# Data management guidelines for students

## Describe your data

* Briefly describe, what kind of data you will collect or produce, and what kind of existing data you will use?
* How much storage space do you need for your data?
* Make sure that the data you collect is of sufficient quality for your research. Collect, for example, a small test data before the actual data collection: conduct a test interview, test the survey questionnaire...
* If you use already existing data, please familiarise yourself with it carefully and assess whether it will enable you to answer your research questions.

Answer the questions and read the following guidance:

Research data refers to the data you use in your thesis and which your thesis is based on. The data can be collected by yourself and it includes, for example, surveys, interviews, measurement results, field notes... In addition to these, "already existing data" can be used as a research data. This can include, for example, archive sources, audio recordings, YouTube videos, photos, literature as a research subject, movies, websites, forum threads, medical imaging, simulations, etc. You may also use research data obtained from your supervisor or data archive. You can search for survey and interview data, which you can use in your own [thesis,](https://www.fsd.tuni.fi/" \t "_blank) for example in the Data Archive.

In addition, your data may also include classifications, categorisations, tables, databases, visualizations, notes, etc. based on the data listed above. In this context, research data do not mean source literature or research publications.

## What IPR, licenses or other rights are related to your data?

Answer the question and read the following guidance:

Maria Rehbinder from Aalto University describes intellectual property rights and data protection issues related to research data in the [YouTube video.](https://www.youtube.com/watch?v=DtqSEMx-IDI) In particular, observe any copyrights or terms of use related to the data you have received elsewhere. Remember also that not all data available online may be used as such in your own work. Read the terms and conditions of the service provider to determine whether research use of the data is permitted.

If your data contains [images](https://libguides.aalto.fi/imagoa_eng), texts or other works, they are likely to also involve copyrights. Especially, if you upload images or texts to your computer, you should make sure that you are entitled to do so. Sometimes the terms and conditions for images and texts are [defined in licenses](https://libguides.tuni.fi/researchdatamanagement/cc). If you use social media messages as your data, make sure that you are allowed to copy messages to your computer if necessary.

If you receive the data from your supervisor, it is advisable to agree on who is allowed to use the data and what to do with the data after the thesis has been completed in order to avoid any ambiguity. What if you cannot finish the thesis? Similarly, if you are doing your thesis in cooperation with a company. Will the data remain at the company's disposal, will you be allowed to use the material later, or will the material have to be disposed of?

## Do process personal information or personal identifiers?

Answer the question and read the following guidance:

Personal data is any information that makes a person identifiable, either directly or indirectly.For example, when conducting an interview, you will almost certainly process personal data, as the voice will be considered as identifiable information. In addition, the participation consents collected from the research subjects are personal data. So is the contact information of the research participants.

Answers to the survey questionnaire are easily identifiable. Background information, such as occupation, city of residence, age and gender can make a person recognisable when these information are combined. Note that there can be identifiable information in file or folder names also.

Use quotes or citations carefully, especially if you use social media data or data collected from forums, for example. Direct quotations are likely to constitute identifiable information. It may also come as a surprise to the user of the service that quotes are extracted from their messages and presented in a research publication or thesis, taken out of the original context. Marko Ahteensuu discusses the ethical problems of social media data in his article published in vastuullinentiede.fi [website](https://vastuullinentiede.fi/en/doing-research/do-you-use-social-media-data-your-research" \t "_blank).

Read more about the definition of personal data in [FSD’s Data Management Guidelines](https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymisation-and-identifiers/).

#### Data protection

Data protection is the process of protecting personal data. Data protection is a fundamental right and safeguards the rights and freedoms of data subjects when their personal data is processed. Data processing laws set out the principles for the lawful processing of personal data. The processing of personal data must always be based on the law.

#### Risk assessment

A basic risk assessment must always be carried out before the processing of personal data. Risks are assessed from the perspective of the research participant. The risk assessment shall be documented.

A data protection impact assessment (DPIA) must be carried out if the risk assessment indicates that the processing of personal data poses a high risk to the participants. See [model template and guidelines](https://www.tuni.fi/en/research/responsible-research/data-protection" \l "expander-trigger--52f291f7-6b41-4f8d-ae47-a6b51f562a97) for risk assessment and impact assessment.

#### Informing research participants

Informing the research participants about your research and requesting ethical consent (informed consent) is part of good scientific practices. Also, participants should be informed about the processing of their personal information as required by [the Data Protection Act (1050/2018).](https://www.finlex.fi/en/laki/kaannokset/2018/20181050" \t "_blank)

#### Privacy Notice

The purpose of the Privacy Notice is to inform the research participants about the processing of their personal data.

* You can find the [model template and instructions for the Privacy Notice](https://www.tuni.fi/en/research/responsible-research/data-protection" \l "expander-trigger--15950926-4420-44b7-9696-0dcf562d02c0) in data protection path of research
* See also [the privacy package for thesis authors and supervisors](https://intra.tuni.fi/content/news/23028" \t "_blank) (only in Finnish)
* [Instructions for students concerning data protection](https://intra.tuni.fi/en/handbook?page=2771) (intra)

#### Information sheet

[Information sheet](https://www.tuni.fi/en/research/responsible-research/data-protection" \l "expander-trigger--15950926-4420-44b7-9696-0dcf562d02c0) will provide the research participants with basic information about the project. It is essential that it clearly indicates what the participant will agree to.

#### Informed consent

Prior to participation in the study, the research [participants should give you an ethical consent](https://www.tuni.fi/fi/tutkimus/vastuullinen-tiede/tutkimuksen-tietosuoja" \l "expander-trigger--a366303a-e4cb-40dd-ab2c-610afc0c439c" \t "_blank). In human sciences, this can be a signed consent form, an email confirmation or, oral consent given at the beginning of an interview message, for example. If the investigation interferes with physical integrity or processes [sensitive personal data,](https://tietosuoja.fi/erityisten-henkilotietoryhmien-kasittely" \t "_blank) we recommend a signed consent form. Please note that ethical consent is not the same as [consent as a basis for processing personal data.](https://tietosuoja.fi/en/consent-of-the-data-subject)

#### Further use of data

In general, the research participants are asked to participate only for a limited research project. The use of the data is only allowed for research covered by the demarcation. If the data is to be archived and used for other research purposes, the [participants must be informed of this](https://www.fsd.tuni.fi/en/services/data-management-guidelines/informing-research-participants/" \l "informing-research-participants-about-archiving" \t "_blank).

#### Research permits

If your thesis deals with employees or students in an organization, you probably need a research permit from that organization. For example, cities and municipalities typically have their own practices for research permits. Learn more about research permits from [the University's intra](https://intra.tuni.fi/handbook?page=2288" \t "_blank).

**How do I do this in practice?**

* Before collecting the data, plan how to implement the informing of the research participants (information sheet and privacy notice) and how to ask a consent for a participation. For example, you can attach your information sheet to an e-mail message, as part of a questionnaire, or submit it as a paper form. The style is free as long as the research participants receive the relevant information about your project.
* You can also request a consent in different ways. For example, in interviews, consent can be requested orally at the beginning of the interview. In the survey, the user can be asked to tick the check box that they have read the information sheet.
* Avoid processing direct identifiers, such as names, for example by pseudonymizing the data. Remove direct identifiers or replace names with codes.
* Do not use identifying information in file names either.

## Documentation and metadata of the data

**Describe the collection and processing of data.**

* What do the folders and data files and contain?
* When and where was the data collected?
* Why was the data collected?
* How has the data been processed or modified?
* In tabular data, describe the column and row labels, data encodings (including data from missing data), units of measure, etc.
* Describe the abbreviations you are using.

Answer the questions and read the following guidance:

The data, as well as its collection and processing, should be described with the accuracy that someone other than you understand what the data is about. It is advisable to describe the data so accurately that if for any reason you had to stop doing your thesis for a year or two, you could still reasonably safely continue your thesis on where you had left off.

File naming and folder structure

* Pay attention to the names and order of your files and folders.
* Smart file names are short, informative, and consistent. For example, mark dates as standard YYYYMMDD.
* Use version numbering in file names and keep the original file separate from editing files.
* Already one interview easily generates three files: the original audio file, the text transcription file, and the interview notes.

## Where do I store my data?

Answer the question and read the following guidance:

* Online survey tools: [O365 Forms ja LimeSurvey](https://intra.tuni.fi/en/handbook/2687/2731/12248) (intra)
* [Data security of voice recorders and cameras](https://intra.tuni.fi/en/handbook/2687/2731/16353) (intra)
* [Zoom](https://intra.tuni.fi/en/handbook/2677/2743/2824?page=14364" \t "_blank) and [Teams](https://intra.tuni.fi/en/handbook/2677/2743/2824?page=2732" \t "_blank) (intra) are compatible with the Data Protection Act and the EU General Data Protection Regulation (GDPR) [according to the intranet's](https://intra.tuni.fi/en/handbook?page=20847" \t "_blank)  instructions when logging in with TUNI id. The recording is stored on the interviewer's computer.
* [Security on your own home computer](https://intra.tuni.fi/en/handbook?page=21924) (intra).
* External hard drives: Protect external hard drives at least with a password.
* Use university [storage solutions](https://www.tuni.fi/it-services/handbook/2691/18895?) (intra), such as OneDrive for Business, Tuni Groups, and a personal network drive.
* Commercial cloud services for consumer use are not always suitable for the processing of confidential information, such as personal data.

## What do I do with my data after the thesis is completed?

Answer the question and read the following guidance:

Data containing personal information is usually destroyed after the thesis has been completed and approved. Inform the research participants if you intend to use the data after your thesis, for example in a postgraduate degree, or if you write an article based on your thesis. It may be necessary to keep the data for the purpose of verifying the results. If you work in a company or other organisation, consider the potential interests of the client of your thesis in the data.

In order to share the data for further use, you need a permission from your research participants. The sharing of the data is done in archives, such as the [Finnish Social Science Data Archive](https://www.fsd.tuni.fi/en/) (FSD) or [Zenodo](https://zenodo.org/). In order to archive your data in FSD, the data should be sufficiently extensive for the needs of comparative research, methodological education or new research. Typically, data collected for master thesis do not meet these criteria. Read more [about the materials received by the Data Archive](https://www.fsd.tuni.fi/fi/palvelut/arkistointi/ohjeita-aineiston-arkistoijalle/" \t "_blank).

The EU-funded [Zenodo](https://zenodo.org/) suitable for storing many types of data. If necessary, the visibility of the data can be set so that only you can access it, but it is also possible to share your data through Zenodo.

Anticipate the further use or use of data even before the data collection. For example, contact the Data Archive and ask about the possibilities of sharing. It is essential to inform research participants about the further use of the data. Access to the university's storage solutions ends after you have graduated, so take care of the data if you do not destroy it.