, check the usage rights.

You must clarify in the data management plan who else has the right to process the data and who else owns the data, in addition to the author. The thesis with its annexes is a complete work. Conclude an agreement on ownership rights and their transfer, for instance to the sponsor, when you sign a thesis contract. In principle, it is recommended that the usage rights are transferred to the student in which case the student has the right to use the results. Conclude an agreement on the coverage of access rights.

9. Could the collected and processed data be useful later?

Talk with your thesis supervisor whether to store the data for further use after completing the thesis, or whether you should destroy the data. When you plan to give open access to a dataset, think if it is reliable and of quality and could serve other users. Are there any reasons that would prevent open access, such as privacy or data protection rules, commercial use, protection of commercial rights? Do the financial conditions or other agreements restrict open access? Could the data be used for profit? Have you already protected the rights? Are you sure the data has been successfully rendered anonymous? If not, do not give open access to the data.

You can use Creative Commons licenses. Before publishing your data with the licenses, make sure you have the right to do so. Did you agree with a certain research team or companies that you would not publish the data? Does your data contain confidential information?

10. Check the data's retrievability, referencing and access rights

Long term storage and open access enable the data to be found again and used again. Select the best way to store the data with your thesis supervisor. Metadata must contain the data ownership and legal restrictions for data usage according to general standard licenses. Open Science and Research project recommends Creative Commons 4.0 (CC BY 4.0) license for datasets and results, and CC0 license for metadata. More information:
https://creativecommons.org/share-your-work/cclicenses/

11. If Open access is not possible, destroy the data in accordance with the archiving, data destruction and protection rules of your UAS immediately after it is no longer necessary to store the data to complete your thesis/ research or to check the regularity of results.