Appendix 4. Data management in theses – what needs to be taken into account?

1. What kind of data is collected or reused in the thesis?

Think about how the data will be collected and in what format it will be (e.g. survey, interview, observation, picture/sound, automatic measurement result, modelling, physical sample, etc.). If you reuse data collected by someone else, you will avoid double work. Observe then the terms of use of the data and mention, the origin, authors and sources in accordance with good research practice.

2. How is the data described?

The description and systematic storage of data is important so that the data will be intelligible and easy to find in the future, too. The description of data includes, among other things, the terms used, the names of the variables, the codes, and the abbreviations. Also think of how the data will be organised during the study, e.g., by describing the folder structure, as well as the naming of files and version control.

3. How is the quality of the data assured?

Reliable and high-quality data are ethical, flawless, internally consistent, comprehensive, up-to-date, and intelligible. In the case of quantitative data, it is important to make sure that the values of the variables are within permitted limits and that there are no missing or undefined values in the data.

4. How is the data stored and backed up?

Make sure that the data is safely stored during the thesis process and after it, and that the necessary backup copies are made of it. Note that the use of a protected network disk is a safer option than that if a hard disk or USB memory.

5. How is the data protection of the data assured and how are outside users prevented from accessing the data?

During the thesis process, it is important to make sure that the data is protected. Confidential and classified data must be safeguarded, and data protection must be deployed. Consider if the collection of identifiers, classified or ethically suspicious data is essential for your study. Remove the identification information or anonymize the data as soon as it is reasonably analysable without identifiers. A commonly used way of anonymizing is pseudonymization, or replacing real names by pseudonyms. Remember that it is prohibited to save personal or other confidential or classified data in the cloud.

6. Does your data include personal data?

Personal data refers to data of a natural person, their properties or living conditions that can be recognized as referring to the person in question, their family, or the people living in the common household with them. The definition of personal data covers personal data allowing the direct recognition of the person, but also data allowing the indirect recognition of the person in question. The processing of personal data always requires a basis for their processing, as stipulated in the Personal Data Act (523/1999). If you collect personal data, you create a personal data register,
which requires the writing of a register data description. In such a case, you need to plan beforehand the collection, storage, processing, possible handover, removal, and deletion of the personal data and to describe them in the register data description. Models for register data descriptions can be found on the website of the Office of the Data Protection Ombudsman at http://www.tietosuoja.fi/fi/index/materiaalia/lomakkeet/rekisteri-jatietosuojaselosteet.html.

7. How do you take the research ethical matters into account?

In all research, the instruction titled Responsible conduct of research and procedures for handling allegations of misconduct in Finland (2012), by the Finnish National Board on Research Integrity (TENK), must be observed. When collecting and processing data related to a person, also the ethical instructions for research in humanistic, social and behavioural sciences, by TENK, must be observed (Eettinen ennakkoarviointi ihmistieteissä, [ref. 28 June 2017]), as well as the legislation on the use and processing of personal data.

8. How are the immaterial rights and copyright taken into account?

Define who owns the data. If you cooperate with a company, make sure that the agreement covers these matters. If you use ready-made data collected by someone else, make sure you have a permission to do that. Also remember to refer to the original author of the data.

9. Where will the data be stored after the thesis has been completed and can others exploit it later on?

Think about where the data will be stored after your thesis has been completed or whether it will be destroyed. This information is also needed when the targets of the study are informed. If you consider your data to be useful for other people, too, you can have it stored at SeAMK’s databank or a public data archive. Also note that, based on Section 14 of the Personal Data Act, a personal register must be destroyed or archived, or its data must be converted into such a form that the person to whom the data relates cannot be recognized when the personal data are no longer necessary for the implementation of the study or for ensuring the appropriateness of its results.

For further information about data management:
Research Manager Seliina Päälysaho (seliina.paallysaho@seamk.fi)
Information Specialist Jaana Latvanen (jaana.latvanen@seamk.fi)