Research Data Management: Archiving and Publishing Research Data

Research Data Management at the University of Paderborn

Dr. Daniela Hartmann

January 15, 2026

Research Data Management Contact Point

The Research Data Management Contact Point at Paderborn University

We

- are the first point of contact for RDM
- advice from application to publication
- provide training (methods, tools, etc.)
- provide services and tools

E-Mail: kontakt-fdm@imt.uni-paderborn.de

https://www.uni-paderborn.de/forschung/foerder-und-antragsberatung/forschungsdaten#ansprechpersonen

Our Workshops

- 30.10.2025: Overview of Research Data Management Topics
- 27.11.2025: Data Management Plans
- 11.12.2025: Data Organization
- 15.01.2026: **Archiving and Publishing Research Data**
- 05.02.2026: Electronic Lab Notebooks, eLabFTW

Where can I find help?

- **Contact us!** The Contact Point Research Data Management at Paderborn University
  - Visit our website: https://www.uni-paderborn.de/en/research/funding-and-application-advice/research-data
  - E-Mail: kontakt-fdm@imt.uni-paderborn.de

- **forschungsdaten.info**
- **forschungsdaten.org**
- **Wiki of the ETH Zürich**: https://unlimited.ethz.ch/display/DD/FDD

- The **NFDI** and their consortia: https://www.nfdi.de/

Team and Expertise


Dennis Friedl


Sina Gantenbrink


Dr. Daniela Hartmann


Daniel Röwenstrunk


Dr. Gabriel Simoneit

Expertise:
Data management plans, electronic laboratory notebooks, data organisation and backup, data publication and long-term availability, legal aspects and licences

Agenda

1. **What is Research Data Management?**
2. **Backing up and archiving data**
3. **Documenting data**
4. **Publishing data**
5. **Finding and reusing data**

What is Research Data Management?

PADERBORN UNIVERSITY

“Research data includes measurement data, laboratory values, audiovisual information, texts, survey or observation data, methodological test procedures and questionnaires. Compilations and simulations can likewise constitute a key outcome of academic research and are therefore also included under the term research data.”

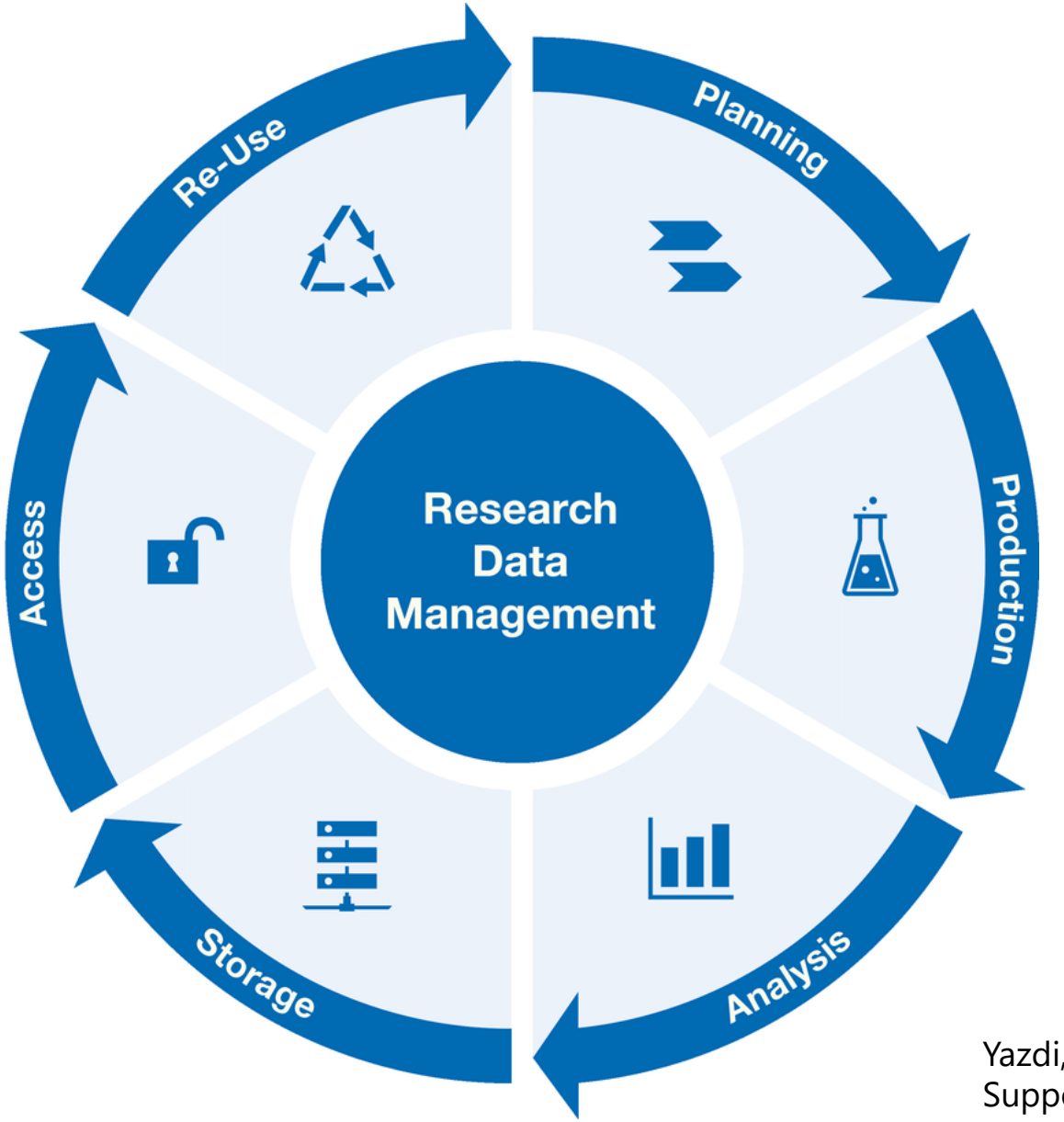
“Handling of Research Data”, DFG, 2023
https://www.dfg.de/en/research-funding/funding-initiative/research-data

What is Research Data Management?

"Research data management encompasses the processes of transforming, selecting and storing research data with the common goal of keeping it ***accessible, reusable and verifiable in the long term, regardless of the individuals involved****. To this end, structured measures can be taken at all points in the* ***data life cycle*** *that are suitable for preserving the scientific significance of research data, maintaining its accessibility by third parties for evaluation and analysis, and securing the chain of evidence."*

Translated from: https://forschungsdaten.info/themen/informieren-und-planen/was-ist-forschungsdatenmanagement/

Research data lifecycle



Yazdi, M.. (2020). Enabling Operational Support in the Research Data Life Cycle.

FAIR Data

Findable
- Persistent Identifier, Metadata, searchable directories

Accessible
- Open and free protocol, accessibility of metadata

Interoperable
- (Meta) data uses a generally applicable language for knowledge representation, references to other (meta) data.

Reusable
- Data usage license, information about its creation

Advantages of Reusable and verifiable data

- Advantages for **researchers**
  - Efficiency (easier work)
  - Traceability
  - Visibility and reputation, citation
  - Legal certainty
  - Long-term usability
- Advantages for **teams**
  - Collaboration
  - Knowledge retention
- Advantages for **science**
  - Global cooperation
  - Innovation
- Advantages for **research institutions**
  - Quality assurance
  - Resource conservation
  - Attractiveness for collaborations
  - Long-term archiving
- Advantages for **funding organizations and policymakers**
  - Verifiable use of funds
  - Promotion of open science
  - Driver of innovation
- Advantages for **society and the economy**
  - Knowledge and technology transfer
  - Education

Backing up and archiving data

Backing up data – UPB network storage

- ZIM storage
- High security standards, GDPR compliant
- Personal storage and project or institute storage
- Synchronization
- Backups and snapshots
- Plenty of storage space on request

- https://hilfe.uni-paderborn.de/Netzwerkspeicher

Backing up data – Sciebo

- Non-commercial cloud storage service (Nextcloud for North Rhine-Westphalia)
- Also suitable for binary data
- Synchronization and collaborative work
- Sciebo project boxes (https://hochschulcloud.nrw/de/hilfe/projektboxen/)

- Additionally: Calendar, contacts, surveys, LaTeX editor (Overleaf)

- https://uni-paderborn.sciebo.de/

Backing up and archiving data – backup vs. archive

- Backup – backing up data correctly so that
  - it is not lost in the short term
  - it can be traced
  - it can be worked on collaboratively
- 3-2-1 rule: 3 copies, 2 media, 1 external copy
- @UPB e.g. file server, Sciebo project boxes

- Archive – proper storage of data so that
  - it remains accessible long after the project has ended
- Open file formats (csv, xml, ASCII, txt…)
- @UPB e.g., Coscine

Archiving data – Coscine

- Management of data sets in projects
- Storage via Datastorage.nrw
- GitLab integration
- As a central storage service in North Rhine-Westphalia
- Central description with metadata
- Applications for storage space possible

- https://coscine.rwth-aachen.de/

Documenting data

Documenting data – Overview

- Describing data with **metadata**
  - README files
  - Citation File Format (CFF)
  - Code books, naming conventions
  - Electronic lab notebooks
  - Metadata standards: Dublin Core, DataCite Metadata Schema
- Catalogs
  - FAIRsharing.org
  - RDA Metadata Standards Catalog

AI-generated with Perplexity AI (Pro), January 2026

Documenting data – Metadata

- **Data about data** (data alone is usually not self-explanatory)
- Metadata provides information about the **content**, **origin**, **format**, **context of origin**, and **conditions for use** of data.
- It contributes significantly to the findability, comprehensibility, and reusability of data.
- Metadata is intended for researchers who reuse it, but also for yourself.
- Metadata is based on (subject-specific) standards so that it can be understood by both humans and machines.

Documenting data – Metadata

A distinction is made between:

- **Bibliographic metadata** such as titles, authors, descriptions, or keywords enable the citation of data and code and help with findability and thematic narrowing.
- **Administrative metadata** on file types, locations, access rights, and licenses help with the management and long-term preservation of data.
- **Process metadata** describes the steps and actions, including the methods and tools used, that were applied to create and process the data.
- **Descriptive metadata** can vary greatly depending on the discipline and provides additional information on the content and origin of the data.

Documenting data – Metadata standards

Examples of **general metadata standards** include:

- DublinCore from the Dublin Core Metadata Initiative
- DataCite Metadata Schema

Documenting data – Metadata standards

DublinCore

title: "Perfect Chocolate Cake - Optimized Recipe v2.1"
creator: "Dr. Anna Müller, Engineering Test Kitchen"
subject: "chocolate cake; baking protocol; recipe optimization; food engineering"
description: "Complete sensor-monitored baking data from 12 test batches. Temperature, timing, ingredient ratios, and structural analysis for perfect results."
publisher: "University of Paderborn Recipe Repository"
contributor: "Test Kitchen Team: J. Schmidt (baking), M. Weber (documentation)"
date: "2026-01-14"
type: "Recipe Dataset"
format: "CSV (sensor data), PDF (protocol), 150 MB"
identifier: "https://doi.org/10.1234/upb.cake.v2.1"
source: "Conventional oven, digital scale, temperature probes"
language: "en"
relation: "Part of Food Engineering Test Series 2026"
coverage: "Paderborn Test Kitchen; temperature-controlled environment"
rights: "CC-BY 4.0 (Attribution required - share your improvements!)"

AI-generated with Perplexity AI (Pro), January 2026

Documenting data – Metadata standards

Examples of **general metadata standards** include:

- DublinCore from the Dublin Core Metadata Initiative
- DataCite Metadata Schema

In addition, a large number of **subject-specific schemes** have been developed that must be taken into account.
Subject-specific diagrams can be found under:

- Metadata Standards Catalog
- fairsharing.org

Documenting data – Metadata for data on file systems

- Information about the data in “README.md” files
- No fixed structure
- Markdown as formatting syntax
- Documentation on Markdown at GitHub
- GitHub Flavored Markdown Spec

- Citation File Format (CFF)
- Metadata for software and data

```yaml
cff-version: 1.2.0
message: "If you use this software, please cite it as below."
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/1234-5678-9101-1121
title: "My Research Software"
version: 2.0.4
identifiers:
  - type: doi
    value: 10.5281/zenodo.1234
date-released: 2021-08-11
```

Publishing data

Publishing data – Why?

- Good scientific practice: reproducibility of research (research integrity)
- DFG Guideline 13
- Reusability of data for other research questions
- Complex global challenges (e.g., climate change) require worldwide, interdisciplinary efforts
- Publication “forces” researchers to be careful

- Guidelines and procedures for ensuring good scientific practice at Paderborn University
- Open Access Policy of Paderborn University
- Open Educational Resources (OER) Policy of Paderborn University
- Guidelines for handling research data at Paderborn University

“Guidelines for Ensuring Good Scientific Practice” (09/2019)

Guideline 13: Making research results publicly accessible

"As a matter of principle, scientists contribute all their results to the scientific discourse. (...) Once a decision has been made to make results publicly accessible, scientists describe them in a complete and comprehensible manner. This also includes, as far as possible and reasonable, making available the research data, materials, and information underlying the results, the methods used, and the software employed, and providing a comprehensive description of the work processes. Self-programmed software is made publicly available with the source code. Scientists provide complete and accurate references to their own and others' preliminary work."

Publishing data – Publication channels

- As a supplement to a peer-reviewed article (extended)

- Data journals
  - Publication of detailed descriptions of the data (metadata)
  - Possibly peer-reviewed
  - https://www.forschungsdaten.org/index.php/Data_Journals

- Independent data publication in a research data repository

Publishing data – Repositories

- **Disciplinary repositories**, e.g., Datorium of GESIS, Pangea, FDZ at ZPID, ...

Directory: http://www.re3data.org, https://risources.dfg.de

- Repositories of scientific journals/data journals
- **Institutional data repositories** (not yet available at UPB)
- **General data repositories**, e.g., https://zenodo.org, https://figshare.com
- Subject-specific or project-specific portals (data portals)
- Own website (note: not recommended, as it is not reliably available on a permanent basis)

Publishing data – Seal of approval for repositories

Quality and trust in research data management

What are seals of approval for repositories?

- Certificates or awards given to repositories for meeting quality standards
- signal reliability, long-term availability, and good data management practices

Why are they important?

- increase the confidence of researchers and funding agencies
- ensure that data is correctly archived, documented, and remains accessible
- support in selecting a suitable repository

Publishing data – Seal of approval for repositories

Quality and trust in research data management

Well-known quality seals and standards:

- DIN 31644 / nestor certificate (German standard for digital long-term archives)
- DINI-Zertifikat (German seal of approval for scientific repositories: Focus on open access, long-term availability, and user-friendliness)

Tips for use:

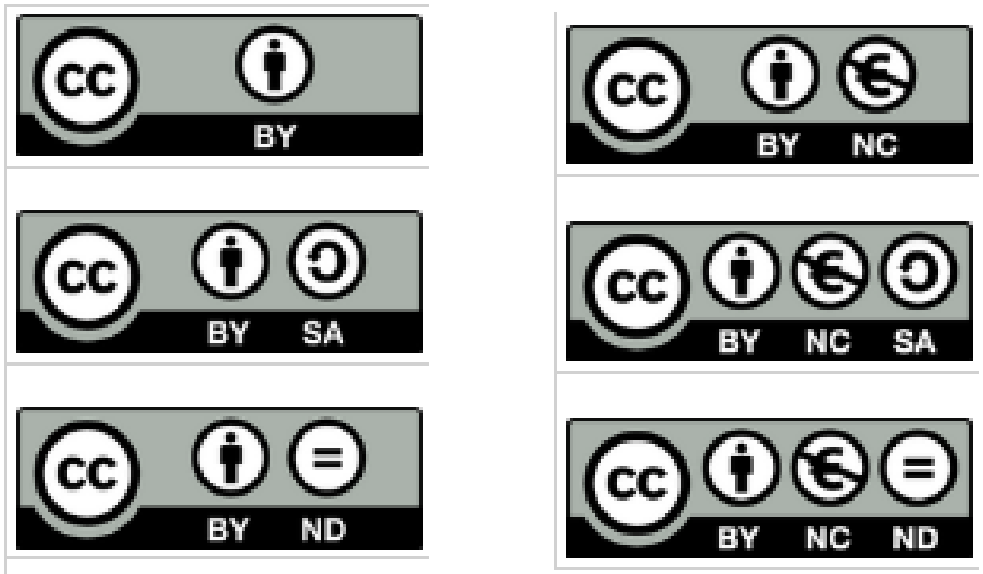
- Before submitting data, check whether the repository is certified.
- Pay attention to transparency regarding data protection, access conditions, and metadata quality.
- Use quality seals as a decision-making aid for long-term data archiving.

Publishing data – Licenses and rights

- License = contractually agreed right of use
- This allows the rights holder to permit their contractual partner to use a work in various ways (e.g., to copy, store, or make it digitally accessible).
- In many cases, rights holders charge a license fee for this. In addition to such commercial licenses, free licenses are also available. These allow the work to be used free of charge.

Publishing data – Licenses and rights

- Creative Commons(CC):
  - Easy to use
  - Modularly combinable
  - Offer security during publication and use
  - Freely available, standardized
  - National and international versions



https://de.creativecommons.net/was-ist-cc/

Publishing data – Persistent identifiers – DOI

- Long-term findability and avoidance of dead links, even in the event of name changes
- DOI for data (ISO standard 26324)
  - DOI = Digital Object Identifier
  - Unique, persistent (permanent) identifier for electronic publications
  - Enables data to be cited
  - Free of charge for scientific institutions in Germany

Example:

Schumacher, C. J., & Schmitt, M. (2025, Juli 8). Zeitzeugengespräch: "Internationalisiertes Geschäft Die Nixdorf Computer AG von Paderborn über Hannover bis Hongkong" (1.0). Zenodo. https://doi.org/10.5281/zenodo.15834060

Finding and reusing data

PADERBORN UNIVERSITY

Finding and reusing data

- Directly in subject repositories: https://www.re3data.org/
- In research data centers: https://www.konsortswd.de/datenzentren/alle-datenzentren/
- Via meta search engines Academic Search Engine “Base”: https://www.base-search.net/Search/Advanced
- Google Dataset Search: https://toolbox.google.com/datasetsearch
- B2FIND: http://b2find.eudat.eu
- gesisDataSearch: https://datasearch.gesis.org/start
- Elsevier Data Repository: https://elsevier.digitalcommonsdata.com/research-data/
- at Datacite (DOI Registrar): https://search.datacite.org/
- via references in publications
- Data on public administration (statistical data, geodata, traffic data, etc.) see the German federal government's “Open Data Act”: https://www.bmi.bund.de/DE/themen/moderne-verwaltung/open-government/open-data/open-data-node.html

Finding and reusing data – Legal information

- **Copyright (urheberrecht)** arises by operation of law and does not need to be registered.
- **Moral rights (urheberpersönlichkeitsrecht):** Relationship between creator and work.
  - Cannot be assigned and ALWAYS remain with the author (until 70 years after the author's death).
- **Exploitation rights (Verwertungsrecht):** Protection of economic interests.
  - Cannot be assigned.
  - Includes reproduction rights, distribution rights, rights of public access, etc.
- **Rights of use (Nutzungsrecht):** Determination of use by others
  - Can be granted
  - Exclusive to simple rights of use
  - Limited in terms of space, time, or content

PADERBORN UNIVERSITY

Summary

Summary

- Back up data – archive – document – publish – reuse
- FAIR is the guideline
- Repositories + PIDs + licenses = key to success

Help resources

Overview of help pages: Forschungsdaten.info, re3data.org
UPB: FDM contact point, Speicher, Sciebo, GitLab, Coscine
Tools: DMPonline, Zenodo, metadata generators

Time for questions and discussion

UNIVERSITÄT PADERBORN

Research Data Management Contact Point

Mail kontakt-fdm@imt.uni-paderborn.de

Web https://www.uni-paderborn.de/forschungsdaten