Research data management in a nutshell

9.10.2020

Jari Friman, Saila Huuskonen, Kaisa Kylmälä
Tampere University library, Open science services
In this presentation

• What is research data?
• Why manage research data?
• Research data management
  • Research data management plan
  • How do I manage research data during the research project?
• Document your research data
• Opening and sharing 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.
- Metadata is also part of research data.
What is research data?

• Data collected or produced by you or your research group
• 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 kind of data do you have?

What kind of data do you
• collect
• use or
• reuse?
Why research data management

• Practical benefits
  • Streamlines your work and helps you save time and money.
  • Helps you organize, save and find your data.

• It is good research practice!
  • Demonstrates that you are working ethically and responsibly.
  • Makes your research more transparent, reliable and reproducible.

• It helps you to open your data and make your data reusable.
  • Gives you more visibility (citations, getting merits).
  • Reduce overlapping work.
Data Management Plan

- Describes how you will collect, organise, analyse, preserve and share your data
- Start writing your DMP as soon as you start planning your project and keep it up to date.
- Use DMPTuuli – tool for writing Data Management Plan
  - Tampere University guidance, example data management plans,
  - Share and co-write your plans with colleagues, leave comments, ask feedback from 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)
Ownership and user rights

• Make an agreement on usage rights of your data
  • Who is allowed to use the data during the project?
  • What are you planning to do with the data after the project?
  • What happens to data if someone leaves the project?
• Are there any patents or copyrights related to your data?
  • Third party data (e.g. data archives, companies, registry data, open data)
• Use licenses to define terms and use of your data
  • License selector
  • How to choose a Creative Common license?
Keep notes of your data > Document
Why should you 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 publicly available.
Metadata in open science policies

- **Open science and research in the Tampere higher education community**

  “The essential metadata describing the research data is always open. The metadata include information on the structure of the data and details concerning its production, producers, owners, and terms of use, as well as the unique persistent identifier of the data.”

- **Funders’ requirements: Academy of Finland**

  “If the research data cannot be made openly available, the metadata must be stored in a Finnish or international data finder.”
What to describe?

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…
- 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?

How should I document my data?

- README-file example

Example of topics in a README file:

- TITLE: Name of the dataset or research project that produced it.
- CREATOR: Names and addresses of the organization or people who created the data.
- DESCRIPTION: of the data package and folder overview.
- LOCATION: Where the data relates to a physical location, record information about its spatial coverage.
- METHODOLOGY: How the data was generated, including equipment or software used, experimental protocol, other things you might include in a lab notebook.
- INFORMATION ABOUT DATA FILES:
- IDENTIFIER: Number used to identify the data files, even if it is just an internal project reference number.
- LOCATION: Where to find data files and additional information such as the data dictionary explaining variables used.
- DATES: Key dates associated with the data, including project start and end date, data modification date, data release date, and time period covered by the data.
- SUBJECT: Keywords or phrases describing the subject or content of the data.
- FILE FORMATS: What file formats have been used.
Keep your files organized

- Organized and consistent folder structures

Source: FSD
File naming

• Keep your file names
  • Short
  • Informative – describing the content
  • Consistent

• Caution! No personal information in file names

• Include version numbers in file names
  • For example v01, v02, v03

• Use standardized date and time format
  • Date → 20200924_datapilot.docx
**Qvain** – Research Dataset Description Tool

- **Qvain** is a tool to create the metadata for your research dataset.
- It offers you a simple form to fill-in and save the metadata.
- Part of the [Fairdata.fi](https://www.fairdata.fi) – service provided by CSC.
- Files in Fairdata IDA can be linked to your description.
- You can link your description not only to your files in Fairdata IDA, but also files in other data archives and services.
FSD3316 Welfare and Inequality in Finland 2017-2018

Authors
- Kainulainen, Sakari (Diaconia University of Applied Sciences)
- Niemelä, Mikko (University of Turku, Department of Social Research)

Keywords
costs, housing, income, satisfaction, social inequality, social security benefits, unemployment, well-being (health)

Abstract
The study charted Finnish opinions on welfare and inequality. The study was as part of a project entitled Tackling Inequalities in Time of Austerity (TITA) funded by the Strategic Research Council of the Academy of Finland (decision number: 293103).

First, the respondents were asked about their life satisfaction, mood, ability to achieve things, perceived status in society and trust in other people.

Regarding income and personal finance, net income of the household and the ease of covering usual expenses with the income was surveyed as well as the monthly costs incurred by debt. The respondents were asked whether their parents had been unemployed or received social security benefits when they were in their teens and whether the respondents themselves had received social assistance in the previous year.

The respondents were asked how much they cared about the well-being of different groups (e.g. the homeless, immigrants, refugees and asylum-seekers, children in poor families, the elderly). The respondents were presented statements about whether the differences in income, health, neighbourhoods and education were too high in Finland. A number of statements about social assistance (the social security benefit of the last resort) were presented. Finally, the respondents were asked to what extent they agreed with statements about their own welfare and well-being.

Background variables included the respondent’s gender, year of birth, household composition, housing tenure, education, economic activity and choice of political party.

Permanent Link to this dataset:
http://um.finnunib.fi/sid/T:FSD3316

Direct link to this tab:
Data level metadata

FSN3316 Welfare and Inequality in Finland 2017-2018

Overview Detailed description Variables Publications Download data

Select variable

[k1] FSD study number
[k2] FSD edition number
[k3] FSD case id
[k4] Gender
[k5] Respondent's year of birth
[k6] Thinking about your current life situation in general, how satisfied are you with your life?
[k7] At present, how would you describe your mood on average?
[k8] People want different things in life. At present, how well do you achieve the things that you want?
[k9] Where would you place your own social standing in Finnish society?
[k10] Do you mostly feel
[k11] Would you say that people can be trusted or that you can’t be too careful in dealing with people?
[k12] Did your parents or one of your parents receive social assistance (social security benefit) when you were a teenager?
[k13] Were your parents or one of your parents unemployed or underemployed for extended periods of time when you were a teenager?
[k14] Have you received social assistance within the past year?
[k15] What is the total net monthly income of your household when you take into account all income of all household members? If you do not know the exact sum, please give an estimate.
[k16] What are the total monthly expenses of debts in your household (interest, debt repayments/installments, other) if you do not know the exact sum, please give an estimate.
[k17] When you take into account all income in your household, how well are you able to make ends meet, that is, pay your usual expenses?
[k18] How much do you care about the well-being of people belonging to the following groups? Other people in general
[k19] How much do you care about the well-being of people belonging to the following groups? The homeless
[k20] How much do you care about the well-being of people belonging to the following groups? People with substance abuse problems (alcohol, drugs)
Tips for writing quality metadata

• Pay attention to titles
  • An example: Greater Yellowstone (where) Rivers (what) from 1:126,700 (scale) U.S. Forest Service (who) Visitor Maps (1961-1983) (when)

• The goal of a metadata record is to give the user enough information to know if they can use the data without contacting the dataset owner.

• Be detailed: there’s no such thing as too much metadata!

• Select keywords wisely. Use thesauri for keywords whenever possible.

• Remember: a computer will read your metadata.
  • Do not use symbols that could be misinterpreted by software: Examples: ! @ # % { } | / \ < > ~

• When copying and pasting from other sources, use a text editor (e.g., Notepad) to eliminate hidden characters

Source: https://www.dataone.org/education-modules Lesson 07: Metadata
Data opening
Why to open: benefits for a researcher

• More visibility
• More citations:
  • Colovizza et al. (2019) "We also find an association between articles that include statements that link to data in a repository and up to 25.36% (±~1.07%) higher citation impact on average, using a citation prediction model."
  • Drachen et al. (2016) "The examined papers in the dataset with datalinks received in total 40.5 citations per year on average, whereas the papers without links to data received correspondingly fewer citations; 35.3 per year."
• Finding collaborators
• Research data as a merit in your CV—new template for researcher’s CV (TENK)

14. **Scientific and societal impact**
   • Promoting open science and research, for example the production and responsible distribution of research material and datasets
Why to open: benefits for the society and scientific community

• Solving problems together → benefits for the society
• Enables replication and validation of the research
• Possibility to conduct meta-analysis
• Ethical choice – saving time of the research participants
• Economical choice – no overlapping work (reuse of resources)
Why to open: required by many stakeholders

• Several peer reviewed scientific journals require that research data must be opened because of transparency and verification of the findings.
  • PLOS data availability
  • Nature Research data availability statement and data citations policy: guidance for authors
  • BioMed Central's Policy on Open Data

• Funders’ requirements
  • Academy of Finland and EU Horizon 2020
    • by default data must be opened keeping in mind "as open as possible, as closed as necessary"

• Tampere Higher Education Community
  • Tampere Higher Education Community Research strategy
  • Tampere Higher Education Community Open Science and Research Policy
  • Tampere University Action Plan on Open Science
Where to open: 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
  - Examples of data archives:
    - Finnish Social Science Data Archive (FSD)
    - Language Bank
    - Zenodo

- Note! Data archives curate and love your data.
- Data requested from a primary investigator is not really open.
  - Broken e-mails and obsolete storage devices were the main obstacles to data sharing.

Illustration credit: Ainsley Seago. doi:10.1371/journal.pbio.1001779.g001
How to open: plan and prepare before hand

• Inform research participants and ask their permission for data opening
  • Consent for participation
  • Privacy notice (intra)

• In your research group, make an agreement about data opening.
• Describe your data in the way that the others can understand it.
• Use data formats which are appropriate for a long term storage and can be opened without commercial software.
• Anonymize your data.
How to open: licencing and terms of use

• Licensing data
  • CC-lisences
    • License selector
    • Guide for choosing Creative Commons license

• Terms of use in Finnish Social Science Data Archive (FSD)

Dataset availability:
(A) openly available for all users without registration,
(B) available for research, teaching and study,
(C) available for research only (including Master’s, doctoral and Polytechnic/University of Applied Sciences Master’s theses),
(D) available only by permission from the data depositor/creator.

• Language Bank
  • accessible without signing in
  • accessed with credentials issued by your own institute or other organization
  • require personal access rights in addition to signing in
Check list for data opening

• In the beginning of the research project
  • Make agreements about data opening in your research group
  • Think which part of your data can be opened
  • Make a plan about timing of data opening
  • Check ethical, legal and intelectual property right issues set limits for data opening
  • Check funders’ requirements
  • Ask research participants consent for data sharing
  • Choose a data repository and check their requirements

• In the end of the research project
  • Check that data protection issues are ok
  • License your data
  • Publish metadata of your research data
  • Promote your data for example in social media
  • Add information about opened data in your CV

• See also:
  • Responsible research: Researcher's check list for publishing research data
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.