

Research Data Management From planning to archiving



RDM Competence Centre
Heinrich-Heine-Universität Düsseldorf





► *RESEARCH DATA*

Research data might include measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections, or samples that were created, developed or evaluated during scientific work (DFG Guidelines on the Handling of Research Data, 2015).

► *RESEARCH DATA MANAGEMENT (RDM)*

RDM is the systematic effort to make this data usable over the entire data life cycle. This ranges from the collection, the evaluation, (further) processing to the archiving and, if necessary, the publication of the data.

► *FUNDING*

Projects are based on collaboration and central access to data simplifies cooperation in collaborative projects.

For the approval of third-party funds, proof of a structured data management is becoming increasingly important – have an answer ready!

► *ADVANTAGES: RDM ...*

- reduces the risk of data loss
- makes data available and reusable
- prevents overabundant data and redundancy
- promotes the implementation of ethical standards and principles of good scientific practice
- creates legal certainty
- improves data exchange within research groups (e. g., data transfer between generations of doctoral students)

“Quality-assured research data form a cornerstone of scientific knowledge and can [...] serve as a basis for further research. The sustainable safeguarding and provision of research data therefore does not serve only the examination of previous results, but to a large extent also the achievement of future results.”

From the preamble of the “Principles for the Handling of Research Data”, Alliance of Science Organisations in Germany, 2010



IS YOUR RESEARCH DATA FAIR?



Findable

Your research data can be found and cited.



Accessible

Your research data is accessible.



Interoperable

Your research data is technically reusable (formats, software).



Reusable

Your data is understandable and reusable.

... further information at
forschungsdaten.info

RDM Competence Centre at HHU

... supports you in all matters relating to research data management with the following services:

GUIDANCE ON

- Project proposal
- Data organisation
- Data usage
- Data publication

TOOLS

- Software for the handling of data
- Advice in dealing with data

TRAINING

- Courses in data management
- FAIR principles for data
- Data publications

Data Management Plans

How to plan, structure and coordinate data



DMP at HHU

Simply fill in online –
and off to the proposal!

What is a data management plan (DMP)?

A DMP supports you by structuring the handling of research data personally and for collaborative projects.

DMP serve as checklists and for ongoing documentation: from collection to long-term storage or publication of the data. Some EU or BMBF funding programs require a DMP as part of an application - have you already created a DMP?

What information does a DMP contain?



Create your DPM here:
<https://rdmo.hhu.de>



ADVANTAGES

- structures arrangements on data exchange for collaborations
- facilitates documentation for reporting obligations
- simplifies the reuse of your own data
- reduces the risk of data loss

Storage Media

Where to put your data



Margaret Hamilton with the printed source code of the software she developed for the Apollo mission, 1969

Storage Media

No backup – No mercy

Spinning hard drives aren't built to last. Sometimes they don't even last long enough to keep research data for ten years, in accordance with good scientific practice.



• STORAGE MEDIA LIFE

- - Hard disks: 2 to 10 years
 - DVD: up to 30 years
- Flash drive: 10 to 30 years

3... 2... 1... Backup!

Data loss - research loss. You can prevent this disaster with backups, but where and how? Clouds are practical, but there are also ambiguities when using them: where is the data or what happens if the provider is hacked or even quits the service. Secure your data by choosing a reliable platform (e.g. Sciebo) and an additional storage method (e.g. the university's servers).

• 3-2-1-RULE

- At least **3** backup copies on **2** different storage media, with **1** backup copy at an external location.

Storing – and in the longterm?

Hard disks get lost – repositories don't. Store valuable data in a way that keeps it accessible and safe at the same time. Suitable options are repositories or university services such as HHU ResearchData. (<https://researchdata.hhu.de>).

Persistent Identifiers

Finding data and authors –
and being found!



PID and ORCID

How do you find something reliably and permanently on the internet?



> *Persistent Identifier (PID)*

are flexible signposts that transparently link an unchangeable name with the flexible address of a file. In the event of a technically necessary change of address, the PID guarantees permanent accessibility of the desired file.

Publications are referenced and findable with the help of DOI (Digital Object Identifier). At HHU, DOIs can be assigned by the University and State Library.



> *ORCID iD*

Multiple names, different spellings or a name change can lead to the fact that authors are not clearly assignable.

There are other PIDs and the FDM Competence Center will be happy to inform you about them. This is where a PID for authors helps – e.g. the Open Researcher and Contributor ID (ORCID iD). ORCID (ORCID.org) provides a persistent digital identifier that distinguishes you from all other researchers - like a fingerprint. The ORCID iD can be integrated into important research processes such as the submission of manuscripts and project proposals.

There are other PIDs which the RDM Competence Center will be happy to inform you about.

Data Publication

Data speaks for itself



Data Publication

Sharing is caring

“Sharing is caring“ – this is also true for research data. Your data is not only valuable for your own research, but can provide important impulses after your project is finished.

ADVANTAGES

- Publications with research data are cited much more frequently
- Datasets or data publications will continue to be used and then also cited
- Research results become comprehensible and reusable
- It is easier to compare results
- Interdisciplinary research is supported
- Meta-analyses are made possible

How do I publish my data properly?

If you publish your data in a repository that is accessible online, your data is safely stored, findable and usable at the same time.

- > **Discipline-specific repository:** In many scientific communities, there are established services. <https://re3data.org>
- > **Generic repository:** e. g. Zenodo, Radar
- > **University repository:**
HHU ResearchData <https://researchdata.hhu.de>
- > **Data Journals:** Focus on the description and methodology of data generation
- > **Discipline-specific journals:** Supplementary to text publications

Metadata

Knowing what you are
looking for



Metadata Standards

Data about data

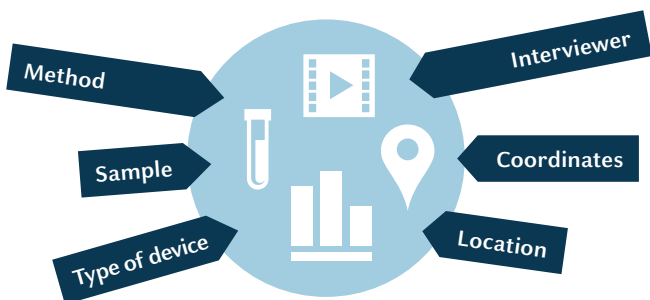
The label on a can indicates what's inside. In the same way, metadata ensures that digital data and objects can be found and used. That's why it's important to "label" your research data.

Types of metadata

> *Bibliographic or administrative data* is rather general and less community-specific. It includes information about the origin and management of the entire data set.

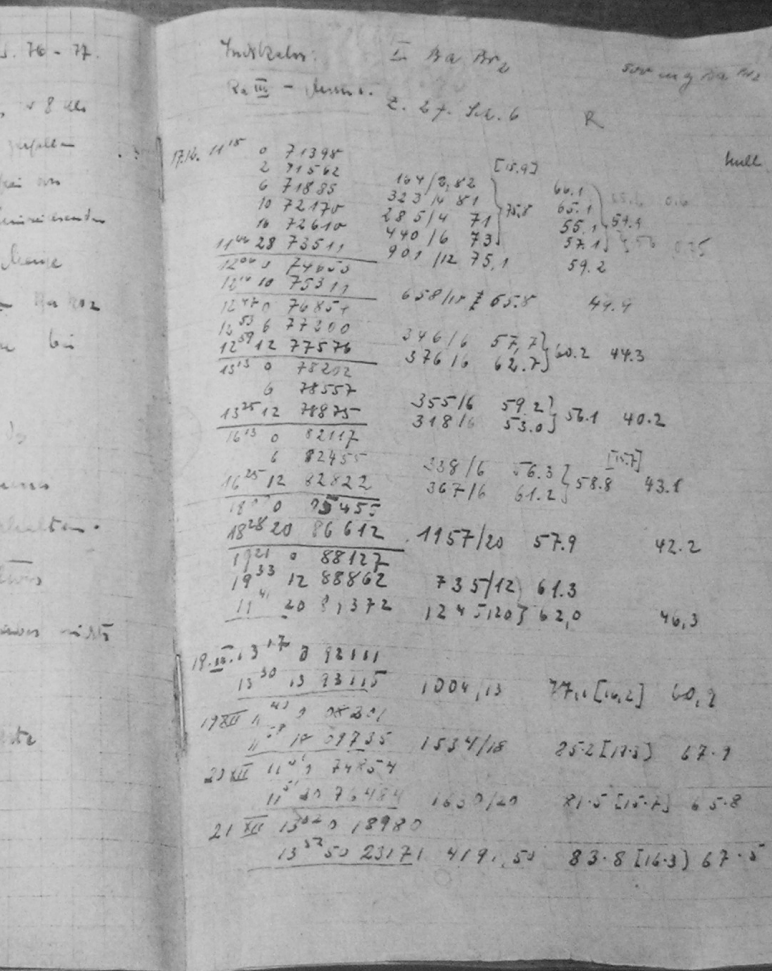


> *Content-descriptive or subject-specific data* describes individual aspects or data sets in more detail and provides additional information. Depending on the discipline, metadata is structured quite diversely. Many disciplines already have their own metadata standards.



Electronic Lab Notebooks

The end of your jotter chaos



Electronic Lab Notebooks

Improve your lab workflows digitally

Notebooks are part and parcel of the day-to-day business of research in the natural sciences: this is where measurements, sketches of experimental setups or evaluations end up. But what if your data is digital? Print everything out and glue it into your notebook?

Probably not. Instead, consider using electronic laboratory notebooks (ELN).



WHY USE AN ELN?

- Work collaboratively
- Full text search in all contents
- Data security, access control
- Keep book with PC, mobile phone or tablet
- Ensure data integrity, create verifiability through time stamps
- Import or link files
- Connect to other systems (API)
- Export to PDF (and other formats)



Overview

<https://fdm.hhu.de>

Archiving Images and Audio-visual Files

The format matters



Archiving Images and Audio-visual Files

... including „born-digitals“

Without a player, your tapes are plastic waste. Technical evolution also impacts digital file formats: In a worst case scenario, your research data can no longer be accessed or edited.

To sidestep any issues, use the following formats to archive your data.

FIT FOR THE ARCHIVE: FORMATS

► Raster graphics

- Tagged Image File Format (TIFF) – uncompressed > *.tif
- Portable Network Graphics (PNG) > *.png
- JPEG2000 > *.jpg, *.jpeg

► Vector graphics

- Scalable Vector Graphics (SVG) > *.svg, *.svgz

► Computer-Aided Design (CAD)

- AutoCAD Drawing > *.dwg
- Drawing Interchange Format, AutoCAD > *.dxf
- Extensible 3D, X3D > *.x3d, *.x3dv, *.x3db

► Sound, Audio

- Waveform Audio File Format (WAV) – uncompressed > *.wav

► Video

- FFV1 Codec in Matroska Container > *.mkv
- Motion JPEG 2000 (ISO / IEC 15444-4) > *.mj2
- AVI - uncompressed > *.avi

Personal Data

Operation: undercover



(Research-)Data Protection

Keep personal info personal

General Data Protection Regulation (GDPR), ethical guidelines, federal or state data protection laws, written consent, principles of purpose or "broad consent", pseudonymisation or anonymisation: anyone who wants to handle personal research data in a responsible and legally compliant manner must observe many regulations.



WHY DATA PROTECTION?

- Compliance with ethical standards
- Strengthen confidence in research
- Putting long-term storage, use and/or transfer on a legally sound foundation
- Adhere to the specifications of funding bodies



Further information

<https://hhu.de/datenschutz>