

Introduction to research data management (RDM) – NatLife

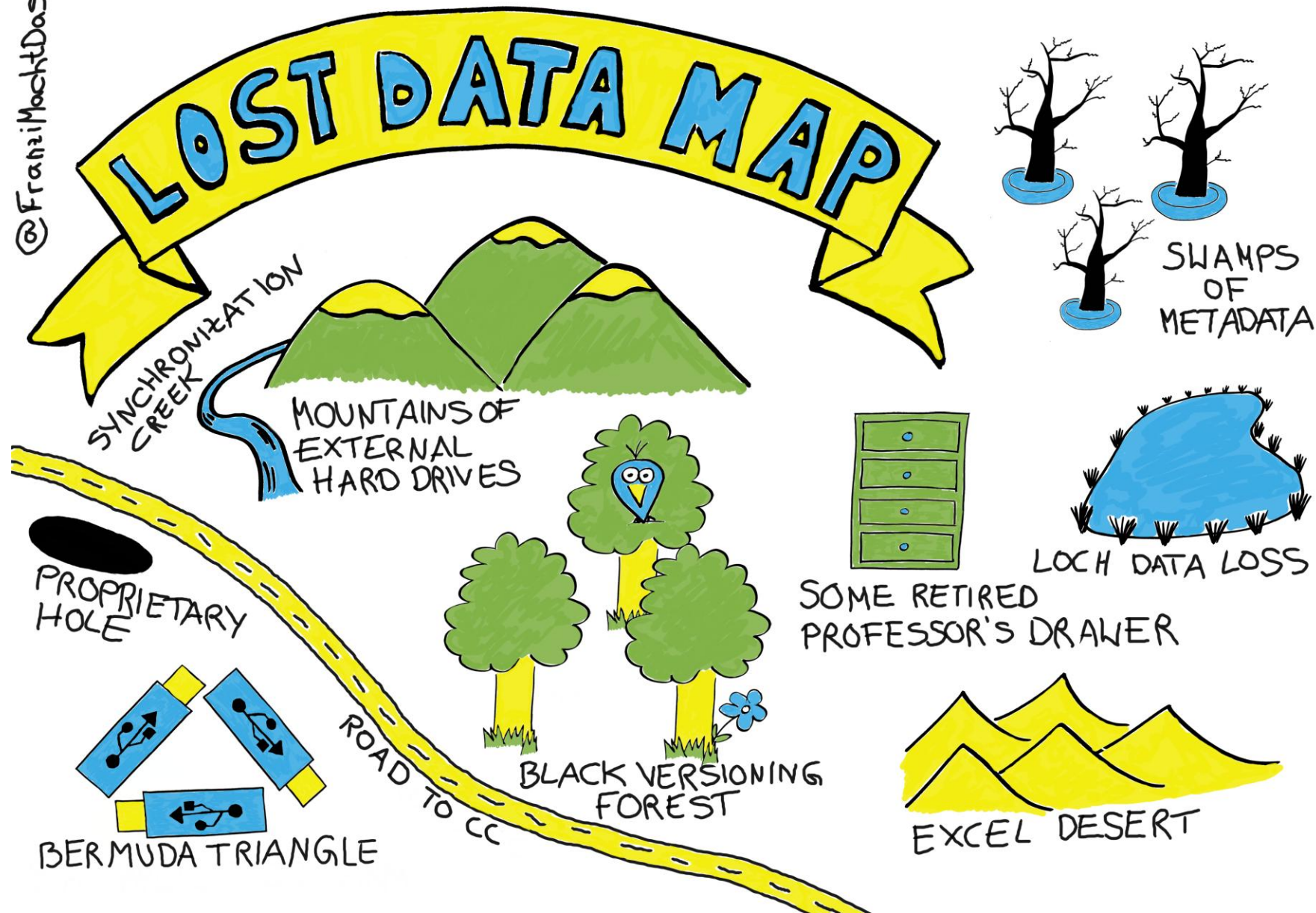
Dr. Anne Vieten
October, 7th 2026

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

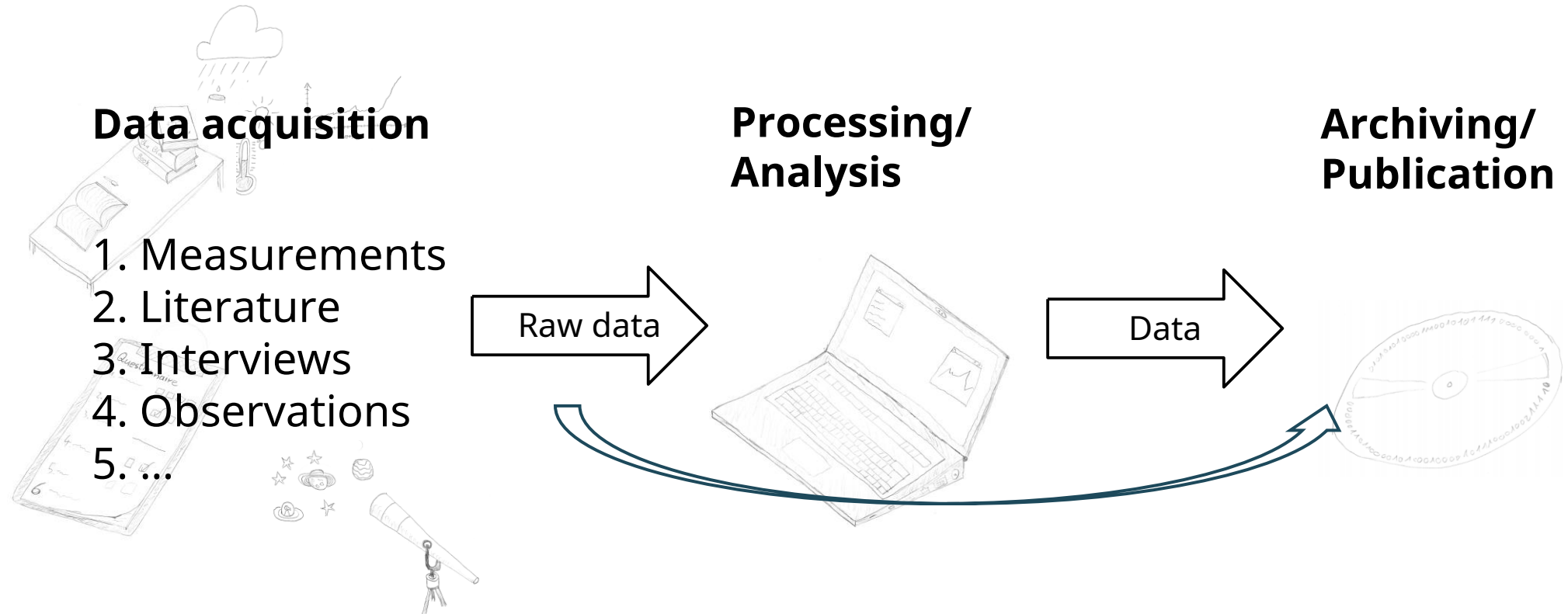
JG|U

Dieses Werk ist unter einer [CC-by 4.0 Lizenz](https://creativecommons.org/licenses/by-sa/4.0/) veröffentlicht



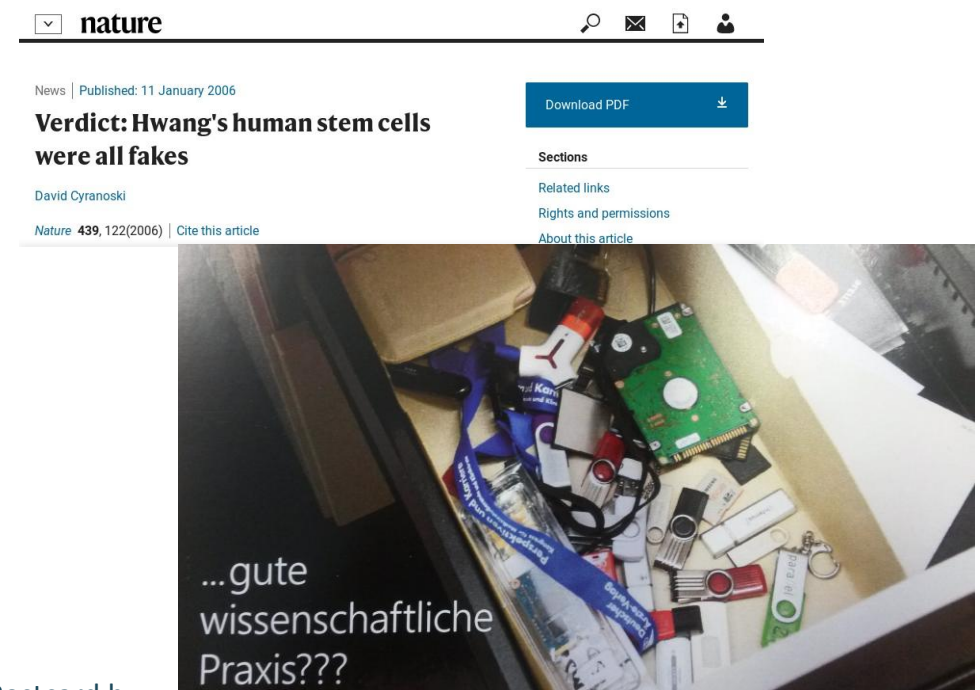


RESEARCH DATA, A DEFINITION



WHY RDM IS IMPORTANT FOR YOU?

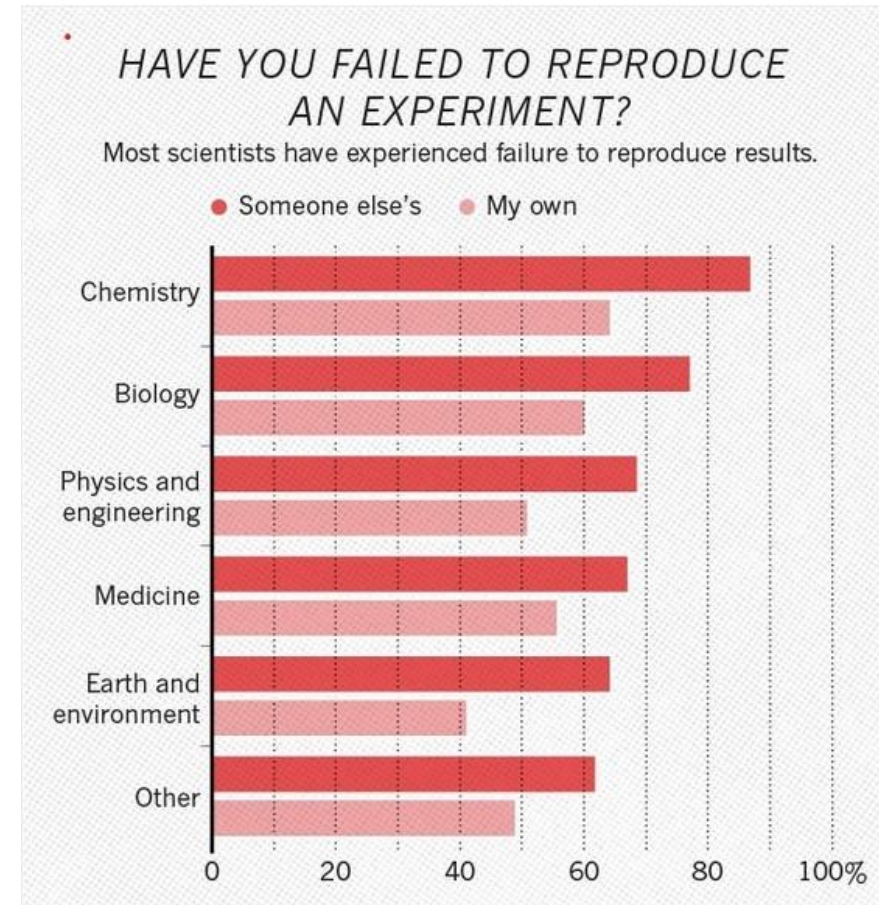
- Prevention of data loss
- Reproducibility, verifiability and transparency of research results
- Reduction of scientific errors
- Faster retrieval of data and information
- Long-term availability and reusability of research data
- Better scientific reputation



Postcard by:
Landesinitiative NFDI der digitalen Hochschulen NRW - Forschungsdatenmanagement

REPRODUCIBILITY OF RESEARCH RESULTS

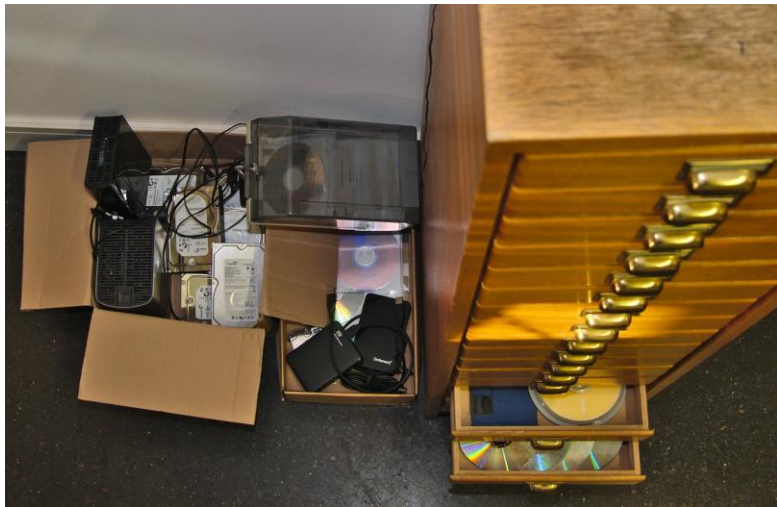
...often fails



[Baker, M.; Nature; May 25th 2016](#)

REMEMBER: ONLY A GENIUS CAN HANDLE CHAOS ...

... but what happens if the genius leaves?



FURTHER REASONS: REQUIREMENTS OF THIRD-PARTY FUNDING BODIES AND JGU

- [Regulations for ensuring good research practice at JGU; Presentation](#)
- DFG: [Leitlinien zum Umgang mit Forschungsdaten](#); [RDM-Checklist](#)
- Horizon Europe: [Open Science in Horizon Europe \(europa.eu\)](#)
- [BMFTR](#)
- VW-Foundation:
https://www.volkswagenstiftung.de/sites/default/files/documents/Open_Science_Policy_und_Umsetzung_VolkswagenStiftung.pdf
- [Leitlinie zum Umgang mit Forschungsdaten der JGU](#)

RDM REQUIREMENTS IN THE REGULATIONS FOR ENSURING GOOD RESEARCH PRACTICE AT JGU

Section 6: Handling of research data

- ... research data underlying a publication should be published, if this is feasible and reasonable.
- The publication of research data is based on the **FAIR Guiding Principles** for scientific data management and stewardship.
- Research data must be published in a form that enables others to replicate the findings.
- If possible, a publication platform that is well established in the discipline should be used – for instance, a research data repository or a data journal.

DATA SCARYTALE: BACKUP IS THE KEY

Because they had forgotten their key they had to go through another year of "physical" hardships.

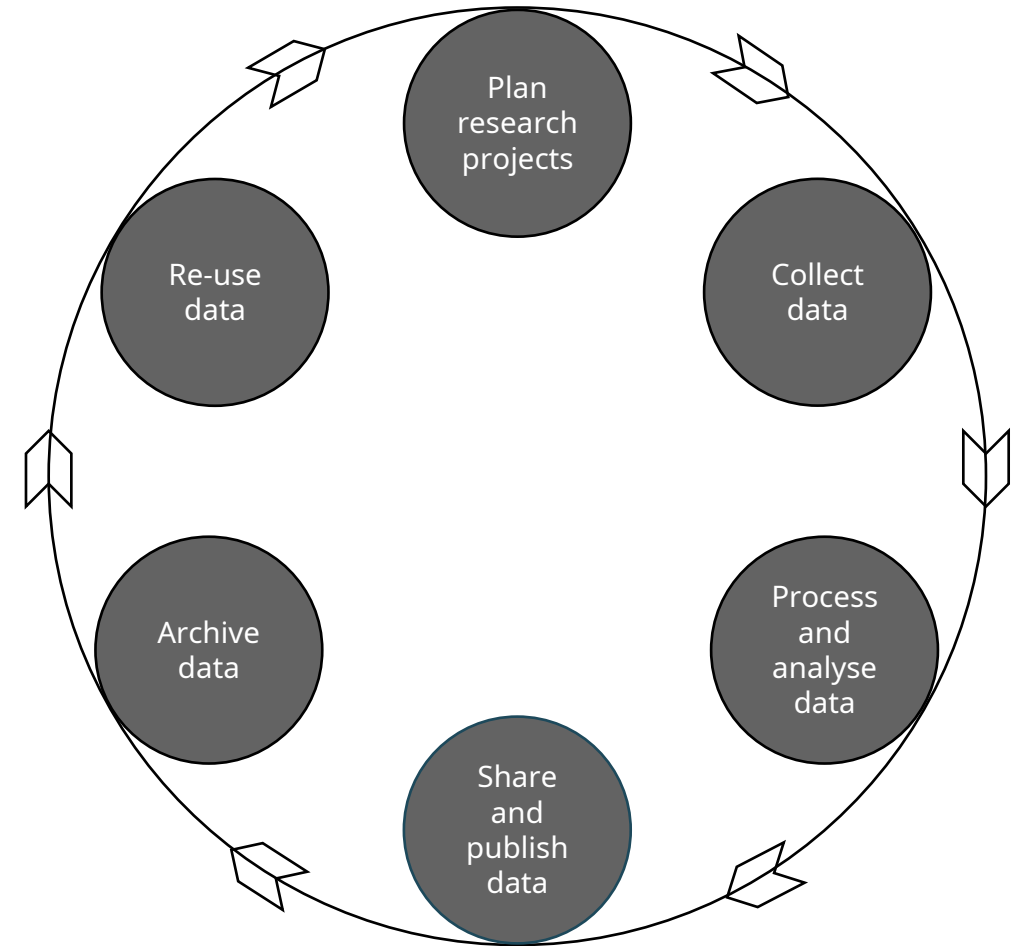
Ideas what happened?

[Source: Research Data ScaryTales](#)

DOI10.5281/zenodo.1746339210.5281/zenodo.17463392

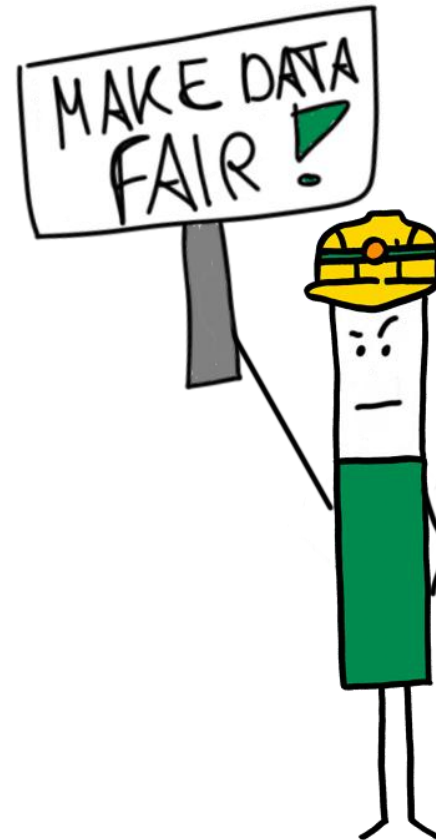
THE RESEARCH DATA LIFE CYCLE

... describes all steps of
research data management



HOW DO I IMPLEMENT RDM IN PRACTICE?

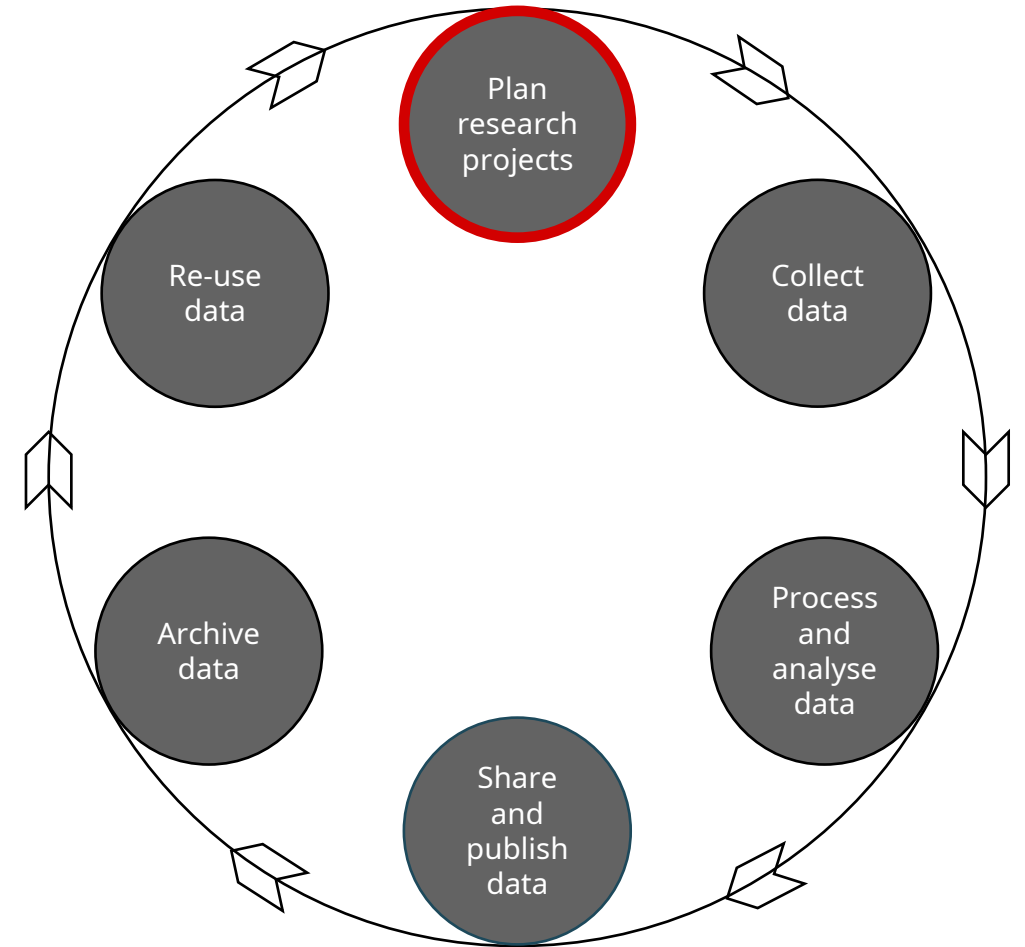
???



[Abb.: Sketchnotes & Sketches – FranzMachtDas _CC_BY-SA 4-0.png](#) CC-BY-SA 4.0 Franziska Mau

PLAN RESEARCH PROJECTS

- Design/planning of the research process
- Selection of software and data formats
- Choosing the storage location
- Organisation of backups
- Preparation of an initial data management plan (DMP)
- Data organisation and standardised file naming
- Definition of data documentation (metadata, etc.)
- Regulations for sharing data within the project
- Clarification of legal issues (data protection, copyright)
- etc.



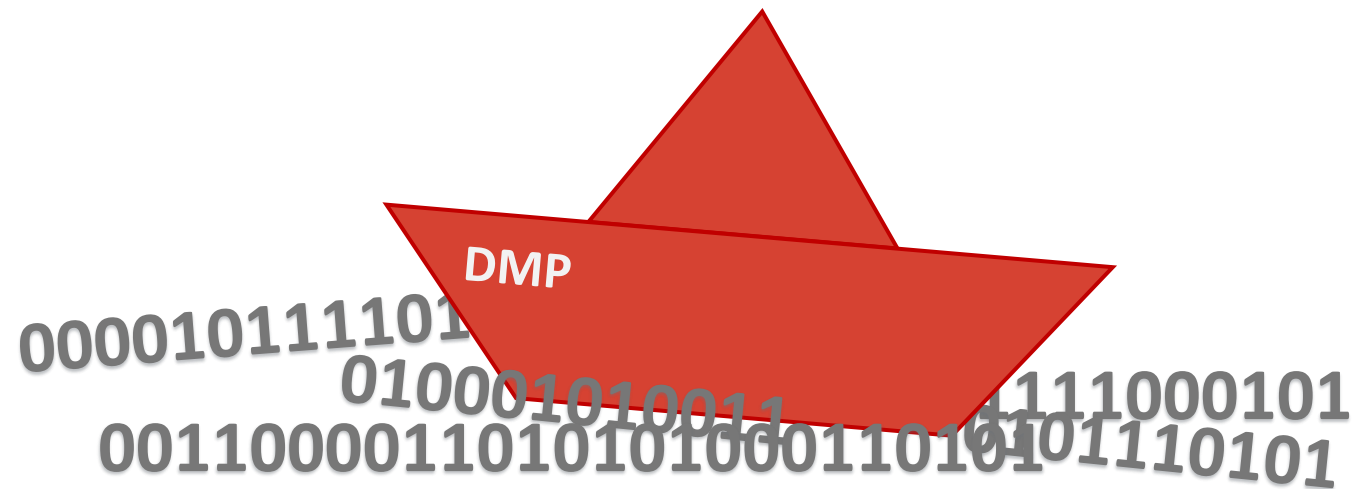
DATA MANAGEMENT PLAN (DMP)

Typical contents:

- 1) Administrative information
- 2) Documentation and metadata
- 3) Ethical and legal issues
- 4) Storage and back-up
- 5) Data selection and preservation
- 6) Sharing data with others
- 7) Responsibilities and resources

Tasks:

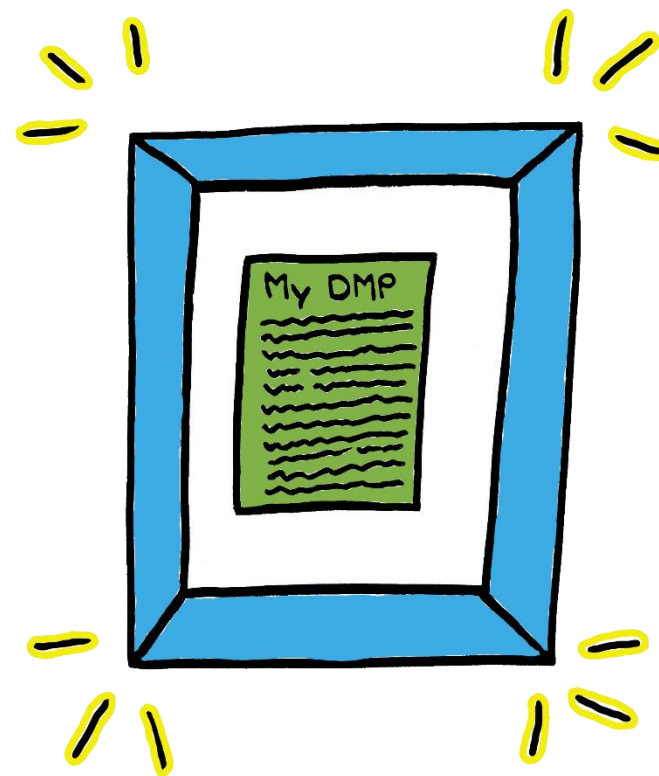
- Management of research data
- Definition of RDM steps
- Dynamic RDM tool



Vergl.: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf

WHY IT MAKES SENSE TO PREPARE A DMP?

- Saves time in the long run!
- Through RDM planning: Avoidance of errors in data handling
- Project management tool for research data
- Better long-term comprehensibility of data sets
- Improved long-term retrievability of data



https://filedn.com/lmWGYHN8VkubygwHQ/U0BO5k/Sketches/framed_dmp_co_CC_BY-SA_4-0.png CC-BY-SA 4.0 Franziska Mau

DMP TOOLS

General: RDMO: RDMO ZDV/Uni-Mainz or <https://rdmo.rhein-main-universitaeten.de/>

Biology: <https://www.gfbio.org/plan>

Chemistry: RDMO NFDI4Chem

Engineering: RDMO NFDI4ING

Medicine: <https://rdmo.publisso.de/>

RDMO

A tool to support the planning, implementation, and organisation of research data management.

Welcome to RDMO

The aim of the RDMO project is to deliver a web application to assist structured planning, implementation and administration of the data in a scientific project. Additionally, the gathered information can be cast into textual forms suitable for funding agencies requirements or for reports.

This is a prototype of the software, for demonstration purposes.

For more information about the project visit rdmorganiser.github.io.

Login

Login with Shibboleth

Log in with your
university account

<https://rdmo.zdv.uni-mainz.de/>

Create new project

Title

The title for this project.

Description

A description for this project (optional).

Catalog

The catalog which will be used for this project.

☐ SFB 1551

This catalog is a revised version of the Molecular Biology catalogue for the students of the SFB 1551.

☒ Molecular Biology v2

This catalog is a revised version of the Molecular Biology catalogue.

☐ RDMD

☐ Brief questionnaire

☐ Horizon Europe

This catalog is intended for research projects funded by EU Horizon Europe. It is compiled according to the [Data Management Plan Template, Version 1.0 from May 5, 2021](#). Some of the EU questions are broken down into sub-questions. In addition, for many questions there is the option of answering the question on a dataset-specific basis. There is a suitable view for the output of the data management plan, which

- outputs the data management plan in the deliverable form desired by the EU;
- uses the original EU questions;
- automatically inserts references to questions that have already been asked in a similar form.

☐ DFG checklist

Catalog according to the DFG checklist of December 21, 2021. Updated June 24, 2025.

☐ JGU checklist

☐ Software Management Plan for Researchers

This catalogue is for the management of scientific software projects. It supports scientists in the development and project organisation of software developments through fifty questions in different topic blocks. This is version 3.0 of the catalogue.

☐ Offboarding checklist JGU

Document key information about existing data when a project ends or a scientist leaves the research group.

Parent project

The parent project of this project.

Create project

Cancel

FDM-Kurs

Description	No description available.	
Catalog	Brief questionnaire	

Tasks

Tasks are generated automatically from the answers given in the project. On the page of each task you can see which of your answers lead to the activation of the task.

No active tasks found.

Views

Views are created using the answers given in the project and can then be exported in various formats. Initially, all views are empty. Please answer some questions by visiting Answer Questions (at the top of the sidebar).

View	Description	
CITEC DMP	DMP Vorlage der Universität Bielefeld für CITEC geförderte Projekte.	
Bielefeld	DMP template from the University of Bielefeld.	
CITEC DMP	DMP template from the University of Bielefeld for CITEC funded research projects.	
Cost overview	Overview of the personnel and non-personnel costs	
Cost overview [en]	Overview of the personnel and non-personnel costs [en]	
DMPonline template	Template from DMPonline, online https://dmponline.dcc.ac.uk	
DMPtool template	Template from DMPtool, based on "NSF-GEN: Generic", online: https://dmptool.org	
Horizon 2020 FAIR Data Management Plan template	Template for Horizon 2020, from "Guidelines on FAIR Data Management in Horizon 2020", online: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf	
Horizon Europe	View (export template) for the funding framework program "Horizon Europe" (2021-2027) of the European Commission (version 1.0, 5th May 2021)	
SNF Template	DMP of the SNF (Switzerland)	
List of all variables	for internal control	

Options

- Answer questions
- View answers
- Update project information
- Update project catalog
- Update parent project
- Update project tasks
- Update project views
- Copy project
- Delete project
- Add member
- Create snapshot
- Back to projects overview

Export

- RDMO XML
- CSV (comma separated)
- CSV (semicolon separated)
- JSON

Import values

Import from file

Select file



Members

Here you can see who can access the project and invite additional members. You can use the user roles to manage which rights the users have. Unless you are the last owner, you can leave the project with the button next to your name.

User	E-Mail	Role	
Anne Vieten	vieten@uni-mainz.de	Owner	

Snapshots

Snapshots allow you to save all responses at a given point in time and preserve a certain stage of the project. Later the snapshot can be used to create views, and the project can also be reset to a previous snapshot if needed.

No snapshots found.

Data

The following questions collect information on the data that is produced or used in the project. They also help to estimate the value of the data in terms of potential re-use and long-term preservation.

Before data is newly created, it is advisable to check if there is existing data that could be re-used. This way, redundant collection or creation of research data is prevented. This saves efforts and costs. Furthermore, in the case of personal data, the [German Federal Data Protection Act](#) allows the collection of personal data only when there are no other reasonable means to clarify the research question (re-use of existing data would be such a reasonable means). Also, there shall be collected no more information than necessary.

The information regarding the data collected, produced or used in the project is gathered along datasets. The definition of these datasets is an important conceptional decision that has to be made individually and carefully for each project.

Please fill in the form for each tab. The same tabs may be used later on other pages. You can add a new tab using the green button. Once created, you can edit or delete tabs using the buttons in the top right corner.

d [+ Dataset](#)



What kind of dataset is it?

Please briefly describe the data type and / or the method used to create or collect the data, for example:

- quantitative online survey
- 3D model / digital reconstruction of a stone age settlement
- software developed within the project



Back

Proceed

Overview

Project: [TestB](#)

Catalog: Brief questionnaire

[Show management panels](#)

[Reload page](#)

[Back to my projects](#)

Progress

2 of 34

Back

Proceed

Navigation

Grey entries will be conditionally skipped based on your input.

[General \(1 of 5\)](#)

[Content classification](#)

→ [Data](#)

[Data origin](#)

[Technical classification](#)

[Data usage](#)

[Metadata and referencing](#)

[Legal and ethics \(1 of 3\)](#)

[Storage and long-term preservation](#)

[My Projects](#) / [TestB](#) / [Technical classification](#)

Data size

Please fill in the form for each tab. The same tabs may be used later on other pages. You can add a new tab using the green button. Once created, you can edit or delete tabs using the buttons in the top right corner.

d

f

[+ Dataset](#)

What is the actual or expected size of the dataset?

- ☐ Less than 1 GB
- ☐ Between 1 GB and 1 TB
- ☐ Between 1 TB and 100 TB
- ☐ More than 100 TB
- ☐ Exact size:

[Back](#)[Proceed](#)

Overview

Project: [TestB](#)Catalog: [Brief questionnaire](#)[Show management panels](#)[Reload page](#)[Back to my projects](#)

Progress

2 of 62

[Back](#)[Proceed](#)

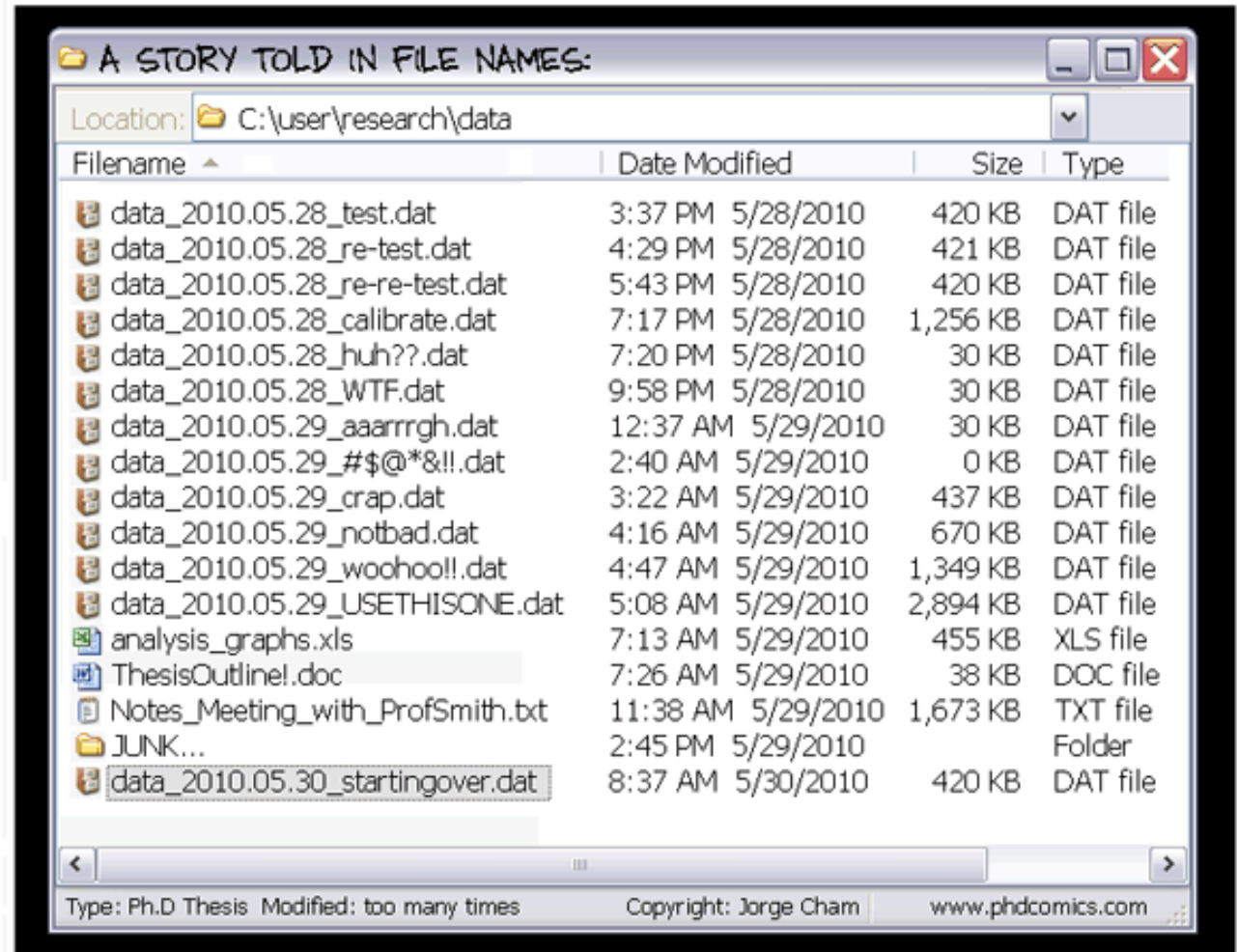
Navigation

Grey entries will be conditionally skipped based on your input.

[General \(1 of 5\)](#)[Content classification](#)[Technical classification](#)[→ Data size](#)[Formats](#)[Tools](#)[Versioning](#)[Data usage](#)[Metadata and referencing](#)[Legal and ethics \(1 of 5\)](#)[Storage and long-term preservation](#)

ORGANISE DATA

...or lose track of things?



Jorge Cham: PHD Comics: A story in file names

BETTER LIKE THAT!

Initial planning is key:

Develop a naming convention for files and folders. Document it!

File naming:

- Meaningful
- Sortable (e.g. fixed order of name components, chronological sorting YYYYMMDD)
- Possible elements: version number, date, content description, project number, etc.;
- Naming should also make sense when moving files to other folders.
- < 32 characters.
- As short as possible and as long as necessary.
- No special characters such as % or ? and no spaces.
- Take versioning into account and make it clearly identifiable.

WHICH FILE NAMES ARE ACCEPTABLE?

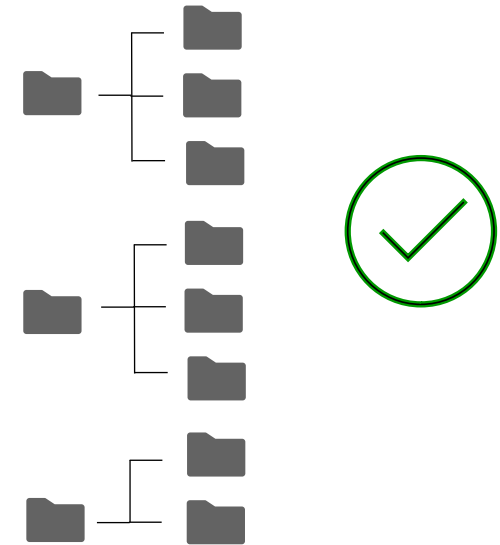
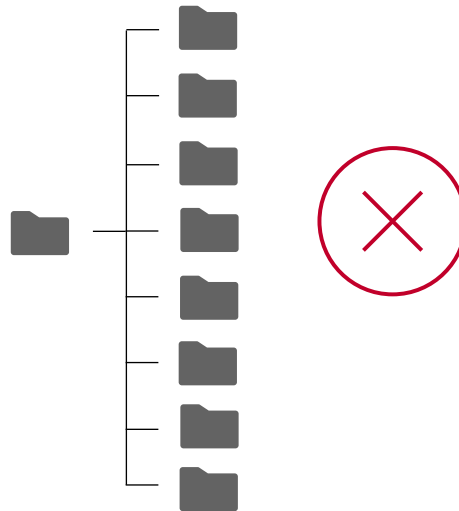
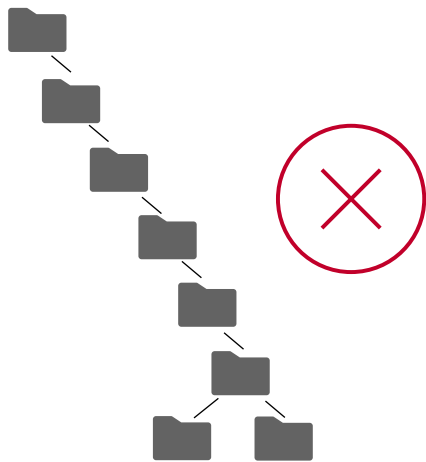
1. Fg23Spat6ev.mp3 ?
2. 005ProjectCityLife_survey1_20250803.csv ?
3. Presentation_research_data_management.pptx ?
4. Sample51 2504 Petra+Uwe new corrected.docx ?
5. 20200918_JGU_RDM_introduction.pdf ?

WHY?

- | | |
|---|--------|
| 1. Fg23Spat6ev.mp3 | no |
| 2. 005ProjectCityLife_survey1_20250803.csv | yes |
| 3. Presentation_research_data_management.pptx | yes/no |
| 4. Sample51 2504 Petra+Uwe new corrected.docx | no |
| 5. 20200918_JGU_RDM_introduction.pdf | yes |

ARRANGEMENT OF FOLDERS

- Do not create redundant folder names
- Follow naming conventions
- Find a balance between nested folders and the number of (sub)folders



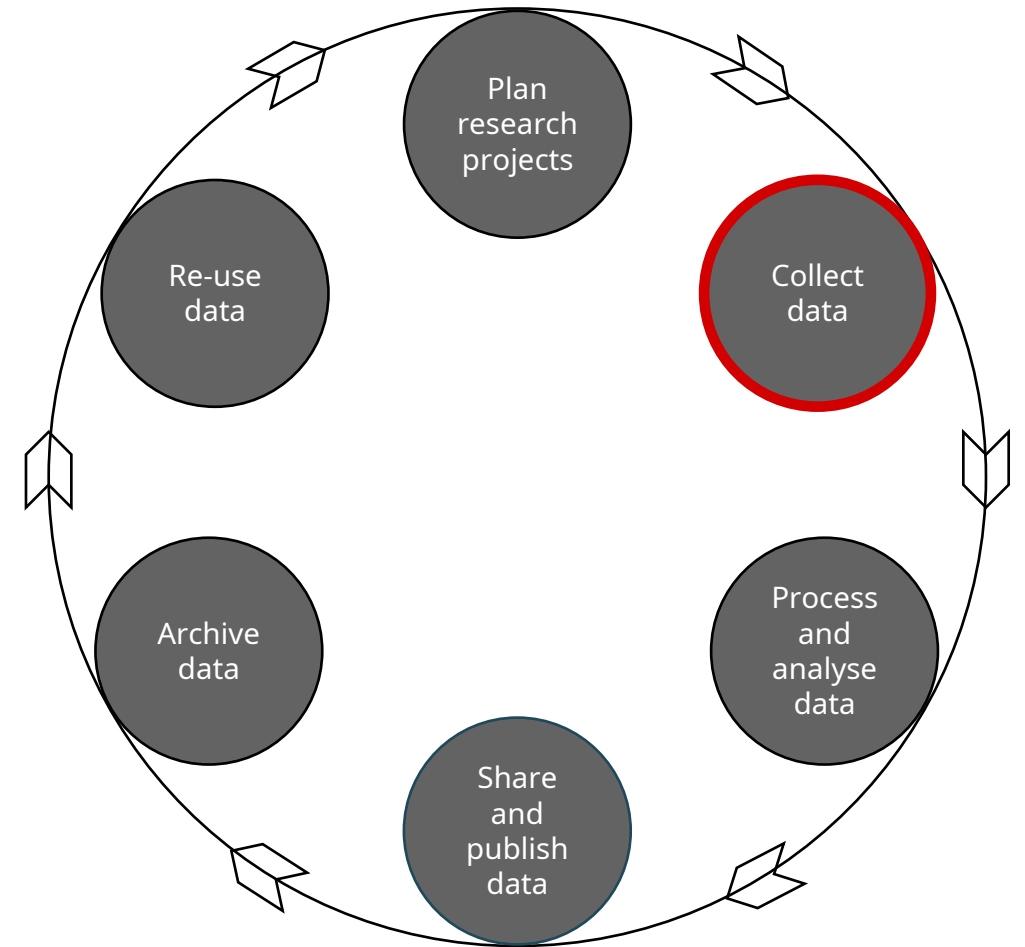
VERSIONING

Options:

1. **Version management as part of the file naming convention.** Version numbers in the file name, e.g. V01, V02, etc. Use of the current version of data in the file name
2. **Use of versioning tools.** E.g. Git-based software such as GitLab (actually for software code, but also applicable to other texts)

COLLECT DATA

- Data entry and transfer
- Digitisation of media, if necessary
- Data verification and evaluation
- Anonymisation of data, if necessary
- Creation of documentation/metadata
- Versioning measures
- Backup
- Assigning access rights
- ...



METADATA

Meaning: Metadata describes/defines other data in more detail.

Example: Metadata for a journal article includes information about the author, year of publication, publisher, etc.

Necessary for:

- In-depth understanding of the actual data.
- Searchability of repositories/databases.

Standards: E.g. Dublin Core, DataCite; Repositories sometimes have their own standards.

Data Cite Basic Scheme

- Identifier
- Creators
- Title
- Publisher
- Publication year
- Subject
- Language
- Resource type
- Descriptions

List of metadata tools:

[RDA list of metadata tools](#)

OTHER METADATA EXAMPLES

Metadata of a photographic picture:

- Geolocation
- Date and time
- Camera settings
- Camera software used
- device used
- ...

Helligkeit	2.58
Lichtquelle	D65
Belichtungsprogramm	Normal
Sättigung	
Schärfe	
Weißausgleich	Automatisch
Fotometrische Interpretation	
Digitalzoom	1
EXIF-Version	0220
GPS	
Breitengrad	49; 59; 34.59109999999...
Längengrad	8; 14; 42.360000000000...
Höhe über Normal-Null	172.9
Datei	
Name	IMG_20260414_071030...
Elementtyp	JPG-Datei

Metadata of a measuring instrument:

- Kind of measuring instrument
- Instrument settings
- Software version used
- Manufacturer
- special features of a measuring instrument
- ...

TYPES OF METADATA

Descriptive/bibliographic: title, author, keywords, year of publication, etc.

Administrative and legal: rights holders, licence conditions, access rules, etc.

Structural: physical or logical relationships between objects (e.g. between tables in a database)

Administrative or preservation-related: conversion and preservation measures, etc.

Technical: information on software or hardware, file formats, character encoding, parameters of measuring devices, etc.

SUBJECT SPECIFIC METADATA STANDARDS

NeXus

NeXus is an international standard for the storage and exchange of neutron, x-ray, and muon experiment data. The structure of NeXus files is extremely flexible, allowing the storage of both simple data sets, such as a single data array and its axes, and highly complex data and their associated metadata, such as measurements on a multi-component instrument or numerical simulations. NeXus is built on top of the container format HDF5 and adds domain-specific rules for organizing data within HDF5 files in addition to a dictionary of well-defined names.

Categories for the Description of Works of Art

Resource Metadata for the Virtual Observatory

Defines metadata terms and concepts necessary for discovery and use of astronomical data collections and services.

The extension is based on extensions. Resource Metadata populated and synchronized Handling). Version 1.12, March Resource Registry Working Group

CSMD (Core Scientific Metadata Model)

A study-data oriented model, primarily in support of the ICAT data management software. The CSMD is designed to support data collected within a large scientific workflow; however the model is also designed to be generic across disciplines.

Sponsored by the Science and Technologies Facilities Council, the latest available is v 4.0, from 2013.

ABCD (Access to Biological Collection Data)

The [Access to Biological Collections Data \(ABCD\) Schema](#) is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data). The ABCD Schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases. It is compatible with several existing data standards. Parallel structures exist so that either (or both) atomised data and free-text can be accommodated.

Sponsored by Biodiversity Information Standards TDWG - the Taxonomic Databases Working Group, the current specification was last modified in 2007.

OpenPMD (Open Standard for Particle-Mesh Data)

OpenPMD provides naming and attribute conventions that allow the exchange of particle and mesh based data from scientific simulations and experiments. The primary goal is to define a minimal set/kernel of meta information that enables the sharing and exchange of data to achieve

- portability between various applications and differing algorithms;
- a unified open-access description for scientific data (publishing and archiving);
- a unified description for post-processing, visualization and analysis.

AIMS

[Home](#) [About](#) [News](#) [In action](#) [Online Courses](#) [FAO AGRIS](#) [AGROVOC](#) [AGOR](#)

Agricultural Metadata Element Set (AgMES)

Genome Metadata

Some metadata on PATRIC consists of 61 different metadata fields, called attributes, which are organized into the following seven broad categories: Organism Info, Isolate Info, Host Info, Sequence Info, Phenotype Info, Project Info, and Others.

METADATA AS JSON FILE

JSON = Java Script Object Notation

- A metadata JSON file contains descriptive information about data objects
- Example for bibliographic metadata:

```
{'title': 'Sample Data', 'author': 'Max Mustermann', 'date': '2026-04-15', 'tags': ['sample', 'data', 'json']}
```
- Advantage: human and machine readable
- Save the JSON file together with the data
- Important tool to make data FAIR

EXAMPLES OF DATA DOCUMENTATIONS AND DESCRIPTIONS

- Laboratory notebooks (paper-based/digital, e.g. eLabFTW) - Experiment protocols
- Code books
- Database schemas
- Software syntax
- JSON files
- Readme.txt file describing the data sets
- ...

WHAT DO I NEED TO DOCUMENT?

Everything that third parties need to understand the data set!

DATA BACK-UP/PROTECTION IN GENERAL

- Perform regular backups (e.g. use University server)
- Use multiple storage media!
- Clear, standardised naming conventions
- Capture metadata
- Observe versioning (e.g. use GitLab)
- Consider data protection!
- Clarify access rights!

BACKUP 3-2-1 RULE

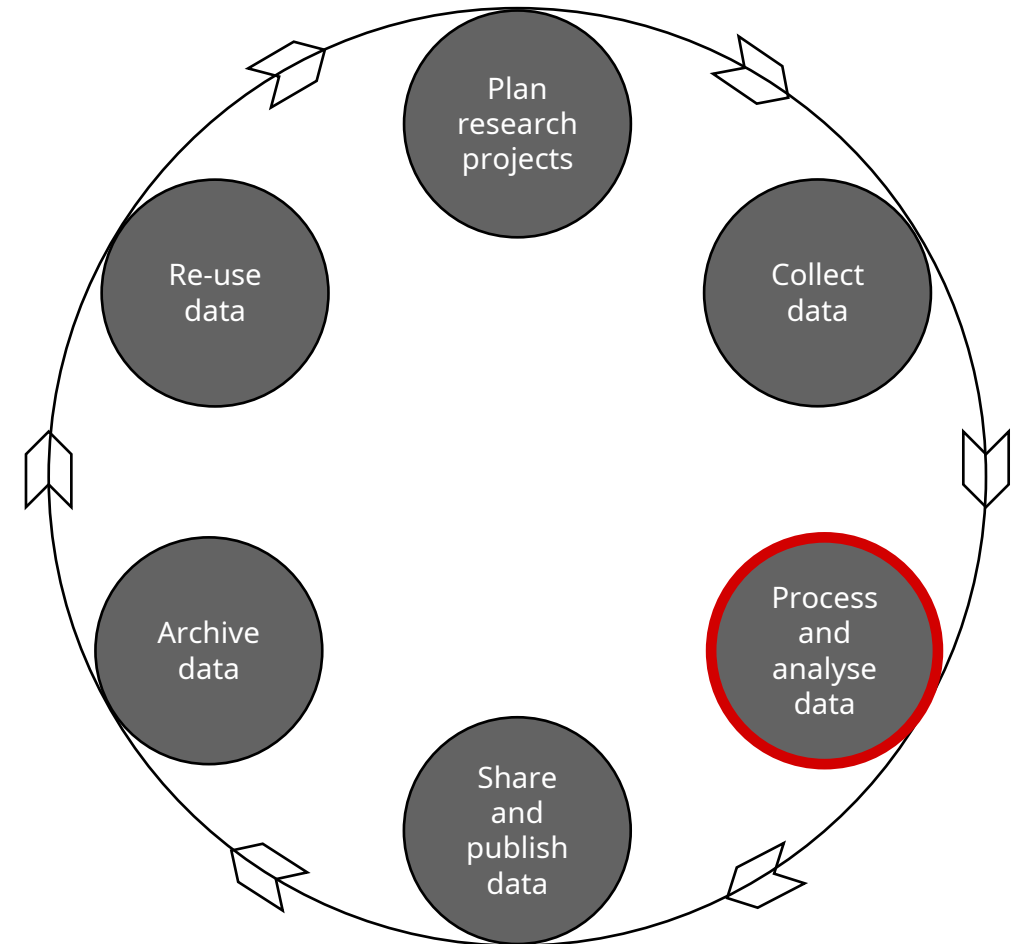
3 = 3 copies

2 = 2 different storage media

1 = 1 copy at a different location from the other copies

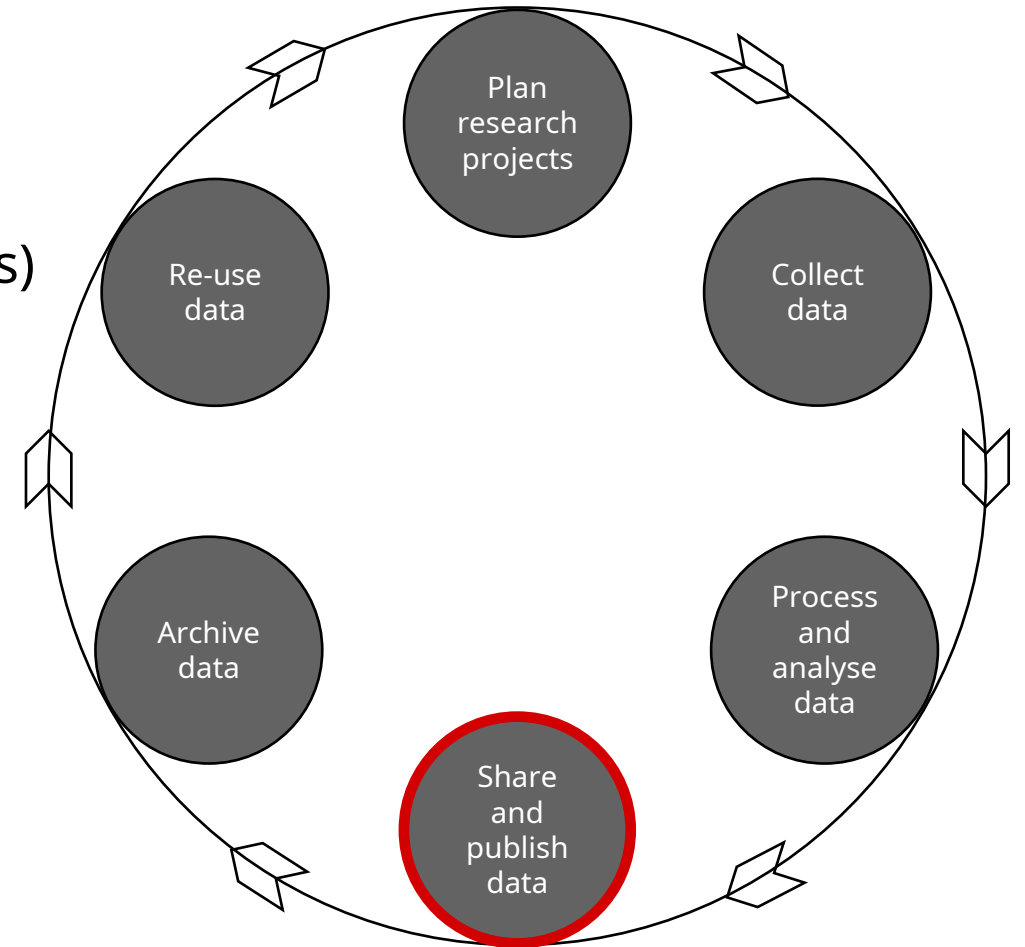
PROCESS AND ANALYSE DATA

- Interpretation of data
- Interlinking of data
- Measures for data plausibility and reproducibility
- Creation of documentation/metadata
- Production of research results
- Quality assurance
- ...



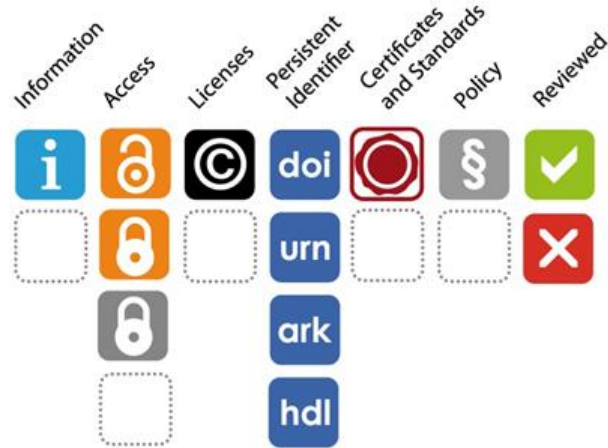
SHARE AND PUBLISH DATA

- Creation of publications in journals, books, etc.
- Publication of data (e.g. via repositories/databases)
- Access control/rights management for data
- Ensure/define copyright protection for data (if possible) or terms of use (e.g. Creative Commons)
- Keywords for searchability in repositories
- Define open or restricted access
- Consider data protection ...



HOW TO FIND SUITABLE REPOSITORIES?

- ... use <http://www.re3data.org/>
Registry of Research Data Repositories
➤ Nearly 3500 entries



http://www.forschungsdaten.org/index.php/Datei:Re3data_icons.jpg

Browse by subject

Graphical

Text

click to zoom into subjects or to select a bottommost subject in the hierarchy as filter for the re3data search page
ctrl + click on a top subject to select it as filter



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Molecular Biology

Search

Filter

Subjects ▾

Content Types ▾

Countries ▾

AID systems ▾

API ▾

Certificates ▾

Citation References ▾

Data access ▾

Data access restrictions ▾

Database access ▾

Database access restrictions ▾

Database licenses ▾

Data licenses ▾

Data upload ▾

Data upload restrictions ▾

Enhanced publication ▾

Institution responsibility type ▾

Institution type ▾

Keywords ▾

Metadata standards ▾

Metrics ▾

PID systems ▾

Profiles ▾

Provider types ▾

Quality management ▾

Repository languages ▾

Software ▾

Syndications ▾

Repository types ▾

Versioning ▾

Molecular Biology

← Previous 1 2 3 4 5 6 7 ... 11 Next →

Found 270 result(s)

European Chemical Biology Database

ECBD

Subject(s)

Bioinformatics and Theoretical Biology Microbiology, Virology and Immunology Basic Research in Biology and Medicine Medicine Chemistry Agriculture, Forestry and Veterinary Medicine Structural Biology Biology Life Sciences Medicine Natural Sciences Agriculture, Forestry and Veterinary Medicine

Repository type(s)

disciplinary

Provider type(s)

dataProvider

Country

European Union Germany

The European Chemical Biology Database (ECBD) is a prototype database, currently being developed for EU-OPENSREEN. The purpose of ECBD is to store and integrate screening data from EU-OPENSREEN partners and contributors.

EMBL-EBI

European Molecular Biology Laboratory - European Bioinformatics Institute

Subject(s)

Basic Research in Biology and Medicine Bioinformatics and Theoretical Biology Biology Life Sciences

Repository type(s)

disciplinary Institutional

Provider type(s)

dataProvider serviceProvider

Country

United Kingdom European Union

The European Bioinformatics Institute (EBI) has a long-standing mission to collect, organise and make available databases for biomolecular science. It makes available a collection of databases along with tools to search, download and analyse their content. These databases include DNA and protein sequences and structures, genome annotation, gene expression information, molecular interactions and pathways. Connected to these are linking and descriptive data resources such as protein motifs, ontologies and many others. In many of these efforts, the EBI is a European node in global data-sharing agreements involving, for example, the USA and Japan.

DEPOD

DEPhOsporylation Database

Subject(s)

Life Sciences Cell Biology Biology Basic Research in Biology and Medicine Bioinformatics and Theoretical Biology

Repository type(s)

disciplinary

Provider type(s)

dataProvider

Country

European Union United Kingdom Germany

DEPOD - the human DEPhOsporylation Database (version 1.1) is a manually curated database collecting human active phosphatases, their experimentally verified protein and non-protein substrates and dephosphorylation site information, and pathways in which they are involved. It also provides links to popular kinase databases and protein-protein interaction databases for these phosphatases and substrates. DEPOD aims to be a valuable resource for studying human phosphatases and their substrate specificities and molecular mechanisms; phosphatase-targeted drug discovery and development; connecting phosphatases with kinases through their common substrates; completing the human phosphorylation/dephosphorylation network.

APID

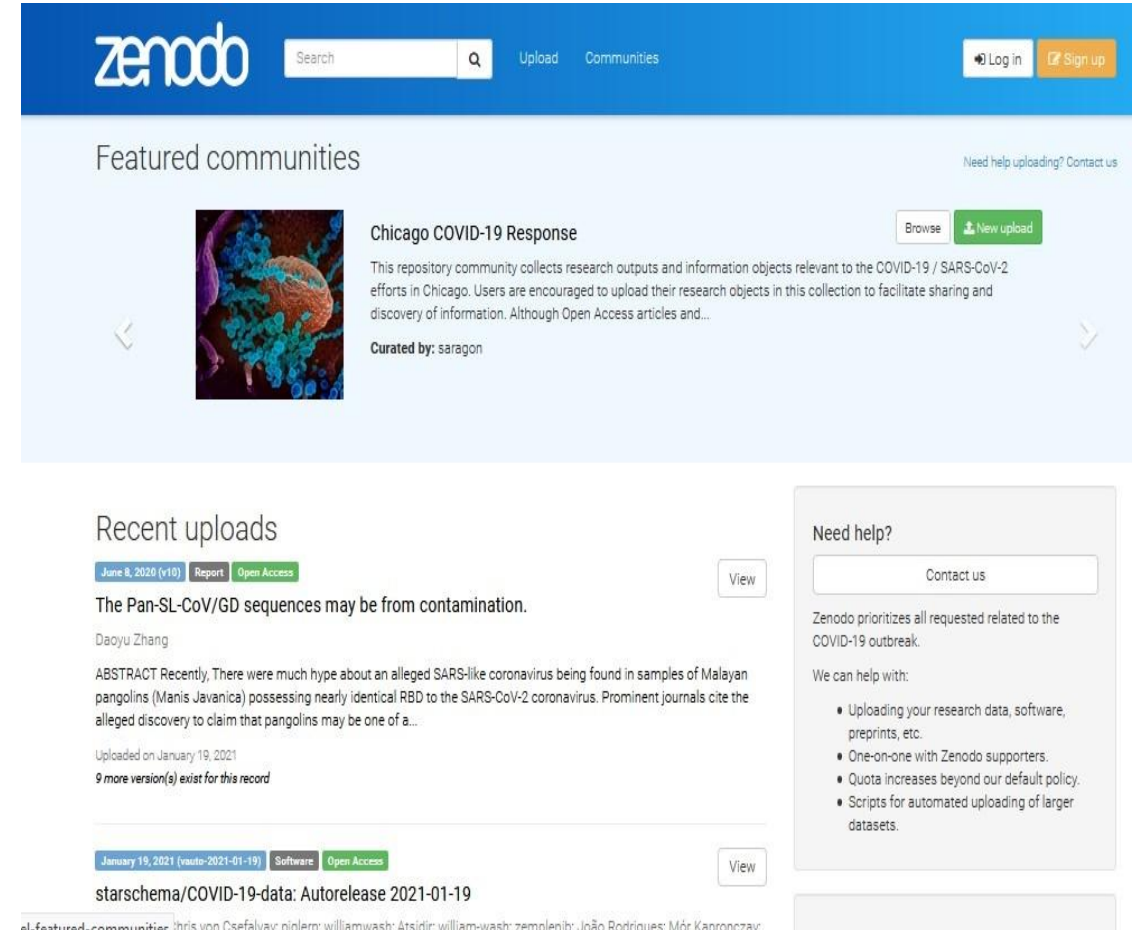
Agile Protein Interactomes DataServer

Subject(s)

Basic Research in Biology and Medicine Cell Biology Biology Life Sciences General Genetics and Functional Genome Biology Bioinformatics and Theoretical Biology Structural Biology

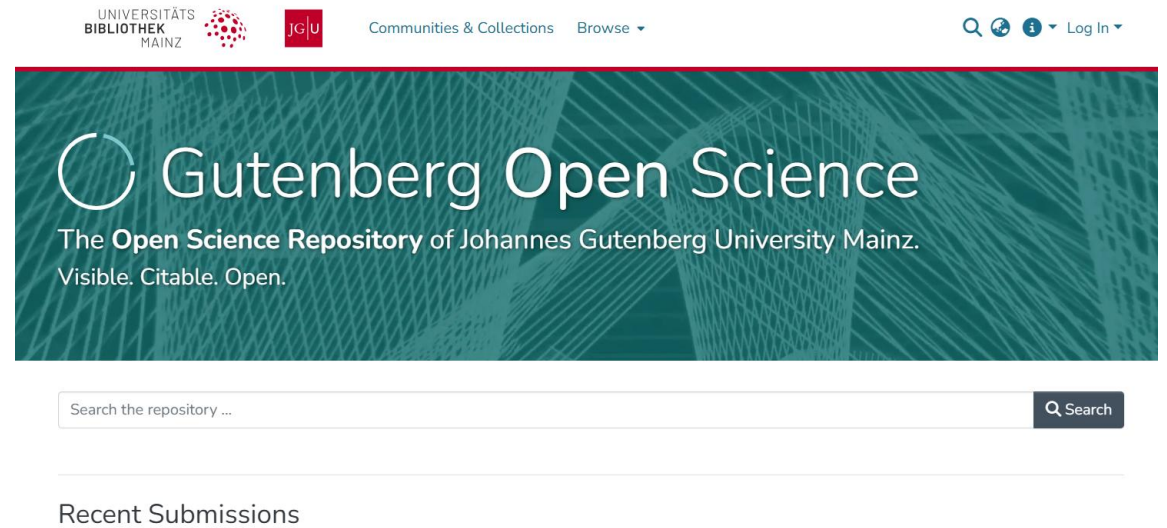
OR USE ZENODO A GENERAL REPOSITORY

- Open to all fields of science
- Textual publications, research data, code
- DOI for all uploads
- Hosted and maintained by CERN
- Flexible licensing
- Upload up to 50 GB per dataset
- Communities
- Access restriction possible (limited in time)



GUTENBERG OPEN SCIENCE JGU

- Institutional Open Access repository
- Pre- und Postprints, doctoral thesis
- Research data (up to 10 gb per dataset)
- Standardised Metadata
- DOI
- Notification in the local research information system
- [Gutenberg Open Science :: Home](#)



LEGAL THINGS: OPEN LICENCES

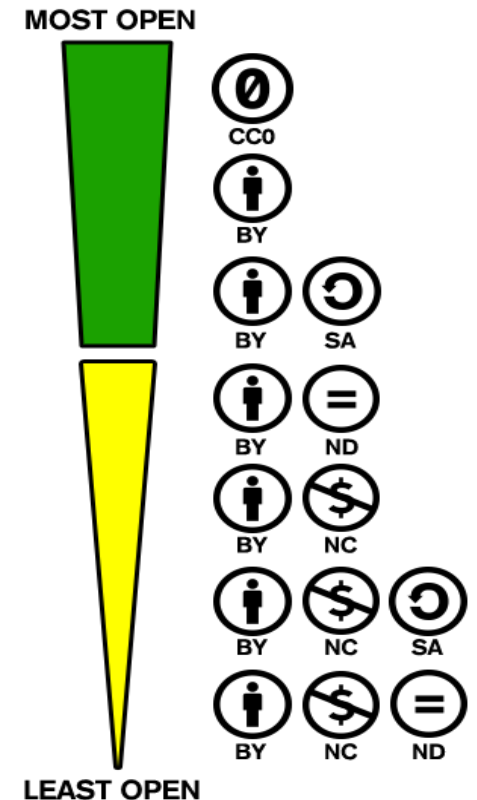
Offer the possibility to set conditions for the reuse of data

CC offers standardised licence agreements:

- CC 0: No rights reserved
- CC BY: Attribution (commonly used in science)
- CC BY SA: Attribution - Share alike
- CC BY ND: Attribution - No derivatives

Open source software: e.g. GNU General Public Licence (GPL)

Data collections: e.g. Open Data Commons (ODC)



creativecommons.org, CC BY 4.0

<https://creativecommons.org/licenses/by/4.0/>

<https://www.forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/>

LEGAL THINGS: GERMAN URHEBERRECHT

Protects intellectual creations (elements: perceptible form, personal creation, intellectual content, personal influence)

- Research data is often not subject to copyright. As a rule, there is no copyright on unstructured measurement data.
- Clarify legal issues in advance (e.g. rights to data in cooperation projects)
- Data publication: Assignment of Creative Commons licences.
- When using third-party data/programmes:
 - Are there any restrictions on the use of the data/programmes?
 - May data/programmes be included in archiving/publication?

<https://www.forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/>

YOUR RESEARCH DATA SHOULD BE FAIR!

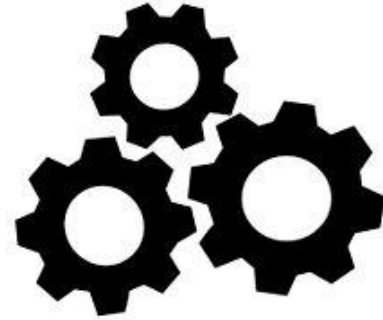
F_{indable}



A_{ccessible}



I_{nteroperable}



R_{eusable}



FAIR DATA

Short Keyword – High Impact!

WHAT EXACTLY DOES FAIR MEAN?



Findable: Easy to find for both - humans and computers

- Describe research data sets with metadata
- Metadata are indexed in a searchable directory or repository
- Publication of research data, if applicable
- Assign persistent identifiers (DOI, URN, etc.) to research data (metadata)

PANGAEA.

ECOLOGY

18121 datasets found on search in topic Ecology

< 1 2 3 4 5 6 7 8 9 10 >

1. **Chouvelon, T; Munsch, C; Bruzac, S et al. (2021):** Stable isotopes, energy density and elemental composition of deep pelagic crustaceans and fish sampled in October 2017 in the Bay of Biscay, North-East Atlantic
Related to: **Duhamel, E; Pawlowski, L; Garren, F (2017):** EVHOE 2017 cruise, Thalassa R/V. *Sismer*
Size: 2 datasets
<https://doi.pangaea.de/10.1594/PANGAEA.937345> - Score: 1.8
2. **Verdugo, J; Damm, E (2021):** Biogeochemical data during POLARSTERN cruise PS106/1 (ARK-XXXI/1.1)
Related to: **Macke, A; Flores, H (2018):** The Expeditions PS106/1 and 2 of the Research Vessel POLARSTERN to the Arctic Ocean in 2017. *Berichte zur Polar- und Meeresforschung = Reports on Polar and Marine Research*
Nikolopoulos, A; Linders, T; Rohardt, G (2018): Physical oceanography during POLARSTERN cruise PS106/1 (ARK-XXXI/1.1). *Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven*
Rohardt, G (2018): CTD Data Processing Report - RV POLARSTERN Cruise PS106.
Size: 2 datasets
<https://doi.pangaea.de/10.1594/PANGAEA.937242> - Score: 1.8

[FAIR Principles - GO FAIR](#)

WHAT EXACTLY DOES FAIR MEAN?



Accessible: User gets the information how and under which conditions data can be accessed

- Easily accessible (metadata and/or actual data) - (Meta)data are retrievable by their identifier using a standardised communications protocol
- The protocol is open, free, and universally implementable
- The protocol allows for an authentication and authorisation procedure, where necessary
- Defined licence and access conditions
- Metadata also available when actual data is not accessible

Coverage:	Median Latitude: 47.196250 * Median Longitude: 8.468551 * South-bound Latiti
License:	 Creative Commons Attribution 4.0 International (CC-BY-4.0)
Size:	73 datasets

[FAIR Principles - GO FAIR](#)

WHAT EXACTLY DOES FAIR MEAN?



Interoperable: Data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- Controlled vocabulary such as keywords. References to related (meta)data.
- Use of commonly used formats that are as non-proprietary, open and documented as possible.

[FAIR Principles - GO FAIR](#)

WHAT EXACTLY DOES FAIR MEAN?



Reusable: Goal of FAIR! Metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- Non-proprietary, open formats
- Defined usage licence
- (Meta)data comply with professional (subject specific) standards
- (Meta)data are associated with detailed provenance

Empfohlene Formate

Tabellen: CSV, TSV, SPSS portable

Text: TXT, HTML, RDF, PDF/A

Bild: TIFF, JPEG2000, PNG

Multimedia: AVI, WAV, MP4

Data Cite (Metadata) Basic Scheme

- Identifier
- Creators
- Title
- Publisher
- Publication year
- Subject
- Language
- Resource type
- Descriptions

Coverage: Median Latitude: 47.196250 * Median Longitude: 8.468551 * South-bound Latit
License:  Creative Commons Attribution 4.0 International (CC-BY-4.0)
Size: 73 datasets

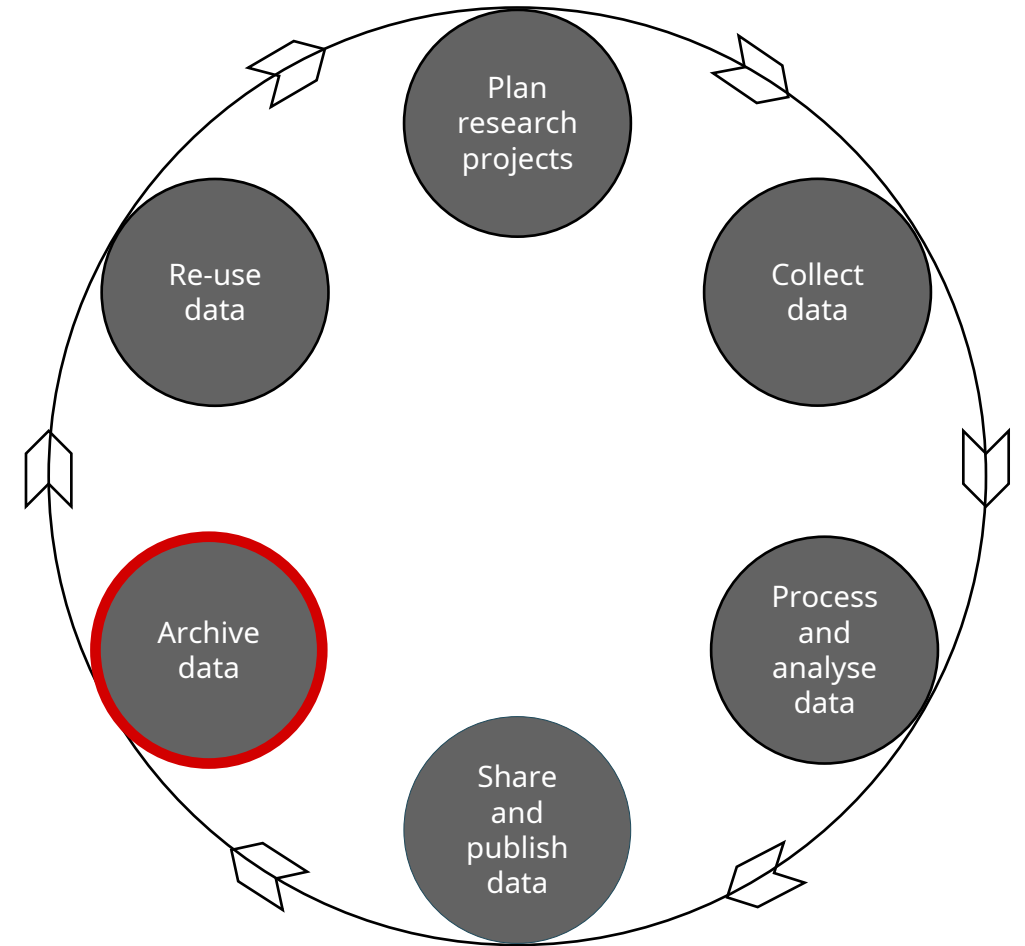
FAIR4RS PRINCIPLES

The FAIR principles have been extended to research software:

FAIR Principles for Research Software (FAIR4RS Principles)

ARCHIVE DATA

- Migration of data to the most suitable archiving format (non-proprietary)
- Metadata and data documentation (preferably directly upon creation)
- Transfer of data to experts for long-term archiving (libraries, data centres, specialist repositories)
- Permanent data archiving in a professional archiving infrastructure



BACK-UP VS. ARCHIVING

	Backup	Archiving
Aim	Actively used data	Final Data
Purpose	Protection against data loss	Long-term preservation and retrieval of information
Timeframe	Short term	Long termj



(LONG-TERM) ARCHIVING – REQUIREMENTS

- Use of widely **established formats**
- Use of formats **with open, documented and non-proprietary standards**
- Avoidance of compression
- **Description with metadata**
- Migration of data to current data formats
- Use of professional infrastructures for archiving

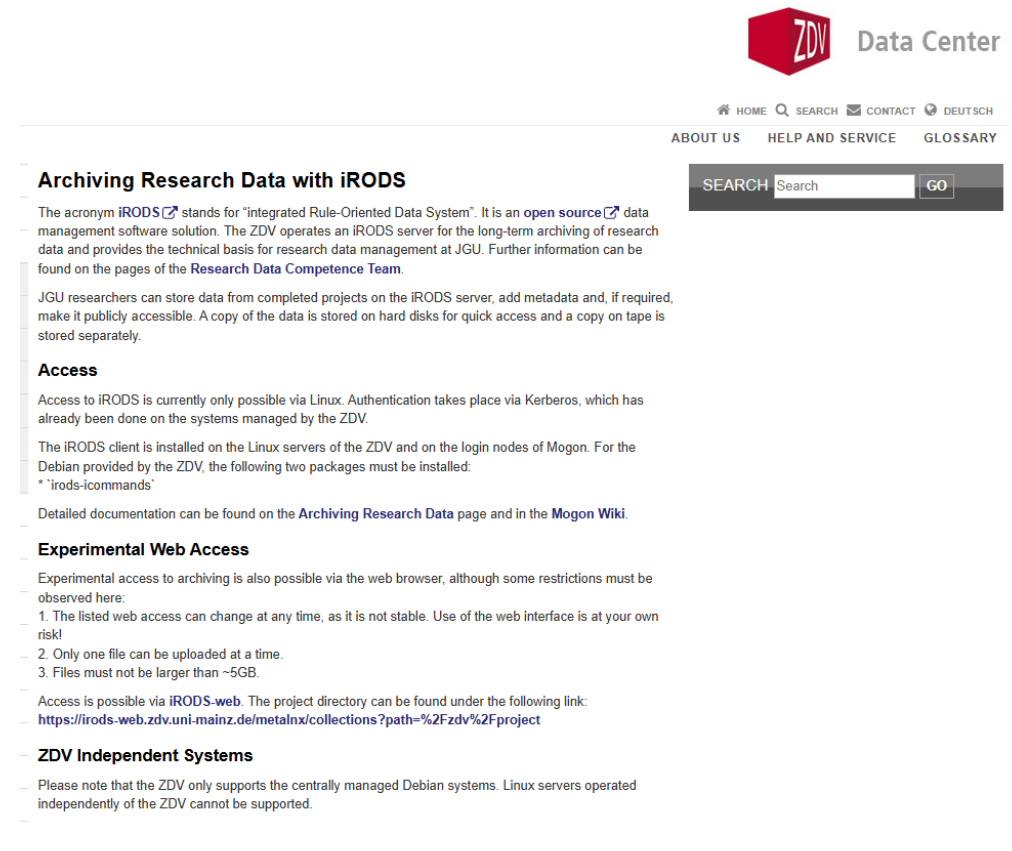
JGU'S RESEARCH DATA ARCHIVE

What is/can iRODS do?

- Based on iRODS software
- Virtual file system
- Searchable metadata
- Publication option (without PID)
- Fine-grained access control
- Automated workflows
- Command line tools

[Archiving Research Data with iRODS | Zentrum für Datenverarbeitung](#)

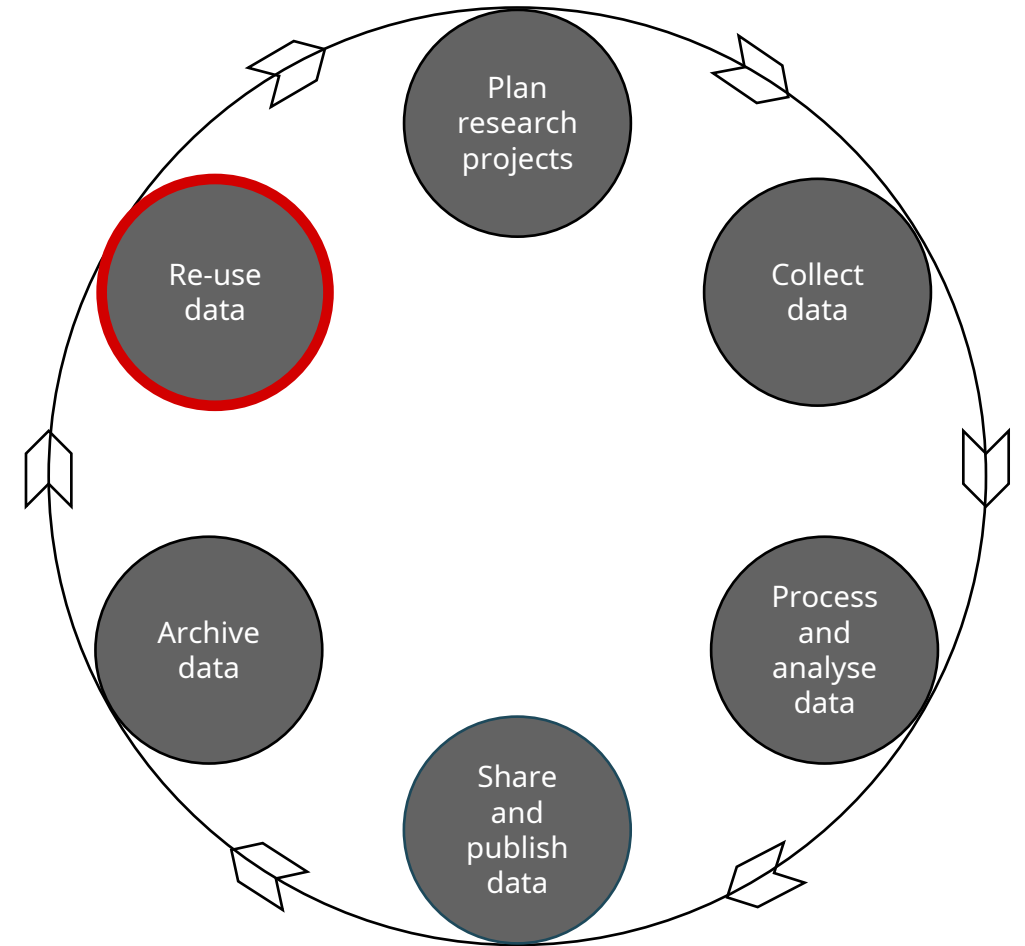
[iRODS Wiki – Forschungsdatenmanagement](#)



The screenshot shows the ZDV Data Center website. At the top right is the ZDV Data Center logo. Below it is a navigation bar with links: HOME, SEARCH, CONTACT, DEUTSCH, ABOUT US, HELP AND SERVICE, and GLOSSARY. A search bar is also present. The main content area is titled 'Archiving Research Data with iRODS'. It explains that iRODS stands for 'Integrated Rule-Oriented Data System' and is an open source data management software solution. It mentions that the ZDV operates an iRODS server for the long-term archiving of research data. Below this, there is a section titled 'Access' which states that access to iRODS is currently only possible via Linux and that authentication takes place via Kerberos. It also mentions that the iRODS client is installed on the Linux servers of the ZDV and on the login nodes of Mogon. For the Debian provided by the ZDV, the following two packages must be installed: '*irods-icommands'. There is also a section titled 'Experimental Web Access' which states that experimental access to archiving is also possible via the web browser, although some restrictions must be observed. It lists three points: 1. The listed web access can change at any time, as it is not stable. Use of the web interface is at your own risk! 2. Only one file can be uploaded at a time. 3. Files must not be larger than ~5GB. It also mentions that access is possible via iRODS-web and provides a link: <https://irods-web.zdv.uni-mainz.de/metalnx/collections?path=%2Fzdv%2Fproject>. Finally, there is a section titled 'ZDV Independent Systems' which states that the ZDV only supports the centrally managed Debian systems and that Linux servers operated independently of the ZDV cannot be supported.

RE-USE DATA

- New research questions using existing data
- Replication of research results
- Questioning results
- Creating research reviews
- Using data for teaching and learning
- ...



MOODLE COURSES ON RDM

Findable on RDM-Website:

[Forschungsdatenmanagement – eine Online-Einführung](#)

[Research Data Management – An online introduction](#)

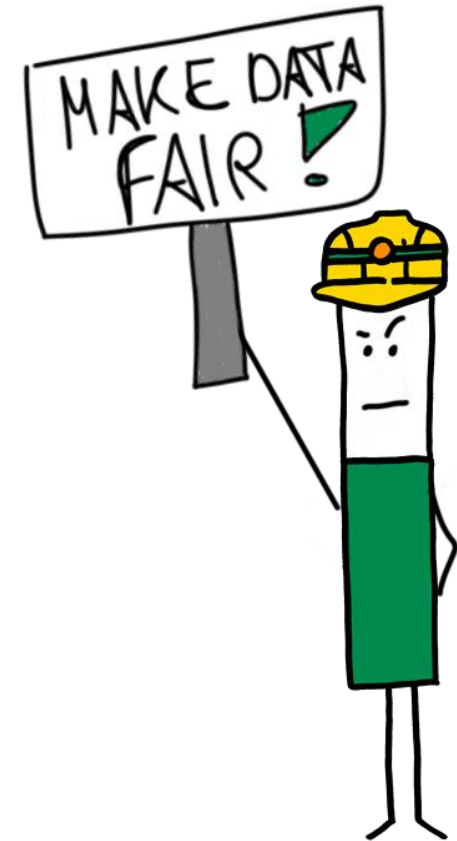
YOUR TEAM OF EXPERTS RESEARCH DATA

- **Dr. Anne Vieten** (Department for Research and Technology Transfer) – General RDM aspects, requirements of funders, DMPs ...
- **Dr. Jörg Steinkamp** (Data Center) – Technical RDM aspects, GitLab, data archiving...
- **Sarah Wettermann** (Data Center) – DMPs ab Research Data Management Organiser, technical RDM aspects, ...
- **Arnela Balic** (core4u) – RDM in social sciences, RDM in Core facilities,
- **Dr. Andrej Berg** (core4u) – RDM in natural sciences, RDM in Core facilities, technical RDM
- **Karin Eckert** (University library) - Metadata, Gutenberg Open Science, ...
- **Dr. Melanie Groß** (University library) – Open Science
- **Esther Reineke and Tina Rotzal** (University library) – Good research practise, academic integrity

Contact: forschungsdaten@uni-mainz.de **Website:** [Kompetenzteam Forschungsdaten \(uni-mainz.de\)](https://kompetenzteam.forschungsdaten.uni-mainz.de)

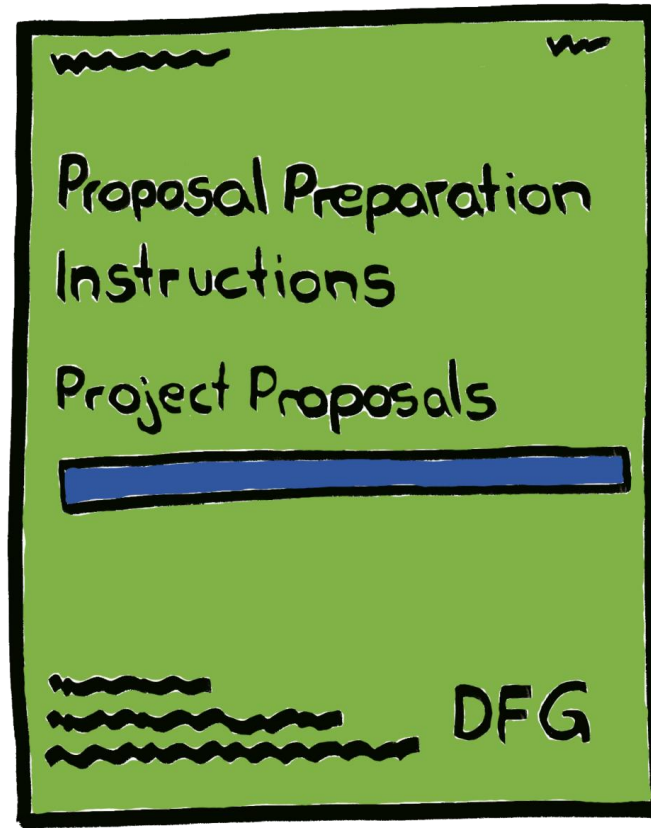
THANK YOU FOR YOUR ATTENTION!

Questions?



[Abb.: Sketchnotes & Sketches – FranzMachtDas _CC BY-SA 4-0.png](#) CC-BY-SA 4.0 Franziska Mau

RDM REQUIREMENTS OF THE THIRD-PARTY FUNDERS



CC BY-SA 4.0
Franziska Mau
[FranziMachtDas –
Sketches, Zeug & der
tägliche Wahnsinn](#)

RDM REGULATIONS BY THIRD PARTY FUNDERS

- DFG: [Handling of research data ; RDM Checklist](#)
- Horizon Europe: [Open Science in Horizon Europe \(europa.eu\)](#)
- BMBF: [Aktionsplan Forschungsdaten – BMBF](#)
- [Volkswagenstiftung](#)
- [JGU: Leitlinie zum Umgang mit Forschungsdaten](#)

- Code of conduct: „[Good Research Practice](#)“ - Implementation into JGU regulations until summer 2023

DFG- RDM CHECKLIST

1. Data description
2. Documentation and data quality
3. Storage and technical archiving
4. Legal obligations and conditions
5. Data exchange and long-term data accessibility
6. Responsibilities and resources

DFG CHECKLIST IN DETAIL

1. Data description

- Process of data creation
- Data formats (e.g. image data, text data or measurement data)
- Type of data processing
- Expected data volume

2. Documentation and data quality

- Metadata or documentation standards or ontologies
- Quality assurance/control measures
- Necessary methods and tools (software) for data use

DFG CHECKLIST IN DETAIL

3. Storage and technical backup during the course of the project

- Storage location
- Data protection/backup
- Access and usage management for sensitive data

4. Legal obligations and framework conditions

- Legal specifics in the project
- Impact of specifics on publication/accessibility
- How to take into account ownership and copyright issues
- Existing scientific codes or professional standards

DFG CHECKLIST IN DETAIL

5. Data exchange and permanent accessibility of the data

- Data suitability for re-use (other context)
- Criteria for data selection for publication
- Archiving data in suitable infrastructure, where?
- Time of publication
- Embargo periods for data publications

6. Responsibilities and resources

- Responsibility for adequate handling of research data
- Resource requirements (costs, time or other)
- Responsibility for data after the end of the project

DFG FUNDING FOR RDM

Eligible costs:

- Personnel costs for the preparation of data for subsequent use or for the transfer to existing infrastructures can be raised.
- Costs for the use of relevant infrastructures (e.g. user fees for external repositories)
- Funding for (project-specific RDM hardware and software)

HORIZON EUROPE - RDM

- Data management plans mandatory!
- Projects generating research data MUST manage their data responsibly and in accordance with the **FAIR** principles
- Open access to research data. "as open as possible, as closed as necessary", i.e. there may be exceptions to open access to research data. (personal data protection, patenting etc.)

HORIZON EUROPE - RDM

- Create and regularly update a DMP (first after 6 month)
- Publish data in a trusted OA repository
- Deposition and open access as early as possible
- Partial obligation to use repositories with EOSC connection
- CC BY or CC 0 (or equivalent) license required
- Justify exceptions to Open Data in the DMP (e.g. data protection)
- Information on the repository on other research results or other tools and instruments needed to reuse or validate the data;
- Requirements for metadata: CC0 and PIDs
- Costs for RDM are eligible

[Open Science in Horizon Europe \(europe.eu\)](https://europe.eu)

BMFTR

- Take account to **FAIR data principles**
- When data collections are funded, obligation to process and store the data standardized and systematically,
- Data publications in accordance with FAIR standards are considered as scientific achievement

Funded projects:

- Necessary expenses for RDM are eligible for funding
- DMPs sometimes obligatory (e.g. educational research)