Data management and data protection in a nutshell

11.10.2022
Research Data Services, Tampere University Library
Jari Friman, Katja Fält
Goals of the training

• Understand the basic principles of data management at different stages of research.

• Be able to apply the principles of the data protection and open research data in your own research.
  • Identify data protection responsibilities.
  • Identify possible data security issues in your data management software and solutions.
  • Anticipate and act in such a way that the data can be made openly available
  • Knows the services suitable for making the data openly available and knows how to utilise them.
Today

- Research data, data life cycle and FAIR principles
- Data management planning
- Intellectual property rights, authorship and licenses in research data
- Personal information and data protection in research
- Data storage and security
- Documenting and describing research data
- Sharing research data
What is research data?

• All the materials, methods and results produced and used in the research

• Research data can be
  • Collected by your self
    • Interviews, surveys, observation diaries, drawings, photos etc.
  • Reused data
    • Data collected by other researchers, for example surveys
    • Search datasets: FSD, Language Bank of Finland, Etsin
  • Data, which exist without research or a project
    • For example, Open Data Tampere, European Data Portal, Findata
  • New data which is created during the research process
    • Data which is based on collected data and/or already existing data
Data life cycle

AFTER THE RESEARCH:
- Publishing and sharing data and metadata
- Preserving data for verification
- Preserving the data for your future projects
- Data disposal

BEFORE RESEARCH:
- Research plan
- Data management plan
- Ethical review
- Research permits

DURING RESEARCH:
- Data collection
- Organizing and documenting
- Processing and analysing
- Storing and protecting
- Back-up
FAIR principles

• Fair principles require careful data management as it is difficult to make research data FAIR-compliant retrospectively.
• That's why data management planning is very important.
• FAIR data is not the same as open data.
• The openness of metadata is essential.
• The actual research data should be as open as possible and as closed as necessary.

https://libereurope.eu/article/fairdataconsultation/
Maximise the reuse of your data.

1. Plan ahead! The planning helps to make the data as open as possible (and also to consider the restrictions for sharing). Also, update your Data Management Plan (DMP) during your project.

2. Describe your data throughout the lifecycle of your project. Describing the data may require a lot of work and it may be impossible to find time for it if you do it afterwards.

3. Prefer open file formats.

4. Select the data archive or repository for your research data. Search for suitable archives already in in the beginning of your project. Archives usually assigns your data with DOI.

5. Publish your metadata. Metadata should be openly available even if the data itself is not.

6. Make sure that the data and metadata receive a persistent identifier (DOI, for example).

7. Specify the terms of use of your data with licenses (or terms of use provided by the archive or repository).
Why open research data?

• Openness of science is a prerequisite for critical evaluation of research and the scientific progress.
  • Verifiability of research results and replicability of research.
  • Reproducibility crisis: [https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970](https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970)

• Preserving your research data and making it available to other researchers is one way of ensuring open science.

• Finding new collaboration opportunities with researchers, business or other communities

• More visibility and citations to your research.

• Credit to the researcher. Also, a merit for the student for working life.

• New kinds of research questions.

• Overlapping work is reduced, and it’s also a matter of research participants’ time.

• You meet the requirements of our higher education community, several funders and scientific publishers.
Data management plan in practice

• Create a data management plan (DMP) already when you are planning your project.

• You can create your plan by DMPTuuli-tool.

• There are templates for students and researchers

• National DMP template is as follows:
  1. General Description of Data
  2. Ethics and Legal Compliance
  3. Documentation and Metadata
  4. Storage and Backup during the Research Project
  5. Opening, Publishing and Archiving the Data after the Research Project
  6. Data management responsibilities and resources

• Keep your DMP updated!
Copyrights, authorship and licenses in research data
Roles and responsibilities

• Who can use the data collected in the project?
• What about data obtained from outside the project?
• What happens to data if someone leaves the project?
• Agree on roles and responsibilities regarding your data.
  • Agree on who may use the data during the project
  • Agree on the further use of the data
  • Agree on the terms which apply to reuse of the data
Intellectual property rights in research materials

Intellectual property rights, which includes copyrights, related to research are mainly influenced by the source of funding and researcher's employment relationship.

• In the case of computer software, inventions or databases, rights is usually held by the university. The University typically has interests in software that are intended for commercial use.

• The research data may also contain images or writings produced by research subjects and those may be protected by copyright.

In academic research, where there are no commercial interests, more relevant than IPR issues is that researchers agree among themselves (and both with project partners and with the research subjects) on the (further) use of the material.
Authorship and moral rights

• The creator of a work holds moral rights, which cannot be transferred.

• Authorship related to research data have to be defined case-by-case. This is influenced by different practices of the disciplines and the principles of research ethics.

• The Finnish Advisory Board on Research Integrity (TENK) recommends that researchers should agree on authorship in the beginning of the project.

• When researchers agree on authorship concerning research data and publications made based on the data, authors become visible in data citations and references.

Licence your research data

• Use licenses to define how your research data can be reused.
• The Creative Commons (CC) license requires that others who use your work, must give you credit the way you request.
• CC0 waiver is the most efficient way of facilitating reuse of your data. This is typically used with metadata.
• A typical license for open research data is Creative Commons 4.0.
• For codes used in academic research we recommend MIT or other commonly used open software license.
• How to select a Creative Commons license?
• Data and software licensing wizard by EUDAT
Personal data and data protection
What is personal information

• 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 data: direct and indirect identifiers

#### 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
Special categories of personal data

- 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
Anonymous 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.

Processing personal data

• When you collect information on research participants it is likely that you process personal information.

• Examples of processing personal data (under GDPR):
  • collection, recording
  • organisation, structuring, storage
  • adaptation or alteration
  • retrieval
  • consultation, use
  • disclosure by transmission
  • dissemination or otherwise making available
  • alignment or combination
  • restriction, erasure or destruction
Informing about processing personal data

Documents you need to inform research participants

• **Information sheet**: research topic, description of how the data will be collected, potential risks to research participants…

• **Ethical consent form** and permission for data archiving.

• **Privacy notice** about processing personal data.

Informing can be done, for instance: face-to-face, in the beginning of a survey form, e-mail, web page of the project…
Data controller

• Data controller determines the purposes and means of the processing of personal data.
• The data controller can be a natural or legal person, or a public authority, for example.
• Data controller in research:
  • When the principal investigator is in employment contract with Tampere University, the data controller is usually Tampere University Foundation.
  • When the principal investigator has an individual grant or similar funding, with no employment contract with Tampere University, the researcher is usually the data controller.
  • When your research is conducted on behalf of a company (customer-funded research) or if you are employed by a company, it is possible that the company acts as the data controller.
  • Many data controllers, which may result in joint controllership in collaborative projects with other universities, companies or other organisations.
• In students’ thesis:
  • students are data controllers.
  • However: When thesis is conducted as part of a preplanned university project, it is possible that the university acts as the data controller for the thesis.
Lawful basis for processing personal data

- Processing personal information requires always a lawful basis.
- In scientific research, the lawful basis is usually:
  - Public interest or the exercise of official authority: Scientific or historical research purposes or statistical purposes.
- In student’s thesis, the lawful basis is usually:
  - Voluntary, specific, informed and explicit consent provided by a data subject.
- Other lawful basis:
  - Exercise of the legitimate interests of the data controller or a third party if it is possible based on a so-called balance test.
- In principle, the processing of data concerning sensitive personal data is prohibited. The exemption criteria are:
  - processing is necessary for archiving purposes in the public interest or for scientific and historical research purposes or statistical purposes
  - the data subject has given his or her explicit consent to the processing of such personal data
  - the processing concerns personal data explicitly made public by the data subject
Meanings of consent

In context of research, the word consent has three distinct meanings:

a) consent to participate in non-medical research in compliance with applicable ethical standards (see the guidelines provided by the Finnish National Board on Research Integrity TENK)

b) consent to participate in medical research (Finnish Medical Research Act, 6 §, 9.4.1999/488)

c) consent as a lawful basis for processing personal data (the EU's General Data Protection Regulation)

• An informed consent form signed (to participate in a study) by research participants does not necessarily mean that consent is the lawful basis for processing their personal data.

• If consent is used as a legal basis, you should make sure that the consent meets the requirements defined in the law.
Risk assessment of processing personal data

• Risk assessment
  • Must be done before processing personal data.
  • Assess the risks from the perspective of the data subject.
  • Document the risk assessment.

• DPIA = Data protection impact assessment
  • DPIA is mandatory under the EU’s General Data Protection Regulation (GDPR) if the processing of personal data poses a high risk to people’s rights and freedoms.
  • See the checklist for determining whether a Data Protection Impact Assessment (DPIA) is required

• Risk assessment and DPIA instructions and templates

Photo by Marija Zaric on Unsplash
Accountability & required documentation

• Document all processing activities. Store the documentation and take care of information security. Keep documents updated.

• Examples of data protection documentation:
  • Data management plan
  • Research plan
  • Privacy notice
  • Risk assessment from the perspective of the data subject, Data Protection Impact Assessment (DPIA)
  • Record of processing activities (often part of research data management plan or research plan)
  • Information sheet about the research
  • Consent forms
  • Contracts and guides for data processors (data processing agreement, non-disclosure agreement)
Useful links

• Data protection path of research (also templates for privacy notice and consent form)
• Research in the field of human sciences – responsible conduct of research (also information sheet template)
• Informing research participants about processing their personal data and about archiving (FSD)
• Data protection passport for students (intra)
Data storage and security
Storage solutions

• 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! Tampere University have no contract with Dropbox, Google, etc.

Photo by Billy Huynh on Unsplash
Data storage services at Tampere universities

OneDrive for Business

• For staff and students
• Capacity: 1000 GB.
• Use with your TUNI account

Tuni Groups

• Similar to OneDrive, but it is especially useful for group working
• Data can be shared with users outside our university

Personal (P) and shared (S) network drives

• Capacity 50 GB for Staff and 10 GB for students
• Shared network drives available for research groups.
• Data can not be shared outside the university.

Protect your files and folders with encryption

• Boxcryptor is a centrally supported software that gives added security against the unlawful use of data.
• Cryptomator is a free software that partly contains similar characteristics. It is also available to your computer directly from the Cryptomator.org site.
• Boxcryptor requires a TUNI id which makes Cryptomator possibly more easily accessible to other partners.

• Read more about personal storage space (TUNI intra) and external website.
Online video interviews

Zoom and Teams

• GDPR compliant but only if logged in with tuni account
• Zoom: recording is stored at the interviewer's computer
• Teams: recording is stored at OneDrive
• Security guidelines for setting up remote meetings and interviews using Teams or Zoom:
  • Online survey tools O365 Forms ja LimeSurvey (intra)
  • REDCap system for collecting and managing research data
  • Data security of voice recorders and cameras (intra)
Documentation and metadata
Why should you document your data?

• Documentation is an integral part of FAIR principles.
• Helps other researchers understand how the data has been collected and managed.
• Describes the necessary technical details, how the data is collected and processed.
• Original researchers who wish to return to their data some time later, or new users who want to use data, need sufficient contextual and explanatory information to make sense of those data.
• The goal of a metadata record is to give users enough information to know if they can use the data without contacting the dataset owner.
Documentation in practise

• Write down what you do with your data.

• Use README files, codebooks, laboratory notebooks, field notes, diary → what is the common practice in your field?

• To prepare your data for secondary research
  • explain what you have done in practice when you collect your data
  • explain the objectives and methodology of the research
  • explicitly describe the meanings of variables and codes used. Additionally, they should describe any derivation, transformations, de-identification (pseudonymisation/anonymisation) or data cleaning carried out.
  • Remember: Naming files and organising folders is also part of metadata.
Public description - metadata

• Public description includes information
  • about the data collection
  • about the content of the data
  • on the authors of the research data
  • about the terms of reuse and licenses
  • permanent identifier (DOI, URN)

• Metadata tool Qvain can be used to describe the research data. Metadata is then published in research data set finder Etsin. Both tools are provided by Finnish Fairdata services.

• If you store your data in an archive, the archive defines a metadata format. The archive also provides permanent identifiers.
What to describe?

- For what purposes was the data created?
- 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?

Sharing your data
When planning data sharing, consider these:

- **The consortium agreements** define the responsibilities related to the research project involving several institutions. If you share personal data with your collaborators from other organisations, 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 (**privacy notice**).

- **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 (e.g. Findata).

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

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

- **Funders’ requirements** for making your research data publicly available. “As open as possible, as closed as necessary”
Where to open and share data: a 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))

Illustration credit: Ainsley Seago.
doi:10.1371/journal.pbio.1001779.g001
Where to open and share data: a publication

Peer-reviewed scientific publication

• Data as a supplementary data
• If you have to pay in order to get the article and the data, the premise of open science is not supported.
• Typically data is archived in a data repository (many journals have recommendations about suitable repositories).
• When drafting a publication plan check the journal's data policy --> consider when to publish the data.

Data journal

• A data journal publishes articles that focus on data quality or data collection methods, etc.
  • Data is deposited in a data repository
• Examples of data journals:
  • Springer Nature: Scientific Data
  • Elsevier: Data in Brief
  • Brill: Research Data Journal for the Humanities and Social Sciences
Data sharing in practice

• Inform the participants of the research and request permission to make the data openly available, i.e. an information sheet, a privacy notice and a consent to participate to study are required.

• Make sure there are no ethical or legal restrictions to data sharing.

• Describe your data adequately.

• Use file formats that can be opened without a special software.

• Anonymize your data. Check FSD’s guidance on anonymisation.

• Contact data archive, in advance.

• Examples of data archives:
  • FSD,
  • Language Bank,
  • Zenodo (general-purpose open-access repository funded by EU and operated by CERN)
  • Search for repositories Re3data.org
Links and guides

- [Research Data Management -guide](Tampere University Library)
- [Data Management Guidelines](Finnish Social Science Data Archive)
- [Data protection path of research](Tampere University)
- [Data protection passport](TUNI intra)
- [Quick guide to information security](TUNI 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](TAU intra)
- [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.

Contact us: researchdata@tuni.fi