

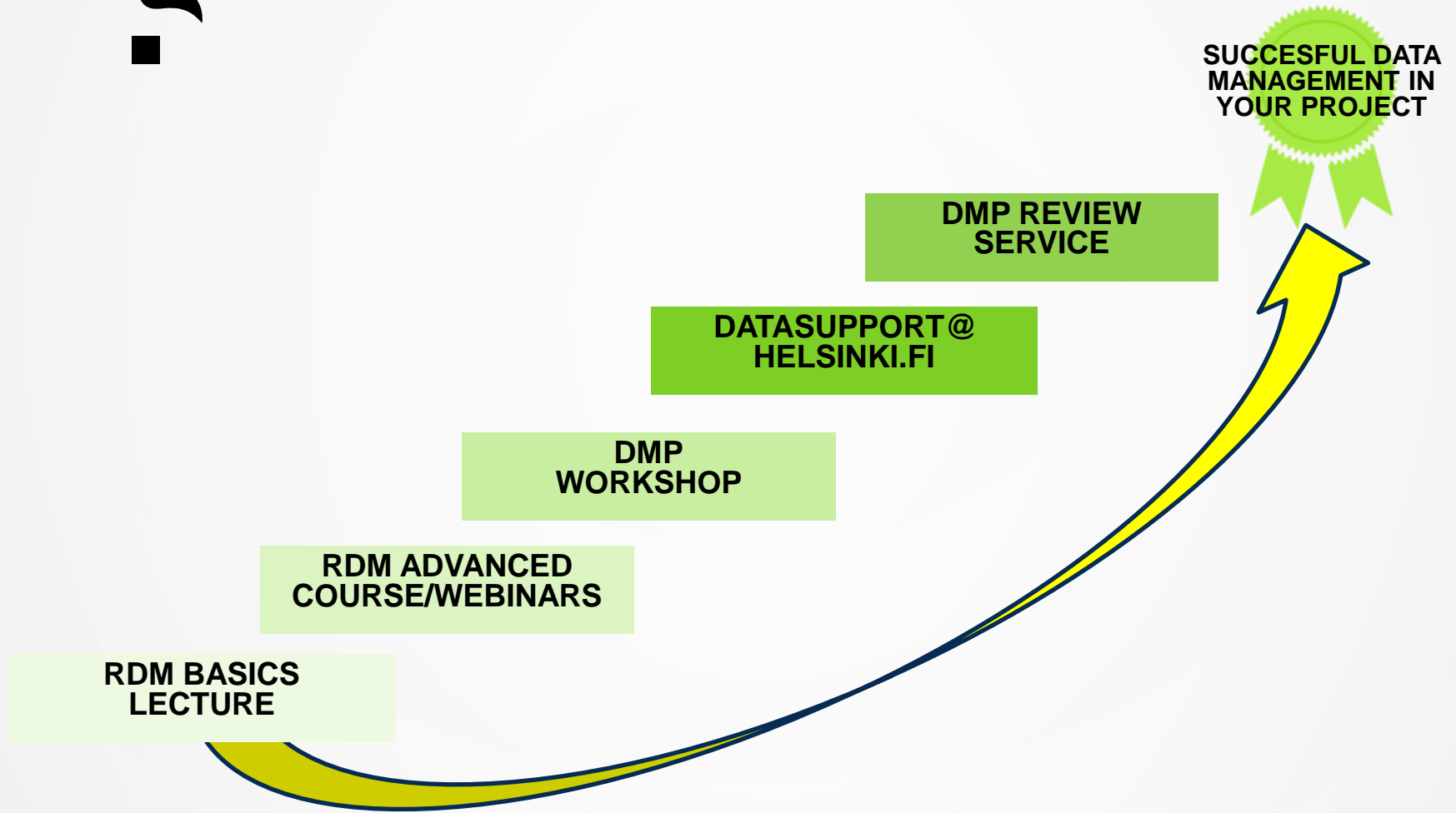


RDM BASICS –

RESEARCH DATA MANAGEMENT BASIC CONCEPTS AND WORKFLOWS

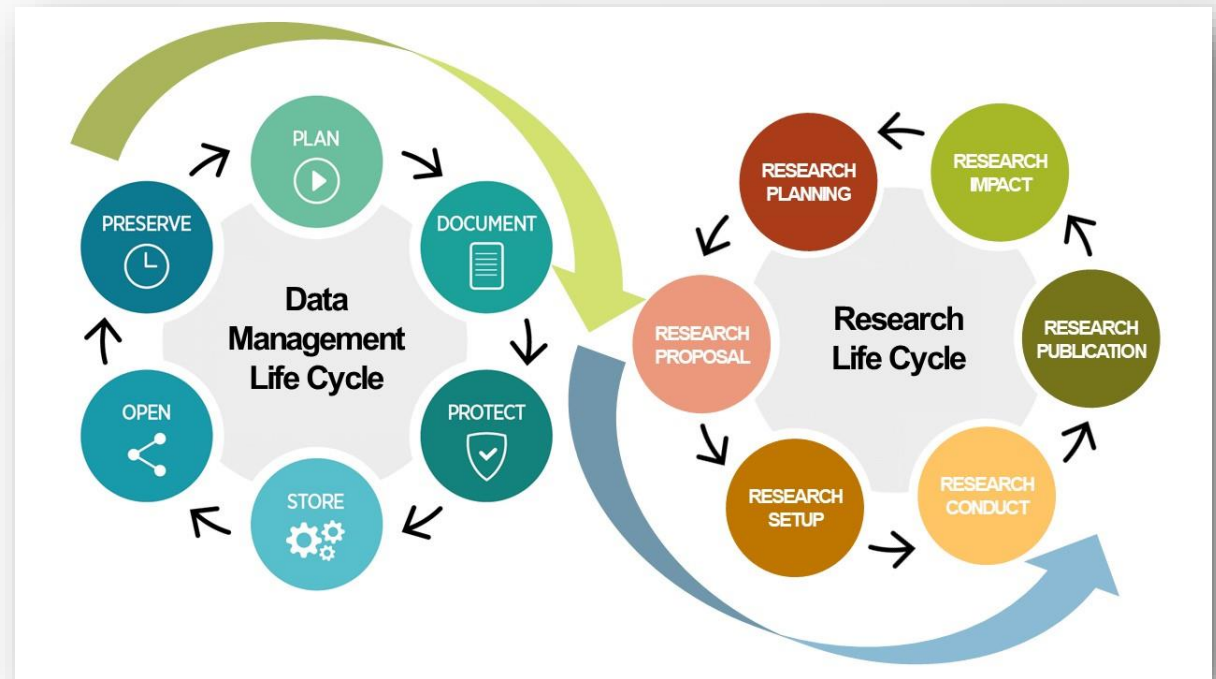


FROM RDMBASICS TO SUCCESSFUL DATA MANAGEMENT





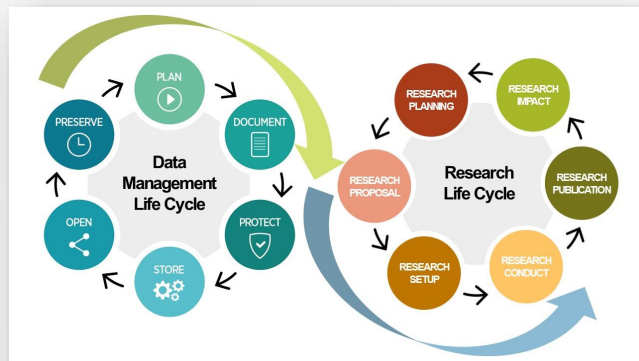
WHAT IS RDM AND WHY NOW?



- Two life cycles
 - RDM and the conduct of research should be synchronized
- An old issue with new points of view
- Digital files can be distributed globally very fast
- Poses new threats



WHAT IS RDM AND WHY NOW?



- RDM is based on your research task and research questions
- Many of the RDM issues and especially those discussed in the DMP are directed by the task and questions, for example:
 - -> minimum data collection requirement
 - -> what can be collected, what can be archive, what should be destroyed
 - -> reflect your own project; do not use example sentences
- Generic researcher skill!

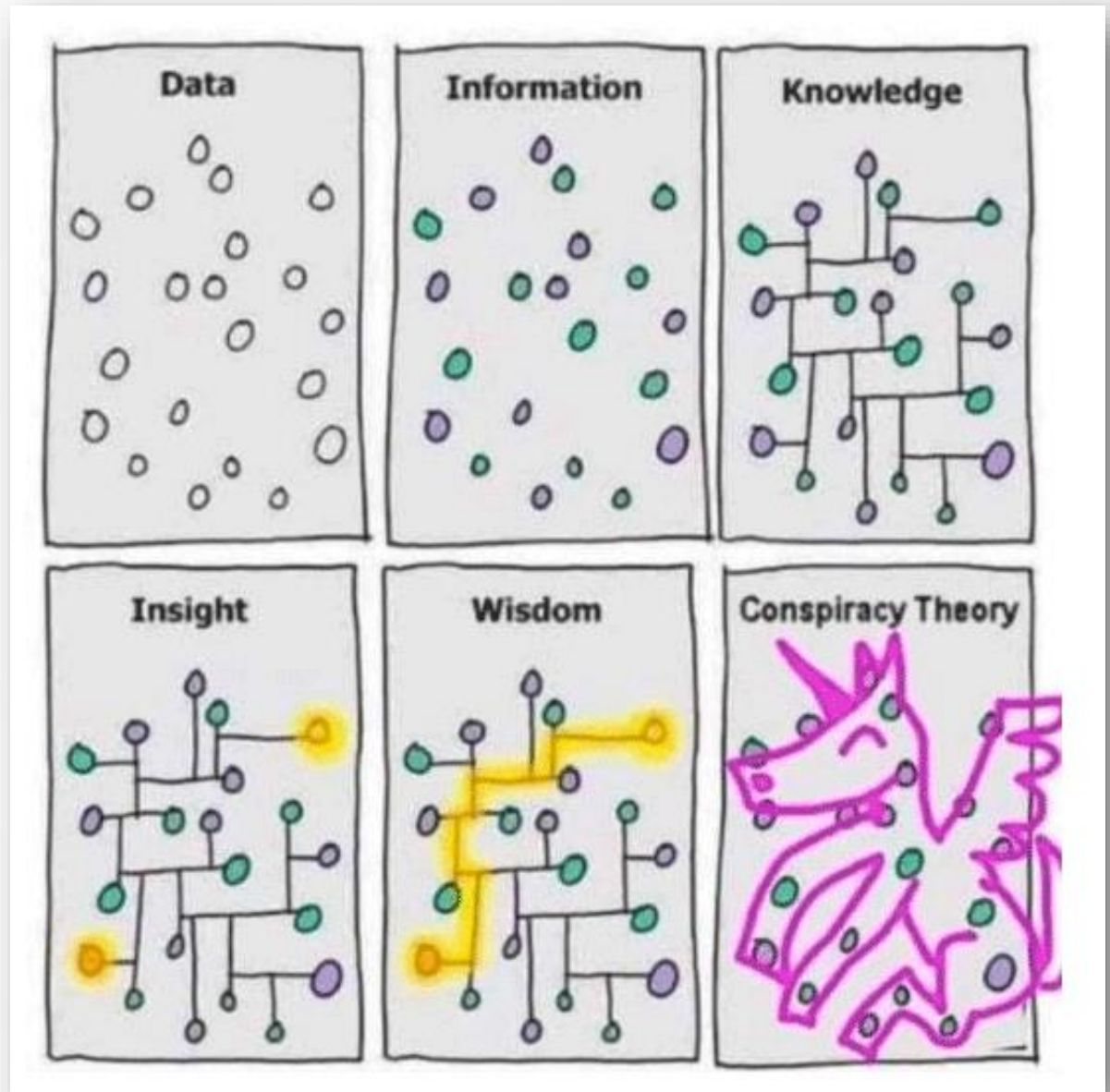


WHAT IS RDM AND HOW TO MEET THE NEW CHALLENGE = DMP!

- **Plan ahead! Train yourself in RDM issues!**
- DMP = Data Management Planning
- **DMP is YOUR TOOL!**
- Similar role in your project than you research plan
 - Risk assesment document – points for example towards DPIA (Data Protection Impact Assessment)
 - A timely plan is a prerequisite for a successful project
 - National DMP guide (<https://wiki.helsinki.fi/x/ZB5zEw>); based on Science Europe guidelines
 - In line with organizational data policies



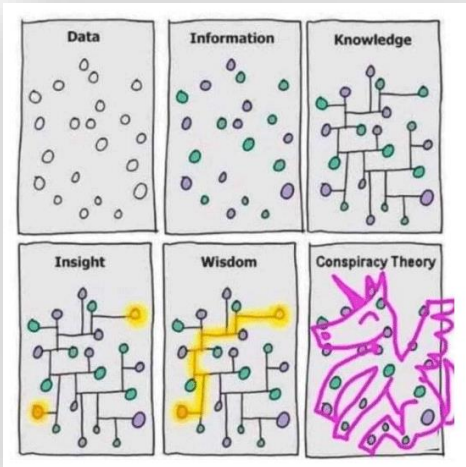
WHAT IS DATA?





- In the RDM context ***data is understood as a broad term*** including:
 - research ***material*** (such as any kind of physical artifacts)
 - research ***sources*** (such as various archive material)
 - data produced during the research (such as ***digitized copies*** of the aforementioned physical artifacts)
 - data ***collected by various methods*** (such as surveys, interviews, measurements, imaging techniques etc.)
 - ***curated collections***
 - ***annotation and coding*** of the material on various levels
 - all ***revisions of a data set*** produced in/for the analysis process
 - fieldnotes, diaries
 - physical and electronic ***lab journals***
 - ***source codes, algorithms*** and ***software***
 - etc.

WHAT IS DATA?



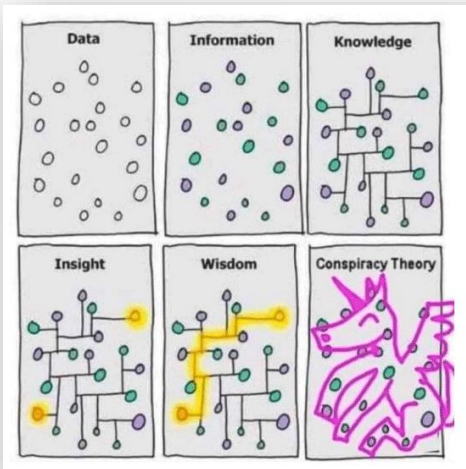
→ Everything that your research (results) is based on!

“All information that is needed to replicate a study should be preserved, and everything that is potentially useful for others.”
– Sarah Jones /DCC



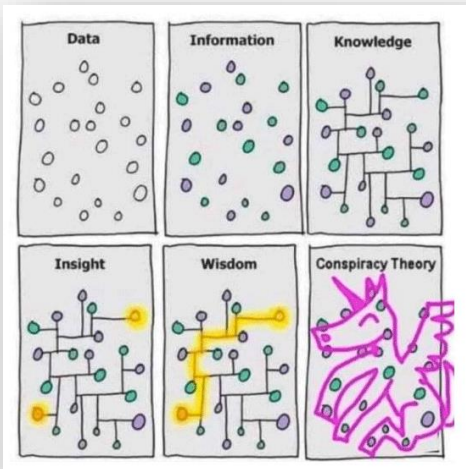
- What is data? Especially in RDM context?
 - "I do not have data, I have research material"
- What is the difference between research data and research material?
- Which part of the data are you responsible for?
- Which part has value after your project?
- Which part can be / should be destroyed?
- How to define the value?
- How much the preservation and archiving costs?
Could it be easier and cheaper to produce or collect the data again?

WHAT IS DATA?





WHAT IS DATA?



- General description of data
 - *What kinds of data are collected or reused? What file formats will the data be in?*
- Describe/list all datasets and material, e.g.
 - A) data collected by yourself
 - B) data reused in your project
 - C) data produced during your project
 - D) managerial documents, agreements, contracts etc.
- The categorization help you to understand the ownership and user rights issues



DEFINITIONS OF RESEARCH MATERIAL AND RESEARCH DATA

- Research material = Data
 - Used as synonyms
 - Data can be understood as a general term which requires a prefix such as qualitative, quantitative, numeric, computational, interpreted etc.
- Research material \neq Data
 - In many cases data only refers to digital material
 - Data is considered as computational, quantitative, machine-readable (e.g. Big data)
- Data produced in one project can be reused as raw data in another
 - From raw material to data
 - From data to analysis
 - From analysis to interpretation
 - From interpretation to research results
- Data can be duplicated, copied, enhanced, developed etc. from one project to another



CATEGORIZATION OF DATASETS

- **General description of data**

- *What kinds of data are collected or reused? What file formats will the data be in?*

- Here, outline a structure which is followed throughout the whole DMP

= describe/list all datasets and material which are discussed later in the plan, e.g.

A) Data collected by yourself

various locations; raw, non-catalogued private collections
various types

B) Data reused in your project

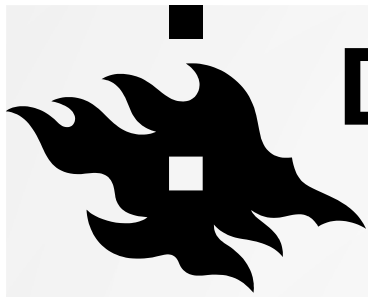
ready-made dataset in an archive
remember to cite the original creator or collector in your work!

C) Data produced during your project

notebooks, research diaries, field notes, comments, annotations, coding, and so on register a
PID for your datasets so you can be cited!

D) Managerial documents, agreements, contracts etc.

- Can be presented in a table together with file formats to save space!



DATASETS TABLE EXAMPLE

Data type	File format	Personal or sensitive data	Storing data and backups during the project	Documentation and metadata	Ownership and Agreements	Opening or publishing data after the project
Measurements	.xls .csv	No	Personal storage at UH (home folder)	readme.txt codebooks	UH and LUKE Agreement	Opening via publication at DRYAD or Zenodo
Gene sequences	.txt fasta	No. Collecting only from plants.	Group storage space	.FASTA	PI	NCBI Genome
Programme codes	.xml ASCII R-code	No	GitLab & Shared network drive hosted by UH	GitLab & readme.txt	Co-ownership of the research group	Via publication and Zenodo
Microscopy images	.tiff	No	Server storage space	OME-TIFF	PI	Electron Microscopy Public Image Archive (EMPIAR)
Lab notes	.doc .txt .pdf	No Patenting or commercializing?	Electronic lab notebook Scinote Cloud service	Programme generates metadata by itself	PI and me	No
Samples (applying from THL Biobank)	.xls	Anonymization will be done by Biobank	Freezer at the Institute of Biotechnology (PI's lab)	Unique identifier code	THL Biobank-licence Research agreement DMP	Samples discarded one year after publishing the results.
Questionnaire forms	Paper forms	Yes Data Controller UH	Locked filing cabinets in PI office.	codebook readme.txt	PI Informing participants	No, only metadata will be open in FSD. Forms discarded 2 years after project ends.
Spatial data about land use and forest stand	.tiff, Coloured	No, open data	Datacloud at UH (service coming soon)	Supplements at Etsin	National land survey of Finland: license CCBY	Processed data at Zenodo



RESEARCH DATA AND OPEN SCIENCE



OPEN SCIENCE

- Open science principles **do not demand opening everything** but teach researchers how and on what grounds publications, data and processes can be opened
 - Open Access to Research Publications, Data and Methods
 - to make research results available to everyone
 - to verify research results
 - to prevent scientific fraud
 - to make research data available for reuse
 - **Opening should not be an accident, but a controlled process!**
- **”As open as possible, as closed as necessary”**
– European Commission

Source: https://ec.europa.eu/research/openscience/pdf/openaccess/ord_extension_faqs.pdf



OPEN SCIENCE IS NOT ABOUT OPENING EVERYTHING

- For example, four stages of dataset availability in the Finnish Social Science Data Archive AILA

The screenshot shows the Aila website interface. At the top, there is a navigation bar with the Aila logo, 'Data Search Help' links, and user options like 'Suomeksi', 'Hello visitor!', 'Login', and 'Register'. Below this is a breadcrumb trail: 'Home / Data catalogue'. The main heading is 'Study descriptions in English', followed by 'In total 1273 datasets.' and a link to 'Display the data catalogue in Finnish.' A filter sidebar on the left includes 'Availability' (with options A, B, C, D) and 'Type' (Qualitative, Quantitative). The main content area explains that users can filter studies by availability status, data file language, or type, and can sort by ID number, title, availability, data type, or publication date. A red circle highlights the 'Dataset availability' section, which lists: (A) openly available for all users without registration, (B) available for research, teaching and study, (C) available for research only (including e.g. Master's, licentiate and doctoral theses), and (D) available only by permission from the data depositor/creator.



OPEN SCIENCE AND OPEN DATA

- Research is publicly funded -> the collected/created data should benefit the whole society
- Funders want that the funded research gets as wide an impact as possible and that the data is reused
- Not everything can be opened! > you can still open the metadata of your data



RESEARCH DATA AND FAIR PRINCIPLES

FAIR DATA: WHAT YOU NEED



F

FINDABLE

- persistent identifiers
- metadata



A

ACCESSIBLE

- where the data can be found and under what access conditions
- doesn't necessarily mean open
- open formats



I

INTEROPERABLE

- standards
- ontologies



R

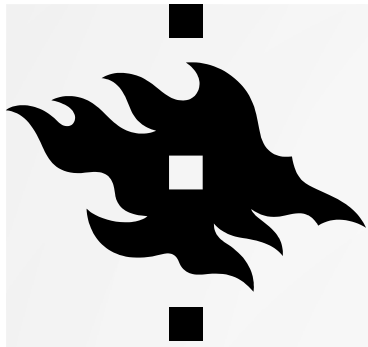
RE-USABLE

- licenses
- documentation



FAIR PRINCIPLES IN YOUR DATA





EVERYTHING STARTS WITH PLANNING

DMP – DATA MANAGEMENT PLAN

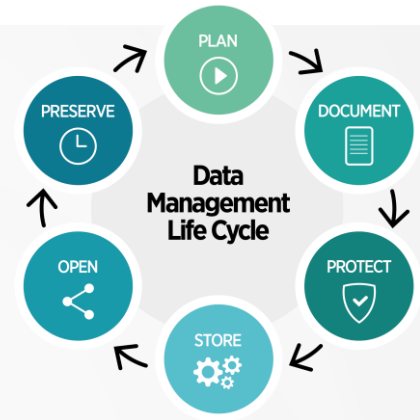


WHY YOU SHOULD WRITE A DATA MANAGEMENT PLAN (DMP)

- Helps you to consider **the whole research life cycle in advance!**
- It is good research practice and helps you to save time and money
- You will reduce the risk of losing your data.
- You will be able to anticipate complex ownership and user rights issues in advance.
- DMP supports making your data FAIR: Findable, Accessible, Interoperable and Re-usable. This will increase data reuse as well as visibility of your project.
- You will meet your funder's requirements.
- Your DMP reflects your researcher skills.



DMP AND THE DATA MANAGEMENT LIFECYCLE



- Writing a DMP helps you to get a full picture of the lifecycle of your data and all the things you need to do are more manageable
- DMP is a check list for your data
- You need some information even before you start collecting your data
 - Before informing research participants you need to know where you will open the data or how/when you will destroy the data
 - Data repositories may require certain kind of metadata, it's a good idea to follow that practice already during the project
- Do you need special storage solutions for your data? E.g. You have a huge amount of data or your data is very sensitive
- Careful planning is needed on how to manage sensitive data particularly in field conditions



DATA MANAGEMENT PLAN (DMP)

- DMP is an integral part of good research practice and ensures research integrity and quality. It helps you to manage your data, meet funder requirements and help others to use your data if shared
- DMP is a living document – it is updatable and reviewable. Create your data management plan early and review it regularly throughout the research project.
- DMP length: 1-2 pages
- DMP is not a research plan – it complements it. Avoid redundancy/overlapping with the research plan!

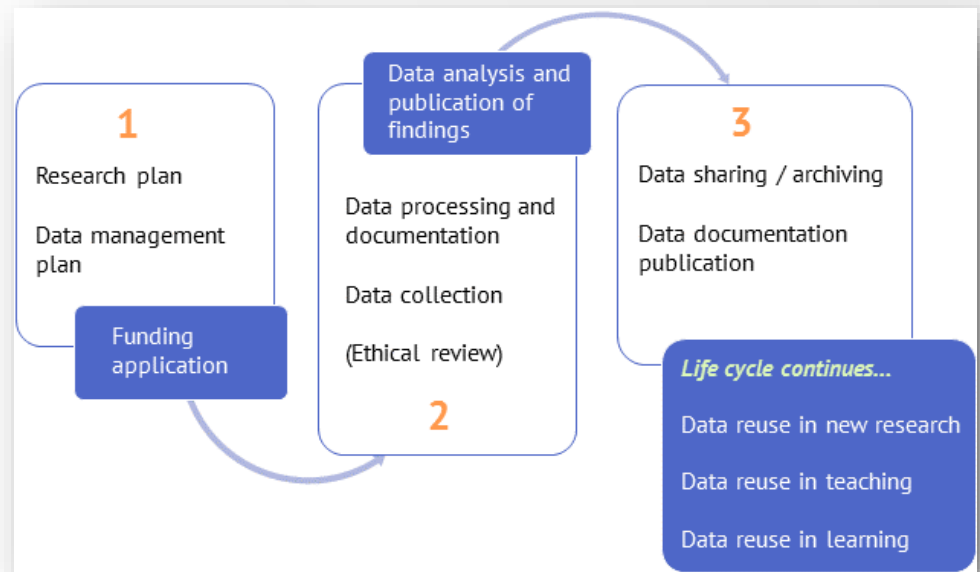


Figure:

<http://www.fsd.uta.fi/aineistonhallinta/en/why-are-research-data-managed-and-reused.html#research-data-life-cycle>



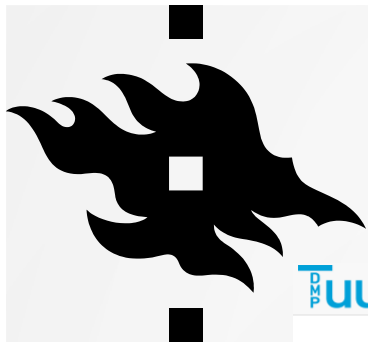
DATA MANAGEMENT PLAN (DMP)

- Write only something you yourself understand!

THAT IS: Reflect your own project and data; DO NOT COPY&PASTE
EXAMPLE ANSWERS!

”...I will use XML...” and XML was never used or even mentioned elsewhere.

- A good DMP is clearly written and consistent (within itself and as a part of the whole plan)
- Advice on how to write out the different parts of the DMP is available from our [RDM webpage](#)! The topics in the page are organized in the same order as in the Academy of Finland DMP.



Welcome

Data management planning tool DMPTuuli helps you to create, review, and share data management plans that meet institutional and funder requirements.

Join the growing number of researchers that have adopted DMPTuuli:



10 725 Users



12 612 Plans



50 Organisations

You can download funder templates without logging in, but from DMPTuuli you will find tailored guidance from many research organisations, including universities and service providers like the Finnish Social Science Data Archive. Why not sign up for an account and try it out?

[Sign in](#) [Create account](#)

* **Email**

* **Password**

[Forgot password?](#)

Remember email

[Sign in](#)

- or -

[Sign in with institutional credentials \(HAKA only\)](#)

- DMPTuuli is an easy way to do a data management plan! Go to → www.dmptuuli.fi
- Click on 'Create account' and register for a new account **or** use HAKA registration
- You'll need your ORCID-number and the plan number from the Academy of Finland electronic application portal
- In Tuuli you will find a number of templates for different funders and institutions and generic and specific guides to help you answer the DMP questions



KEY COMPONENTS OF RDM

KNOW YOUR DATA!

1. **Know your data – documentation and metadata**
2. Ethical and legal issues
3. Solutions for storing data during the research project (data in dynamic state)
4. Publishing and archiving (static phase) data (after the research project)



KNOW YOUR DATA!

DOCUMENTATION AND METADATA

- What kind of data do you have?
- What software and file formats do you use?
- How much data do you have?
- How do you describe what your data is about?
- Do you have a system for file naming?
- Can you locate the latest version of your data?



DOCUMENTATION AND METADATA

- **Description of your data = Metadata = data about your data**
- What data is needed about your data to make it understandable to yourself and to others?
- **Data about data:** file names, file formats, version control, where are the files located (directory structure), variable descriptions etc.
- **Project metadata** (the context of your data): How, where and when was the data collected? Subject, keywords.
- **Administrative metadata:** administrative documents, and other research related descriptions
- Metadata can be opened e.g. in Research Data Finder Etsin (<https://etsin.avointiede.fi/en/>)
- **Data level documentation is part of the research plan!**



SOME TIPS

- Tag your files with keywords for easier discovery
- File naming system: create a simple and systematic naming of your files. Follow it consistently!
- **Make your research group & collaborators use the file naming system.**
- Directory structure: create a folder structure that suits the needs of your project
- Version control: use automatic version control if possible
- Have README-files together with your data: provide information about data files so that they are understandable
- Data dictionaries and codebooks: Dictionaries explain the variables used in a dataset and codebooks are collections of codes, algorithms and calculations used

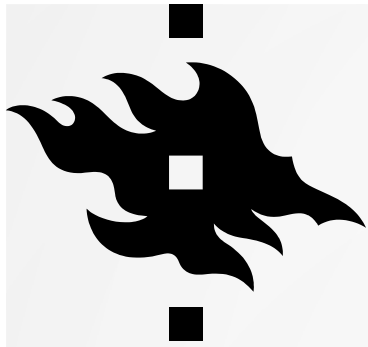
Documentation guide:

<https://www.helsinki.fi/en/research/guide-for-data-documentation>

PDF-guide:

<https://doi.org/10.5281/zenodo.1683181>





EXAMPLE: NAMING TIPS FOR FILES

- Balance with the amount of **elements** in the name: too few making it too general vs. too many hinder understandability. Limit the name to 32 characters or less.
- Use meaningful **abbreviations**
- Order the elements **from general to specific**
- Use the underscore (_) as element delimiter and hyphen (-) or capitalizer to delimit words within an element. Don't use special characters: & , * % # ; * () ! @\$ ^ ~ ' { } [] ? < >.
- **Time** should be ordered: year, month, day (YYYYMMDD) or more specifically if needed: hours, minutes, seconds (HHMMSS)
- For **version control** use the letter V followed minimum by two digits (V06), and extend it if needed for minor changes (V06-02). Remember the leading zeros to make sure files sort correctly.
- Write a **read me file about the naming system** and explain abbreviations



METADATA STANDARDS

- Use metadata standards if possible
- Describes data in a controlled way
- Field and disciplinary specific
- Typically repositories use a specific metadata standard and they usually assist you with the documentation.
- Minimum requirements for data documentation, see for example [Qvain Light](#).
- [Disciplinary Metadata](#) / Digital Curation Centre DCC
- [Metadata Standards by Subject](#) / Research Data Alliance RDA
- [General Research Data](#) / Digital Curation Centre DCC.

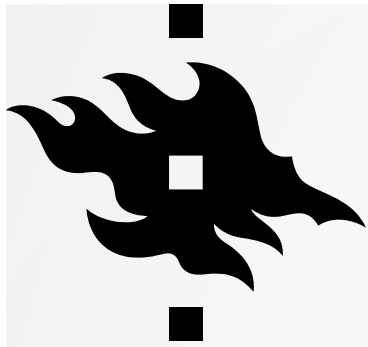
Examples (by Stanford University Libraries)

[Basic Metadata](#)

[Advanced Metadata](#)

An example of dataset in FSD's Aila:

https://services.fsd.uta.fi/catalogue/FSD2926?tab=description&lang=en&study_language=en



KEY COMPONENTS OF RDM

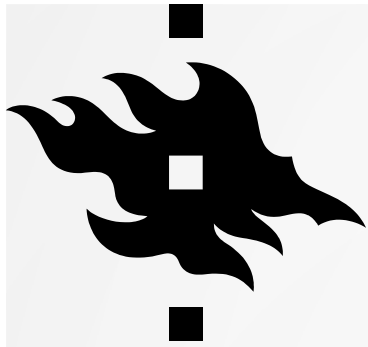
KNOW YOUR DATA!

1. Know your data – documentation and metadata
- 2. Ethical and legal issues**
3. Solutions for storing data during the research project (data in dynamic state)
4. Publishing and archiving (static phase) data (after the research project)



ETHICAL AND LEGAL ISSUES

- Do you know who has rights to your data?
- Does your research contain personal data?
- Do you know if your data is sensitive or not?



ETHICAL AND LEGAL ISSUES

Ownership and usage rights?

- Who owns your data?
- Who has usage rights to your data during your research project?
- Who has the right to issue re-use of your data?

→ **Make agreements with your organization and your research team as early as possible!**

Handling personal and sensitive data?

- Do you recognize personal data? Sensitive data? Sensitive personal data?
- How do you inform your research participants?
- Who is the controller when processing personal data?
- Anonymization and pseudonymization! There is a difference.

Ethical review

- What it is and when do you need it?



WHAT IS PERSONAL DATA?

- **Data related to an identified or identifiable living natural person**
 - Identifiable: Directly or indirectly, for example by combining data from multiple sources
 - Example: **A123**, **a 45-year-old man**, **lives in Smalltown** (a village with 50 residents), **favourite color: green**
 - Data may be personal data even if you do not know a person's name, social security number, contact details or other direct identifiers
- **It is very likely that your project involves processing of personal data**



COMMON MISTAKES RELATED TO DATA PROTECTION

- **Using free consumer versions of cloud services and online survey services**
 - It is not possible to conclude the agreements required by law, risk of insufficient level of information security
- **Anonymisation vs. pseudonymisation**
 - The threshold of anonymisation is very high. In most cases, the data is actually pseudonymised, not anonymised.
- **Wrong definition of "personal data"**
 - This may lead to problems if you, for example, promise to delete all personal data while actually meaning that direct identifiers will be deleted.
- **Missing agreements**
 - An agreement is required whenever personal data is provided to recipients outside the University of Helsinki: please contact the legal services
- **Informing participants**
 - Data subjects (people whose personal data are processed) must be informed about the processing of their personal data when informing is reasonably possible – remember to include data protection information in the documents you provide to participants
 - Simple solution: use the data protection notice template on Flamma/Yammer



SENSITIVE DATA AND SPECIAL CATEGORIES OF PERSONAL DATA

- Sensitivity of data often needs to be determined on a case to case basis
- According to the GDPR special categories of personal data includes:
 - Racial or ethnic origin
 - Political opinions
 - Religious or philosophical beliefs
 - Trade union membership
 - Genetic data
 - Biometric data for the purpose of uniquely identifying a natural person
 - Health
 - A natural person's sex life or sexual orientation
 - Criminal convictions or offences or related security measures
- Anonymisation prerequisite for making sensitive data completely open
- For data that is not made completely open requirements vary by archive :
 - The Finnish Social Science Data Archive does not archive sensitive data unless it has been anonymized
- Anonymisation can be tricky — the Finnish Social Science Data Archive has to further anonymize about 80 percent of data they receive as anonymized



ANONYMISATION AND PSEUDONYMISATION

- **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.
- After anonymisation not possible for a research participant to withdraw from a study

Source: <http://www.fsd.uta.fi/aineistonhallinta/en/anonymisation-and-identifiers.html>



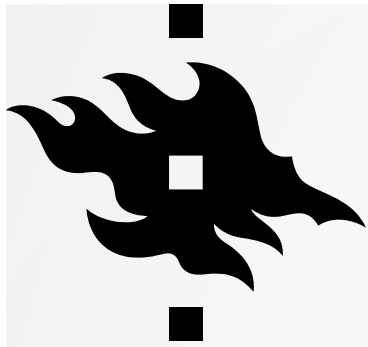
WHO IS THE CONTROLLER WHEN PROCESSING PERSONAL DATA

When processing personal data in research

- Usually the data controller is the University
- The data controller may be the researcher if doing research as an individual rather than as a representative of the university
 - e.g. master's thesis or dissertation research if not employed at the University or otherwise working in close connection with the University

Contact Univ of Helsinki Legal Affairs – researchlawyers@helsinki.fi - if any uncertainty about who is the controller

See (Privacy Notice for Scientific Research v4_2019 **Privacy Notice for Scientific Research v4_2019**;
in **Flamma** in the **Data Protection guide for researchers** page
<https://flamma.helsinki.fi/en/group/tutkimuksen-tuki/tutkimuksen-tietosuoja-asiat>)



DATA OWNERSHIP

- No simple guidelines: Discuss with the PI
- Affected by factors such as:
 - Source of funding
 - Cooperation with other parties
 - Employment relationship
- A transfer of rights undertaking must be signed by all participating researchers in projects that involve external funding paid to the University or cooperation with parties outside the University:
<https://flamma.helsinki.fi/fi/group/tutkimuksen-tuki/sopimukset-sopimusohje-ja-neuvottelut>



ETHICAL AND LEGAL ISSUES

Data protection guide for researchers

<https://flamma.helsinki.fi/en/group/tutkimuksen-tuki/tutkimuksen-tietosuoja-asiat>

Ethical Review Boards in the University of Helsinki

<https://www.helsinki.fi/en/research/ethical-review-board-in-the-humanities-and-social-and-behavioural-sciences>

<https://www.helsinki.fi/en/research/viikki-campus-research-ethics-committee>

National Committee on Medical Research Ethics TUKIJA

<https://tukija.fi/en/frontpage>

The Finnish National Board on Research Integrity TENK

<https://tenk.fi/en>



KEY COMPONENTS OF RDM

KNOW YOUR DATA!

1. Know your data – documentation and metadata
2. Ethical and legal issues
- 3. Solutions for storing data during the research project (data in dynamic state)**
4. Publishing and archiving (static phase) data (after the research project)



STORING SOLUTIONS BY THE IT DEPARTMENT

- How much disk space do you need?
- How do you back-up your data?
- Do you process your data on a daily basis?
- Do you only use it yourself or do you need to share it with others?



STORAGE SOLUTIONS

Storage

- Typically during the active phase of your project (vs. archiving and publishing)
- Where to store?
 - Storage solutions at UH table: <https://wiki.helsinki.fi/x/kgV5FQ>

Back up and version control

- Do not rely on manual back ups
- How do you control versioning in your filing system?

Access control

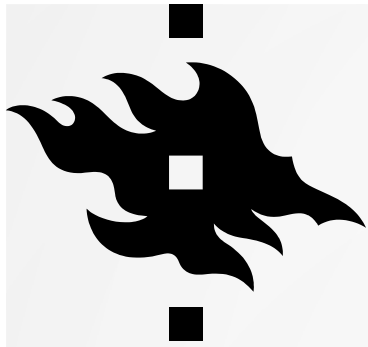
- How do you keep your data secured?
- How do you share it safely with your colleagues?
- Sharing vs. publishing?



STORAGE SOLUTIONS

Encryption

- When transferring sensitive data, encryption is recommended.
- Several options are available for different needs:
 - 7-Zip (Encrypting local files and folders)
 - Cryptomator (Encrypting local files and folders)
 - S/MIME, GNUPG (Encrypting e-mail)
 - Microsoft Word & Excel (Encrypting documents)
 - Funet filesender (Encrypting a file transfer)
- TIP: When encrypting files using a password, use long (20 characters) and complex passwords in order to protect your data! Don't use any words found from general languages.
- The drives of UH Windows 10 computers are encrypted by default. For other operating systems Helpdesk will provide more information on encryption options



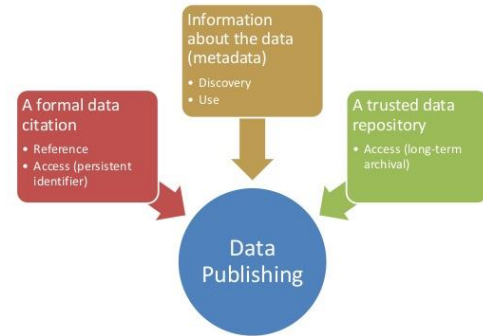
KEY COMPONENTS OF RDM

KNOW YOUR DATA!

1. Know your data – documentation and metadata
2. Ethical and legal issues
3. Solutions for storing data during the research project (data in dynamic state)
- 4. Publishing and archiving (static phase) data (after the research project)**

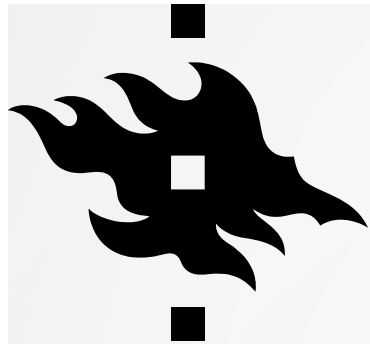


Data Publishing needs to support data discovery, reference, access, and use



PUBLISHING AND ARCHIVING DATA

- Is your data valuable for others?
- Where do you archive your data?
 - Selecting a data repository: curated – non-curated; PIDs?
- How do you prepare your data for archiving?
 - Opening data or metadata?
- Licensing your data



PUBLISHING AND ARCHIVING DATA: CERTIFIED ARCHIVES



- When publishing data, it is advisable to use curated data archives
- Curated, Core Trust Seal certified archives in Finland:
 - [The Language Bank of Finland](#)
 - [The Finnish Social Science Data Archive](#)
- Note that Academy of Finland also accepts many other data archives, such as IDA, Zenodo, Figshare etc. for publishing and archiving data.



REGISTRY OF RESEARCH DATA REPOSITORIES


- re3data.org is a global registry of research data repositories that covers research data repositories from different academic disciplines.
- It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions.
- A tool for the easy identification of appropriate data repositories to store research data. Prefer curated archives if there are those in your field of research.
- <https://www.re3data.org/>





Qvain and Etsin

- [Qvain](#) is a research dataset metadata tool, provided by CSC. With the tool you can publish your metadata to [Etsin](#) which is a database for documentation of datasets.
- User guides available here: <https://www.fairdata.fi/en/qvain/qvain-user-guide/>
- Qvain Can use files from IDA and other repositories and includes a metadata auto checking tool.

Fairdata.fi  **Qvain** [User Guide](#)


Qvain

Research Dataset Metadata Tool

By using Qvain the user agrees:

- That he or she has asked consent from all persons whose per...
- the Terms of Usage

[Login](#)

Fairdata.fi  **Etsin** [HOME](#) [DATASETS](#) [Help](#) [Add dataset](#) [FI](#) [Login](#)

Search datasets

Search term

9903 datasets	5517 keywords	37 fields of science	57 projects
-------------------------	-------------------------	--------------------------------	-----------------------

What is Etsin?

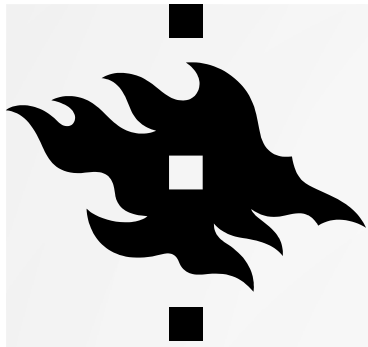
Etsin enables you to find research datasets from all fields of science. Etsin contains information about the datasets and metadata in the national Finnish Fairdata services. We also currently harvest information from the Language Bank of Finland, the Finnish Social Science Data archive and the Finnish Environmental Institute, and new sources will be included.



CREATIVE COMMONS LICENSES

- Licensing your work makes possible a wide distribution of your work, as the license tells the users how you would like it to be shared. The Creative commons -licenses (latest version 4.0) have become a standard for open publishing (<https://creativecommons.org/>)
- You can choose the license when you publish your work (for example, in Tuhat - publication database or in data-archive of your choice, but not in commercial publishing platforms like Academia.edu or Research Gate).
- University of Helsinki recommendations: CC 0 for research data: author gives up all copyrights, but when material is used in research, the guidelines on good research practice should be followed, that is, the sources and authors should be named. Use CC BY if you want to be mentioned as creator of the data.
- Avoid CC BY NC (non-commercial)! The copyright law in Finland does not determine what is commercial use and the person using the works has to consider each case separately (for example, a work with this license cannot be shared in a commercial or advertise-funded blog, webpage, repository or social media platform)
- On different CC-licenses, see our License guide: <http://libraryguides.helsinki.fi/oa/eng/license>





HOW TO CITE DATA

- Data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.
- **Required information:** **Author** - Who is the creator of the data set? This can be an individual, a group of individuals, or an organization, **Title** - What is the name of the data set, or what is the name of the study? **Publication year**, **Producer** - The organization that made the creation of the data set possible, **Distributor** - The organization that makes the data set available for downloading and use, **Edition or Version** - Is there a version or edition number associated with the data set? **Material designator** - What type of file is the data set?, **Access information** - URL or other persistent identifier like DOI.
- Tracing data : Data citation roadmap for Finland:
<http://urn.fi/URN:NBN:fi-fe201804106446>
- <https://www.force11.org/datacitationprinciples>

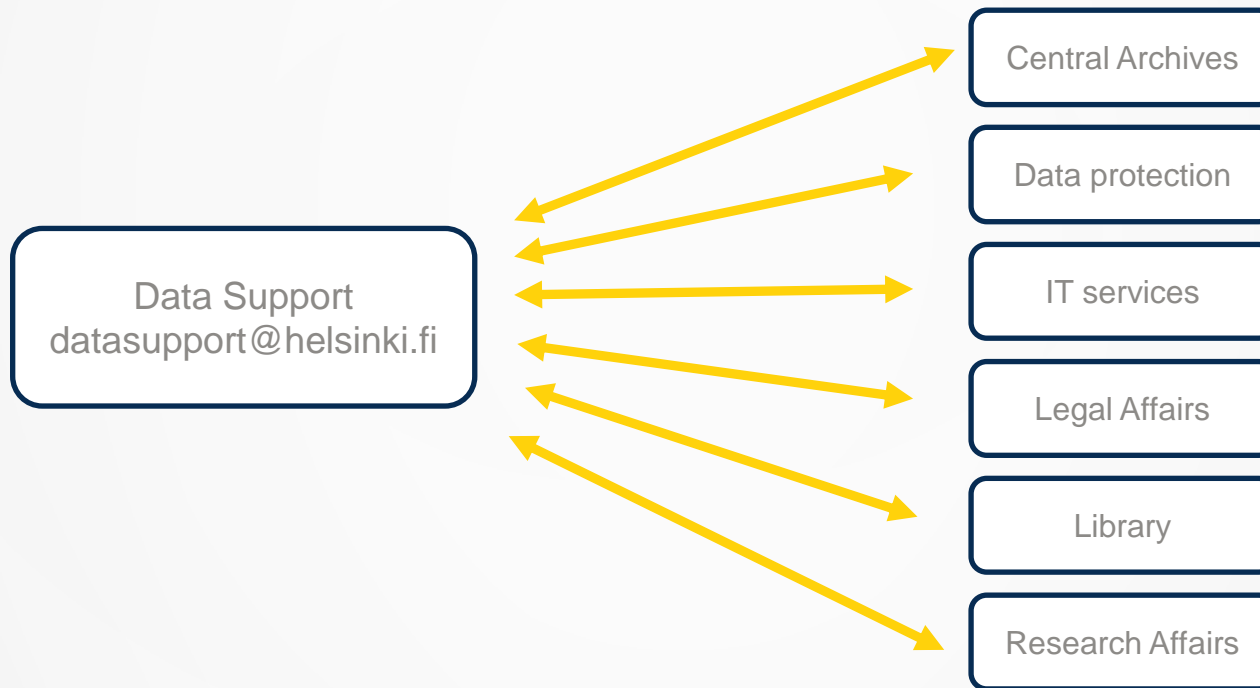
Data Management Guidelines [Online]. Tampere: Finnish Social Science Data Archive [distributor and producer]. <<http://www.fsd.uta.fi/aineistonhallinta/en/>>. urn:nbn:fi:fsd:V-201504200002 (cited dd.mm.yyyy.)



HELP AND SUPPORT



DATA SUPPORT AT THE UNIVERSITY OF HELSINKI



More information: <https://www.helsinki.fi/en/research/university-of-helsinki-datasupport>

RDM RESEARCH GUIDE



- Research Data Management Web-page <https://www.helsinki.fi/en/research/research-environment/research-data-management>
- Might be handy to keep the page open in another window while outlining a DMP with Tuuli
- University of Helsinki Research Data Policy: <http://www.helsinki.fi/kirjasto/en/get-help/management-research-data/research-data-policy/>

RESEARCH / RESEARCH ENVIRONMENT /

RESEARCH DATA MANAGEMENT

We provide assistance in research data management through out the research life cycle including data organization, storing and sharing. On this page you can find guidance to the above mentioned topics and how to answer the general questions of a Data Management Plan (DMP).

- ↓ Data policy and support services
- ↓ Data management planning
- ↓ 1. Research data
- ↓ 2. Ethical and Legal Compliance
- ↓ 3. Documentation and metadata
- ↓ 4. Storing data and access control
- ↓ 5. Opening data and long-term preservation after the research project
- ↓ 6. Data management responsibilities and resources

Data policy and support services



Research data policy	→
Research data service catalogue	→
Data Support at the University of Helsinki	→
Courses & workshops	→

Every member of the University of Helsinki community is responsible for good data management. The University provides research data infrastructures that includes tools and services for supporting the management, use, findability and sharing of data as well as with the capacity for storage, preservation, computing and processing.

Data Support at the University of Helsinki assists researchers in the management of research data. Data Support is a network of experts from the university library, IT Services, Central Archives, Research Affairs, Personnel Services, and Legal Affairs. You can contact us by email: datasupport@helsinki.fi

On this page you can find University of Helsinki guidance for research data management.

Data management services available from Data Support

datasupport@helsinki.fi

Long term preservation (PAS)



LINKS

Instructions:

- Making a research project understandable - Guide for data documentation: <https://doi.org/10.5281/zenodo.1914401>
- Instructions on how to get a group folder at UH: <https://helpdesk.it.helsinki.fi/en/instructions/saving-and-sharing/group-storage-space/group-storage-space>
- Instructions for UH OneDrive: <https://helpdesk.it.helsinki.fi/en/instructions/collaboration-and-publication/office-365/office-365-onedrive-business>
- Information security and terms of use: <https://www.helsinki.fi/en/it/information-security-and-terms-of-use>
- IT Center's service price list: <https://flamma.helsinki.fi/en/group/it-ja-puhelin/it-centers-service-price-list>

Service catalogs:

- IT services at the university: <https://www.helsinki.fi/en/it/it-services-at-the-university>
- CSC's service page: <https://www.csc.fi/services>
- Cloud guide: <https://wiki.eduuni.fi/display/cloudguide/Cloud+Guide>
- EUDAT Services: <https://www.eudat.eu/catalogue>



FINNISH SOCIAL SCIENCE DATA ARCHIVE & THEIR DATA MANAGEMENT GUIDELINES

- FSD's Data Management Guidelines <http://www.fsd.uta.fi/aineistonhallinta/en/>
- Excellent guide on all RDM issues!
- Might be handy to keep the page open in another window while outlining a DMP with Tuuli
- Especially useful on dealing with sensitive data!

The screenshot displays the website's navigation bar with the logo, 'FINNISH SOCIAL SCIENCE DATA ARCHIVE', and 'DATA MANAGEMENT GUIDELINES'. It includes links for 'Home', 'Chapters', and 'Citation', along with a search bar for 'Suomeksi' and 'Search FSD'. The main content area features a large illustration of a blue cube character holding documents, with the heading 'Why Are Research Data Managed and Reused?'. Below this, a grid of six numbered sections is visible: 1. Why Are Research Data Managed and Reused?, 2. Data Management Planning, 3. Copyright and Agreements, 4. Informing Research Participants, 5. File Formats and Software, and 6. Processing Quantitative Data Files. Each section contains a list of key points or sub-topics.

- Table of Contents
- Why Are Research Data Managed and Reused?
- Data Management Planning
- Copyright and Agreements
- Informing Research Participants
- File Formats and Software
- Processing Quantitative Data Files
- Processing Qualitative Data Files
- Anonymisation and Personal Data
- What is personal data?
- What type of information constitutes identifiable data?
- Processing research data containing identifiable data
- Types to be considered
- When are data generated and when anonymised?
- The conditions in order to protect only the necessary
- Anonymisation principles
- Anonymisation of quantitative data
- Anonymisation of qualitative data
- Other topics
- Data Description and Metadata
- Physical Data Storage
- Examples

Anonymisation and Personal Data

What is personal data?

According to the definition given in the General Data Protection Regulation (GDPR), personal data means any information relating to an identified or identifiable natural person. A natural person is considered identifiable if they can be identified by an identifier such as a name, an identification number, address data or other identifier or by one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. (EU) General Data Protection Regulation Article 4, Paragraph 1. By the definition, when it comes to research data, personal data are not limited to information relating to research participants. Research data may also contain identifiers relating to research subjects, family and friends or other third parties. Identifying information relating to these persons also constitutes personal data.

There are no limitations regarding the nature and character of personal data. Any information related to a natural person may be personal data. This includes opinions, opinions, attitudes and value judgments. Personal data may be objective or subjective. The information may be in a verbal or non-verbal form or in electronic form. The information may also be an individual's private or family life, health, physical characteristics, professional activities, and economic or social behaviour.

What kind of information constitutes identifiable data?

Personal data are any kind of data that may be used to identify a natural person or a cluster of persons, such as individuals in the same research. Identification can occur on the basis of all or some factors specific to the physical, psychological, mental, economic, cultural or social identity of an individual or individuals. Data that are not directly about people can also be personal if they contain identifiers. For example, secondary personal data could be the department information on the occurrence of fires, which may include addresses. (Eliot, et al. 2016)

Information that is sufficient on its own to identify an individual includes a person's full name, social security number, email address containing the personal name, and biometric identifiers (fingerprints, facial image, voice patterns, iris scan, hand geometry or manual signature). These types of data are called direct identifiers.

Other information that may be used to identify an individual, but which does not include a personal address, phone number, mailing registration number, bibliographic citation of a publication by the individual, email address not in the form of the personal name, web address that is not a long containing personal data, email address, very rare disease, or profession that is only one person at a time (e.g., chess player in an organisation). A web event can also reveal the identity of an individual. The Finnish Social Science Data Archive (FSD) collects these types of information among indirect identifiers.

All FSD strong indirect identifiers also include the types of codes that can be used to unequivocally identify an individual from among a group of individuals. These include, for instance, a student ID number, insurance or bank account number, IP address of a computer etc.

OUR SERVICES

- Data Management in Aika
- Data Management Guidelines
- FAQs
- FSD Bulletin

