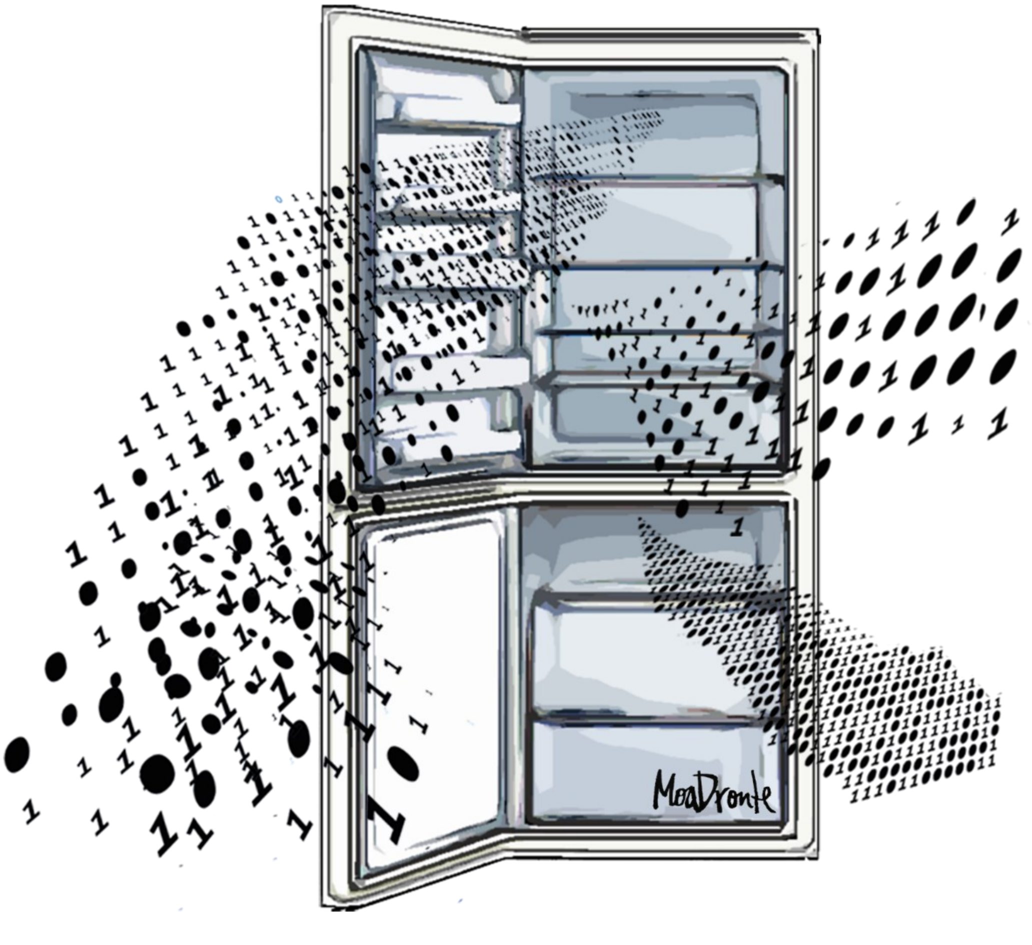
How to become a data preserver

The official University of Helsinki guide to the responsible preservation of research data

Timo Lahtinen, Matilda Mela, Mikko Mäkelä, Niina Nurmi, Mari Elisa Kuusniemi



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

cc BY

Guide to responsible data preservation

Open science – Everyone benefits

This guide has been compiled to support the preservation of research datasets after the collection of data and the active analysis stage. In the guide, we describe the steps to take before and during research to enable preserving the data generated. The guide begins with a **checklist** where we have compiled the key issues related to data preservation. In the following sections, we take a closer look at relevant matters, while at the end of the guide we have compiled links for further information.

While the responsible preservation of research data takes a little effort, it is also rewarding. The efforts to combat climate change and the rapid development of coronavirus vaccines are examples of the **power of collaboration**, which is what this is also about: we utilise research datasets as thoroughly as possible and collaborate.

Reader's guide: To ease reading, we have collected all the links to the end of each chapter. The checklist in the beginning of this guide contains some terms that may not be familiar to all readers. An explainer for those terms has been included in the chapters that follow.

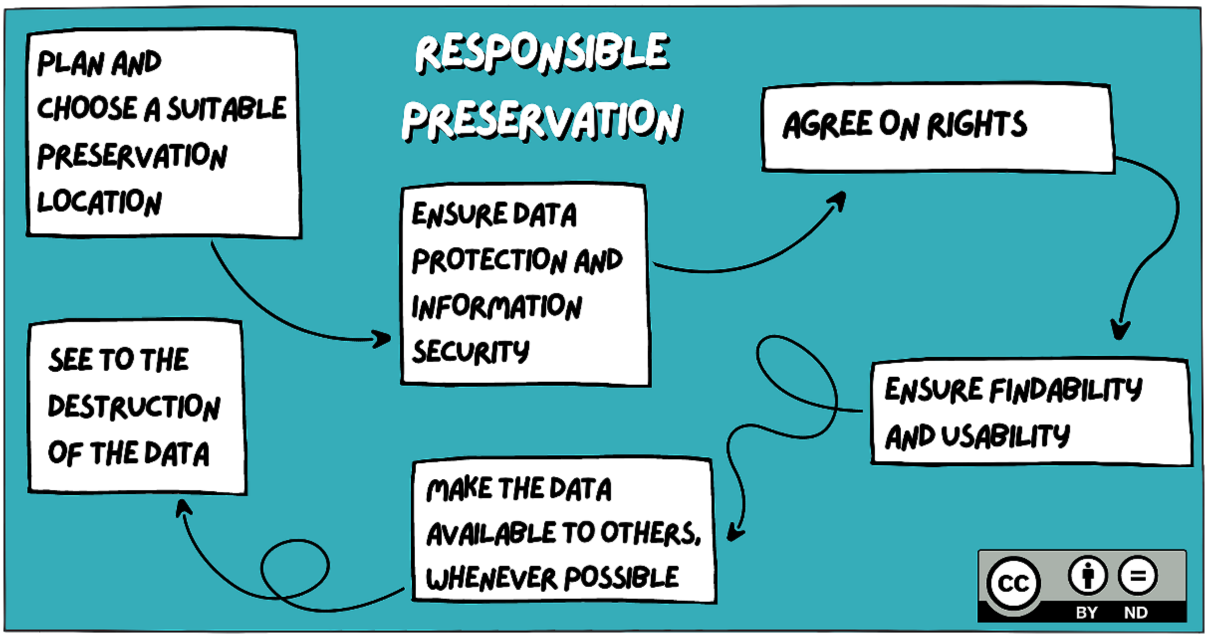

RESPONSIBLE PRESERVATION
PLAN AND CHOOSE A SUITABLE PRESERVATION LOCATION
ENSURE DATA PROTECTION AND INFORMATION SECURITY
AGREE ON RIGHTS
ENSURE FINDABILITY AND USABILITY
MAKE THE DATA AVAILABLE TO OTHERS, WHENEVER POSSIBLE
SEE TO THE DESTRUCTION OF THE DATA
cc BY ND

Image 1: Stages of responsible preservation.

Content

Checklist for the responsible preservation of research data

Where to start?

□ Find a preservation service suited to your data. For help, see the decision tree in section 5 entitled 'Where are data preserved?' or visit re3data.

□ Make sure that you have sufficient access rights to submit the data for preservation. Access rights must be verified, especially if the data originate from other sources or if they have been collected in cooperation with other individuals or parties.

□ In the case of data repositories not suited to preserving sensitive data, make sure that your data do not contain sensitive personal data or other confidential data.

Data preparation

□ Whenever possible, prefer persistent, open, and common file formats.

□ Create a clear folder structure for the dataset.

□ Name files and folders in a consistent and descriptive manner.

Description of data

□ Ensure the understandability and applicability of the data by describing them with comprehensive metadata.

□ Give the dataset a descriptive title.

□ Create a separate README.txt file for the data.

□ If your dataset contains files in table format, define and explain the column headings, the values for variables, any abbreviations and the codes used.

□ Attach documentation that supports the comprehensibility and applicability of the data, such as a description of research methods, questionnaires, a consent form template, or electronic notes.

□ Add links to other research output in the metadata of the dataset, including the DOIs of published articles, project website addresses or links to the code or software produced during the study.

Licences

□ Choose a licence suited to your dataset. For further information, see the licence guide of Helsinki University Library.

What to do after storing the data?

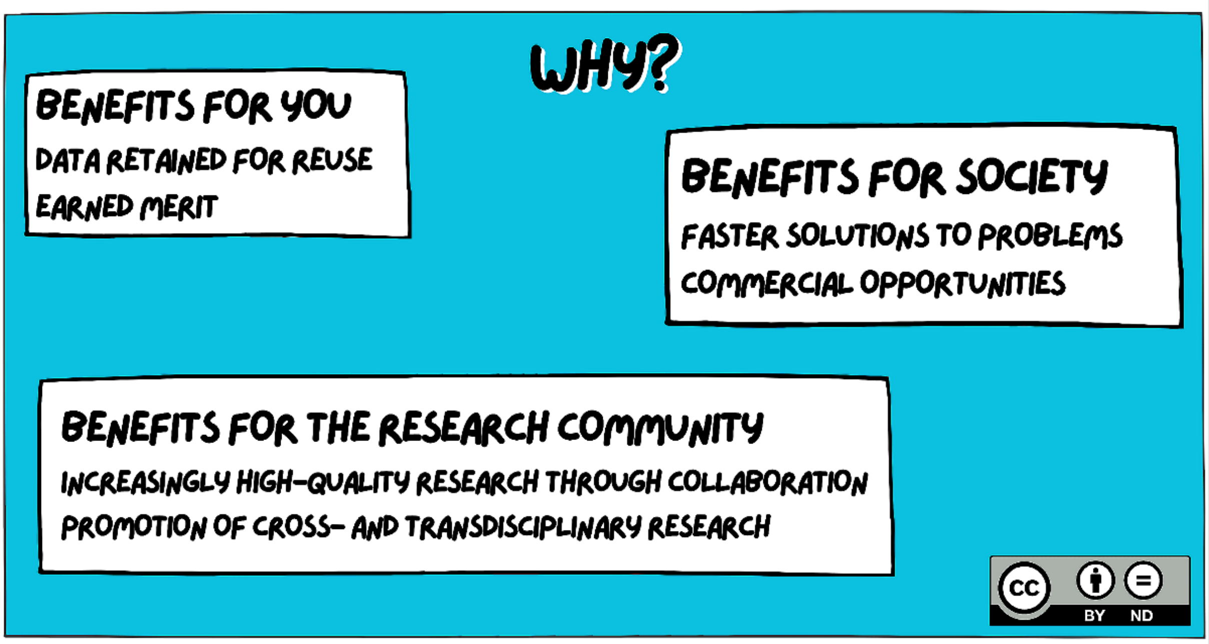
□ If the data repository of your choice provides your dataset with a DOI or other persistent identifier, include it in the data availability statement of the article you publish.

□ Tell others about the data you share! For example, you can add a mention with links to your University of Helsinki TUHAT profile.

Should you have any questions related to data preservation, you can always reach out to Data Support.

1. Why are research data preserved?

The responsible preservation of research data is linked to, among other things, the goals of open science,[1] which the University of Helsinki has pledged to promote. In a world where the amount of research-based knowledge is increasing at an unprecedented rate, it is particularly important to support its systematic preservation and findability. This way, data collected can be utilised as diversely as possible for the benefit of research and society.



WHY?

BENEFITS FOR YOU

DATA RETAINED FOR REUSE

EARNED MERIT

BENEFITS FOR SOCIETY

FASTER SOLUTIONS TO PROBLEMS

COMMERCIAL OPPORTUNITIES

BENEFITS FOR THE RESEARCH COMMUNITY

INCREASINGLY HIGH-QUALITY RESEARCH THROUGH COLLABORATION

PROMOTION OF CROSS- AND TRANSDISCIPLINARY RESEARCH

Image 2: Responsible data preservation has many benefits.

Benefits for you

As a researcher, you will benefit greatly from the preservation of research data:

- Your data will be stored in a secure preservation service, giving you easy access to the data even long after your project has concluded.
- Your research design, publication or funder may require the preservation of the data.
- Obtaining a persistent identifier (DOI) for your dataset makes it easier to refer to it.

[1] Open Science: What is open science? (28 September 2023).

Benefits for the research community

The preservation of data strengthens the research community. By sharing your data, you will

- Provide researchers and students with access to high-quality data
- Facilitate the verification and reproducibility of your research
- Support research across disciplinary and organisational boundaries

Benefits for society

Preserving research data is often less expensive than repeating the research. This is why it is important to preserve and distribute research findings in an appropriate manner to generate wider societal benefits, for example, through innovations based on research.

It is difficult to determine which information will be valuable in the future. The value of research-based knowledge may be recognised only years later. For research to have as broad a societal impact as possible, it is important to preserve the data generated through research and enable future generations to utilise the knowledge accumulated.

The many names of preservation services

Data archives, data repositories and data banks are all related to the storing of data after research. However, they have certain qualitative differences related to their use. In this guide, we use the term '**preservation service**' for all of the above.

1.1. Principles of responsible data preservation

Responsible conduct of research

How is the responsible conduct of research reflected in the responsible preservation of research data? You are likely to have the most accurate knowledge of the matter in your field, as practices vary between disciplines, for example, in terms of the recommended preservation services and requirements as well as the retention period required for the verification of research results. In certain fields, making research data openly available is commonplace, whereas in others it is only now being initiated.

Fortunately, you don’t have to be able to do everything on your own. You can get started by familiarising yourself with the basics of your field and research ethics, and by observing them.

Can sensitive or confidential data be stored?

Sensitive and confidential data means data whose disclosure may cause harm. This is why the preservation of such data is particularly demanding.

The researcher who has collected the data is responsible for the safe preservation of sensitive and confidential data. For the time being, there are few services available that accept such data for preservation on behalf of researchers. Most services require the erasure of all sensitive data before transfer to preservation.

Data protection legislation regulates the preservation of data containing sensitive personal data. Data can be preserved as identifiable data for a long time, provided that the research subjects have been informed and that the data are only disclosed for their original purpose. When collecting such data, it is indeed advisable to anticipate your own future data needs and those of other researchers. In fact, the purpose of use should be defined as broadly as possible. The fact that the relevant legislation is relatively new, and its interpretation is yet to become established means that developing preservation services is not easy.

The preservation of confidential data not including personal data is easier. Of course, this too requires careful consideration of information security. The University of Helsinki offers a service suited to the long-term preservation of confidential data.

Questions to consider with regard to the storage of sensitive data

Consider whether the data will be useful after erasing all sensitive data. If not, find out whether there is a safe preservation location available for storing the data as identifiable.
Check the following before storing sensitive data:

- Have you informed research subjects about data preservation?
- What has been announced to the research subjects about the purpose of the data?

- If you have indicated that the data will be anonymised, have sensitive data as well as personal direct and indirect identifiers been erased from the dataset? Have the metadata been created in a way that precludes the identification of the subjects?

Read more about the responsible preservation of sensitive data:

- Open Science: Opening sensitive datasets in a responsible manner (link in Finnish only)
- Responsible Research: Safeguard the anonymity of research subjects!
- Data Archive: Anonymisation and personal data

1.2. Use cases of data preservation

When a researcher leaves the University of Helsinki

How will your research data remain high quality and applicable for current and future members of the research group after you move on to other things?

At least check the following:

- Make a written agreement on the rights of your research data. For more information on agreements, please contact the University of Helsinki's legal counsels for research at tutkimuksenjuristit@helsinki.fi.
- Make sure that your research data are well organised, the metadata carefully compiled and the file formats valid – in accordance with the requirements of the future storage location. Documentation is sufficient when a researcher opening the data for the first time understands the contents without your help. In such cases, your data are self-explanatory.
- Store your data, including metadata, in a suitable preservation service or database.
- Make sure that your data has a suitable licence.

Publishing delays

Can you restrict access to research data and metadata for the duration of a publishing delay, or embargo? Of course! The embargo will be determined in advance in accordance with the researcher’s wishes. You can easily check the embargoes set by journals, if any, at Sherpa Romeo. Please note that certain research data preservation services only grant embargoes for a well-grounded reason.

Fewer laboratory animals in preclinical trials

Carefully stored data and their reuse reduce the need to repeat the same tests in new laboratory animals, particularly in control groups. Developing computational methods and observation-based modelling and simulations can further reduce the need. However, the fragmentation of research datasets and varying file formats constitute a significant obstacle to this.

Publish non-significant results by preserving the research data

As a rule, all research datasets should be preserved to make it unnecessary to collect them again. This also makes it possible to publish what are known as negative or non-significant findings with the help of research data, which has traditionally not been possible in scholarly articles. By also storing data that did not meet your expectations, you will promote the accumulation of knowledge and future research efforts.

2. How to choose the data to be preserved

1. Identify **valuable** data that can be utilised in other research as well.
2. Identify data that **must** be preserved.

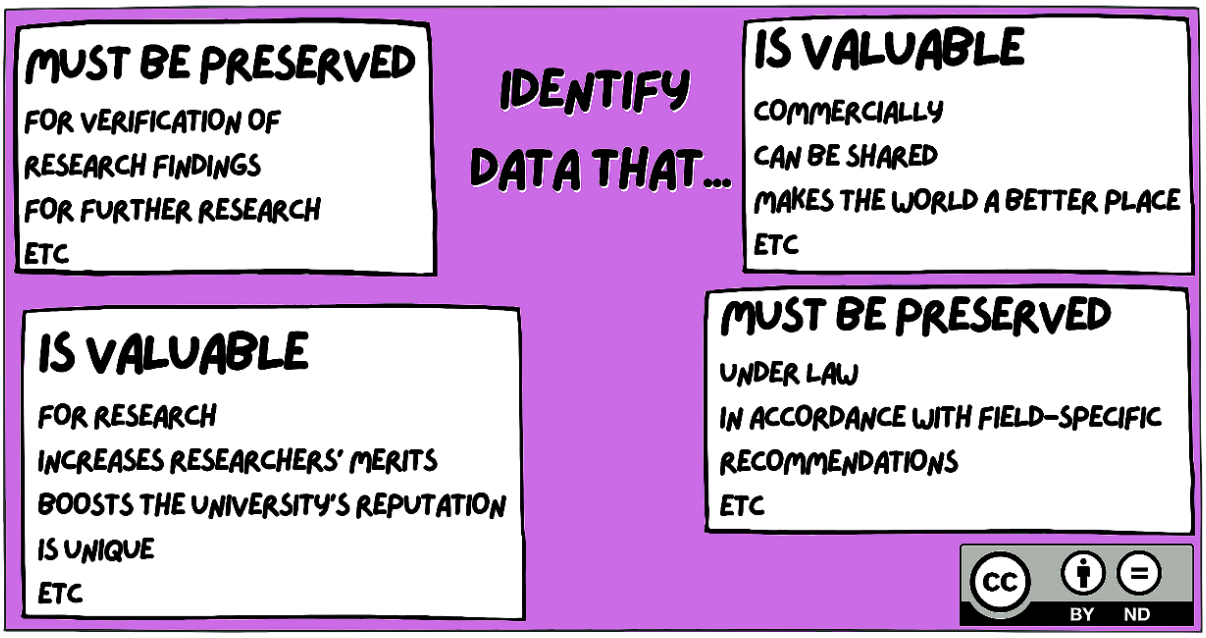

MUST BE PRESERVED
FOR VERIFICATION OF RESEARCH FINDINGS
FOR FURTHER RESEARCH
ETC
IDENTIFY DATA THAT...
IS VALUABLE
COMMERCIALLY
CAN BE SHARED
MAKES THE WORLD A BETTER PLACE
ETC
IS VALUABLE
FOR RESEARCH
INCREASES RESEARCHERS' MERITS
BOOSTS THE UNIVERSITY'S REPUTATION
IS UNIQUE
ETC
MUST BE PRESERVED
UNDER LAW
IN ACCORDANCE WITH FIELD-SPECIFIC RECOMMENDATIONS
ETC
CC BY ND

Image 3: Characteristics of data to be preserved.

Always submit **coherent** data for preservation. It is usual to create copies or versions of datasets during research, including raw data, data generated in the interim stages of analysis and data related to findings. All data are rarely preserved. To arrive at a coherent whole and avoid unnecessary preservation, you have to consider what is to be retained and what is to be destroyed. You will save time and money when preservation is **planned** in advance in the data management plan and the data are already cleaned during the project.

Identify valuable data

Valuable data include

- One-of-a-kind and **unique** datasets
- Datasets that increase the **reputation** of researchers or the University
- Datasets that can be **commercially** utilised
- Interesting datasets that can be **made available** to all
- Datasets that can be used to **make the world a better place**

It may also be possible to utilise the data in **teaching** or upcoming **new projects**: it is possible that new perspectives or methods will be developed, monitoring carried out, or new funding secured.

Identify data that must be stored

There are data that must be stored

- For **statutory** reasons
- For the **verification** of research findings
- For **further** research
- On the basis of **field-specific recommendations**
- On the basis of the informing of **research subjects**

Further information on the selection of the data to be stored:

- ❖ Digital Preservation (Fairdata-PAS): Guidelines for UH Evaluators
- ❖ DCC: Five steps to decide what data to keep
- ❖ UK Data Service: Collections development selection and appraisal criteria

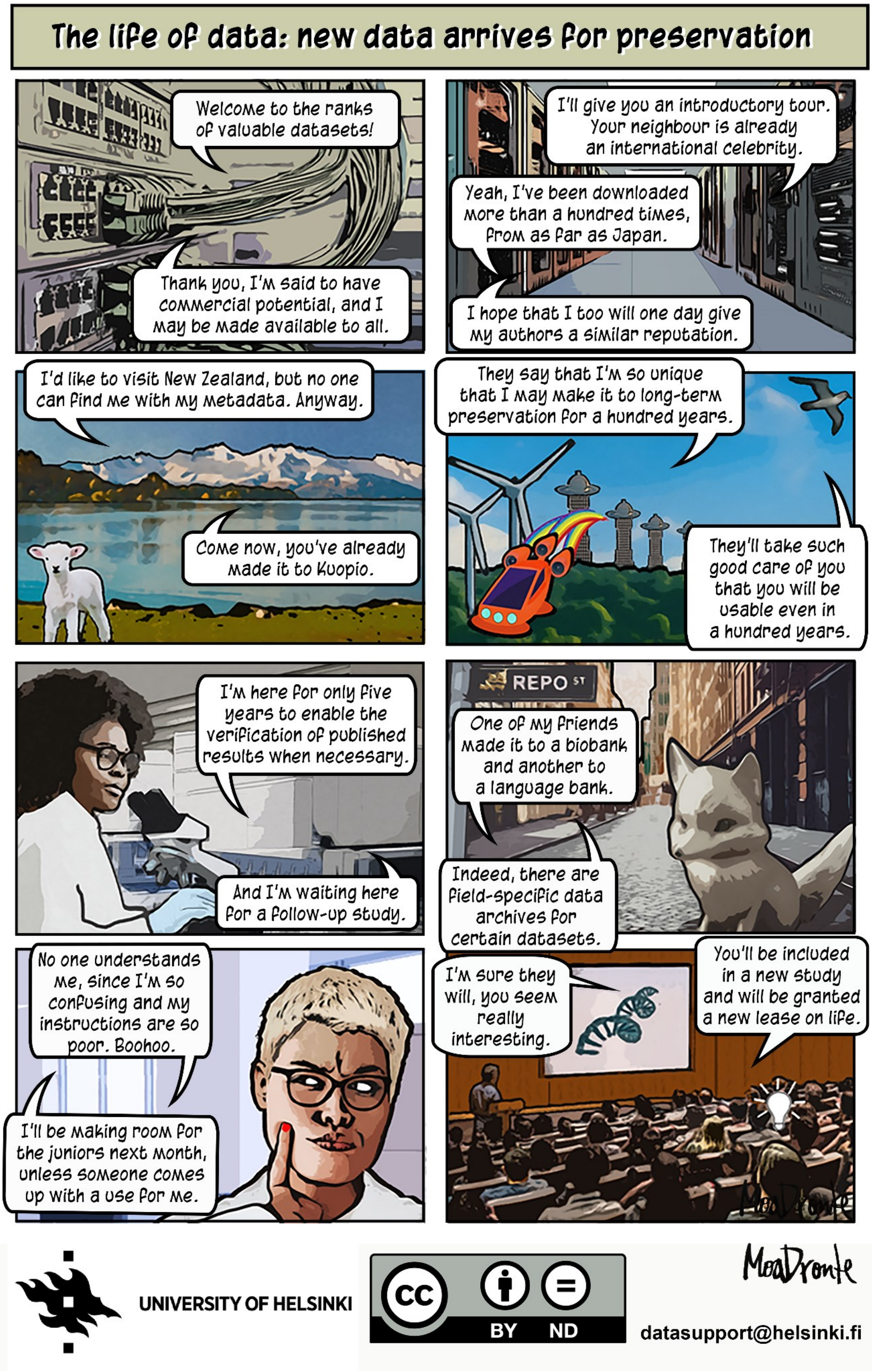

The life of data: new data arrives for preservation
Welcome to the ranks of valuable datasets!
Thank you, I'm said to have commercial potential, and I may be made available to all.
I'll give you an introductory tour. Your neighbour is already an international celebrity.
Yeah, I've been downloaded more than a hundred times, from as far as Japan.
I hope that I too will one day give my authors a similar reputation.
I'd like to visit New Zealand, but no one can find me with my metadata. Anyway.
Come now, you've already made it to Kuopio.
They say that I'm so unique that I may make it to long-term preservation for a hundred years.
They'll take such good care of you that you will be usable even in a hundred years.
I'm here for only five years to enable the verification of published results when necessary.
And I'm waiting here for a follow-up study.
REPO ST
One of my friends made it to a biobank and another to a language bank.
Indeed, there are field-specific data archives for certain datasets.
No one understands me, since I'm so confusing and my instructions are so poor. Boohoo.
I'll be making room for the juniors next month, unless someone comes up with a use for me.
I'm sure they will, you seem really interesting.
You'll be included in a new study and will be granted a new lease on life.
MoaDronte
UNIVERSITY OF HELSINKI
BY ND
MoaDronte
datasupport@helsinki.fi

Image 4: New data arrives for preservation.

3. Retention periods and destruction

1. How **long** will the data be preserved?
2. Also see to the **destruction** of data.

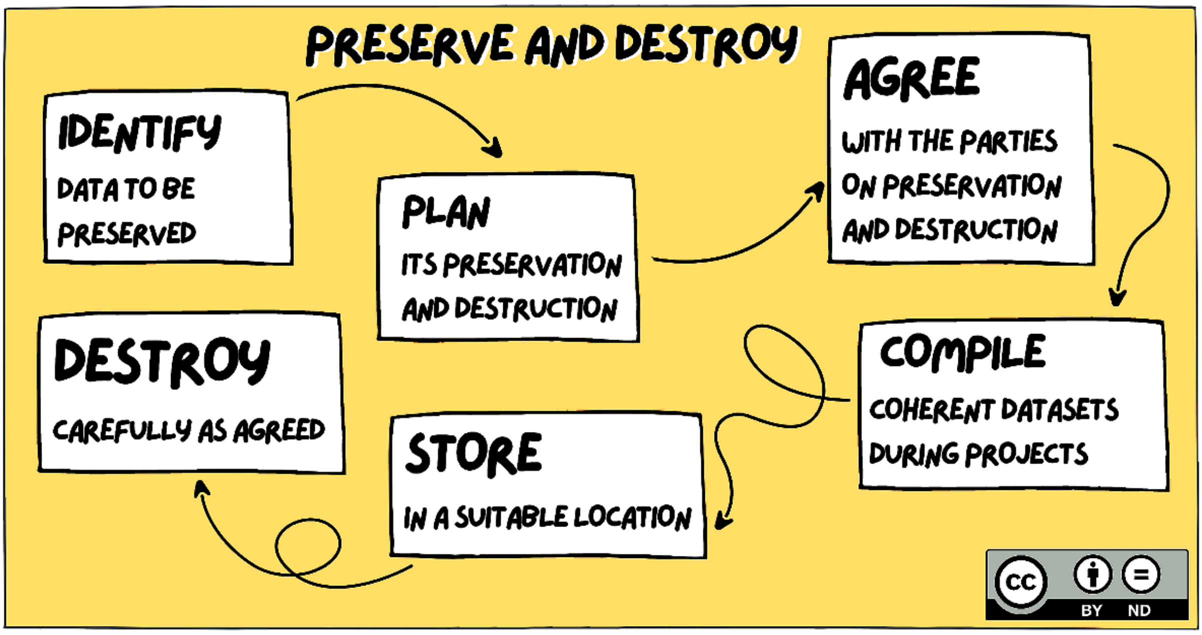


Image 5: Stages of preservation and destruction.

How long will the data be preserved?

While the general retention period for verifying research results is five years, in medicine it can be 15 years, for example, if the research has long-term effects on patient data. In other words, the period varies by dataset and field of science. For the time being, the University of Helsinki has no instructions on the retention periods of research data.

For example, the HUS Helsinki University Hospital provides the following guidelines:

- In accordance with the standard for good clinical practice, research datasets are preserved, as a rule, 15 years after the conclusion of the trial. From the beginning of 2021, this period was extended to 25 years.
- In the case of studies related to marketing authorisations of drugs granted in the EU, data must be stored for at least two years after the last marketing authorisation has been granted in the EU or two years after the conclusion of studies on the product.

Please take into consideration any funder- and field-specific recommendations on the retention period. Also check how long the publisher requires the data related to publication to be preserved. **Agree** on the retention period with the parties in good time. Parties can include research group members, research subjects, funders, and publishers.

Further information on retention periods:

- Office of the Data Protection Ombudsman: Policy for the preservation of personal data
- Deakin University: What data do I need to keep and for how long?

Also see to the destruction of data

Almost all data will eventually be destroyed. Consider and decide when and how. Filling disk space with unusable data is not considered good practice. The retention period and destruction of data should be planned in advance in the data management plan. To avoid unpleasant misunderstandings, it is important to discuss data destruction among the entire research group. Preservation services may determine the date of destruction and also the related grounds.

Sensitive material in particular must be carefully disposed of when its preservation is no longer necessary for the purpose. Pressing the delete-command is generally not sufficient to destroy data as it may still be recovered at a later stage.

Further information on the destruction of data:

- University of Helsinki Helpdesk: Deleting files safely

4. How to prepare data for preservation

Why are preparations needed?

The purpose of preparing research data is to ensure the **legality of the storage as well as the preservation and reuse of the data** in the future.

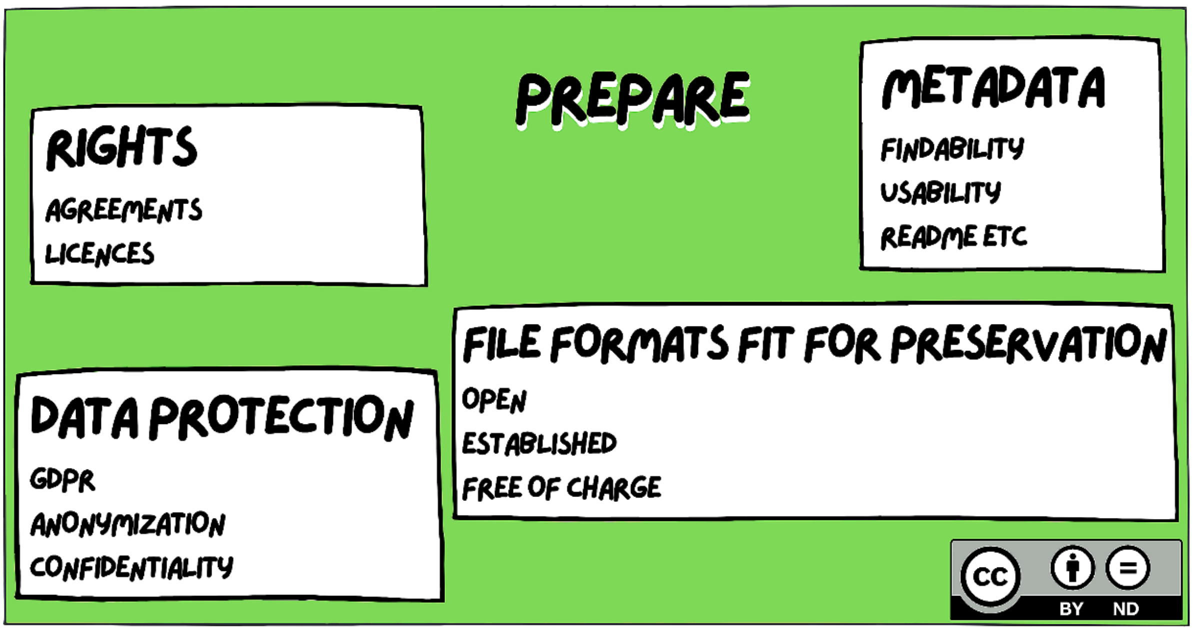


Image 6: Preparations.

4.1. Things to consider in the preparations

Access rights, agreements, and licences

Access rights must be agreed on by all parties involved in the collection of the data. The easiest way to make the agreements is before data collection commences. Instructions on the required agreements can be found on Flamma. If the data have been obtained from elsewhere, the limits of the researcher’s right to retain the material must always be verified. Different preservation services have their own terms of use, which should be reviewed in advance. With these terms, the services ensure that only datasets for which the service is suitable are preserved there. Preservation services often do not accept sensitive personal data or company-owned data. The user is responsible for complying with these guidelines.

A licence will be assigned to the data to be retained to inform others of the rights of use granted for the data. Licences can be used to prevent or enable reuse of the data.

Advanced instructions related to data preparation:

- ❖ Flamma: Agreements on the opening of data between collaboration partners
- ❖ Flamma: Undertaking on the transfer of rights
- ❖ Helsinki University Library: Licences (CC), licences for code and software

Data protection

Before storing or opening data, you must ensure that the data do not contain any personal data or other confidential material that cannot be stored in the selected service. Check at least the following:

- What **personal data** do the data contain? Do the personal data fall under special categories of personal data defined by the GDPR?
- Can the data be **anonymised** or **pseudonymised**?
- How have research subjects been informed? In connection with this, the subjects are informed of the preservation service to be used following the project, the purpose of the reuse of the data, and the parties to whom the data can be disclosed at a later date. These guidelines should be worded in a way that makes it possible to retain the data after the study.
- Do the data contain other **confidential** material, such as trade secrets or the distribution data of certain endangered species?
- If the data cannot be made available to others, the **metadata** should nevertheless be open.

Further information on data protection:

- ❖ University of Helsinki Data Support: ‘Do I have personal data?’ test
- ❖ Helsinki University Library: Copyright and licences
- ❖ Office of the Data Protection Ombudsman: Special categories of personal data

4.2. File formats suitable for preservation

File formats that are **open** or **established in the field** will help in preserving data for several years to come, as well as expand the opportunities for their reuse use. For the reuse of data, it is important that the file formats used can be predicted to remain useful for several years to come. Such file formats are most likely **standardised** and widely used. If the data has to be stored in a format specific to the field, you should ensure that the format is well-established in the field and comprehensively documented.

- Use **open** and commonly used file formats.
- Check from a data preservation service used in your field which file formats they accept.
- You can use specialised formats established in your field if they are comprehensively documented. Sufficient documentation is often available on Wikipedia.

Further information on file formats fit for preservation:

- ❖ Wikipedia: List of open file formats
- ❖ Digital Preservation: File formats suitable for long-term preservation (PDF, recommendation for the preservation of data sets for over 25 years)

Descriptions of and metadata associated with data to be preserved

Comprehensive descriptions help others find, understand, and use data. Try to organise your data so that it constitutes a **coherent and self-explanatory whole**.

- Organise the files in a clear folder structure.
- Use clear and unambiguous names for files and folders.
- Make sure that the variables, abbreviations, and codes used in the data have been explained in an understandable manner.
- Supplement the metadata by drawing up a **README.txt** file
- Choose a preservation service that enables the provision of sufficient metadata along with persistent identifiers, such as DOIs, to support findability of the data.

Further information on describing data:

- ❖ Fuchs & Kuusniemi (2018): Making a research project understandable
- ❖ An example of minimum requirements for metadata: DataCite Schema – Mandatory fields

5. Where are data preserved?

A wide range of preservation services are available for research data, which can be roughly divided into general and discipline-specific services. In addition, organisations may have their own preservation services for data produced by researchers or research infrastructures.

Discipline-specific data archives, data repositories and databases have evolved from the needs of specific research communities or data types. Consequently, they provide a good starting point for choosing a preservation service. Another important starting point is the reliability of the service. In certain cases, a preservation service provided by the organisation may be the most appropriate solution.

From the beginning of 2024, the University of Helsinki will offer a service, University of Helsinki Databank, where research data can be stored for five to 15 years according to need and as per agreement. On application, the retention period can be extended. Zenodo, a commonly used preservation service, offers a 25-year retention period. The University of Helsinki Data archive (Fairdata PAS service) is available for research data that should be preserved for future generations. In the service, data will remain useful for as many as hundreds of years thanks to active data curation.

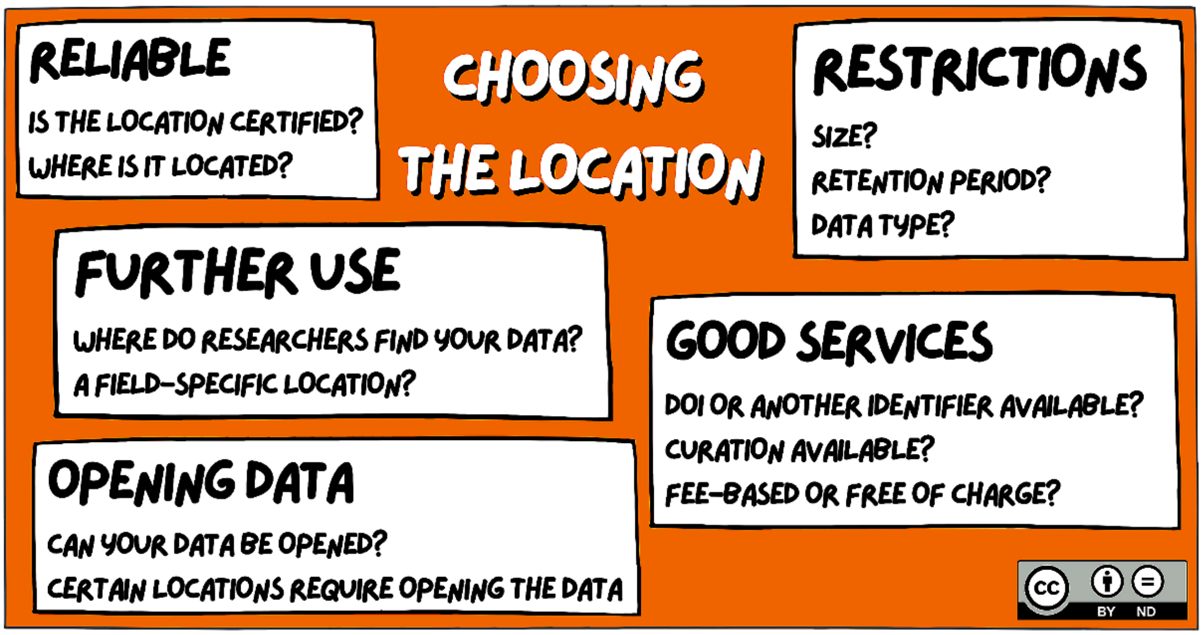

CHOOSING THE LOCATION
RELIABLE
IS THE LOCATION CERTIFIED?
WHERE IS IT LOCATED?
RESTRICTIONS
SIZE?
RETENTION PERIOD?
DATA TYPE?
FURTHER USE
WHERE DO RESEARCHERS FIND YOUR DATA?
A FIELD-SPECIFIC LOCATION?
GOOD SERVICES
DOI OR ANOTHER IDENTIFIER AVAILABLE?
CURATION AVAILABLE?
FEE-BASED OR FREE OF CHARGE?
OPENING DATA
CAN YOUR DATA BE OPENED?
CERTAIN LOCATIONS REQUIRE OPENING THE DATA
cc BY ND

Image 7: Consider these matters when choosing a storage location.

5.1. Principles for choosing a preservation service

Specialists in data preservation divide services into roughly two categories: services that store or preserve research data. Storing services retain files exactly as they were stored in the service, while preserving services have the capacity to maintain the data in understandable form, even though file formats and software change over time. Discipline-specific services that accept data only in a certain format are often preserving services, while general services that accept data in all forms are only able to store the data as such. The passage of time makes it increasingly likely that advances in the software used render the stored files unusable. In other words, if you wish to preserve your data for a very long time, you should choose services specialising in preservation.

As a rule, you should prefer universally recognised, discipline-specific data repositories. Certified data repositories offer another good option. Identifying a suitable preservation service for data is not always a simple matter. To make it easier, you can rely on the principles listed below: reliability, suitability, reusability, retention periods and levels of openness. In addition, you can use the flow chart for choosing a preservation service in section 5.3.

Reliability and services (certificates, curation, persistent identifiers):

- Check whether the preservation service of your choice provides persistent identifiers, such as Digital Object Identifiers (DOI) or accession numbers. Persistent identifiers are machine-readable unique strings of characters that make it possible to find and refer to research data.
- By choosing a curated data repository, you can promote the reuse of your data. In certain repositories, researchers store their data on their own, without quality control. In curated data repositories, the quality of data is ensured, among other means, by checking the functionality of file formats and enriching metadata.
- Remember to check data repository fees, such as a curating fee or surcharges for datasets exceeding the maximum size.
- Check whether the data repository you are considering has a certification, such as CoreTrustSeal. CoreTrustSeal certification is granted to repositories whose

preservation practices have been found to be sufficient by external assessors. Such assessment focuses on, among other things, the level of curation and the guarantees of service continuity.

- In certain cases, the geographical location of the server underlying the data repository may have a bearing on data preservation. Personal data must not be transferred outside the EU unless research subjects have been informed of this.

Reusability (discipline-specificity, metadata):

- Check whether there is a commonly used data repository in your research field. By choosing a discipline-specific repository, you will promote the increasingly efficient reuse and findability of data.
- General data repositories accept data from a range of disciplines and different data types. Such repositories are a good alternative, especially in the case of research fields that do not have established discipline-specific data repositories.
- If the data are not suited to a general data repository, for example, because of their large size, an organisation-specific preservation service may be the most appropriate solution.
- Check in advance the requirements of the preservation service of your choice for describing data (e.g., keywords, methods used, description of variables).

Suitability (size limits, data types, file formats):

- The size of the research dataset affects the selection of data repository – remember to check any size restrictions and surcharges.
- Check whether the repository is suitable for the data type or file format you need. Not all data repositories necessarily accept code or software.

Retention periods (minimum retention periods, long-term preservation):

- Information provided by data repositories on retention periods varies. Check whether the repository has indicated on its website the long-term nature of preservation or minimum retention periods.

Levels of openness (access rights, publishing delays):

- Many data repositories support open licences. Some repositories require a CC0 licence as a condition of storing. The CC0 licence is an open licence type of Creative Commons (CC) licence. In other words, it contains no restrictions on use, with the author waiving all rights to the material.
- If you are unable to make your data open even though you need a long-term preservation location for them, check whether the data repository of your choice can preserve the data in a way that they are only available to others with limited access rights, such as a research permit.
- If you need an embargo for the data before opening it, check that the data repository of your choice provides an option to set an embargo period.

Further information on choosing a preservation service:

- Zenodo data archive: Frequently asked questions
- Dryad data archive: Frequently asked questions
- Software Sustainability Institute: Software Deposit: Guidance for Researchers
- CoreTrustSeal: CoreTrustSeal criteria

5.2. How to identify a suitable preservation service

You should take advantage of your research community when looking for a discipline-specific data repository. Ask your colleagues which repositories they usually use or check key articles in your field to see where their data has been made openly available. You should note that the general practice of opening data merely as supplementary material to articles is not conducive to promoting the findability or reuse of data.

The Registry of Research Data Repositories (re3data) makes it easier to find field-specific data archives.

5.3. Data repository decision tree

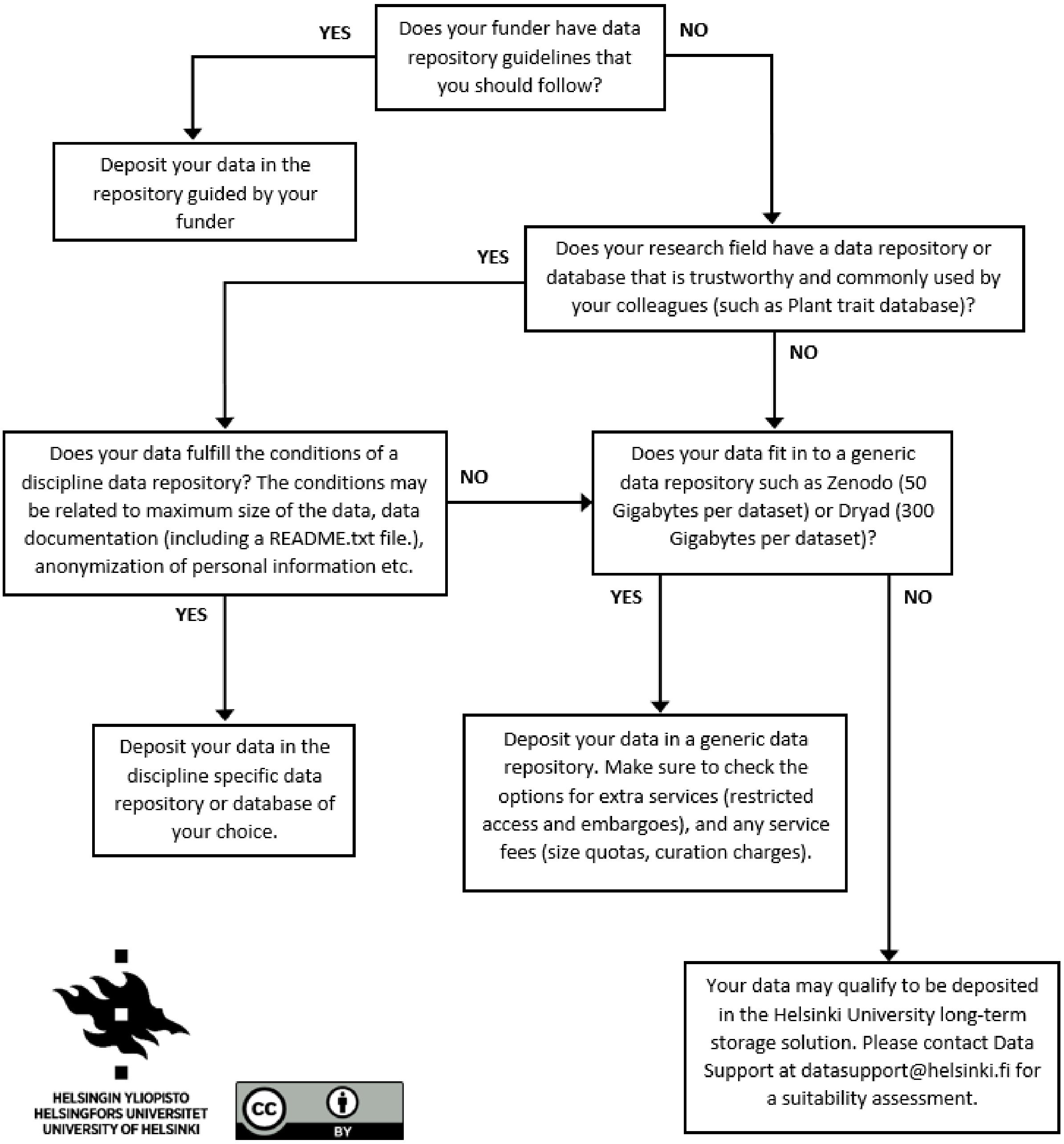

Does your funder have data repository guidelines that you should follow?
YES
NO
Deposit your data in the repository guided by your funder
Does your research field have a data repository or database that is trustworthy and commonly used by your colleagues (such as Plant trait database)?
YES
NO
Does your data fulfill the conditions of a discipline data repository? The conditions may be related to maximum size of the data, data documentation (including a README.txt file.), anonymization of personal information etc.
NO
Does your data fit in to a generic data repository such as Zenodo (50 Gigabytes per dataset) or Dryad (300 Gigabytes per dataset)?
YES
YES
NO
Deposit your data in the discipline specific data repository or database of your choice.
Deposit your data in a generic data repository. Make sure to check the options for extra services (restricted access and embargoes), and any service fees (size quotas, curation charges).
Your data may qualify to be deposited in the Helsinki University long-term storage solution. Please contact Data Support at datasupport@helsinki.fi for a suitability assessment.
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
CC
BY

Image 8: A flow chart guiding the selection of a data repository.

6. Link compilation

1. Why are research data preserved?

- Open Science: What is open science?
- UNESCO (2021): Recommendation on Open Science

Principles of responsible data preservation:

- University of Helsinki: Long-term preservation of confidential data
- Open Science: Opening sensitive datasets in a responsible manner (link in Finnish only)
- Responsible Research: Safeguard the anonymity of research subjects!
- Data Archive: Anonymisation and personal data

Examples of use:

- Legal counsels for research at the University of Helsinki: tutkimuksenjuristit@helsinki.fi
- Publishing delays: Sherpa Romeo

2. How to choose the data to be preserved

Choosing the data to be preserved:

- Digital Preservation (Fairdata-PAS): Guidelines for UH Evaluators
- DCC: Five steps to decide what data to keep
- UK Data Service: Collections development selection and appraisal criteria

3. Retention periods and destruction

Retention periods:

- Data Protection Ombudsman: Storage limitation (with regard to personal data)
- Deakin University: What data do I need to keep and for how long?

Destruction of data:

- University of Helsinki Helpdesk: Deleting files safely

4. How to prepare data for preservation

Advanced instructions related to data preparation:

- Flamma: Agreements on the opening of data between collaboration partners
- Flamma: Undertaking on the transfer of rights
- Helsinki University Library: Licences (CC), licences for code and software

Data protection:

- University of Helsinki Data Support: ‘Do I have personal data?’ test
- Helsinki University Library: Copyright and licences
- Office of the Data Protection Ombudsman: Special categories of personal data

File formats fit for preservation:

- Wikipedia: List of open file formats
- Digital Preservation: File formats suitable for long-term preservation (PDF, recommendation for the preservation of data sets for over 25 years)

Describing data:

- Fuchs & Kuusniemi (2018): Making a research project understandable
- An example of minimum requirements for metadata: DataCite Schema – Mandatory fields

5. Where are data preserved?

Principles for choosing a preservation location:

- Zenodo data archive: Frequently asked questions
- Dryad data archive: Frequently asked questions
- Software Sustainability Institute: Software Deposit: Guidance for Researchers
- CoreTrustSeal: CoreTrustSeal criteria

How to find a suitable preservation location

- A search platform for data archives: re3data