

Research Data Management Micro-Guides

Concise, practical guidance for researchers and students



Dr.-Ing. Fadwa Alshawaf

ORCID: [0009-0004-2091-1802](https://orcid.org/0009-0004-2091-1802)

Research Data Management Services

Humboldt-Universität zu Berlin

Version 1.0 · March 2026

Licensed under **Creative Commons Attribution 4.0 (CC BY 4.0)**


Contents

Support and services at HU	2
Research Data Management: What it is and how it supports good research	3
Data Management Plan (DMP): How research data are handled	5
Data documentation: Making data discoverable, understandable, and reusable.....	7
Metadata: Structured information for discoverability and interoperability	10
Data quality and quality control: Ensuring reliable and reusable research data.....	12
Data Storage: Secure research data during the project lifetime	14
Data archiving and preservation: Retention for at least 10 years	17
Rights, licenses, and ethical sharing: Enabling reuse responsibly	19
Roles, responsibilities, and resources: Who is responsible for research data and what is needed	21
FAIR Data Principles: Making research data findable, accessible, interoperable, and reusable	23
Sharing and citing research data: Making data reusable and giving credit	25



Support and services at HU

 Visit our Website **[Research Data Management](#)**

 Contact Us at **forschungsdaten@hu-berlin.de**

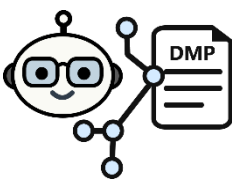
For technical issues, accounts, storage access contact **[CMS User Support](#)**

Is something not working as it should? You may find an answer on the CMS disruption page **[Service Disruptions](#)**.

Key Tools



I am your RDM assistant, ask me anything!



DMP Tool (AI-based data management plan preparation)



Data preservation (Institutional Storage services and data archiving with DOI assignment)



Research Data Management Organizer (Interactive DMP Tool)



Research Data Management: What it is and how it supports good research

What this is

Research Data Management (RDM) is a structured approach to handling research data throughout a project's lifecycle. It covers how data are **planned, collected, documented, stored, preserved, shared, and reused**.

RDM helps ensure that research data remain understandable, secure, reusable, and compliant with institutional and funder requirements.

The data lifecycle

Research data move through a lifecycle that typically includes:

- Planning how data will be created and managed
- Collecting, storing, and documenting data
- Processing and analyzing data
- Publishing data and findings
- Preserving data for the long term
- Reusing shared data



Data Lifecycle

Why RDM matters

Good RDM supports both research quality and research impact.

- Transparency and validation: Keeping clear records allows others to understand, verify, and reproduce your work.

- Impact and visibility: Well-managed and shared data increase the reach and reuse of your research.
- Efficiency for future projects: Your data can serve as a valuable resource for follow-up studies.
- Collaboration: Clear documentation and structure make collaboration within and across research communities easier.
- Long-term preservation: Research data can remain usable for future generations.
- Compliance: Many funders, institutions, and publishers require structured data management.

✅ When to think about RDM

As early as possible.

Ideally, RDM should be considered:

- When planning a research project
- Before data collection begins

Early planning prevents data loss, confusion, and rushed decisions later.

❌ Avoid this

- Delaying the Data Management Plan until the project is already underway
- Treating documentation as an afterthought
- Relying only on personal storage or ad-hoc file naming
- Assuming you will “remember later” how data were created or processed

💡 Practical next step

Start by creating a Data Management Plan (DMP).

A DMP helps you reflect on:

- What data you will create
- How you will document and store them
- How data will be preserved and shared

➡ Learn more

- [How to create a Data Management Plan](#)
- [The FAIR principles](#)
- [Metadata: Structured information for discoverability and interoperability](#)



Data Management Plan (DMP): How research data are handled

What this is

A Data Management Plan (DMP) is a planning document that describes how research data are **created, organized, documented, stored, preserved, and shared** during and after a research project. It helps researchers think through data-related decisions early, align with good scientific practice, and meet institutional and funder requirements. Good planning reduces risks, saves time later, and supports high-quality, reusable research data.

A DMP is typically required:

- At the **proposal stage** for funded research projects
- At the **start of a project**, before data collection begins
- Especially when projects involve:
 - personal or sensitive data
 - large or complex datasets
 - long-term preservation or reuse

A DMP should be a **living document** that can be updated as the project evolves.



Data Description



Documentation



Data Storage



Legal Obligations



Data Exchange



Responsibilities

Elements of a DMP.

Helpful resources

You can create a DMP using:

- [Text-based templates](#)
- [AI-based DMP tool](#)
- [RDMO](#)

These guides support you in writing your DMP:

- [Data description](#)
- [Documentation: Metadata](#)
- [Data quality and quality control](#)
- [Data storage and backup](#)
- [Data archiving and preservation](#)
- [Rights, licenses, and ethical sharing](#)
- [Roles and resources](#)

Avoid this

- Treating the DMP as a purely administrative requirement
- Ignoring the DMP until it is needed only for compliance
- Copying generic text without reflecting actual practices
- Ignoring data management once the DMP is submitted

Learn more

- [Research Data Management](#)
- [FAIR Data Principles](#)
- [Data sharing and citation](#)



Data documentation: Making data discoverable, understandable, and reusable

Documentation includes both **data description** (README, context, variables) and **metadata** (standards, identifiers).

Why it is necessary

Good documentation makes your data discoverable, understandable, and reusable by others, and by you in the future. Undocumented data are effectively lost data.

Without proper documentation, data cannot be reliably found, interpreted, or reused, even if they are stored in a repository.

Data description

Human-readable explanation of the data's content and context. It helps the understanding and interpretation of data.

Typical forms:

- README files
- Codebooks
- Method descriptions
- Variable explanations
- Editorial notes

✓ Do this

- Describe what the data are and how they were created
- Document methods, instruments, software, and versions used
- Explain variables, units, codes, software, and abbreviations

- Record data processing and cleaning steps
- Link documentation clearly to the data files

Data description should be created **at the same time as the data**, not afterwards.

✗ Avoid this

- Relying on memory instead of written explanations
- Using abbreviations or codes without definitions
- Keeping documentation in a separate, unlinked location
- Writing documentation only at the end of the project

What to describe

At a minimum, description should answer these questions:

- **What** does this file contain?
- **What** do the data represent?
- **What** do the variables mean?
- **How** were they generated or processed?

If someone outside your project cannot answer these questions, the data are not findable.

Example

File: interviews_youth-employment_2024.csv

Description:

- Study context: Semi-structured interviews on youth employment experiences
- Population: Participants aged 18–25, recruited in German urban areas
- Data collection: Interviews conducted between March–June 2024
- Language: English
- Variable Q1_status: Employment status at time of interview
 - 1 = employed
 - 2 = unemployed
 - 3 = in education or training
- Transcription method: Manual transcription, anonymized
- Missing values coded as NA

This level of documentation allows both humans and machines to understand the data.

[→ Learn more](#)

- [Metadata: Structured information for discoverability and interoperability](#)
- [How to create a data management plan](#)
- [Data storage](#)



Metadata: Structured information for discoverability and interoperability

What this is

Metadata are structured information that describe data in a way that can be understood by systems and machines.

Metadata enable the discoverability and indexing of data by search engines, repositories, and research information systems. They are essential for making data findable beyond their immediate project context.

Following recognized metadata standards improves the interoperability of data across platforms, disciplines, and services.

Metadata are the foundation of the **F (Findable)** principle of FAIR.

Metadata are usually entered into structured fields when a dataset is uploaded to a repository or archive. These fields are what makes data discoverable and findable in catalogues, search engines, and research information systems.

Typical forms:

- Repository metadata fields
- Metadata standards (Dublin Core, DataCite, discipline-specific schemas)
- Controlled vocabularies
- Persistent identifiers (DOI, ORCID, ROR)

Example metadata:

- Who created the dataset?
- What is the title, date, license?
- How can systems index and link it?

✗ Avoid this

- Leaving required metadata fields empty
- Using abbreviations or project-specific terms without explanation
- Using free-text keywords when controlled vocabularies are available
- Selecting a license without checking that it matches your data
- Treating metadata entry as a final, rushed step

✓ How to prepare metadata in advance

Metadata are usually entered at upload time, still you can prepare the content in advance. This saves time and improves quality. When uploading data, this information can be copied directly into repository fields.

Good practices include:

- Keeping a metadata text file during the project (title, description, creators, keywords, methods)
- Reusing information from:
 - Project descriptions
 - Data Management Plans
 - README files
- Maintaining a simple metadata checklist, such as:
 - Dataset title
 - Creator names and identifiers (e.g. ORCID)
 - Description of the data
 - Keywords or subject terms
 - License
 - Related publications or projects

➡ Learn more

- [Data description](#)
- [How to create a data management plan](#)
- [Data storage](#)



Data quality and quality control: Ensuring reliable and reusable research data

What this is

Data quality refers to how **accurate, complete, consistent, and reliable** research data are for their intended use. High data quality is essential for interpretation, reuse, and trust in research results. It requires planned measures and quality controls throughout the research project.

While good documentation and metadata enable discovery, data quality enables trust and reuse.

✅ What data quality means in practice

- **Accurate:** values correctly represent what was measured or observed
- **Complete:** no undocumented gaps or missing information
- **Consistent:** formats, units, and codes are used uniformly
- **Reliable for reuse:** data can be understood and used beyond the original project

Note that quality requirements depend on the **discipline, methods, and research questions**.

Data quality should be addressed **during data creation and processing**, not only at the end of the project. Measures to ensure data quality include:

- Using standardized methods, protocols, or instruments
- Applying consistent file formats, naming conventions, and units
- Using controlled vocabularies, classifications, or ontologies where appropriate
- Documenting assumptions, limitations or uncertainties in the data

Quality control and validation

Quality controls are the practical checks used to identify errors and inconsistencies, including:

- Plausibility checks (e.g. value ranges, missing values)
- Consistency checks across files or variables
- Version control for data and scripts
- Review or cross-checking by project members
- Logging corrections and changes

Quality controls should specify:

- when checks are performed
- how issues are documented
- Which digital methods and tools were used

✗ Avoid this

- Assuming data quality will be “fixed later”
- Applying checks without documenting them
- Using undocumented transformations or corrections
- Relying on proprietary tools without recording versions or alternatives
- Treating quality control as a one-time final step

➔ Learn more

- [Data archiving and preservation](#)
- [Metadata: Structured information for discoverability and interoperability](#)
- [How to create a Data Management Plan](#)



Data Storage: Secure research data during the project lifetime

What this is

Short-term data storage refers to the methods and locations you use to keep your **active research data safe, accessible, and backed up** while your project is ongoing. Proper short-term storage protects data from loss, protects sensitive data, supports collaboration, and ensures that analyses and documentation remain linked and reproducible throughout the research process.

Researchers are encouraged to use **institutional storage services** that provide redundancy, access control, and backup, rather than personal devices.

Why this matters

Short-term storage is key to:

- Data security: institutional backup protects against hardware failure and loss
- Collaborative access: all team members can reliably access current data
- Documentation integration: structured folders support linking to README and metadata
- These practices also make the later step of *archiving and publishing* far easier.

✅ Do this

- Use institutionally provided storage services rather than USB sticks or personal laptops, see options below.

- Ensure automatic backups by storing data where institutional backup systems are enabled
- Organize data into project folders with clear structure and naming (linked to your documentation guide).
- Set appropriate access rights for collaborators and restrict access for sensitive data.
- Regularly check that your data are accessible and backed up, especially after significant updates.
- Ensure that research data are protected through secure physical access, encryption, and controlled access management.
- Sensitive files should be encrypted, for example using strong password-protected folders in HU-Box or encryption software such as Gpg4Win or VeraCrypt.
- To prevent loss of access, at least two authorized persons should have access to the data.

The 3-2-1 Backup rule

To protect your active research data from loss, carry out regular backups at a specified time and follow the **3-2-1 backup rule**:

- **3 copies** of your data, the working copy plus two backups.
- **2 different storage types**. For example: network storage and institutional cloud storage.
- **1 copy stored off-site**. A location physically or logically separate from your main system.

✗ Avoid this

- Storing your only copy of data on:
 - personal laptops
 - USB sticks
 - unsynchronized external drives

These options are **not backed up and are vulnerable to loss**.
- Using multiple uncoordinated storage locations without clear folder structure.
- Sharing working data via email attachments instead of centralized project storage.

Examples of institutional short-term storage at HU Berlin:

- **HU-Box:** secure cloud storage with sync and share capabilities.
<https://hu.berlin/box>
- **Windows File Service / WebDAV:** network drives with central backups.
<https://hu.berlin/webfiles>
- **Media Repository:** for media files with structured storage options.
<https://hu.berlin/medien>
- **Datenbank-Service:** Databases suitable for storing and managing **large volumes of structured data** during an active research project.
<https://hu.berlin/datenbank>
- Contact information is available under this [link](#).

Learn more

- [Data description](#)
- [Metadata: Structured information for discoverability and interoperability](#)
- [How to create a Data Management Plan](#)



Data archiving and preservation: Retention for at least 10 years

What this is

Archiving and preservation ensure that research data remain secure, accessible, and useful for the long term. Many institutions and funders expect research data (or evidence supporting publications) to be retained for **at least 10 years**. Preservation is not “keeping a copy”, it is maintaining data and documentation so they can still be used in the future.

✓ Do this

- Select a repository intended for long-term preservation.
- Prefer discipline-specific repositories when available.
- Ensure the dataset receives a persistent identifier (typically a DOI).
- Deposit data + documentation together (README/codebook + context).
- Use open file format for storing digital data unencrypted, uncompressed, when applicable, to ensure that data can be accessed and read by a wide range of software applications. Explore the [List of open file formats](#).
- Define access level: Common options are:
 - **Open access** (Freely downloaded by anyone).
 - **Embargoed access** (Data are hidden and become open later).
 - **Restricted access** (Data are requested from a responsible person).
 - **Closed access** (Data are archived but cannot be shared outside the project).
- Sensitive data should be archived with restricted or closed access and accompanied by documentation explaining why they cannot be shared and under what conditions access may be granted.

How to Choose a repository

1. **Discipline-specific repositories:** Best for findability and reuse. They use community metadata standards, domain terminology, and established discovery channels (via [re3data](#)).
2. **Institutional repository:** Strong choice when no suitable disciplinary option exists, or when institutional policy requires institutional deposit ([edoc-server](#), [Media Repository](#)).
3. **Generic repositories:** Use when neither disciplinary nor institutional options fit. Choose reputable services that support metadata quality, persistent identifiers, and clear access controls ([Zenodo](#), [Dataverse](#)).

✗ Avoid this

- Storing “archived” data on temporal storage options.
- Depositing data without sufficient documentation (future users cannot interpret it).
- Uploading proprietary formats when open alternatives exist (future access risk).
- Treating archiving as an end-of-project panic task.

Practical checklist

Before deposit, confirm you have:

- Dataset, title, and description
- Context needed to interpret the data
- Creator names and affiliations (and ORCID where possible)
- Keywords/subjects/controlled vocabularies
- File description (what each file contains)
- Ownership statement
- License (or a clear restriction note)
- Access conditions (open/embargo/restricted)

Learn more

- [Rights, licenses, and ethical sharing](#)
- [Data storage](#)
- [Data sharing and citation](#)



Rights, licenses, and ethical sharing: Enabling reuse responsibly

What this is

Sharing data for reuse requires more than “uploading files.” You must clarify what others are allowed to do with the data and ensure sharing complies with legal, ethical, contractual, and cultural obligations. This is essential for responsible reuse and often required in Data Management Plans.

Why this is necessary

- A license tells users what is permitted (reuse, adaptation, redistribution, commercial use).
- If you do not assign a license (or rights statement), reuse becomes legally unclear and often avoided.

✓ Do this

- Check restrictions: consent terms, contracts, NDAs, data provider agreements, copyright.
- Classify data sensitivity: personal data, confidential data, culturally sensitive data, high-risk locations or species.
- Identify **who holds the rights**: you, your institution, collaborators, or third parties.
- Decide the **appropriate access** route:
 - Open access
 - Embargoed access (time-limited delay)

- Restricted/controlled access (application or mediated access)
 - No sharing (with documented justification)
- Select a **clear license** for reuse when sharing is permitted (e.g., Creative Commons for many research outputs (**CC0 or CC-BY**). Other licenses may be needed for software (**MIT/GPL/Apache**) or databases (**ODC or ODC-BY**).
- Choose a license consistent with:
 - funder/institution policy
 - third-party rights
 - ethical commitments
 - participant consent

✗ Avoid this

- Assuming “publicly available” means “free to reuse.”
- Choosing a license before checking third-party content (images, questionnaires, corpora, proprietary sources).
- Sharing sensitive data openly when access controls or anonymization are required.
- Relying on anonymization as a blanket solution (re-identification risk can remain).
- Ignoring cultural or community governance norms where relevant.



Practical step to save time at deposit

To reduce back-and-forth later, prepare a **reuse decision note** during the project:

- What can be shared (and what cannot)
- Conditions (embargo, restriction, mediated access)
- Rights holder(s)
- Selected license/rights statement and why
- Contact point for access requests (if restricted)



Learn more

- [Data archiving and preservation](#)
- [Metadata: Structured information for discoverability and interoperability](#)
- [How to create a Data Management Plan](#)



Roles, responsibilities, and resources: Who is responsible for research data and what is needed

What this is

Clear roles and sufficient resources are essential for an adequate and responsible handling of research data. Defining responsibilities early helps ensure that data are properly managed during the project and curated after the project ends.

✓ Who is responsible for research data

Responsibilities for RDM are usually shared across several roles within a project.

Principal Investigator (PI) / Project lead

- Holds overall responsibility for the appropriate handling of research data
- Ensures that legal, ethical, and funder requirements are met
- Approves decisions on data sharing, access restrictions, and archiving

Researchers and data creators

- Collect, document, and organize research data during the project
- Follow agreed storage, backup, and documentation practices
- Prepare data and metadata for archiving and sharing

Institutional services and infrastructure providers

- Provide storage, backup, repository, and archiving services
- Maintain technical infrastructure and security
- Offer guidance and support for research data management

After the project ends

- A responsible person or unit must be defined for curating and maintaining the data
- This may be the PI, an institutional repository, or another designated service

✓ Resources needed for adequate data management

These resources should be considered early and, where required, included in the project budget:

- **Time** for documentation, metadata entry, and data preparation
- **Storage and infrastructure** for short- and long-term data retention
- **Expert support** (e.g. data stewards, IT services, legal or ethics advice)
- **Financial resources** for storage, archiving, or specialized services, where applicable

✗ Avoid this

- Leaving responsibilities undefined or implicit
- Assuming that data management happens automatically
- Planning resources only at the end of the project
- Failing to assign responsibility for data after project completion

➔ Learn more

- [Data Storage](#)
- [Data archiving and preservation](#)
- [How to create a Data Management Plan](#)



FAIR Data Principles: Making research data findable, accessible, interoperable, and reusable

What this is

The **FAIR principles** are a set of internationally recognized guidelines for research data management developed to make data **Findable, Accessible, Interoperable, and Reusable**. They describe how research data should be organized and documented so they can be discovered, accessed, combined, and reused by both humans and machines.

Findable

Data and metadata must be easy to discover.

- Assign a persistent, globally unique identifier (e.g., DOI).
- Provide rich metadata that describe the data fully.
- Include the identifier explicitly in the metadata.
- Register data and metadata in a searchable resource (e.g., repository).

Interoperable

Data and metadata should be usable in combination with other data and tools.

- Use formal, shared, and widely applicable formats for metadata.
- Adopt controlled vocabularies and ontologies that follow FAIR principles.
- Include qualified references to related (meta)data so they can be linked.

Accessible

Data and metadata must be retrievable under well-defined conditions.

- enable metadata and, where possible, data retrieval.
- Support authentication and authorisation when necessary.
- Ensure metadata remain accessible even if the data are no longer available.

Reusable

Data need to be described to support future use.

- Rich metadata with accurate and relevant attributes.
- Metadata include clear licences for reuse.
- Metadata include detailed provenance (how and why the data were created).
- Metadata follow community standards relevant to the domain.

The four FAIR principles

Important clarification

FAIR is a **framework**. FAIR does **not** mean data must be **open**.

Data can be FAIR under controlled access if the conditions for access are clear and documented. FAIR does not specify a single technology or format, but it does require clear, structured, and machine-actionable metadata and processes. [Read scientific article](#).

Why FAIR matters

Applying FAIR principles ensures that:

- data are discoverable in repository indexes and search engines
- data can be interpreted and reused by others, including computational systems
- metadata and data remain meaningful long after a project ends
- research impact and collaboration are enhanced

✅ How FAIR connects to your workflows

FAIR is implemented through the practices you already follow:

- Documentation and metadata → supports **Findability**
- Repository selection and persistent IDs → supports **Accessibility**
- Standards and vocabularies → supports **Interoperability**
- Licenses and rights statements → supports **Reusability**

❌ Avoid this

- Treating FAIR as merely open access
- Adding metadata only at the last minute
- Ignoring machine-actionable formats and structures
- Assuming FAIR happens automatically in repositories

➡ Learn more

- [Data archiving and preservation](#)
- [Data sharing and citation](#)
- [How to create a Data Management Plan](#)



Sharing and citing research data: Making data reusable and giving credit

What this is

Sharing research data enables transparency, discovery, reproducibility, and further research. Citing data ensures that data creators receive appropriate credit and supports collaboration.

Ownership, access, and license

Research data are shared according to the ownership, access conditions, and license defined at the time of archiving. These determine who may access the data, where they can be accessed, and what forms of reuse are permitted.

✓ Sharing research data

Good practice includes:

- Depositing data in a discipline-specific repository where available.
- Using institutional or generic repositories if no disciplinary option exists
- Sharing data and documentation together
- Ensuring the repository provides a persistent identifier (e.g. DOI)
- Clearly displaying access conditions and license information

✓ Citing research data

Research data should be cited as independent research outputs, similar to publications. Most repositories provide a recommended citation format that should be reused whenever possible. A data citation typically includes:

- Creator(s)

- Year of publication
- Title of the dataset
- Repository name
- Persistent identifier (e.g. DOI)

Example

Jane Doe (2023). Dataset: Integrated water vapor time series. Environmental Research Institute. Version 2.1. Available at: <https://doi.org/10.1234/climate.2022.2.1>

✗ Avoid this

- Hosting shared data on personal or temporary platforms
- Sharing data without a persistent identifier
- Expecting others to “just cite the article” instead of the dataset
- Reusing data without citing the dataset
- Sharing data without documentation or license information

➔ Learn more

- [Rights, licenses, and ethical sharing](#)
- [Research data management](#)
- [How to create a Data Management Plan](#)