

# RDM BASICS

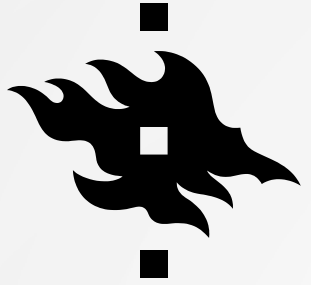
Alexi Peura / Data Support, University of Helsinki

2022–2023

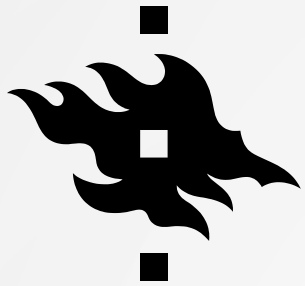


# TODAY'S THEMES

- Data
- Data protection
- Legal and ethical issues
- Metadata and licences
- Data storage
- Opening the data and long-term preservation

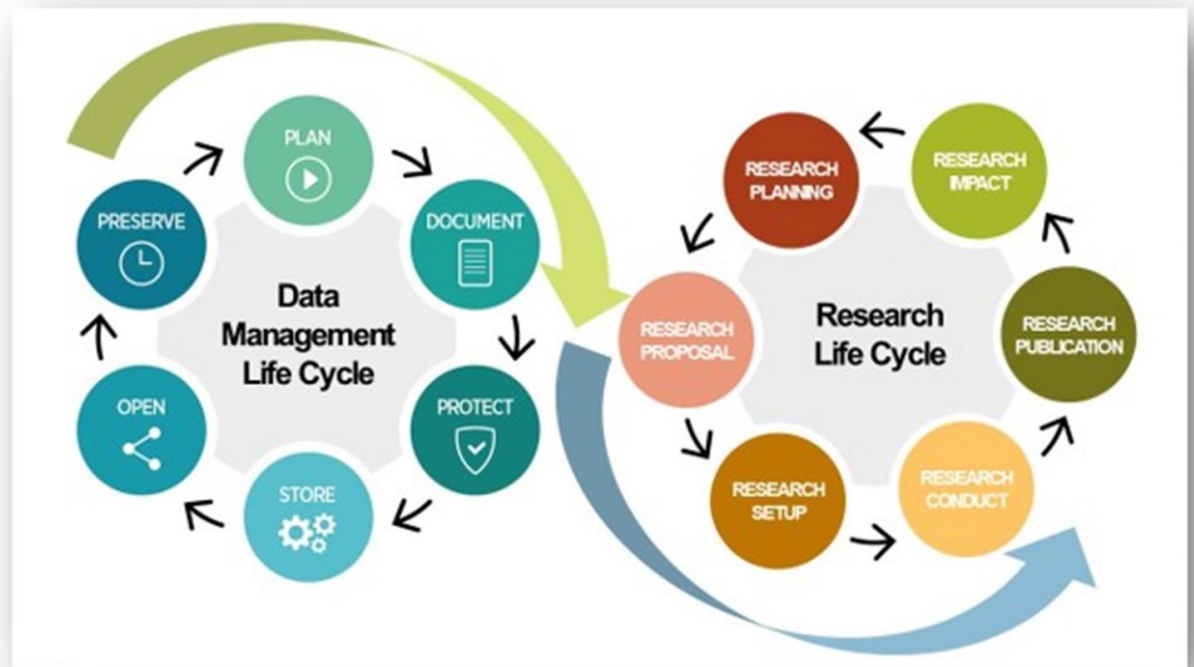


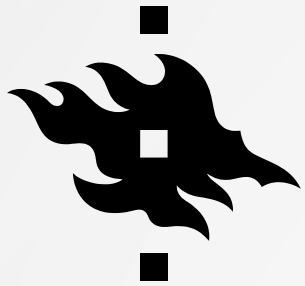
# DATA



# WHAT IS RESEARCH DATA AND ITS MANAGEMENT?

- Research data is any data that forms the **basis for the research's analysis and results.**
- Research data management (RDM) is a **general research skill** that focuses on **responsible, open and sustainable use of research data.**





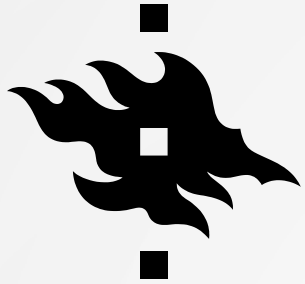
# BASIC PRINCIPLES

- Most things in RDM need to be considered on a **case-by-case-basis**.
- Still, there are some **basic principles** that help understanding and figuring out some more specific issues.
- Therefore, **the better you are acquainted with the basic principles of RDM, the better you can acknowledge and potentially resolve issues relevant to you.**
- You can always contact Data Support ([datasupport@helsinki.fi](mailto:datasupport@helsinki.fi)).
- The key takeaway from this lecture is: **plan ahead!**

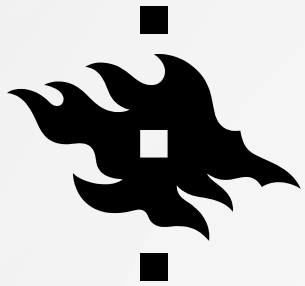


# A FICTITIOUS EXAMPLE OF RESEARCH DATA

- A researcher – let's call her Fatima – is conducting research on **religious beliefs of fans of extreme metal in Finland**.
  - Her main research methods are **interviews**, and therefore the conducted interviews form the majority of her research data.
  - In Fatima's case, she conducts the interviews face-to-face and records them using a dictation machine. Afterwards she transcribes the interviews.
  - After the interviews she also takes **pictures** of the data subjects, focusing on their clothing, tattoos or other accessories.
- ➔ Fatima's research data include at least the **recordings and transcriptions** of the interviews and **pictures** taken.



# WHAT IS YOUR RESEARCH DATA?

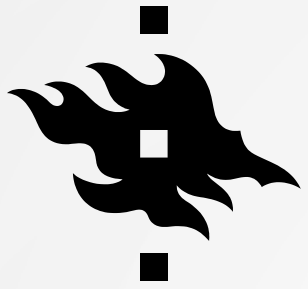


# DATA SHEET

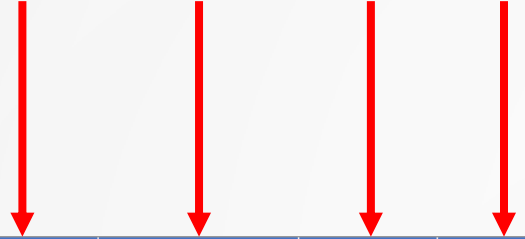
<https://tinyurl.com/datasheettable>

	Data type	Source of the data	File Format	Size estimate	Sensitivity / Personal data + controller	Owner / other agreements?	Documentation	Storage during project	Opening	Long term archiving
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										





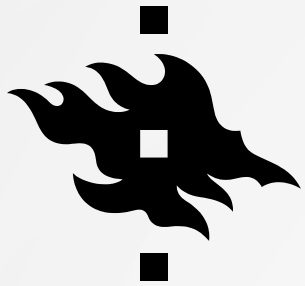
e.g. 100 TB



	Data type	Source of the data	File Format	Size estimate	Sensitivity / Personal data + controller	Owner / other agreements?	Documentation	Storage during project	Opening	Long term archiving
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										



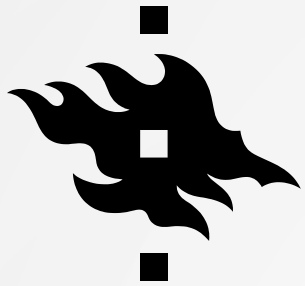
# DATA PROTECTION



# PERSONAL DATA

- **Any data that can be traced back to an individual is considered personal data.**
- Under GDPR collecting, storing, using and sharing personal data is strictly regulated.
- Due to a loose definition of personal data, most of the research projects do use personal data in one way or another.
- Additional question is the **sensitivity** of personal data: if the information deals with subject matters that can be considered tabu, personal in nature or, if leaked, in any way harmful to the data subject, this data is considered sensitive.
  - Different sensitive personal data can have different **risk levels**.

Read more: <https://www.fsd.tuni.fi/en/services/data-management-guidelines/anonymisation-and-identifiers/>



# IDENTIFIERS

- When dealing with personal data, the data subjects can be identified with two principle ways aka **identifiers**:

Can refer only to a particular person

- 1. Direct identifiers:** information that can be directly traced back to an individual.

- Examples of strong direct identifiers: full names, social security numbers, DNA, fingerprints, picture/video of a face, etc.
- Examples of weaker direct identifiers: address, phone number, rare trait (e.g. hobby, disease, occupation), etc.

Can refer only to handful of persons

- 2. Indirect identifiers:** seemingly anonymous information that can be, if taken in bulk, used to figure out the identity of the data subject.

- **As long as indentifying a person directly or indirectly is possible from the data, the data is personal data.**



# EXAMPLE OF INDIRECT IDENTIFYING

- During Fatima's interview, data subject X123 reveals the following information:
  - He's a male between ages 30 and 39
  - He lives in Helsinki
  - He works at a library
  - He has long hair
  - He writes semi professionally
  - He hosts several podcasts
  - He listens to metal music usually 8–12 hours per day





# ANONYMIZATION AND PSEUDONYMIZATION

- Anonymization is the process of **erasing the connection between data and its subject.**
- Absolute anonymization is virtually impossible: theoretically, if a party with near-limitless resources wants to break any anonymization, they will be able to do so.
  - The responsibility of the "anonymizer" is to make this risk **as theoretical as possible!**
- Pseudonymization gets confused with anonymization but these terms are not synonymous.
- In pseudonymization, the direct identifiers (e.g. names) get **replaced with aliases.**
- **Pseudonymous personal data is still personal data!**

1. Original dataset with personal data

name	fav food	shoe size
Matrix, John	sandwich	11,5
Deckard, Richard	noodles	9
Standish, Claire	candy	7
Baines, Lorraine	pizza	7,5
Macready, Randall	ice cream	9,5

2. Pseudonymization. Removing personal data and replacing it with non-personal coding

person 1	fav food	shoe size
person 2	sandwich	11,5
person 3	noodles	9
person 4	candy	7
person 5	pizza	7,5
Matrix, John	cream	9,5
Deckard, Richard		
Standish, Claire		
Baines, Lorraine		
Macready, Randall		

3. Pseudonymized dataset with separate pseudonymization key in a separate location.

name	fav food	shoe size
person 1	sandwich	11,5
person 2	noodles	9
person 3	candy	7
person 4	pizza	7,5
person 5	ice cream	9,5

---

pseudonymization key	
person 1 =	Matrix, John
person 2 =	Deckard, Richard
person 3 =	Standish, Claire
person 4 =	Baines, Lorraine
person 5 =	Macready, Randall



# SENSITIVE DATA, SPECIAL CATEGORIES

- Sensitive data generally refers to GDPR's **special categories of personal data** but is not limited to those.
  - [Special categories](#) include data that reveals e.g. ethnic or racial origin, political opinion, religious or philosophical belief, union membership, health status, genetics, sexual orientation, etc.
  - Not included in GDPR but considered sensitive personal information include: trade secrets, individual behaviour, highly personal writings (e.g. diary entries), financial status, etc.
- A rule of thumb: sensitive data is anything that can be used to oppress or hurt a person, should the information end up in wrong hands.





# SENSITIVE DATA, OTHER TYPES

- In addition, information relating to e.g. **endangered species**, **patents** and matters of **national defence** are considered sensitive.
- Consider, e.g., a following (fictitious) example:

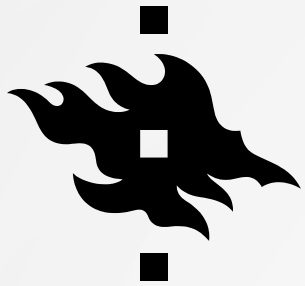
*A landowner is planning to have his forest logged. He finds from a public online database that tracks all endangered species' nests in Finland, and realizes that there are some living in his forest, which would prevent or limit the planned logging. He knows destroying the nests is illegal but does it anyways in hope of financial gain from logging. Even if he gets caught after the fact, the harm to the endangered species is already done.*

- Read more from: <https://laji.fi/about/709>



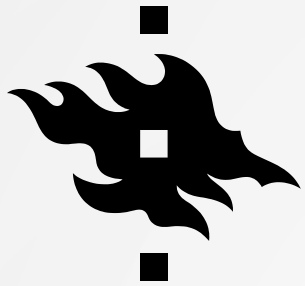
# ROLES AND RESPONSIBILITIES OF DATA PROTECTION

- As previously said, GDPR strictly regulates how, why and by whom personal data can be collected. There are a lot of nuances, and the most important information for a researcher are collected to these [two Flamma-pages](#).
- Some principles:
  - **Data minimization** – collect only the data you need.
  - **Consent** – data subjects must give their consent for the data collection.
  - **Informing** – data subjects should understand to what end and how their data is used.
  - **Controller** – during the research, data controller should be named (usually UH).
    - From [Flamma](#): “The controller is the one who have the ultimate responsibility for the processing of personal data. [...] The University of Helsinki is the controller if the research funding has been allocated to the University. The University of Helsinki is also the controller if the research is performed under an employment contract with the University.”



# INFORMING, CONSENT AND RIGHTS

- Informing should be done **before or during the data collection** and in a way that is easily **understandable, without jargon, and unambiguous**.
  - The most common ways for informing are [data protection statement and privacy statement](#).
- Data subjects should also be able to give their consent for participating in the study → note: **informing and consent are not synonymous!**
  - Consent can be **ethical** (willingness to participate to the study) and **judicial** (as a [legal basis for data collection](#)).
    - It is generally advisable to not use consent as the legal basis for data collection when doing research, but instead **public interest**.
- Data subjects have several [rights](#) that must be taken into account during planning and executing the research – these include e.g. **the right of erasure**.

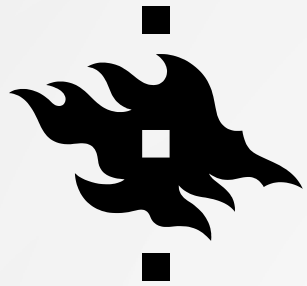


# BASIC PRINCIPLES

- **Most research uses personal data in one way or the other.**
- If you're collecting or using personal data, do:
  - a) Recognize your role as a **controller**
  - b) Determine **legal basis** for data processing
  - c) Assess **risks** and plan **safeguards**
  - d) Create a **data protection statement**
  - e) **Inform** the data subjects
  - f) Ensure **data security** during research
  - g) Sign Material-Transfer Agreements (**MTA's**) if necessary
  - h) **Anonymize** the data before publishing the data (if possible)
- Read more: <https://flamma.helsinki.fi/en/group/tutkimuksen-tuki/tutkimuksen-tietosuoja-asiat>



	Data type	Source of the data	File Format	Size estimate	Sensitivity / Personal data + controller	Owner / other agreements?	Documentation	Storage during project	Opening	Long term archiving
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										



# LEGAL AND ETHICAL ISSUES



# ETHICAL REVIEW

- Ethical review is “an advance scrutiny and evaluation of a research plan in the light of the ethical practices generally followed in that particular discipline of science, with special emphasis on preventing any harm that the research or its results might cause to the research subject.” (from [TENK.](#))
- In UH, there are four committees:
  - [Humanities and social science](#)
  - [Research on animals](#)
  - [Medical research](#)
  - [Natural, biological and environmental sciences and engineering](#)
- Committees issue statements after a review process that, when needed, should be done **before the research has commenced.**



# SOURCE OF THE DATA

- If you are using data generated by someone else than you or your research group, be mindful of the **rights of the person(s)/institution(s) whose data is being used.**
- An example: you are doing historical research and, when visiting an archive, you take pictures of the documents. What rights does the archive have? How should you take these into consideration during research?
- Another example: you complement your research with data from a data repository. Are there clear guidance for the limits of the data's usage? (e.g. a CC-licence) How should the originator of the data be credited?
- Yet another example: you are part of a research consortium in which you use data created by another university. Are there clear agreements in place for the data usage and ownership? Who owns the data and with what terms the using is allowed?





# UNDERTAKING ON TRANSFER OF RIGHTS

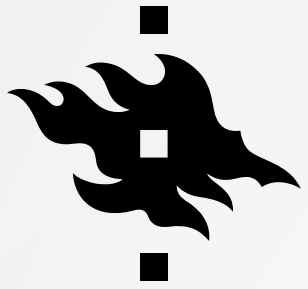
- When doing research within UH, the university demands the researcher to sign an Undertaking on Transfer of Rights -document.
- When doing so, the researcher does **not** give up any of his/hers rights to the data, but instead grant **parallel rights** to the UH.
  - The idea is that even if a researcher leaves mid-project, the project and its other members can still use the data the now-absent researcher has created.
- The two ways for signing:
  1. As an **supplement** to your employment contract (lasts for whole employment)
  2. As a **project-specific** contract (for the duration of the specific project)



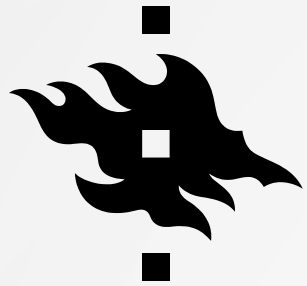
# AGREEMENTS

- **Clear and early agreements benefit everyone involved!**
- The most common agreements relating to RDM are the Material Transfer Agreements aka MTA's.
  - Especially important if data is being **transferred at any point outside EU!**
  - UH's lawyers ([tutkimuksenjuristit@helsinki.fi](mailto:tutkimuksenjuristit@helsinki.fi)) should be consulted and they should read all the agreements.

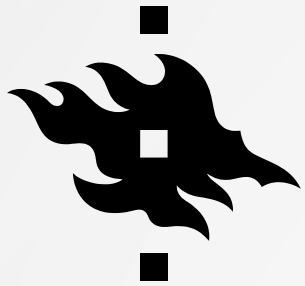
More from Flamma: <https://flamma.helsinki.fi/en/group/tutkimuksen-tuki/tutkimuksen-tietosuoja-asiat#menu9> and <https://flamma.helsinki.fi/en/group/tutkimuksen-tuki/sopimukset-sopimusohje-ja-neuvottelut>



	Data type	Source of the data	File Format	Size estimate	Sensitivity / Personal data + controller	Owner / other agreements?	Documentation	Storage during project	Opening	Long term archiving
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										

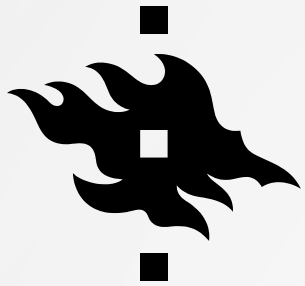


# **METADATA AND LICENCES**



# DATA ABOUT DATA

- Metadata ensures that the data is **understandable, usable and clear** to its user.
- Typical ways of providing metadata are readme.txt-files and codebooks, and .xml-files.
- The most important thing in metadata and documentation is **being systematic** in which ever way you decide to formulate the metadata.
  - E.g. always use the same format for file naming (YYYYMMDD)
- Read more on metadata and documentation here:  
<https://doi.org/10.5281/zenodo.1914401>



# DATA DOCUMENTATION

Make your data describe itself!



## **Understandability** "User manual" of the dataset

Makes the dataset self-explanatory and usable for others.

File naming conventions, explain variables, codebooks, use tags, readme-files + administrative documents, licenses, etc.

## **Discoverability** "Label" of the dataset


Describes what the dataset contains. Should be available even if you cannot open data itself.

Title, description, creator, persistent identifier (PID, e.g. DOI, URN) , etc.

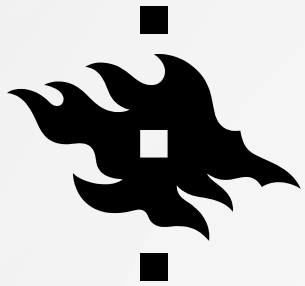
What the data is?  
Where it came from?  
How can it be reused?



# AN EXAMPLE ABOUT THE IMPORTANCE OF METADATA

- A researcher – let's call him Pentti – has **collected/taken 7 000 photographs** from all over Helsinki, but has not construed any metadata about the pictures.
- He has an 11-month break from his research and one of the 7 000 photos looks like this 
- Where was this taken? By whom? When? Why?
- Without metadata answering all these (and similar) questions, **the picture is virtually unusable** to Pentti (and everyone else)!





# READABILITY OF THE METADATA

- Metadata can be construed for a human or for a computer in mind.
  - The ideal **human-readable metadata** should be understandable even to "a moron in a hurry".
  - The **machine-readable metadata** can look very different.
- There are a lot of [metadata catalogues](#), some of which are field specific.
  - Also some data repositories might require the use of a specific format. → **Plan ahead!**

## Tarkemmat tiedot

Nimeke	Uskon ritarit : ristiretkien historiaa
Tekijä	Peura, Aleks, 1990- kirjoittaja. >
Aihe	<a href="#">ristiretket &gt;</a> <a href="#">keskiaika &gt;</a> <a href="#">ristiretkiaika &gt;</a> <a href="#">pyhiinvaellukset &gt;</a> <a href="#">uskonsodat &gt;</a> <a href="#">konfliktit &gt;</a> <a href="#">kristinusko &gt;</a> <a href="#">islamilainen kulttuuri &gt;</a> <a href="#">länsimainen kulttuuri &gt;</a> <a href="#">korståg &gt;</a> <a href="#">medeltiden &gt;</a> <a href="#">korstågstiden &gt;</a> <a href="#">pilgrimsfärder &gt;</a> <a href="#">religionskrig &gt;</a> <a href="#">konflikter &gt;</a> <a href="#">kristendom &gt;</a> <a href="#">islam &gt;</a> <a href="#">islamisk kultur &gt;</a> <a href="#">västerländsk kultur &gt;</a> <a href="#">historia &gt;</a> <a href="#">Eurooppa &gt;</a> <a href="#">Lähi-itä &gt;</a> <a href="#">Europa &gt;</a> <a href="#">Mellanöstern &gt;</a>
Toinen nimeke	Ristiretkien historiaa
Julkaisija	Helsinki : Like
Julkaisuaika	2021
Ulkoasu	272 sivua : kuvitettu, karttoja ; 22 cm
Kieli	suomi
Tunniste	ISBN : 978-951-1-38334-5
Lähde	Helka kirjastotietokanta

```
leader 02940cam a2200613 i 4500
001 9934102706806253
005 20220830064812.0
008 210122t20212021fi ab||e b |00 0|fin c
020 ##$a978-951-1-38334-5 $qkvakantinen
024 8#$a(FI-MELINDA)017487429 $d2021112113835.0
035 ##$a(FI-MELINDA)017487429
040 ##$aFI-BTJ $bfin $erda $dFI-NL $dFI-B $dFI-Hul
041 0#$afin
042 ##$afinb
080 1#$a940 $21974/fin/fennica
080 1#$a950 $21974/fin/fennica
084 ##$a91.6 $2ykl
100 1#$aPeura, Aleks, $d1990- $ekirjoittaja. $0(FI-ASTERI-N)000149603
245 10$aUskon ritarit : $bristiretkien historiaa / $cAleksi Peura.
246 30$aRistiretkien historiaa
264 #1$aHelsinki : $bLike, $c[2021]
264 #3$aKeuruu : $bOtava Kirjapaino
264 #4$c@2021
300 ##$a272 sivua : $bkuvitettu, karttoja ; $c22 cm
336 ##$ateksti $btxt $2rdacontent
337 ##$akäyttävissä ilman laitetta $bn $2rdamedia
338 ##$anide $bnc $2rdacarrier
504 ##$aSisältää bibliografisia viitteitä (sivut 250-254).
650 #7$aristiretket. $0http://www.yso.fi/onto/yso/p7099 $2yso/fin
650 #7$akeskiaika. $0http://www.yso.fi/onto/yso/p2559 $2yso/fin
650 #7$aristiretkiaika. $0http://www.yso.fi/onto/yso/p2557 $2yso/fin
650 #7$apyhiinvaellukset. $0http://www.yso.fi/onto/yso/p7098 $2yso/fin
650 #7$auskonsodat. $0http://www.yso.fi/onto/yso/p5206 $2yso/fin
650 #7$akonfliktit. $0http://www.yso.fi/onto/yso/p7946 $2yso/fin
650 #7$akristinusko. $0http://www.yso.fi/onto/yso/p4668 $2yso/fin
650 #7$aislam. $0http://www.yso.fi/onto/yso/p2082 $2yso/fin
650 #7$aislamilainen kulttuuri. $0http://www.yso.fi/onto/yso/p14178 $2yso/fin
650 #7$alänsimainen kulttuuri. $0http://www.yso.fi/onto/yso/p18650 $2yso/fin
650 #7$ahistoria. $0http://www.yso.fi/onto/yso/p1780 $2yso/fin
650 #7$akorståg. $0http://www.yso.fi/onto/yso/p7099 $2yso/swe
650 #7$amedeltiden. $0http://www.yso.fi/onto/yso/p2559 $2yso/swe
650 #7$akorstågstiden. $0http://www.yso.fi/onto/yso/p2557 $2yso/swe
650 #7$apilgrimsfärder. $0http://www.yso.fi/onto/yso/p7098 $2yso/swe
650 #7$areligionskrig. $0http://www.yso.fi/onto/yso/p5206 $2yso/swe
650 #7$akonflikter. $0http://www.yso.fi/onto/yso/p7946 $2yso/swe
650 #7$akristendom. $0http://www.yso.fi/onto/yso/p4668 $2yso/swe
650 #7$aislam. $0http://www.yso.fi/onto/yso/p2082 $2yso/swe
650 #7$aislamisk kultur. $0http://www.yso.fi/onto/yso/p14178 $2yso/swe
650 #7$avästerländsk kultur. $0http://www.yso.fi/onto/yso/p18650 $2yso/swe
650 #7$ahistoria. $0http://www.yso.fi/onto/yso/p1780 $2yso/swe
651 #7$aEurooppa. $0http://www.yso.fi/onto/yso/p94111 $2yso/fin
651 #7$aLähi-itä. $0http://www.yso.fi/onto/yso/p105025 $2yso/fin
651 #7$aEuropa. $0http://www.yso.fi/onto/yso/p94111 $2yso/swe
651 #7$aMellanöstern. $0http://www.yso.fi/onto/yso/p105025 $2yso/swe
```





# LICENCES

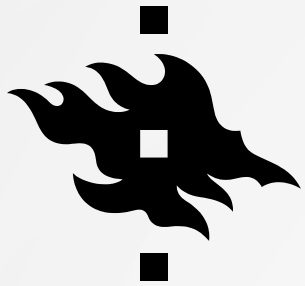
- Licences help set **clear boundaries** for the use of data or other intellectual property.
- The most common licences are the [Creative Commons](#) –licences.
- UH recommends **CC0**-licence for research data.
  - Note: ethics still require you to be attributed as the creator of the data!
- Avoid using CC BY NC (non-commercial), because it can be extremely difficult to determine what is commercial use and what isn't.
  - e.g. Is an advertise-funded blog commercial or not? How about a social media account?
- More about CC-licences: <http://libraryguides.helsinki.fi/oa/eng/license>



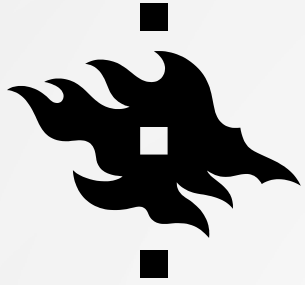


# BASIC PRINCIPLES

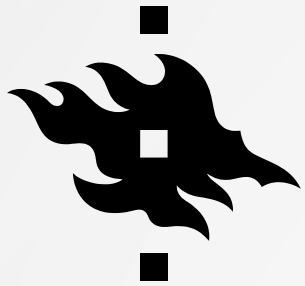
- **Without proper metadata, the data can become useless.**
- The best metadata is comprehensive, and done as soon as possible after the creation of the data.
- Be systematic.



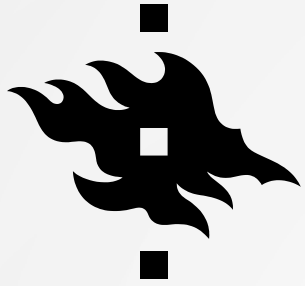
	Data type	Source of the data	File Format	Size estimate	Sensitivity / Personal data + controller	Owner / other agreements?	Documentation	Storage during project	Opening	Long term archiving
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										



# DATA STORAGE



	Data type	Source of the data	File Format	Size estimate	Sensitivity / Personal data + controller	Owner / other agreements?	Documentation	Storage during project	Opening	Long term archiving
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										



# **OPENING THE DATA AND LONG-TERM PRESERVATION**



# OPENING AND ARCHIVING

As open as possible, as closed as necessary!

- Opening and archiving the data is not a binary option: **not everything can or should be opened!**
  - E.g. if personal data is un-anonymizable, it cannot be opened.
  - Also: there can be several options for the degree of openness for the data.
- Be mindful of the costs: some data repositories charge per dataset.
- Even if the data cannot be opened, the metadata (usually) can!
  - You can open metadata in e.g. [Etsin](#).

Home / Data catalogue

## Study descriptions in English

In total 1735 datasets.

Display the data catalogue [in Finnish](#).

You can filter the studies in the catalogue by availability status, data file language or type. Study descriptions can be sorted by ID number, title, availability, data type or publication date. Only one sorting criterion is allowed at a time. [More information](#) on availability and conditions.

Dataset availability:  
(A) openly available for all users without registration (CC BY 4.0),  
(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.

You can browse archived data, for example, by [keyword](#), [topic](#) and [series](#).

[Upcoming data releases](#)

Availability

- A ⓘ
- B ⓘ
- C ⓘ
- D ⓘ

Type

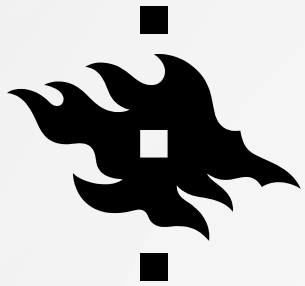
- Qualitative
- Quantitative

Data File Language ⓘ

- English
- Finnish
- Swedish

Display per page 50 datasets.

FSDID	Study title	Availability	Type	Published
<a href="#">FSD3542</a>	<a href="#">Interviews on Peer Support and Experiential Knowledge in Child and Family Services 2019-2020</a>	D ⓘ	Qualitative	2022-09-26
<a href="#">FSD3694</a>	<a href="#">Citizens' Pulse 7/2022</a>	A ⓘ	Quantitative	2022-09-26
<a href="#">FSD3526</a>	<a href="#">Role of Science in Finnish Worldviews and Everyday Thinking 2019-2020</a>	C ⓘ	Quantitative	2022-09-21
<a href="#">FSD3546</a>	<a href="#">Sustainable Policy-Making: Finnish Policy-Maker Survey 2018</a>	B ⓘ	Quantitative	2022-09-15
<a href="#">FSD3560</a>	<a href="#">Physical Activity and Digital Gaming of Sixth-Graders 2015</a>	B ⓘ	Quantitative	2022-09-07
<a href="#">FSD3671</a>	<a href="#">Views of Psychologists and Psychology Students on Digital and Problematic Gaming 2021</a>	C ⓘ	Qualitative	2022-09-06
<a href="#">FSD3514</a>	<a href="#">Interviews with Participants of NASTA Training by Women's National Emergency Preparedness Association 2018-2019</a>	C ⓘ	Qualitative	2022-09-02
<a href="#">FSD3676</a>	<a href="#">Finnish Opinions on Security Policy and National Defence 2022: Spring</a>	B ⓘ	Quantitative	2022-08-26
<a href="#">FSD3495</a>	<a href="#">Accessibility in Basic Education in the Arts 2018</a>	B ⓘ	Quantitative	2022-08-24
<a href="#">FSD3489</a>	<a href="#">Perceptions of Workplace Well-being by Early Childhood Education</a>	B ⓘ	Qualitative	2022-08-18



# OPENING (PUBLISHING) THE DATA

- Opening/publishing the data usually deals in a timespan of **<10 years**.
- Most common ways of opening the data is publishing it in a **data repository** or a **data journal**.
  - Some data repositories are **specific** to some field of science or datatype (e.g. GenBank), some are **general** (e.g. Zenodo).
  - You can browse data repositories here: <https://www.re3data.org/>
- Some things to keep in mind when choosing a data repository:
  - Does it provide a persistent identifier (e.g. DOI, URN)?
  - Does it cost you something?
  - Does it have a [Core Seal Trust](#)?
  - Will the data be curated?
  - What is the physical location of the servers?





# ARCHIVING (LONG-TERM PRESERVATION)

- Archiving / long-term preservation usually deals in a timespan of **decades, even centuries**.
- UH has its own archiving solution (aka [PAS](#)), but the bar for access is rather high: **only the most valuable datasets are accepted** in.
  - The process of getting the datasets in can be long and arduous – but on the other side there's a chance to preserve the research data for future generations!
- Some data repositories have a promise to store the data for decades, but if the data is not curated, the functionality of the data might get compromised (e.g. the data formats could become incompatible with future operating systems).
  - As an example, Zenodo promises the data to be available for as long as Zenodo exists (currently, at least for next 20 years), but "Zenodo makes no promises of usability and understandability of deposited objects over time." (<https://about.zenodo.org/policies/>)



# FAIR-PRINCIPLES

So that anyone can...

**Findable**

Persistent identifiers (e.g. DOI)



...locate the data, even if it has moved around.

**Accessible**

Open and free repositories



...access\* the data without paywalls.

**Interoperable**

Open and free dataformats



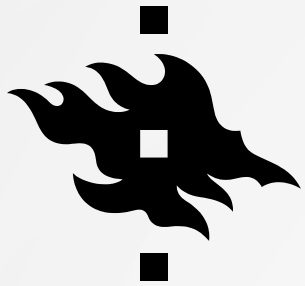
...use the data without specific software (e.g. .xlsx versus .csv).

**Reusable**

Clear metadata and licences



...understand the data and the limits of its useage.

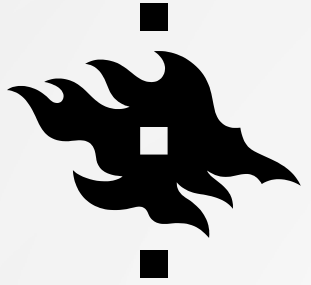


# BASIC PRINCIPLES

- **Not everything can or should be opened!**
  - Opening should **never** be an accident!
- Opening the data generally deals in a timespan of years, archiving in decades or even centuries.
- Costs can rack up quickly! Per gigabyte per year? Per dataset?
- Please **don't** just say "the data will be FAIR" – be more specific!



	Data type	Source of the data	File Format	Size estimate	Sensitivity / Personal data + controller	Owner / other agreements?	Documentation	Storage during project	Opening	Long term archiving
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										



**WHAT NEXT?**



# DATA MANAGEMENT PLAN (DMP)

- When writing a DMP, it is advisable to use [DMPTuuli](#).
  - In DMPTuuli, the main Finnish **DMP templates** are easily accessible.
- A DMP is not meant to be a one-shot document that should be archived and forgotten once done – instead a **DMP is a guide for you, a living document** to be updated during the research project.
- Ideally, a DMP is max. 2-3 pages long.

**DMPTuuli** Home Public DMPs DMP Templates Help Language

## 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:

- 15 797 Users
- 18 840 Plans
- 120 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

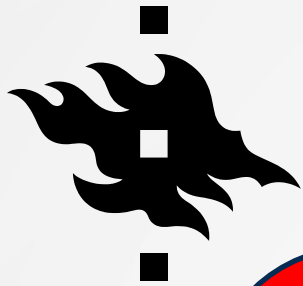
Forgot password?

Remember email

Sign in

- or -

Sign in with your institutional credentials



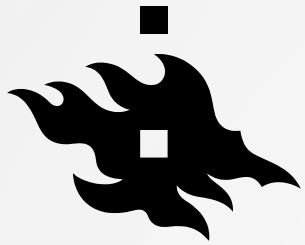
# THE DO'S AND DON'T'S OF A DMP



- ... be vague or general
- ... ramble
- ... re-write your research plan
- ... write a "good DMP" for the sake of bureaucracy
- ... write the DMP for Data Support or AoF
- ... write one thing and do another



- ... be precise and unambiguous
- ... stay on point
- ... focus on the data and its lifecycle
- ... think through and write out your data management process
- ... write the DMP for the sake of your current and future self
- ... practice what you preach



# DATASUPPORT IS HERE FOR YOUR ASSISTANCE!

- [Courses and workshops:](#)
  - RDM Advanced webinars in the spring, recordings from the last year available in [UniTube](#).
  - DMP workshops available usually during late spring and early summer.
  - AoF-webinars in August and September.
  - Course materials available here: <https://wiki.helsinki.fi/x/jb5ZDQ>
- Website: <https://www.helsinki.fi/en/research/services-researchers/data-support>
- Guide: <https://www.helsinki.fi/en/research/services-researchers/data-support/research-data-management>
- Think Open -blog: <https://blogs.helsinki.fi/thinkopen/>
- DMP review: <https://www.helsinki.fi/en/research/services-researchers/data-support/dmp-review-service>
- You can always contact us at [datasupport@helsinki.fi](mailto:datasupport@helsinki.fi)





**THANK YOU FOR ATTENDING!**