

Research data management for thesis supervisors

29.10.2020

Katja Fält, Kaisa Kylmälä

Tampere University library

Open Science Services (Research Support team)



Today

- What is research data
- What is research data management
- Personal data
- Research data management
 - Planning
 - During the project
 - After the project
- Jari Yli-Hietanen: Storing and processing data

What is research data?

- Research data is any information that has been collected, observed, generated or created to validate original research findings.
 - research data also includes non-digital formats such as laboratory notebooks and diaries
 - It can be also biological or physical
- Metadata (information describing the content, structure, collection procedures, etc. of the data) is also part of research data

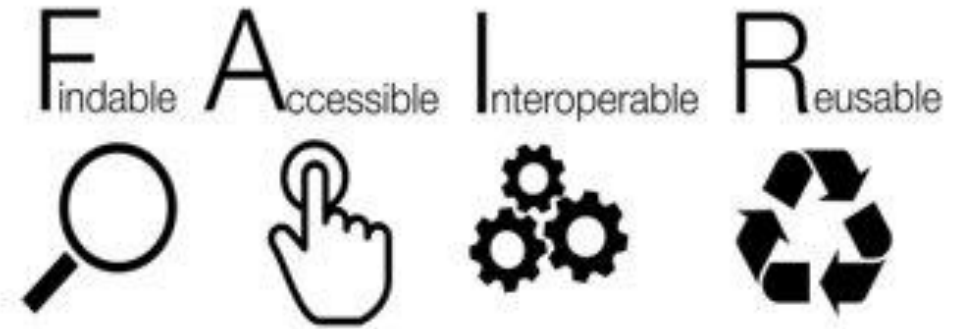
What is research data?

- Data collected or produced by you or your research group
 - Interviews, surveys, observation diaries, drawings, photos etc.
- Data collected by other researchers
 - Can be obtained directly from researchers who have collected the data
 - From data archives or repositories such as [FSD](#), [Language Bank of Finland](#), [Zenodo](#)...
- Data collected for other purposes, such as registry data, archive materials, government materials
- New data which is created during the research

What is Research data management?

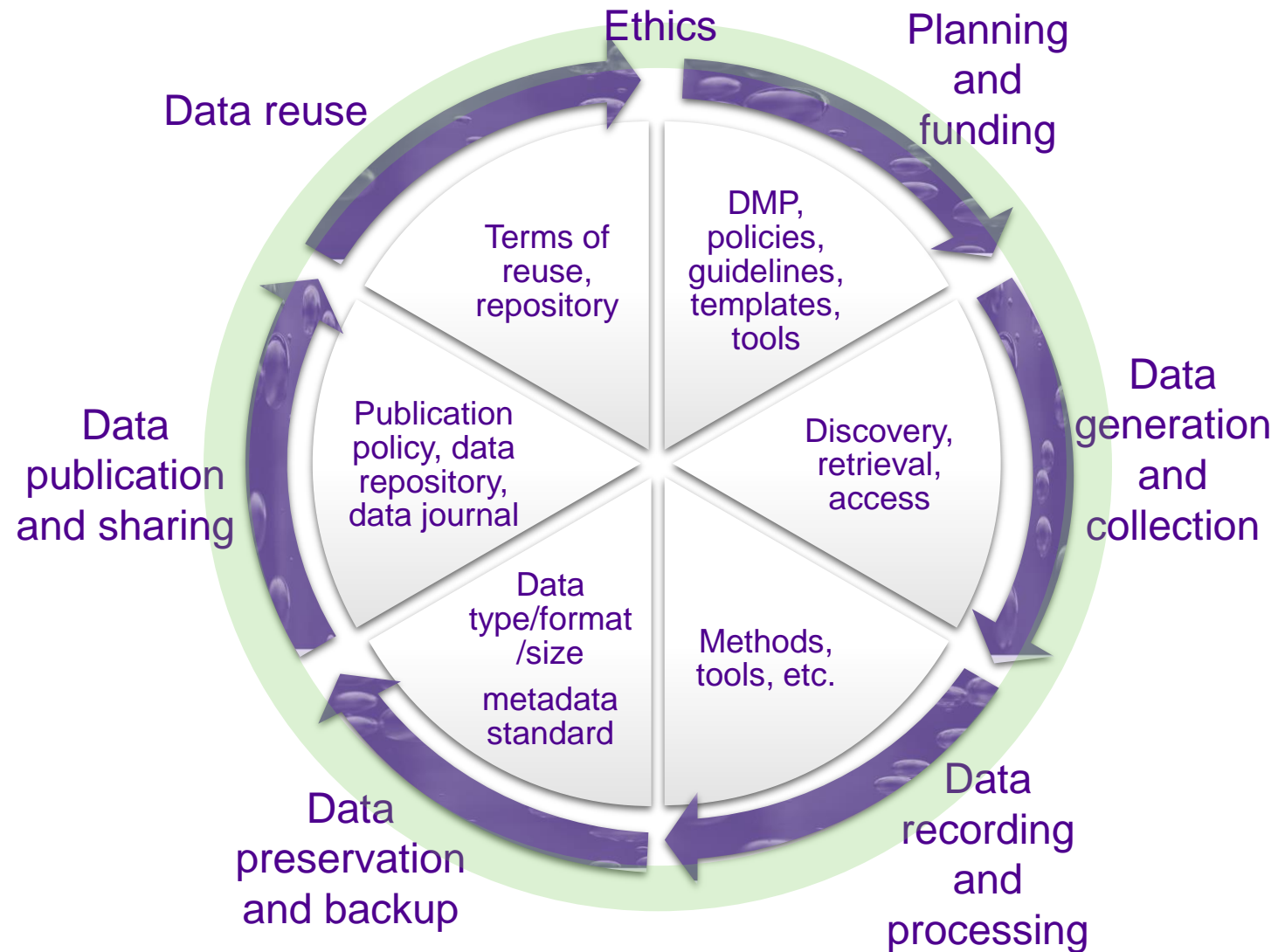
- Research data management is a part of good scientific practice
 - Secures responsible and ethical handling of the data throughout the project
 - Helps to validate research results
 - Transparency, reliability, replicability and verifiability of research
- Saves time and eases the workflow
- Organising, preserving and finding data becomes easier
- Enables the opening and sharing of data
 - Visibility and findability of the data and the researcher (data citation, merit)
 - Reduces overlapping work
 - Makes reusing the data easier
- Data management plan (DMP) helps in data management: a living document where you describe your data management processes
 - Read more: <https://libguides.tuni.fi/researchdatamanagement/plan>

FAIR principles



- The principles emphasise the reuse of data and good data management
- Information, publications and data, metadata and methods used and produced in the project are easily available for reuse
- The aim is to maximise the reuse of research data
- IT perspective: Data and metadata need to be machine-actionable and machine-readable
- Data can be FAIR even though there are limitations in its accessibility
- The principles make sure that data can be found, understood and reused
- Read more: <https://www.fairdata.fi/en/why-fairdata/fair-principles/>

RESEARCH DATA LIFECYCLE



Responsible and open science

- The methods applied for **data acquisition** as well as for research and evaluation, conform to scientific criteria and are ethically sustainable.
- When publishing the research results, the results are communicated in **an open and responsible fashion** that is intrinsic to the dissemination of scientific knowledge.
- The researcher complies with the standards set for scientific knowledge in planning and conducting the research, in reporting the research results and in **recording the data obtained during the research**.
- Research ethics is an essential part of researcher skills and ethical questions should be considered throughout the project
- Open science as part of good scientific practice and researcher skills
- Tampere University: Responsible research: <https://www.tuni.fi/en/research/responsible-research>
- Responsible research: <https://vastuullinentiede.fi/en>
 - A guide to the best practices of research ethics and science communication
 - combines the best practices of research integrity and science communication with the wider framework of open science culture and responsible assessment of research outputs

"as open as possible, as closed as necessary"

Consider in thesis

What are the ethical issues connected to the data (for example)

- Collecting and processing sensitive information
- Protecting the identities of the research participants
- Consent for sharing information

Ownership, copyright and immaterial property rights (IPR)?

- Do ownership, user rights, or other (for example trade secrets, patents) limit the use, reuse and sharing of research data?
- Should you make agreements concerning the data?

Personal data



What is personal data?

- Personal data refers to *all information relating to an identified or identifiable natural person*. Natural persons are considered identifiable, if they can be identified, *directly* or *indirectly*, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, an opinion, a job title, image or audio, or one or more factors specific to their physical, physiological, genetic, mental, economic, cultural or social identity.

EU's General Data Protection regulation

Personal information

Includes all identifiers from which the person is identifiable *directly* or *indirectly*.

Direct identifiers

- Name
- phone number
- personal identity code
- Picture
- Voice
- Fingerprint
- dental chart.

Indirect identifiers

- Gender
- Age
- Education
- Profession
- Nationality
- Work history
- system log history
- marital status
- residence information
- IP address,
- car license number.

Remember that spatial data can also contain enough information for identification of individuals.

Special categories of personal data

Some personal data is sensitive

- Racial or ethnic origin
- Political opinion
- Religion or beliefs
- Trade union membership
- Genetic data, or biometric data processed for the purpose of uniquely identifying a person
- Health information
- Sexual behaviour or orientation
- Criminal convictions and offences

Identify and describe situations when collected personal data are [sensitive](#) as well as the legal basis for processing sensitive personal data. If sensitive personal data are processed in the research, the processing must be based on [Article 9](#) of the General Data Protection Regulation (GDPR).

Principles of processing

- The scope and applicability of the principles must be evaluated on a case-by-case basis *already when you are planning activities that involve the processing of personal data* (e.g. in a research project)
- Careful planning will prevent problems later on.
- Risks must be assessed from *data subjects' perspective*.
- The plan presents the main risks of managing data and how to manage risks.
- Compliance must be demonstrated -> the importance of proper documentation
- **Data processing plan is incorporated into the Data Management Plan and/or the research plan.**
- Do you need to transfer personal data from EU to non EU country, or vice versa?
- If you transfer personal data from non-EU country to EU, you must comply with GDPR.

Privacy notice

- When personal data is processed a register is formed
 - This means that the participants (informants) must be informed
 - For what purpose are their data being collected
 - How their personal data are being collected, processed, used, stored, disseminated and made available
- A privacy notice is compiled in order to inform the informant
 - The form is given to the informant together with an information sheet
- Further information
 - [Informing research participants](#) (FSD's Data Management Guidelines)
 - [Privacy notice](#)
 - [Data protection path of research](#)
 - Research Data Services: [training materials](#) (slides, session recordings)

Data controller

- *Data controller* is the individual, company, public authority or community that determines the purposes and means of the processing of personal data.
- The data controller is responsible for compliance with data protection laws throughout the data lifecycle.
- The role of data controller is assigned to impose the [responsibility for GDPR compliance](#) on the party that has the power to influence processing activities.
- The starting point of data protection policy
 - The student is the data controller of the personal data collected for thesis, unless otherwise agreed on
- Data controller is:
 1. The thesis is compiled as own, individual work --> the student
 2. The thesis is connected to a project, no contract --> case by case
 3. The thesis is being done within a project --> contract --> the university
 4. The thesis is a commission --> the commissioner (or the student)

Handling sensitive data

- Collect the data for specified, explicit and legitimate purposes
- The collection of sensitive data must be based on the voluntary, specific, informed and explicit consent provided by a data subject
- Minimise the amount of personal data (do not collect excess amount of personal data or data you do not need)
- Minimise the time you store the data
- Protect the privacy of the informant by anonymising the data

Principles in practice?

- When planning (and processing), consider at least the following questions:

1. **Do I need to collect personal data?**
2. **Why** is personal data collected and for **what** (lawful) **purpose(s)**?
3. What is the **minimum amount** of personal data that I need?
4. How do I collect the data?
5. Do I need to ask for consent? How do I inform the subjects?
6. How do I store and maintain the data securely?
7. How long do I need to store personal data (in an identifiable form)?
8. How and when do I destroy or archive the data?

Planning



Agreements and research data

- Copyright and user rights of research data
 - Make agreements if you are using somebody else's data (ownership, user rights)
 - Make agreements if your thesis has a commissioner (principal)
 - Read more about agreements and copyright from [Data Management guidelines](#) (FSD)
- Agreements in a project
 - To determine the tasks, responsibilities and rights of all project participants
 - To create mutual understanding in a project
 - To demonstrate responsibility and respect (for informants)
 - To clarify administrative roles within the project
 - To clarify rights to the data (authorship)
 - To avoid disputes and conflict of interest
 - To show responsible conduct of research and good data management skills

Permission to carry out research

- If your research or thesis topic concerns the activities of an organisation, a company or the public sector (such as a city or municipality) or you are planning to interview their staff, you must apply for research permission from the company, city or other organisation
- When do I need a permission from the University?
 - Research permissions are generally granted for scientific studies, theses or development projects that are related to the operations and development of Tampere University. If your research concerns University-level activities, the Vice President may grant you the research permission
 - Deans grant the permission for studies and thesis projects that concern an individual faculty
 - If you need to access to personal data stored in the University's databases for research purposes, you must submit a special form to request access to the data
 - Read more: <https://www.tuni.fi/en/research/procedure-managing-research-permissions-tampere-university>

Remember these

- Assess the need for personal data already in the planning phase
- Inform the research participants
 - Where and how: face to face, by email, in the beginning of the questionnaire
 - Read more: <https://www.fsd.tuni.fi/en/services/data-management-guidelines/informing-research-participants/>
 - Information sheet: the topic of the research, what the research is about
 - [Consent to participate in the research](#) and the permission for data reuse
 - Privacy notice: how personal data is being processed
 - Read more: <https://intra.tuni.fi/en/handbook?page=13212>
- Always remember data protection and data security at every step!

During the research



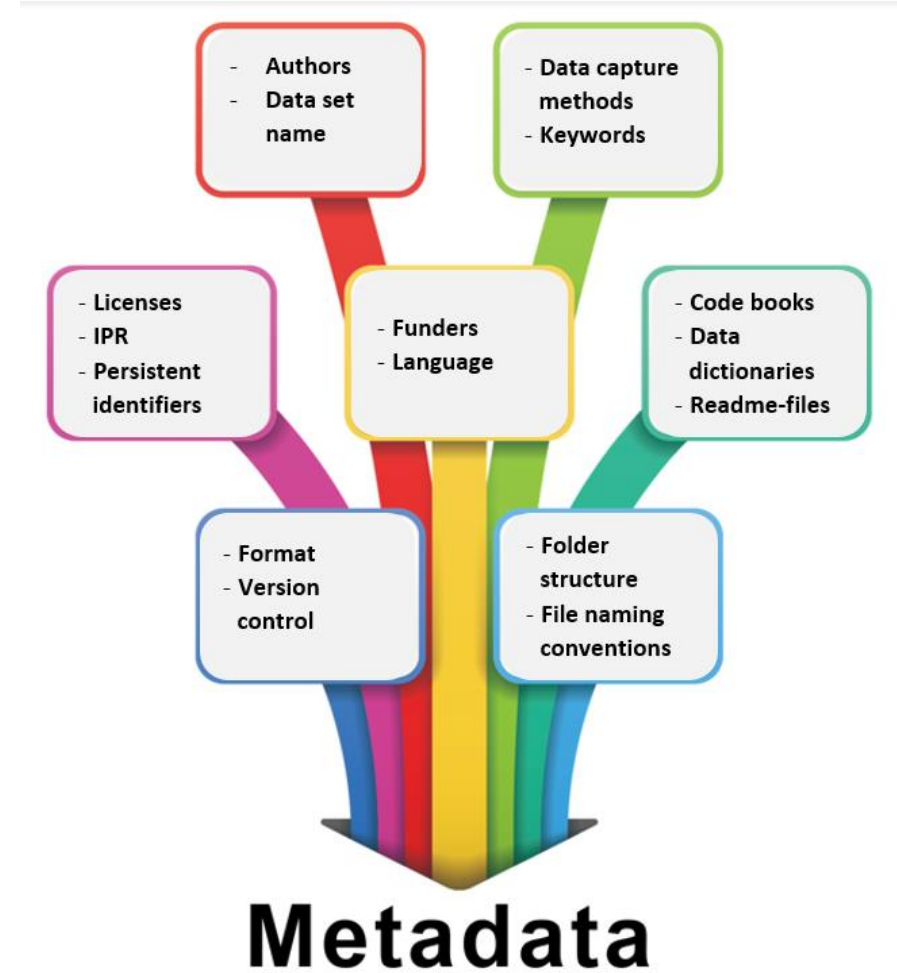
Why should you document your data?

- Carefully described and documented content, data collection procedures and variables of research data is essential to ensure the usability of data
 - Helps you and others understand your data
 - Easier to report your research findings
 - Helps other researchers understand how the data has been collected and managed
 - Makes your research more reliable
 - Makes your research data easier to find and reuse
 - [FAIR-principles](#)
 - Enables opening your data publicly available

What to describe?

- For what purposes was the data created?
 - Background information of your project
- What does the dataset contain?
 - Interviews, questionnaires, pictures...
- How was data collected?
- Who collected the data and when?
- How was the data processed?
- What possible manipulations were done to the data?
 - Is the data anonymised?
 - Are there new versions?
- What were the quality assurance procedures?
- How can the data be accessed?

Source: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>



Kuva 2. Tutkimusaineiston metadata

Published metadata

- [Qvain](#) – Research Dataset Description Tool > Create and publish your metadata
- Example of public metadata in The Finnish Social Science Data Archive (FSD) :

FSD3013 Vitality 90+ Survey 2014

[Overview](#) [Detailed description](#) [Variables](#) [Publications](#) [Download data](#)

Authors

- Jylhä, Marja (University of Tampere. School of Health Sciences)

Keywords

ageing, care of dependants, care of the elderly, elderly, health, home help, old age, physical mobility, residential care, social interaction

Abstract

The survey studied longevity and the oldest-old by charting the care, everyday life, and physical activity and capability of people aged 90 and over living in Tampere.

The respondents who lived at home were asked what kind of housing they lived in (e.g. ordinary home, sheltered housing), who they lived with, whether someone helped them at home, who helped them the most with everyday tasks, and whether a housekeeper or home helper visited them regularly.

The dataset is (D) available **only by permission from the data depositor/creator.**

[Download the data](#)

Study description in other languages

- [in Finnish](#)

Related files

- No other files available

When the thesis is ready – what about then?

- What was the plan? Which matters did you ask a permission from informants? What did you promise for informants?
- Preserve your data for futher use
 - Do you have permission to preserve your data?
 - Postgraduate degree – your interest
 - Principals interest
 - Verifying of your research
- Publish your data
 - Do you have permission to publish your data?
 - Repositories, archives – [Finnish Social Science Data Archive](#) (FSD)
 - Plan publishing before you collect your data
 - Contact archives / repositories in the planning pahase

Good to know

- Guides and instructions
 - [Research Data Management -guide](#) (Tampere University Library)
 - [Data Management Guidelines](#) (Finnish Social Science Data Archive)
 - [Data protection path of research](#) (Tampere University)
 - [Quick guide to information security](#) (TAU intra)
- Tampere Higher Education Community's policies
 - [Open Science and Research Policy](#)
 - [Data Protection Policy](#) (TAU intra)
 - [Information Security Policy](#)

Help available



Research Data Services comprehends

- Library
- IT-services
- Research services
- Record management
- Legal services
- Data protection office
- Finnish Social Science Data Archive (FSD)

- We organize research data management trainings.
- We provide instructions and resources about data management.
- We comment data management plans.

- researchdata@tuni.fi