Storage and Publication of Research Data
What is a Repository?

*<latin>* storage

Repository system:

hardware & software combination to store data in order to make the **accessible**

Archive

the purpose of an archive is **long time storage**
Research data repositories provide
Research data repositories provide

• a searchable catalogue of research data
Research data repositories provide

- a searchable catalogue of research data
- data saved in a format that can still be opened in the future
Research data repositories provide

- a searchable catalogue of research data
- data saved in a format that can still be opened in the future
- combined information packages of data and documentation
Research data repositories provide

- a searchable catalogue of research data
- data saved in a format that can still be opened in the future
- combined information packages of data and documentation
- traceable data
How to choose a repository

Discipline-specific repositories are usually the best place for your data:

https://datorium.gesis.org
Other examples of discipline specific repositories

- **arXiv**: physics, math, computer science
- **CLARIN-D**: linguistic and language data
- **VerbundFDM**: Research and Information in Education
- **GenBank**: genetic sequence database ...

*List of accredited data centres for social and economic data*

https://www.ratswd.de/forschungsdaten/fdz
Trusted repositories: Data Seal of Approval
Trusted repositories: Data Seal of Approval)

- Data can be found on the Internet
Trusted repositories: Data Seal of Approval

- Data can be **found** on the Internet
- Data is **available** (unique rights and licenses)
Trusted repositories: Data Seal of Approval

- Data can be **found** on the Internet
- Data is **available** (unique rights and licenses)
- Data can be **reused** (format)
Trusted repositories: Data Seal of Approval

- Data can be found on the Internet
- Data is available (unique rights and licenses)
- Data can be reused (format)
- Data is reliable/trustworthy
Trusted repositories: Data Seal of Approval

- Data can be found on the Internet
- Data is available (unique rights and licenses)
- Data can be reused (format)
- Data is reliable/trustworthy
- Data have PIDs so that they can be referenced
Research data repository @ UHH
FDR @ UHH
FDR @ UHH

Open

When uploading, you decide whether your files are made available as open, restricted or closed, or whether they are provided with an embargo period.
FDR @ UHH

Open
When uploading, you decide whether your files are made available as open, restricted or closed, or whether they are provided with an embargo period.

Citable
Each entry in the repository is provided with a persistent DOI (Digital Object Identifier) and can be found and cited via this.
FDR @ UHH

Open
When uploading, you decide whether your files are made available as open, restricted or closed, or whether they are provided with an embargo period.

Citable
Each entry in the repository is provided with a persistent DOI (Digital Object Identifier) and can be found and cited via this.

Secure
sets of up to 50 gigabytes can be uploaded via the web interface (more upon request) and are securely stored at the RRZ of the Universität Hamburg (S3 memory).
FDR @ UHH

**Open**
When uploading, you decide whether your files are made available as open, restricted or closed, or whether they are provided with an embargo period.

**Citable**
Each entry in the repository is provided with a persistent DOI (Digital Object Identifier) and can be found and cited via this.

**Secure**
Sets of up to 50 gigabytes can be uploaded via the web interface (more upon request) and are securely stored at the RRZ of the Universität Hamburg (S3 memory).

**Sustainable**
The retention period of your files is at least 10 years.
FDR @ UHH

Open
When uploading, you decide whether your files are made available as open, restricted or closed, or whether they are provided with an embargo period.

Citable
Each entry in the repository is provided with a persistent DOI (Digital Object Identifier) and can be found and cited via this.

Secure
sets of up to 50 gigabytes can be uploaded via the web interface (more upon request) and are securely stored at the RRZ of the Universität Hamburg (S3 memory).

Sustainable
The retention period of your files is at least 10 years.

Flexible
All materials that are available in digital form are suitable for long-term storage in the repository. Any file format is possible.
Publication steps

1. Register with B identifier (Shibboleth)
2. Prepare & Upload Data
3. Enter Metadata
4. Save and publish
Prepare files
Prepare files

- File format: as open as possible
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
Prepare files

- File format: as open as possible
- File selection:
  - check for hidden files
  - check the first in order (this will show in the preview)
  - check for sensitive files / information you might not want to publish
  - delete obsolete backup or versioning information
  - check for big files (max 50GB / file)
  - unzip if necessary
- once published the data can not be changed, every change creates a new version (DOI)
Prepare the documentation
Prepare the documentation

✓ Research project description
Prepare the documentation

✓ Research project description
✓ underlying hypotheses
Prepare the documentation

✓ Research project description
✓ underlying hypotheses
✓ information on the collection of data
Prepare the documentation

✓ Research project description
✓ underlying hypotheses
✓ information on the collection of data
✓ measures for data cleansing or data weighting
Prepare the documentation

- Research project description
- underlying hypotheses
- information on the collection of data
- measures for data cleansing or data weighting
- explanations for codes, labels and data versions
What is good metadata?

# Metadata types

<table>
<thead>
<tr>
<th>Metadata type</th>
<th>Objective</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive metadata</td>
<td>Identification and traceability</td>
<td>Title, Author, Abstract, Keywords...</td>
</tr>
<tr>
<td>Administrative metadata</td>
<td>Creation and storage</td>
<td>Rights management and access rights, storage...</td>
</tr>
<tr>
<td>Structural metadata</td>
<td>Structure of the Resource</td>
<td>Chapters, sections, structure, files...</td>
</tr>
<tr>
<td>Technical metadata</td>
<td>technical characteristics</td>
<td>Scope, coding, technical structure, formats</td>
</tr>
</tbody>
</table>
Exercise “Metadata”

- Get into two groups
- Choose an object
- Describe the metadata (10 min)
- Each group then briefly presents the results
Why is metadata important?

- Metadata ensures long-term data usage and interpretation
- Well-documented data are also understandable by people who were not involved in the research
- Structured metadata can be read and processed by machine for fast, easy and reliable searches
Overview of discipline-specific metadata standards

http://www.dcc.ac.uk/resources/metadata-standards
http://rd-alliance.github.io/metadata-directory/subjects/

- Keywords
  - Data Documentation Initiative (Social Sciences)
  - GESIS; Datorium or SSOAR (Social Science Open Access Repository)
  - Dublin Core (bibliographic data format)
  - Marc21 (bibliographic data format)
Required Metadata

**DOI**
enter DOI if your files have already been published; if empty a new DOI will be assigned

**Publication Date**
enter publication date if your data have already been published; current date as default

**Authors**
specify as many authors as you like, preferably with ORCID

**Description**
enter abstract if you publish a scientific paper; for files describe the data (or use the abstract of the data journal publication)

**License**
Open | Embargoed | Restricted | Closed Access
Recommended Metadata

Keywords
  keyword that will help to find your data

Funding
  the FDR is reporting to the European Commission via OpenAIRE

Related Identifiers
  specify identifiers of related publications
What license should I choose?
Can you copyright data?
• There is no copyright on ideas, concepts, principles or discoveries
  ▪ raw data
  ▪ factual information
  ▪ news
Copyright applies to original creative expression (not the result of hard work)
Databases and enriched data

- EU and German Urheberrecht (as well as US copyright) protects compilations of data (database), provided that a significant investment has been made
- the protection applies to the compilation, the relation not to the data itself
- enriched data may also be protected by intellectual property rights
Legal issues and research data

- Research data in general are free of rights, i.e. for geological data, personal data, surveys, bibliographic data etc. there is no ownership and therefore copyright does not apply here.
- Derivates are protected, like texts in which the data are described, evaluated or visualized can be protected.
- Also descriptions, pictures or other data which go beyond pure information can be protected.
- The freedom of information is an essential principle of the intellectual property rights.
Contractual and other restrictions

- for personal data the personal protection law applies regardless of copyright
- contractual conditions may restrict the use of data
- In contrast to intellectual property rights, the terms and conditions of the contract apply only to the person with whom the contract has been concluded
- funding agencies usually require a statement on the handling of data in advance (Research Data Management Plan)
Literature

FDR @ UHH
Describe and publish your research data at UHH data notes

http://uhh.de/fdm-datanotes

- Must be a member of UHH to publish
- Data has to be stored in long term repository (e.g. FDR) which provies a DOI
- Free of charge
Zentrum für nachhaltiges Forschungsdatenmanagement

Monetastr.4
20146 Hamburg

www.fdm.uni-hamburg.de

Marie Ryan marie.ryan@uni-hamburg.de Tel.: +49 40 42838-9598

Dr. Iris Vogel iris.vogel@uni-hamburg.de Tel.: +49 40 42838-8862
Helpful links

- Centre for Research Data Management (https://www.fdm.uni-hamburg.de/)
- FDM Blog @UHH https://forschungsdaten.blogs.uni-hamburg.de/blog-beitraege/
- Research Data Repository https://fdr.uni-hamburg.de
- DMP Online http://dmp.fdm.uni-hamburg.de/ (coming soon)
- www.forschungsdaten.org
- www.forschungsdaten.info