Data management in
a cooperation project

Jari Friman & Saila Huuskonen
Tampere University Library, Research Data Services
Jari Yli-Hietanen
Anna Rytivaara
8.6.2021
In this presentation

1. Plan research data management
2. Personal information in research
3. Agreements and authorship in research data
4. Documentation and metadata
5. Storage and backup
6. Reusing data
Plan research data management
Have you written a data management plan?
Data management plan (DMP)

- DMP
  - describe how you will manage data during the research life cycle
  - is a part of research plan
  - should be made before the actual research process starts.
  - should be updated during the research project.

- DMP can be made
  - as part of funding application,
  - as part of ethical review or
  - to support your own work.

- DMP is a tool for risk evaluation. It helps you to
  - reduce the risks of losing your data
  - anticipate complex ownership and user rights issues in advance
  - control ethical and legal issues related to your data
  - meet the funder's requirements.
RESEARCH DATA LIFECYCLE

- Ethics
- Planning and funding
- Data generation and collection
- Data recording and processing
- Methods, tools, etc.
- Discovery, retrieval, access
- DMP, policies, guidelines, templates, tools
- Terms of reuse, repository
- Publication policy, data repository, data journal
- Data type/format/size, metadata standard
- Data preservation and backup
- Data publication and sharing
- Data reuse
DMP in a cooperation project

• Different data, different needs for
  • storing and storage period
  • data security and data protection
  • access control. Do all parties or team members have similar rights?

• Common practices
  • Do different organisational practices have an effect on, e.g., the quality of the data, data protection or data security?
  • All members of the research group follow mutually agreed practices and are able to anticipate the risks associated with data quality.

• Different roles
  • In consortium projects, the leader is usually responsible for drawing up the plan.
  • How will the data of the partners’ sub-projects be described in the plan?
  • Who is responsible for the implementation of good data management in practice?
  • Who grants access (read and edit rights) to the data?
Data management planning tool

• Use DMPTuuli – tool for writing Data Management Plan.

  • Tampere University DMP guidance,
  • examples of data management plans,
  • share and co-write your plans with colleagues,
  • leave comments,
  • ask feedback from Research Data Services.

• More information: Research Data Management - guide

1. General description of data (0 / 2)
2. Ethical and Legal Compliance (0 / 2)
3. Documentation & metadata (0 / 1)
4. Storage and backup during the research project (0 / 2)
5. Opening, publishing and archiving the data after the research project (0 / 2)
6. Data management responsibilities and resources (0 / 1)
Personal information in research
What counts as personal data

**Direct identifiers**
- Social security number
- Full name
- Audio file
- Email (with name)
- Video file displaying person(s)
- Photograph of person(s)

**Strong indirect identifiers**
- Email
- Phone number
- Date of birth
- Rare occupation
- Vehicle registration number
- Student ID number

**Indirect identifiers**
- District/part of town
- Municipality of residence
- Age
- Household composition
- Occupation
- Education
- Mother tongue
- Nationality
Anonymoyus and pseudonymous data

• **Anonymous data:** An individual data unit (person) cannot be re-identified with reasonable effort based on the data provided or by combining the data with additional data points. Completely anonymous data do not exist, but with well-executed procedures one can achieve a result where individual persons cannot be identified with reasonable effort. **Anonymisation** refers to the various techniques and tools used to achieve anonymity.

• **Pseudonymous data:** An individual data unit cannot be re-identified based on the pseudonymised data without additional, separate information. **Pseudonymisation** refers to the removal or replacement of identifiers with pseudonyms or codes, which are kept separately and protected by technical and organisational measures. The data remain pseudonymous as long as the additional identifying information exists.

→ Pseudonymous data is also personal data

Processing of personal information

• If you collect information from research participants or about research participants, assume that you will collect personal information -> You must follow Data Protection Act.

• Identify and document the risks related to processing of personal data.

• Protect the identity of participants.
  • Remove personal identifiers by anonymizing and pseudonymizing your data.

• Identify the need for an ethical review.
  • Research protocols requiring ethical review.
  • Some research funders and publishers may require an ethical review
  • More information: TENK guidelines, Tampere Region Ethics Committee for Human Sciences
Informing research participants

- Where do you inform – face-to-face, at the beginning of the questionnaire, in email, remotely via Teams or Zoom…?

- **Information sheet**: what the study is about?

- **Ethical consent** and **permission to archive the data for reuse**.

- **Privacy Notice**: how personal data is processed

- Read more: [Data protection path of research](#)
Agreements and authorship in research data
What kind of agreements have you made regarding research?
User rights in research data

• Uncertainty can occur regarding, for example,
  • the rights to use research data collected for the project or
  • how to use data obtained from outside of the project.

• What happens to data if someone leaves the project?

• From an archiving point of view, it is important to define who has the right to decide on the handover of research data for reuse and to decide the terms which apply to such reuse.

• Overall, in a research project, it is essential to clearly determine the tasks, responsibilities and rights of all project participants.
Authorship in research data

• Researchers can agree on their roles and rights between themselves.

• As part of this allocation of tasks and responsibilities, they can also agree on authorship concerning research data and publications made based on the data.

• According to The Finnish Advisory Board on Research Integrity's (TENK) guidelines researchers should agree on principles relating to authorship before starting on research or recruiting researchers.

• A research publication is always a work, but research data are not always protected by copyright.

• Regardless of whether the data are protected by copyright, researchers should at the start of the research come to an agreement amongst themselves on the use and archiving of the data.

• You should always refer to archived research data and the original researcher or researchers using the same principle as when referring to written sources.

When is it a good idea to draw up a separate agreement about data use?

• Often the need for a separate agreement is greater the more extensive the research project is.

• Some of the most significant aspects of the research project may have been agreed on as part of the research plan and data management plan. A separate agreement is needed if the research or data management plan is not sufficiently comprehensive.

• Sometimes project's organisation or a research cooperation agreement has already been drawn up. In such cases, a supplementary agreement can be drawn up if previous agreements or plans are insufficient.
How should the agreement be drawn up?

• The agreement can be drawn up in a free format, such as verbally. An e-mail is also sufficient.

• When drawing up an agreement, it is essential that all parties commit to the agreement voluntarily.
What should the agreement contain?

• The contractual parties and the validity period of the agreement
• The tasks and responsibilities of parties concerning the project
• Rights of use to data in the original research:
  • Researchers' rights of use to data in accordance with the original plan
  • Rights of use to data in situations wherein funding cannot be secured for all researchers for the entire duration of the project and/or a researcher moves to another project
  • Validity period of the right of use (especially thesis completion)
• Specifying the data that the agreement covers
• Naming of authors of the data
• The time period for handover for reuse
• Restrictions set on reuse and their validity
• Can the right to reuse data be transferred onwards to third parties before data are archived?
• Who can make an agreement on behalf of the other researchers concerning the archiving of research data?
• Who will grant permission to use the data if necessary? (one researcher or a separate group)
When defining your data use, consider these.

- The consortium agreements define the responsibilities related to the cooperation project. If you share personal data with your collaborators, agreement on joint controllership may also be needed if it is not included in the consortium agreement.

- The processing of personal data is agreed with the research participants.

- Research permits from target organisations define the collection and use of data.

- The use of register data and official data in the study is agreed with the authority concerned.

- The funding terms and conditions may specify the use of the research data. For example, setting requirements for data sharing.

- Employment contracts.
  - See: Tampere Universities’ [transfer Agreement for contract research](https://intra) (intra)

- Data archives use licenses or other terms of use to define reuse of research data.

- Copyright may restrict the use of research data if data produced (e.g. photographs, texts…) by research participants are used for research.
Documentation and metadata

who
when
what
why
how

How do you document your data?
Why to document your 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.
What to document?

The project-level documentation explains:

• For what purposes was the data created?
• Background information of your project
• What does the dataset contain?
  • Interviews, questionnaires, pictures, for example.
• 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?

Shared understanding about documentation

• It is good to agree on common practices:
  • Background information on data collection
  • Naming files
  • Folder structure
  • Document all the changes/edits in the data
  • Version control
  • Generating public description about data (metadata)
Storage and backup
## Categories of personal data

**Please note!** The classification may change following the combination of data from multiple sources.

<table>
<thead>
<tr>
<th>General data</th>
<th>Personal data</th>
<th>Examples</th>
<th>Description</th>
<th>Security measures</th>
</tr>
</thead>
</table>
| Restricted   | 1 R           | Special category data (sensitive data; the use of data may interfere with an individual's fundamental rights and freedoms)  
- The processing of data relating to racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership or genetic and biometric data  
for the purpose of identifying an individual, and  
the processing of data concerning a person's health, sex life or sexual orientation is prohibited unless there is a lawful basis for the processing activities.  
- HR: medical certificates and health data.  
- HR: data relating to salaries, financial circumstances and payments.  
- Processing of data about criminal convictions or offences when the processing activities are carried out under the supervision of public authorities or otherwise allowed for in applicable legislation. | Requires heightened security measures and the auditability of processing activities. |

| Confidential | 2 A           | Personal data subject to specific processing conditions  
- Personal identity number.  
- Data relating to a minor.  
- An individual's payment transactions and debit/credit card details.  
- Contact details that are subject to an order of non-disclosure for personal safety reasons and other data that is classified as confidential in applicable legislation to protect the privacy of individuals.  
- The results of polling. | 1. Personal data subject to specific processing conditions.  
The risks associated with processing activities can be controlled without heightened security measures but special care must always be taken when processing data (such as extended risk assessment or DPIA). |

| Domestic     | 3 Y           | Basic personal data  
- Identifying information, email, phone number, name-address.  
- Identifiers, such as a Facebook username, Active directory username, etc.  
- Data that can be combined with internally or externally available data to identify an individual (such as the time and place of events).  
- Data that is directly related to an individual and is collected from or about the individual (purchase history, service usage history, audit trails, information about consent provided by withdrawn by the individual).  
- An IP address that enables the identification of an individual.  
- Basic employment records maintained by the HR unit. | 2. Basic personal data.  
Controlling the risks associated with processing activities does not require heightened security measures. |

| Internal     | 4 G           | Internal personal data  
- Data that is classified as internal based on a decision taken by the organisation or the individual concerned.  
- Example: contact information available on the intranet. | 3. Non-personal data.  
Requires no special security measures. |

| Public       | 5 W           | Publicly available personal data  
- Data that is classified as public based on a decision taken by the organisation or the individual concerned.  
- Example: contact details published on the public website. | 4. Non-personal data.  
Requires no special security measures. |

| Non-personal data  
- Technical log data that contains no personal data; product information; anonymous data  
- May be public, internal, confidential or restricted. | 1. Non-personal data.  
Requires no special security measures. |

---

1. In case data is processed using technical methods that enable the identification of a natural person.  
2. Health data such medical records, the results of medical examinations, physicians' notes and statements and other information relating to treatment or medical care.  
3. Processing must be managed in compliance with special legal requirements and a risk-based approach to data protection. When data is processed, other guidelines that are in force must always be observed, such as the organisation's data classification and handling policy.  
4. Non-personal data.  
Requires no special security measures.
TUNI Groups (Teams space)

• Data can be shared with users outside our university
  • access right management
  • collaboration support tools
• Applicable for processing personal data

OneDrive for Business

• 1000 GB storage space for each user
• Files can be shared with users outside our university
• Applicable for personal data
  • N.B! OneDrive for Business used with TUNI account is not the same service as the OneDrive for consumers.

Cloud services

• It is not essential if it is a cloud or not
• The key is the security of the selected solution
  • How it is built and maintained?
  • Probabilities for data losses and data leaks?
  • Terms of use?
• Tampere Universities have a contractual agreement with Microsoft that the O365 cloud services are e.g. GDPR compliant
  • N.B! Universities have no contract with Dropbox, Google, etc.
Reusing data
Have you made your data openly available, and where?
What to do with the data in the end of the project?

Agree about the following:

• Which data are to be released for further use?
• Which data will be destroyed?
• What data is kept by the researchers, for example to verify the results?
• If one or more researchers have brought previous research data in the research project, will this be included in the data to be released for further use?
• When can the data (or parts of it) be released for further use? (for example, a predetermined schedule or completion of publications)
• For what purposes is the data released (only for research or also teaching and studying)?
• Do you want to set conditions for further use?
• Who has the right to sign in an archiving contract for research data?
• If the permission for data reuse is required, who grants a permit?
Opening of the data

• Responsible data opening is required by funders, organizations and publishers.

• The terms of use of the data are determined by:
  • Data protection responsibilities
  • How the research participants have been informed
  • Research permits
  • Licenses
  • How the use of the data has been agreed between the research group -> the author of the data is agreed
Setting out terms for data reuse

Example from Finnish Social Science Data Archive (FSD)

- Dataset available for all users: the datasets are freely available under the CC BY 4.0 license and can be downloaded without registration. The source must be cited in publications and presentations where the dataset has been used.
- Dataset available for research, teaching and study
- Dataset available for research and for Master’s, licentiate and doctoral theses.
- Dataset available only by permission from the depositor and agreement to comply with the general terms and conditions of data use.

- Select the appropriate license for the data
  - License Selector and a guide to choosing a Creative Commons license
Where to open: in a data repository

• Check that:
  • Repository is appropriate for your data (contains similar datasets, fits requirements of a funder/publisher etc.).
  • Repository assigns persistent identifiers (PID), such as DOI or URN, to your data
  • Repository publishes machine-readable metadata and uses a known metadata standard.
  • Repository has clear policies about accessing and using data.
  • Repository has a certificate indicating trustworthiness (for example Core Trust Seal or ISO 16363 standard)
  • Search for repositories Re3data.org

• Examples of data repositories:
  • Finnish Social Science Data Archive (FSD)
  • The Language Bank of Finland
  • Zenodo

• Note! Data requested from a primary investigator is not really open. Broken e-mails and obsolete storage devices may disable sharing. Data archives curate and love your data!
Where to open: in a journal

• In a **scientific peer reviewed journal**, when you publish an article:
  • Data as supporting material/appendix of an article (risk that data will be behind the pay wall)
  • Data opened in a data archive (journals may have recommendations about archives)
  • When you make a publication plan, check journal’s data policy → think when you want to open your data.
  • Data availability statement may be required.

• Metadata can be published in a **scientific data journal**.
  • Data published as a part of the article.
  • Data opened in a data archive.
  • Examples of data journals:
    • Springer Nature [Scientific Data](https://www.springernature.com/en-gb/journals/47575)
    • Elsevier [Data in Brief](https://www.elsevier.com/products/services/data-management/data-in-brief)
    • Brilin [Research Data Journal for the Humanities and Social Sciences](https://www.brilin.eu/rdjhs)
Responsibilities and resources in collaborative projects

• Who is responsible for data management tasks during the project lifecycle? Name persons for each task.

• Evaluate and justify the resources required in the data management plan:
  • financial
  • working hours
  • workload related.

• Are resources outside the research team needed?
  • e.g. transcription, analysis, disk space, data manager services
Good to know

Guidelines

- 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)
- Responsible conduct of research (Tampere University)
- Researcher’s check list for publishing research data (Responsible research)

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)

- Contact us: researchdata@tuni.fi
Questions?

• My purpose for attending this session is to learn more about data management in cooperation projects with partners from organisations based elsewhere in Europe, as well as non-EU countries. I am in my first year of the PhD in Tampere University, and I work for another EU-based research organisation; I work on a project that requires data collection beyond the EU. I also may soon enter collaborative projects (if funding is granted), therefore, I look for more information about practical as well as legal aspects of data processing, storing, sharing, etc. in this regard.

• Miten saan käyttööni tietokannan ja koituuko tästä kustannuksia projektille?