Research Data Management Cheatsheet

Steps Towards Great Data Management

*Nowadays, funders **require** that you tell in the data management plan how you are going to handle the data during the project, and after the project is completed. Planning, informative documentation, secure project fitted storage and clearly expressed organising makes your day-to-day work **smooth**.*

Plan

DMP (Data management plan)
is required by most of the funders. Use DMPTuuli or benefit from its instructions (https://www.dmptuuli.fi/). Update your plan when necessary. Make it a practical document that fits exactly to the needs of your data.

Foster collaboration – remember the security
Plan together and document how you will promote the data related collaboration. E.g. how the data can be accessed during the project (cloud services and eDuuni provided by Aalto, national Fairdata.fi, etc.).

Take care of ethical and privacy issues
Consider if you should take into account:
- ✓ Legal requirements, e.g. General Data Protection Regulation (GDPR) and Finnish Personal Data Act.
- ✓ Guidelines for ethical research conduct.
- ✓ University's practice of ethical evaluation, e.g. request a statement from the university's Research Ethics Committee **before** the research activities.

Know your IP (Intellectual Property) Rights
Consider all possible ways to benefit from your data. Recognise also the business opportunities. Get professional help and contact Innovation Services in time to safeguard your IPR.

Describe

Document and provide metadata to ensure findability, accessibility, interoperability and re-usability of your data.

Provide proper metadata
Metadata is information about the data itself, a systematic characterization of a data set. Standardisation of the metadata makes it machine-readable and facilitates e.g. the comparison of datasets.

Metadata describes data in data catalogues
Catalogue metadata is an example of structured metadata that often follows international standards. Common metadata includes: Title, Description Creator, Funder, Keywords and Affiliation. See the example of ICPSR Guide to social science data about the important metadata elements.

Documentation bounds your data to context
Makes reference to data in the context of their use in specific systems, applications, settings. It describes the data and the methods so that the suitability for re-use can be assessed. Documentation can be embedded like code, field and label descriptions, or provided as a supplement of data as a readme-file.

Documentation examples
- Readme-file, Laboratory notebooks & experimental protocols.
- Information about equipment settings.
- Provenance information about sources of digitized data.
- Descriptions of qualitative data and its organising principles, abbreviations or codes used in the dataset
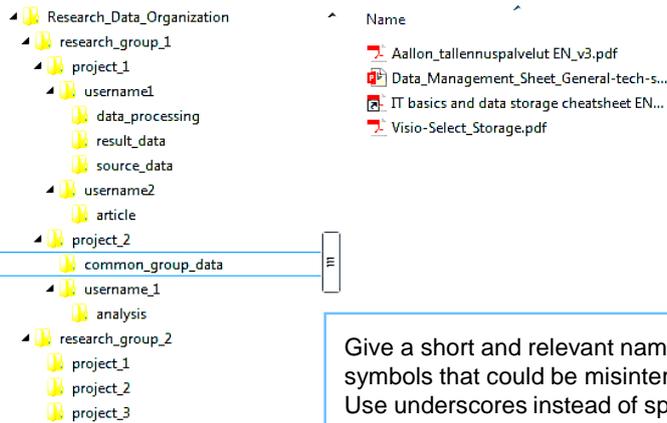
Store & Share

Select the storage and sharing solution so that the day to day work during the project is easy. Consider especially security and confidentiality issues. Examples:

Service	Pros (+)	Cons (-)
Aalto Network Storage Services	Has automatic backups, good for confidential data. Supports internal sharing.	Sharing outside Aalto is not supported.
Aalto Cloud services (Google Drive, OneDrive)	Easy access and sharing. Commenting and reviewing efficient compared to email attachments.	Not suitable for confidential data. Take care of the visibility settings and access.
Aalto eDuuni solution	For collaboration between the organisations within the sector of education. For confidential data.	Aalto IT only provides basic support and creates simple workspaces. Additional services from CSC.
National Fairdata.fi service	Store and use data during the project, publish data and add metadata, long-term preservation.	All parts are not yet ready (02092018).

Organise

Decide on folder structuring, file naming conventions, and version control practices. Separate working files from data at rest with project and archive folders. Wise organising ensures an easy access to the data during the project and helps preparing your data for re-use.

Structure folders hierarchically, starting with broader and creating specific folders within.

Name folders suitably and descriptively to avoid confusion in shared workspaces.

```
▲ 📁 Research_Data_Organization
   ▲ 📁 research_group_1
      ▲ 📁 project_1
         ▲ 📁 username1
              📁 data_processing
              📁 result_data
              📁 source_data
         ▲ 📁 username2
              📁 article
      ▲ 📁 project_2
           📁 common_group_data
         ▲ 📁 username_1
              📁 analysis
   ▲ 📁 research_group_2
        📁 project_1
        📁 project_2
        📁 project_3
```

Name
- 📄 Aallon_tallennuspalvelut EN_v3.pdf
- 📄 Data_Management_Sheet_General-tech-s...
- 📄 IT basics and data storage cheatsheet EN...
- 📄 Visio-Select_Storage.pdf

Use a numbering system that takes into account major changes and minor changes, e.g. major v01; minor v01_01. Also consider using the initials or name of the last editor in the file name where appropriate.

Use standard file formats when possible.

Give a short and relevant name. Do not use symbols that could be misinterpreted (! @ #). Use underscores instead of spaces.

Publish

Data is on the way to be recognized as a research output. So, make your data open, or at least its metadata. The best visibility you get by publishing your data in a repository that provides a proper entry, landing page, for your data.

The ways making your data visible
The best data publishing channel depends on your field's practices and the needs of your research. You may publish your data in the research article, as a supplement of the article or in the data repository. You can also write an article that describes your data and submit it to the data journal.

Publishing in a repository ensures findability and citability
A good data landing page in a repository includes:
- ✓ Option to define access rights, e.g. make the data available under CC-BY licenses or provide access by requests.
- ✓ Machine readable metadata.
- ✓ DOI (or other persistent identifier).
- ✓ Documentation where necessary for re-use.

Repositories
Select a discipline specific repository where possible. E.g. in social sciences the Finnish Social Science Data Archive (FSD). Ask your colleagues about their experiences, check the options from re3data.org. Submit your data to a general-purpose repository, e.g. EU-funded Zenodo and EUDAT B2SHARE, or national Fairdata.fi, if no community-recognized one available.

Provide adequate metadata in the repository
e.g. title, creator, year of creation, description. A complete title is unique and includes: What, Where, When, Who, and Scale. Don't copy the abstract from a research paper, but still, provide a brief description of the data and its context.

Ensure your data is recorded in ACRIS
Your data should be identifiable like your articles. Send the link to your published data to acris@aalto.fi, and the metadata of your data will be deposited to ACRIS. Note! You may also publish the metadata of the confidential data.

Archive

Assure that your data is re-usable after a long time. Many of the requirements to be fulfilled, overlap with the steps of storage and publish data. However, much more effort might be needed to prepare the data for long-term preservation.

Select the data for long term preservation
Assess your data and think over what data or part of it deserve to be preserved for a long period (e.g. over 10 years). To select the data consider does your data have continued scientific value in your own discipline or in the other fields.

Ensure the sufficiency of documentation
Re-check that you have provided enough information about your data and methods. The future users might need a lot of information to be able to reanalyse and understand the data within the new context.

Keeping data in good shape – Data curation
Data curation means special services that are developed for long term preservation of data, often provided by data repositories/archives. Curation consists of both data-specific and general practices. These can mean maintaining and managing the metadata of data, organizing, describing and cleaning data; also converting file formats to the current ones.

Solutions for archiving
Fairdata.fi service provides a national solution, and in social, and related sciences, use the Finnish Social Science Data Archive (FSD).