

Introduction to Research Data Management

Dr. Anne Vieten and Jürgen Windeck

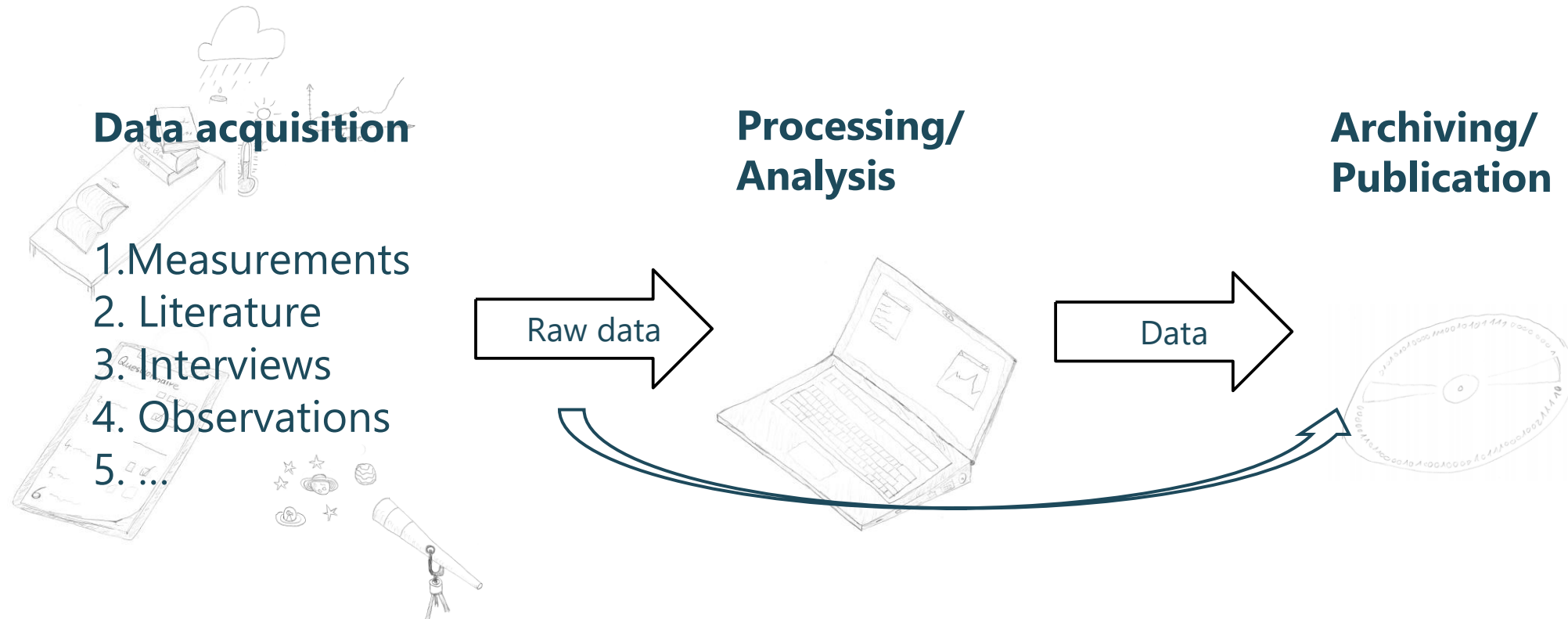
17.06.2026

| CONTENT

1. Motivation
2. Planning RDM
3. Collect data
4. Process and analyse data
5. Share and publish data
6. Archive data
7. Re-use data
8. RDM requirements of funding organisations

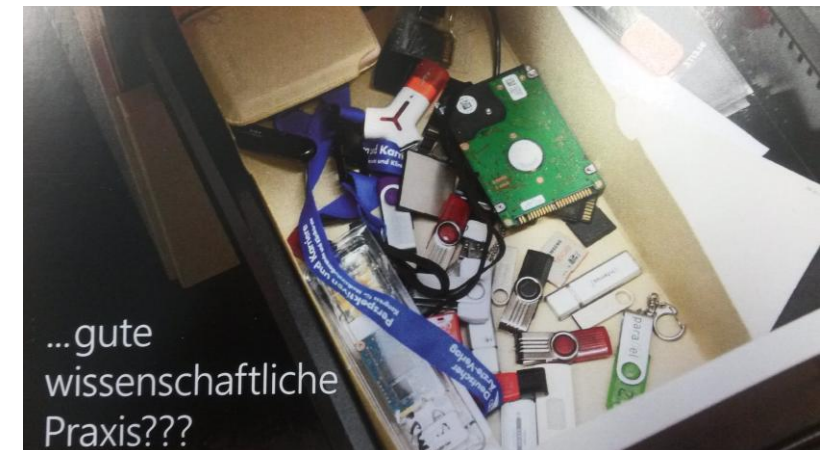
1. MOTIVATION

RESEARCH DATA, A DEFINITION



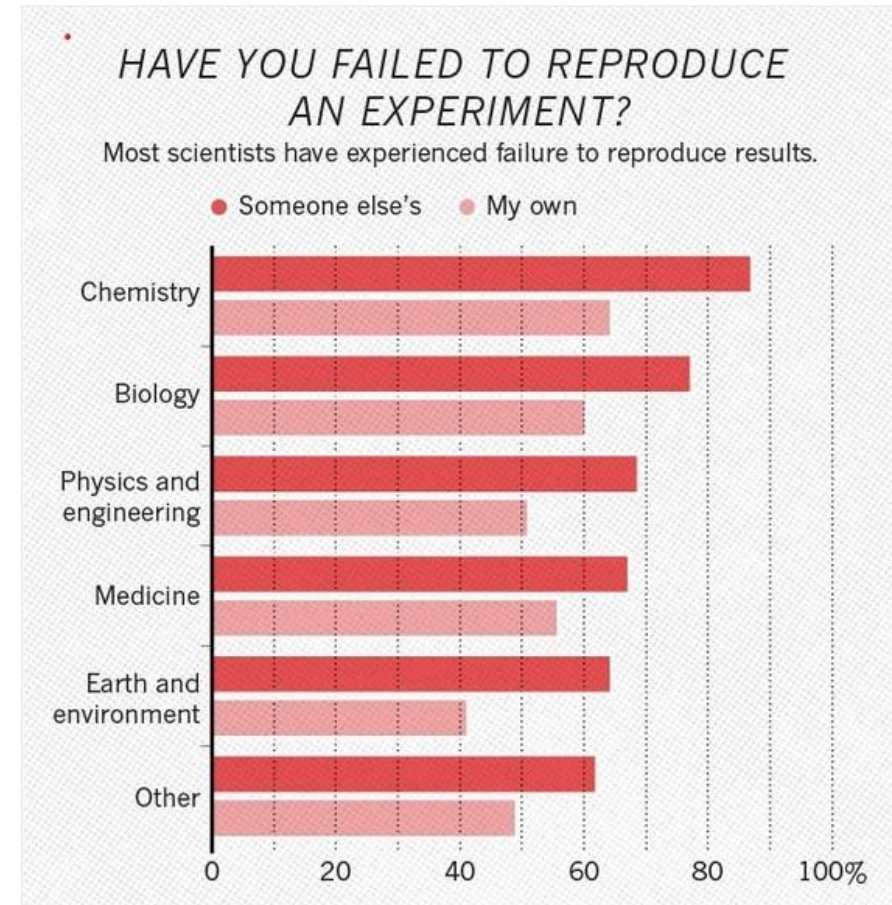
WHY RDM IS IMPORTANT FOR YOU?

- 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



REPRODUCIBILITY OF RESEARCH RESULTS

...often fails



REMEMBER: ONLY A GENIUS CAN HANDLE CHAOS ...

... but what happens if the genius leaves?



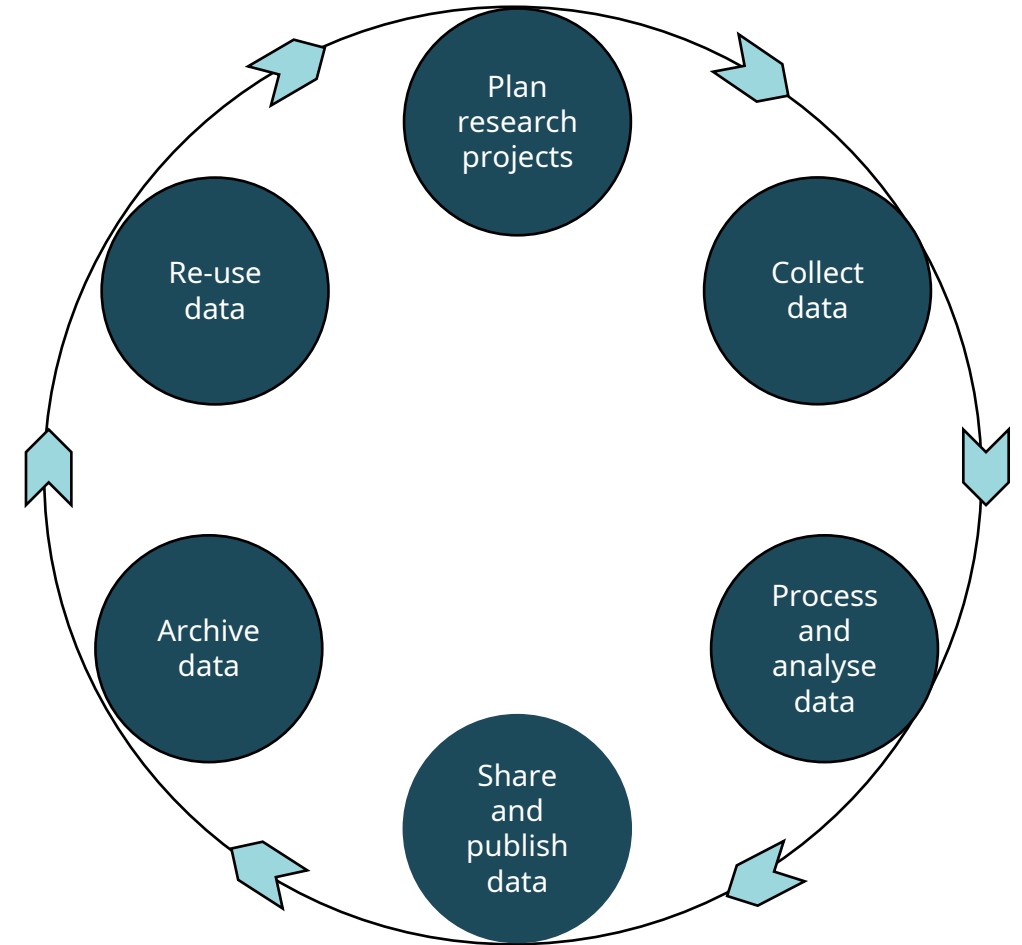
FURTHER REASONS: REQUIREMENTS OF THIRD-PARTY FUNDING BODIES AND RMU

- DFG: [Guidelines for Safeguarding Good Research Practice. Code of Conduct](#)
- [Forschungsdaten-Management-Policy der Goethe-Universität](#)
- [Leitlinie zum Umgang mit Forschungsdaten der JGU](#)
- [Leitlinien zum Umgang mit digitalen Forschungsdaten an der TU Darmstadt](#)
- 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

2. PLANNING RDM

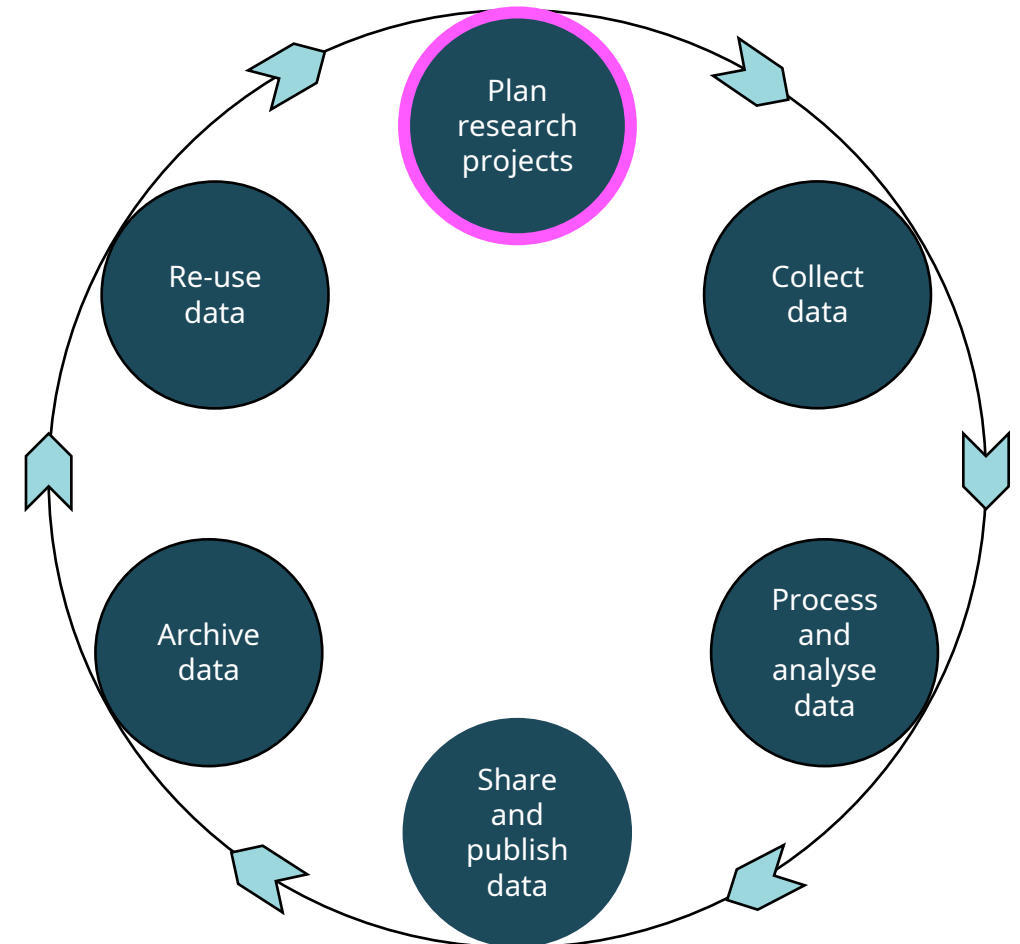
THE RESEARCH DATA LIFE CYCLE

... describes all steps of
research data management



AS IS SO OFTEN TRUE, GOOD PLANNING IS THE KEY TO SUCCESS!

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

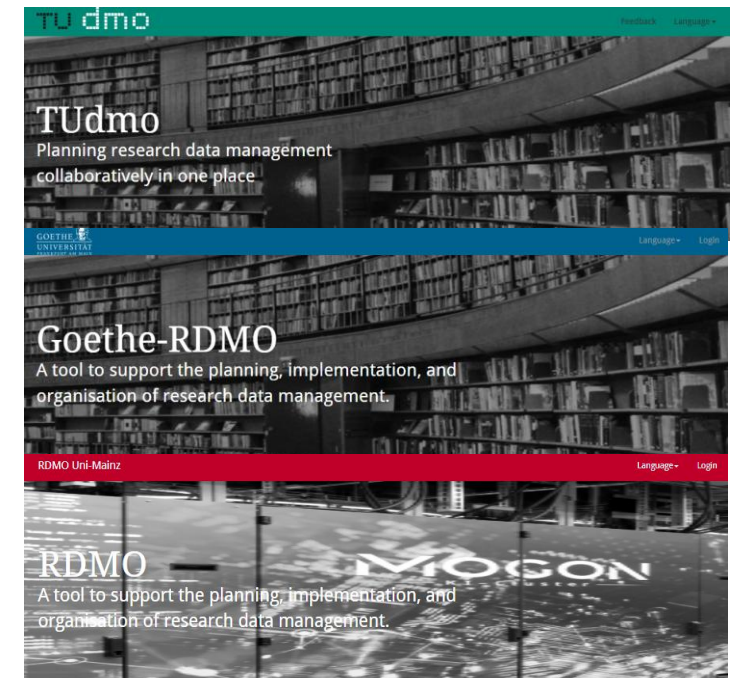


DOCUMENT AND PLAN YOUR RDM WITH A DMP

The DMP-Tool Research Data Management Organiser (RDMO) will be presented in the next session!

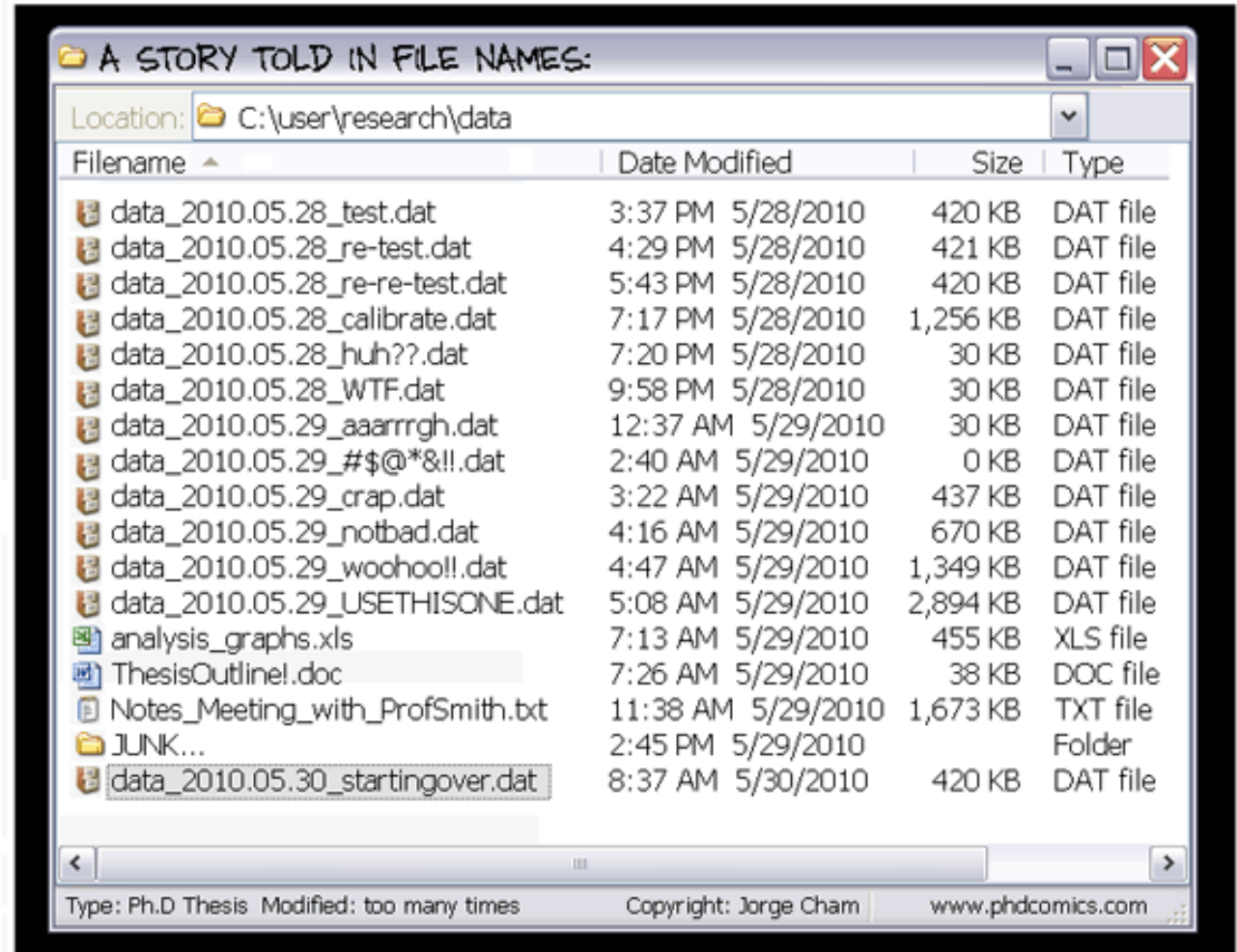


<https://rdmo.rhein-main-universitaeten.de>



ORGANISE DATA

...or lose track of things?



BETTER LIKE THAT!

Initial planning is key:

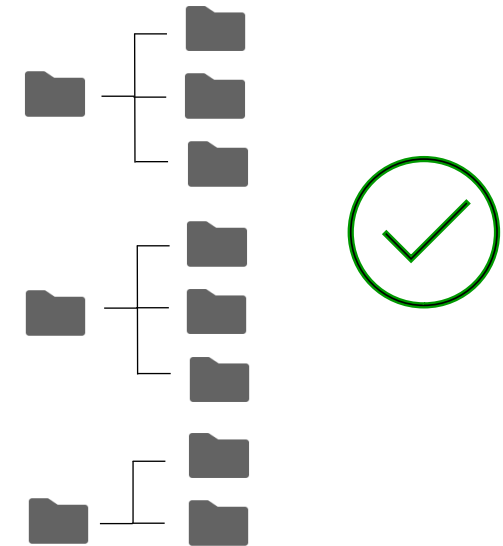
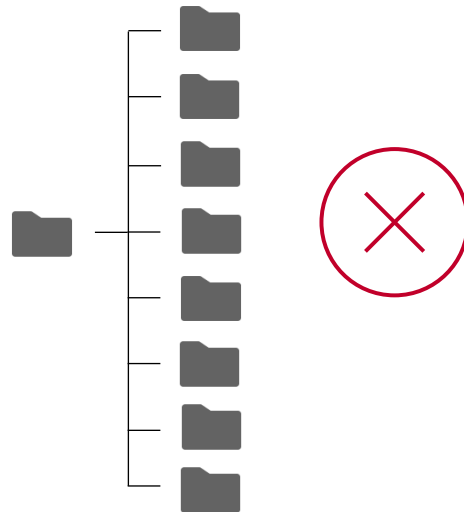
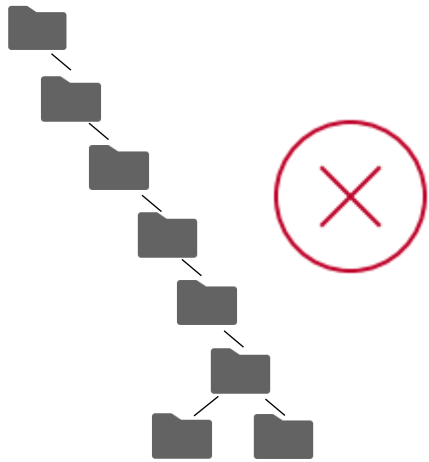
Develop a naming convention for files and folders. Document it!

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.

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:

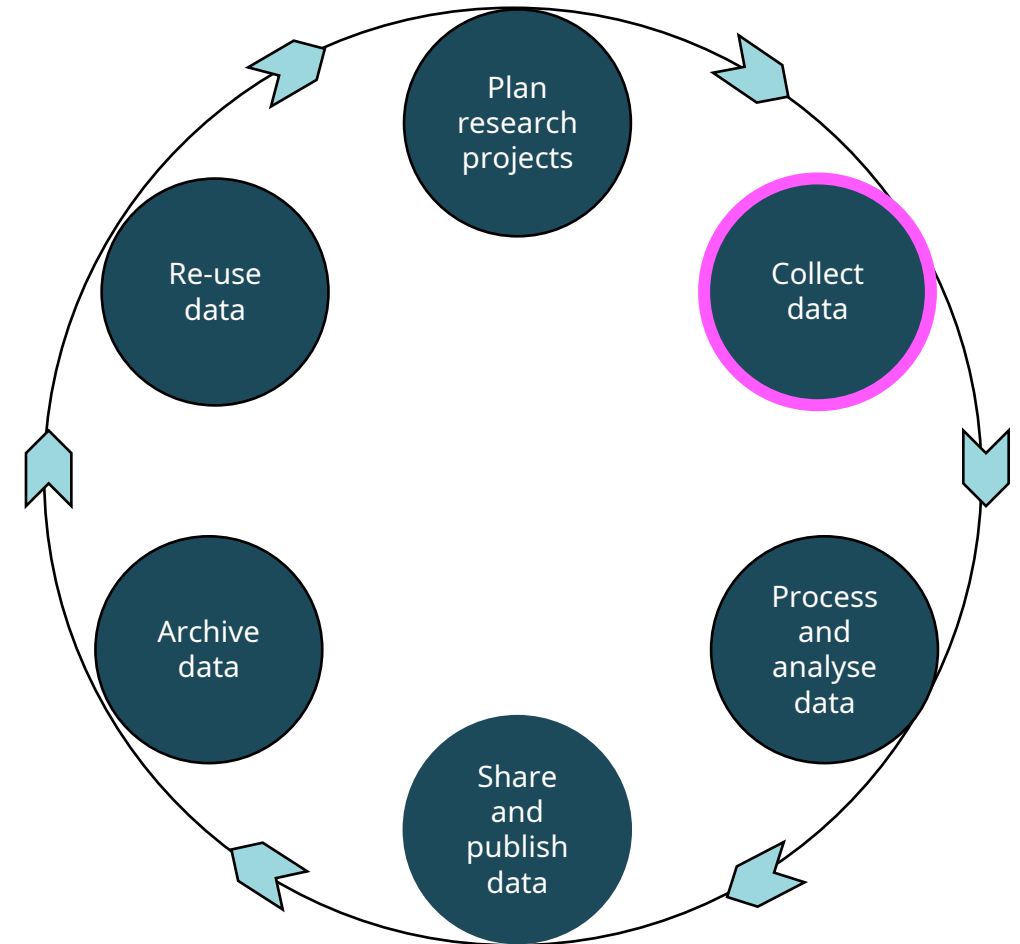
- 1. Version management as part of the file naming convention.** Version numbers in the file name, e.g. V01, V02, etc. Use of the current version of data in the file name
- 2. Use of versioning tools.** E.g. Git-based software such as GitLab (actually for software code, but also applicable to other texts)

 GitLab introductory course today from 2:00 - 5:00 pm

3. COLLECT DATA

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

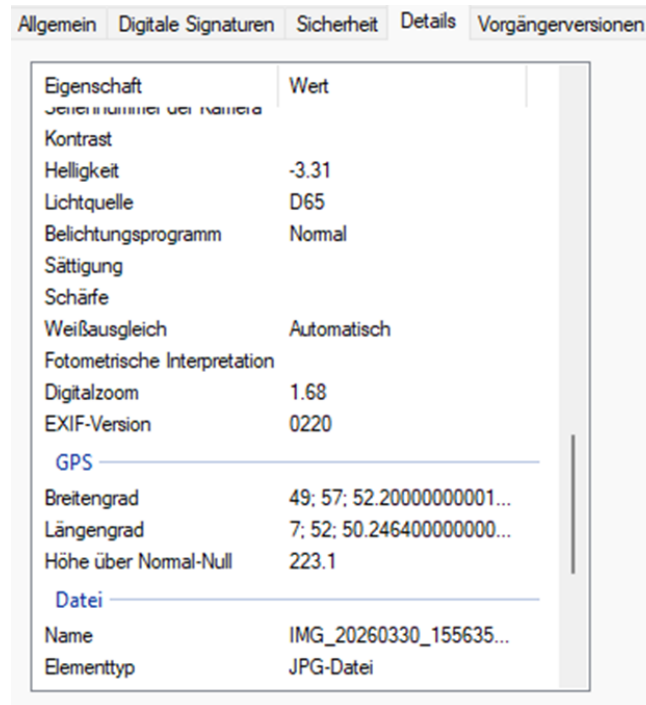
- Identifier
- Creators
- Title
- Publisher
- Publication year
- Subject
- Language
- Resource type
- Descriptions

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
- ...



Eigenschaft	Wert
Jehennunnen uei nannere	
Kontrast	
Helligkeit	-3.31
Lichtquelle	D65
Belichtungsprogramm	Normal
Sättigung	
Schärfe	
Weißausgleich	Automatisch
Fotometrische Interpretation	
Digitalzoom	1.68
EXIF-Version	0220
GPS	
Breitengrad	49; 57; 52.20000000001...
Längengrad	7; 52; 50.246400000000...
Höhe über Normal-Null	223.1
Datei	
Name	IMG_20260330_155635...
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.

METADATA AS JSON FILE

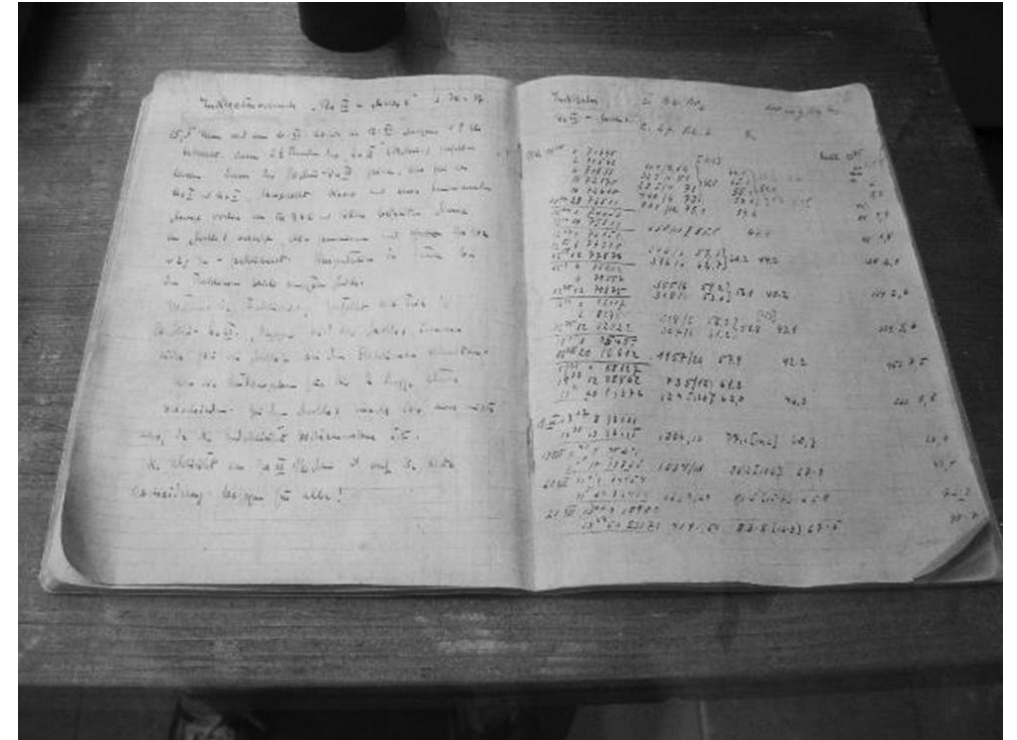
JSON = Java Script Object Notation

- 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) for experimental protocols
- ➔ eLabFTW introduction today from 11:30 - 12:30
- Code books
- Database schemas
- Software syntax
- JSON files
- Readme.txt file describing the data sets
- ...



Otto Hahns Laborbuch 1938, CC BY-SA 2.0 Deutsches Museum München



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 the 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

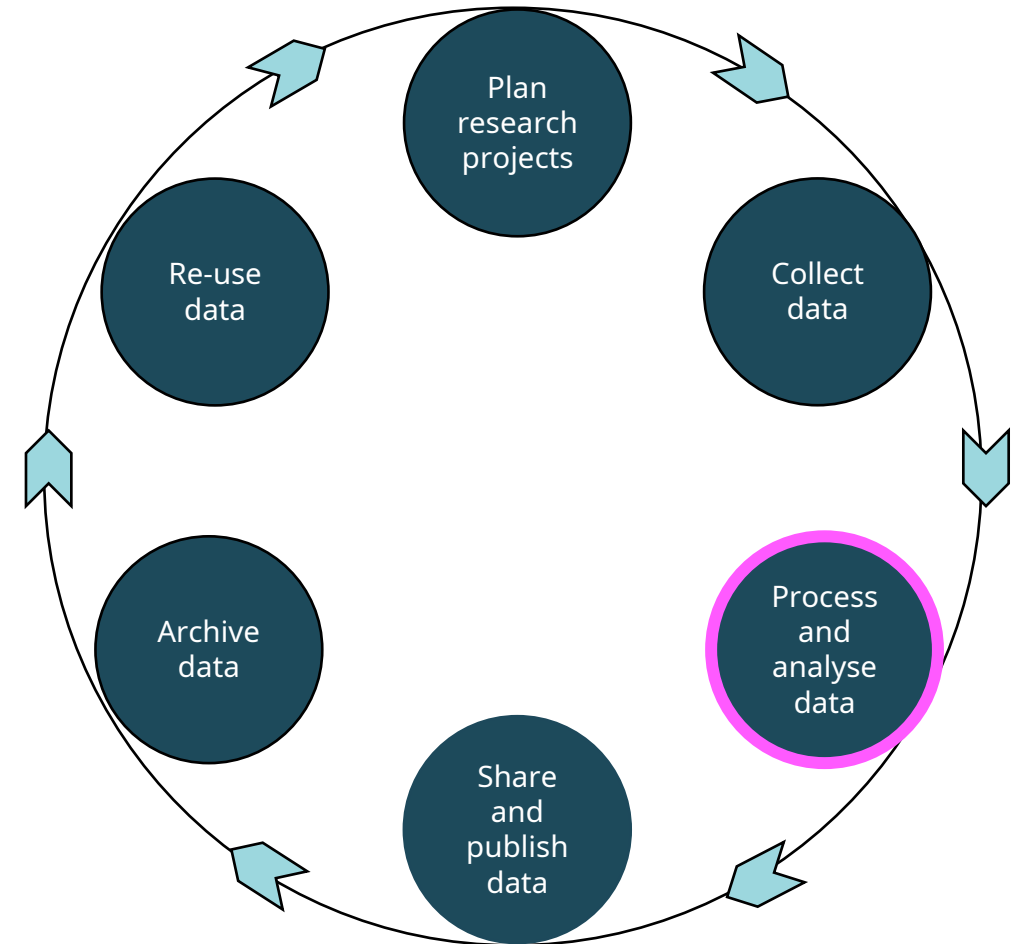
2 = 2 different storage media

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

4. PROCESS AND ANALYSE DATA

PROCESS AND ANALYSE DATA

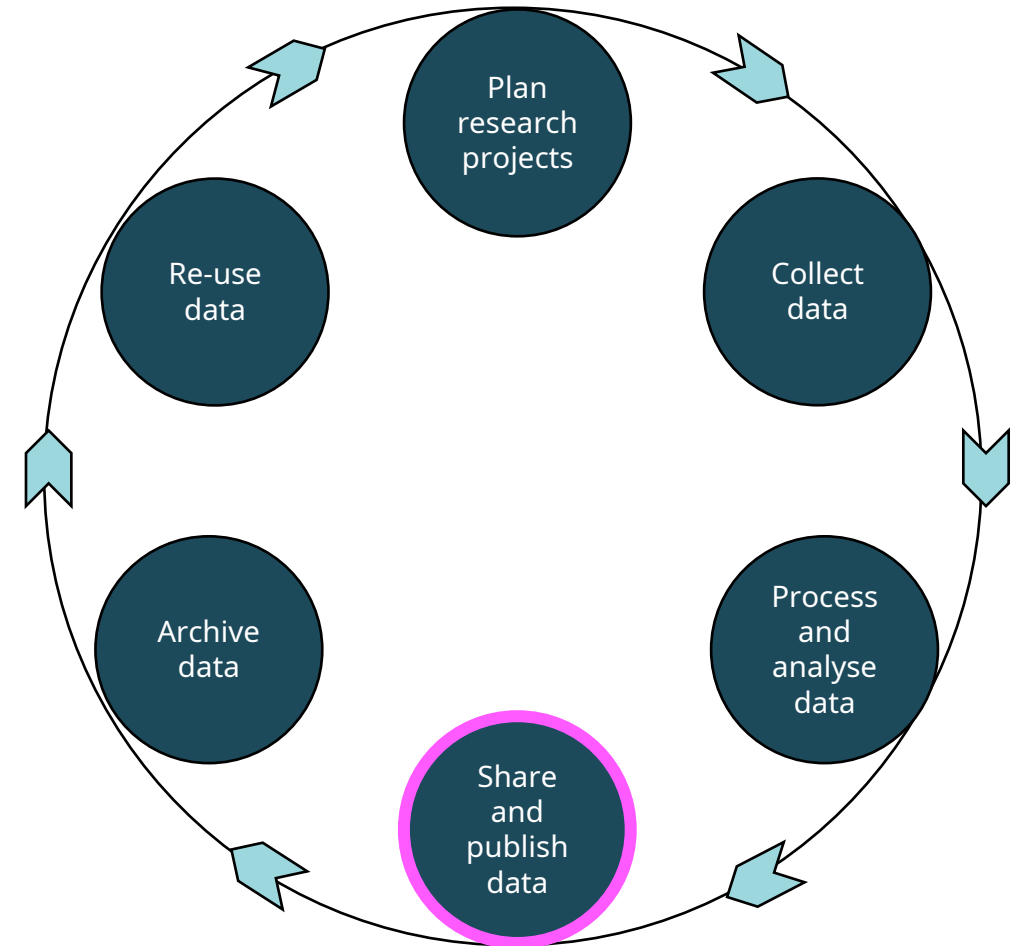
- Interpretation of data
- Interlinking of data
- Measures for data plausibility and reproducibility
- Creation of documentation/metadata
- Production of research results
- Quality assurance
- ...



5. SHARE AND PUBLISH DATA

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
- ...



PUBLISHING DATA AT GUF:

GUDE

GOETHE UNIVERSITY DATA REPOSITORY (GUDE)

- For all members of GUF
- Permanently store research data
- (for at least 10 years)
- Accordance with FAIR criteria
- Archiving or publishing with a DOI
- Operated by GUF Library

- <https://gude.uni-frankfurt.de/home>

PUBLISHING DATA AT JGU:

GUTENBERG OPEN SCIENCE

GUTENBERG OPEN SCIENCE REPOSITORY

- 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](#)

PUBLISHING DATA AT TUDA:

TUDATALIB

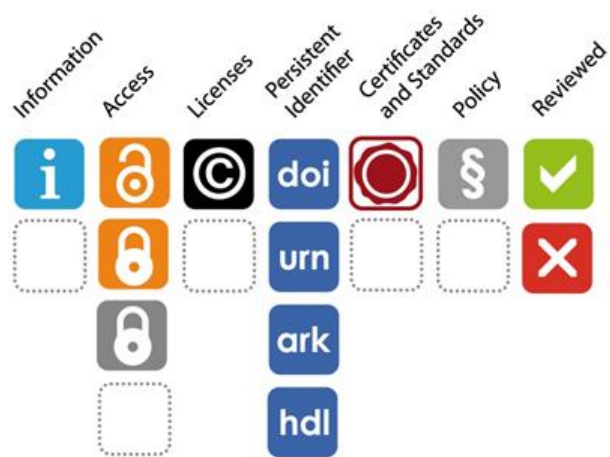
TUDATALIB

- **data sets** can be
 - ... archived (10 years and longer),
 - ... shared with others,
 - ... published including DOI and permanent storage.
- very detailed rights and role management based on TU-ID and affiliation to TU Darmstadt's institutes
- collaboration with non-TU researchers
- <https://tudatalib.ulb.tu-darmstadt.de>

HOW TO FIND OTHER 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

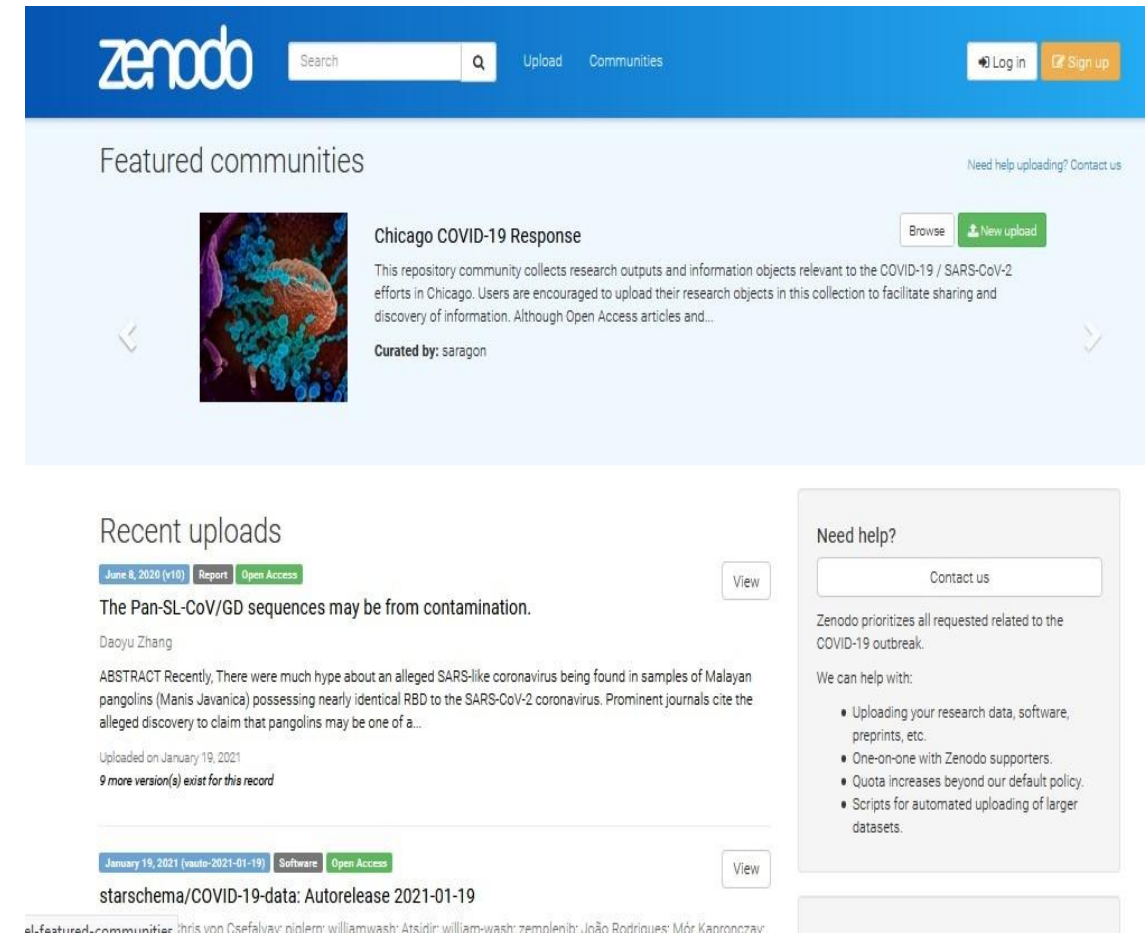
Text

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



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)



The screenshot displays the Zenodo website interface. At the top, there is a blue header with the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. On the right side of the header, there are 'Log in' and 'Sign up' buttons.

The main content area is titled 'Featured communities'. It features a card for the 'Chicago COVID-19 Response' community. The card includes a thumbnail image of a virus particle, the community name, a 'Browse' button, and a 'New upload' button. Below the name, there is a brief description: 'This repository community collects research outputs and information objects relevant to the COVID-19 / SARS-CoV-2 efforts in Chicago. Users are encouraged to upload their research objects in this collection to facilitate sharing and discovery of information. Although Open Access articles and...'. It also mentions 'Curated by: saragon'.

Below the featured communities, there is a section for 'Recent uploads'. It shows a list of uploads with details such as the date, version, and access status. One upload is highlighted: 'The Pan-SL-CoV/GD sequences may be from contamination.' by Daoyu Zhang, uploaded on January 19, 2021. It is marked as 'Software' and 'Open Access'. Below this, there is a link to 'starschema/COVID-19-data: Autorelease 2021-01-19'.

On the right side of the page, there is a 'Need help?' section with a 'Contact us' button. Below this, there is a text block stating: 'Zenodo prioritizes all requested related to the COVID-19 outbreak. We can help with:' followed by a list of services: 'Uploading your research data, software, preprints, etc.', 'One-on-one with Zenodo supporters.', 'Quota increases beyond our default policy.', and 'Scripts for automated uploading of larger datasets.'

LEGAL THINGS: OPEN LICENCES

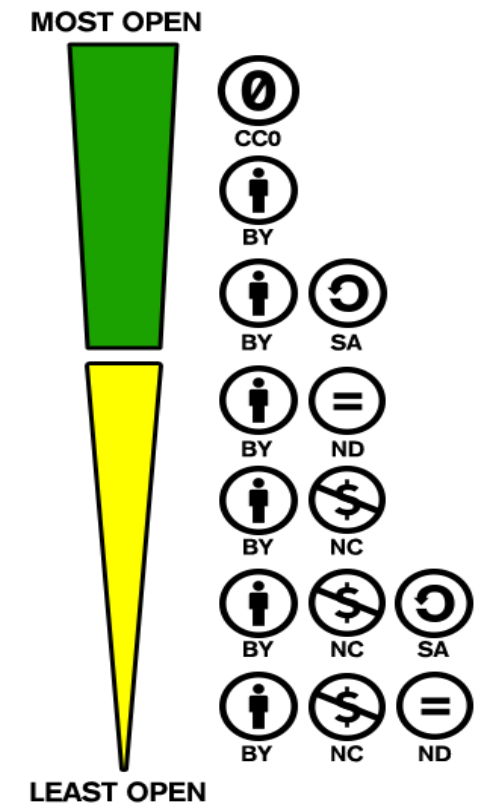
Offer the possibility to set conditions for the reuse of data

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), MIT, ...

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



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?

YOUR RESEARCH DATA SHOULD BE FAIR!

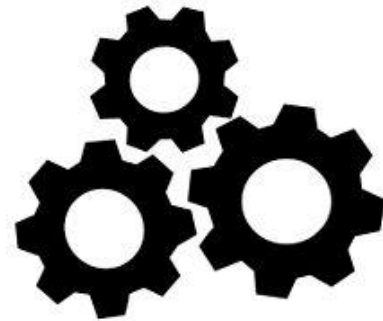
F
indable



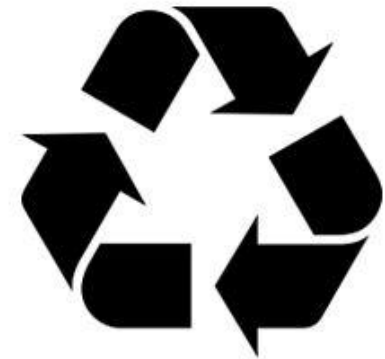
A
ccessible



I
nteroperable



R
eusable



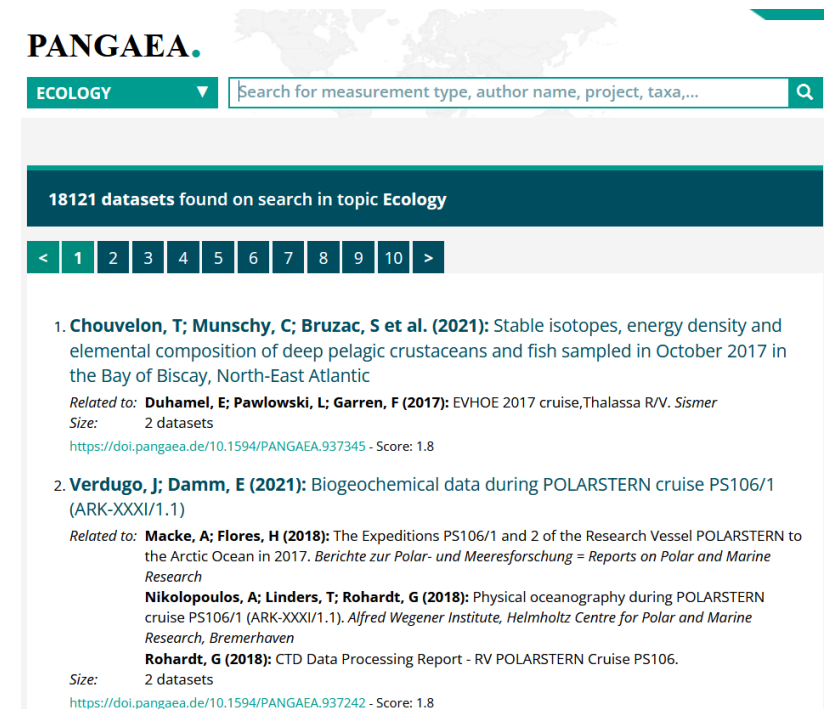
WHAT EXACTLY DOES FAIR MEAN?

[FAIR Principles - GO FAIR](#)



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


Screenshot [Pangaea](#) data repository

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 Latit
License:	 Creative Commons Attribution 4.0 International (CC-BY-4.0)
Size:	73 datasets

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.

Data Cite (Metadata) Basic Scheme

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

Find suitable standards, e.g.via <https://fairsharing.org>

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 Latitu
License:	 Creative Commons Attribution 4.0 International (CC-BY-4.0)
Size:	73 datasets

FAIR4RS PRINCIPLES

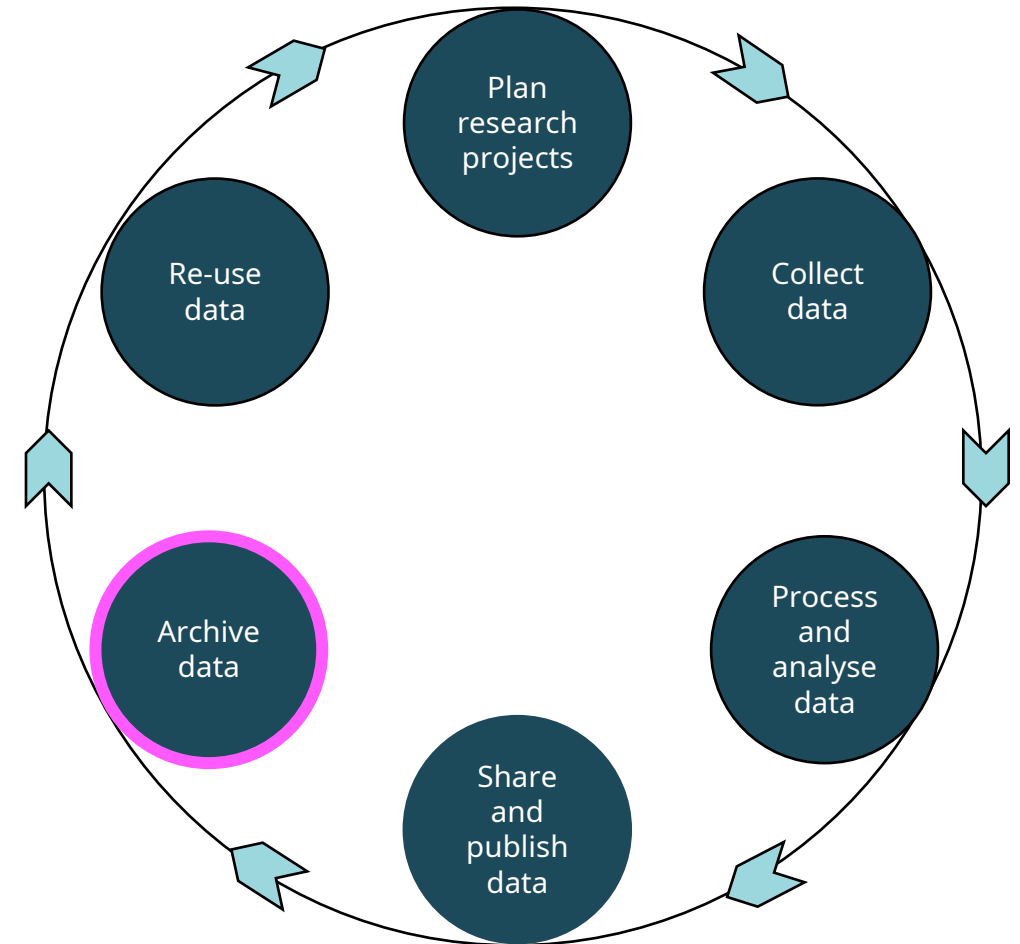
The FAIR principles have been extended to research software:

[FAIR Principles for Research Software \(FAIR4RS Principles\)](#)

6. ARCHIVE DATA

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



(LONG-TERM) ARCHIVING – RECOMMENDATIONS

- 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



ARCHIVING DATA AT GUF

BACKUP AND ARCHIVING SERVICE

- IBM Spectrum Protect software
- no additional cost to users
- licence fees paid by the HRZ
- Backup for 90 days
- Archiving for 10 years

- For more Information:
- https://www.rz.uni-frankfurt.de/134715246/Backup_und_Archivservice



ARCHIVING DATA AT JGU: IRODS

WHAT IS IRODS?

- Virtual Filesystem
- Attached searchable Metadata
- Publishing
- Fine grained Access Control
- Workflow automation
- Commandline tools

PUBLISHING AT TUDA:

TUDATALIB



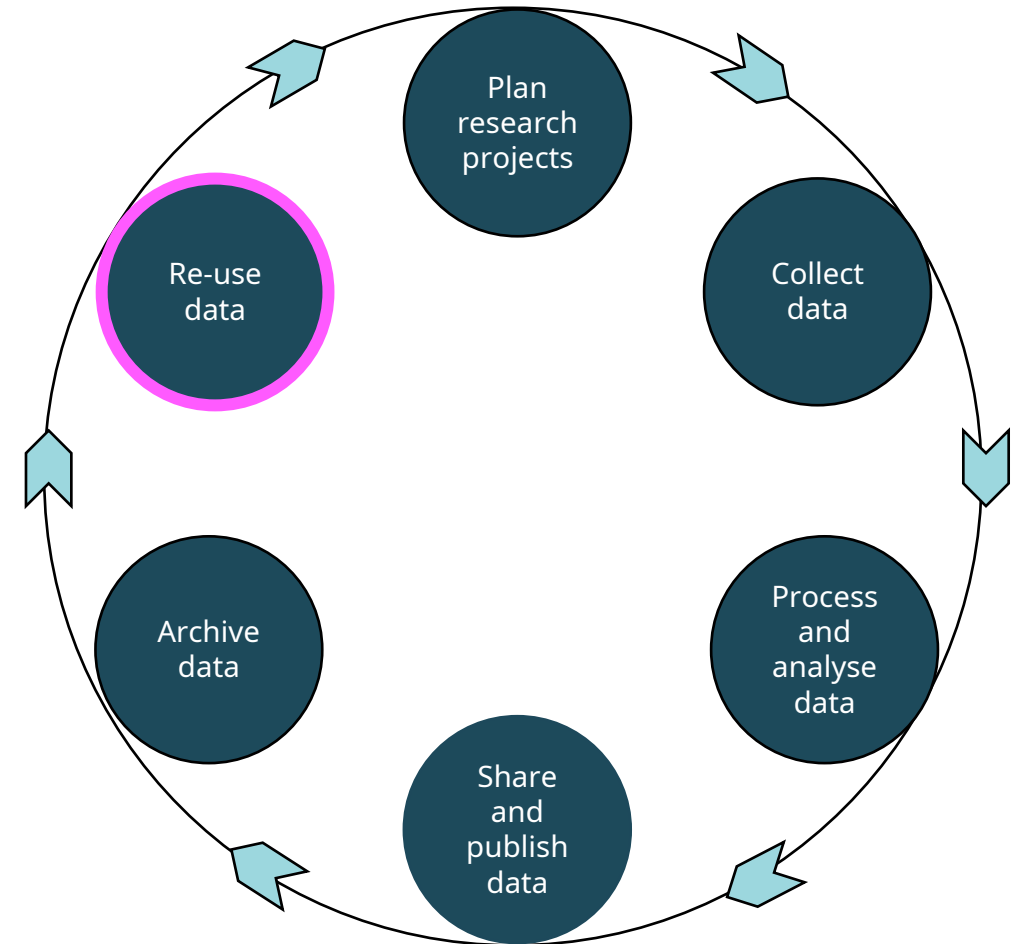
TUDATALIB

- **data sets** can be
 - ... archived (10 years and longer),
 - ... shared with others,
 - ... published including DOI and permanent storage.
- very detailed rights and role management based on TU-ID and affiliation to TU Darmstadt's institutes
- collaboration with non-TU researchers
- <https://tudatalib.ulb.tu-darmstadt.de>

7. RE-USE DATA

RE-USE DATA

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



8.

RDM-REQUIREMENTS OF FUNDING ORGANISATIONS

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

Checklist Regarding the Handling of Research Data

1. Data description

How does your project generate new data? Is existing data reused? Which data types (in terms of data formats like image data, text data or measurement data) arise in your project and in what way are they further processed? To what extent do these arise or what is the anticipated data volume?

2. Documentation and data quality

What approaches are being taken to describe the data in a comprehensible manner (such as the use of available metadata, documentation standards or ontologies)? What measures are being adopted to ensure high data quality? Are quality controls in place and if so, how do they operate? Which digital methods and tools (e.g. software) are required to use the data?

3. Storage and technical archiving the project

How is the data to be stored and archived throughout the project duration? What is in place to secure sensitive data throughout the project duration (access and usage rights)?

4. Legal obligations and conditions

What are the legal specifics associated with the handling of research data in your project? Do you anticipate any implications or restrictions regarding subsequent publication or accessibility? What is in place to consider aspects of use and copyright law as well as ownership issues? Are there any significant research codes or professional standards to be taken into account?

5. Data exchange and long-term data accessibility

Which data sets are especially suitable for use in other contexts? Which criteria are used to select research data to make it available for subsequent use by others? Are you planning to archive your data in a suitable infrastructure? If so, how and where? Are there any retention periods? When is the research data available for use by third parties?

6. Responsibilities and resources

Who is responsible for adequate handling of the research data (description of roles and responsibilities within the project)? Which resources (costs; time or other) are required to implement adequate handling of research data within the project? Who is responsible for curating the data once the project has ended?

https://www.dfg.de/research_data/checklist

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.)



Ref. Area(2023)7009200 - 16/10/2023



Horizon Europe

Data Management Plan Template

Version 1.0
05 May 2021

Horizon Europe [DMP Template](#)

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 \(europa.eu\)](https://europa.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)

CONTACT INFORMATION

Goethe Universität Frankfurt:

forschungsdaten@ub.uni-frankfurt.de

Johannes Gutenberg-Universität Mainz:

forschungsdaten@uni-mainz.de

Technische Universität Darmstadt:

tudata@tu-Darmstadt.de



THANK YOU FOR ATTENTION

Questions?

RMU-GESCHÄFTSSTELLE

RMU gGmbH

+49 69 247477600

gs@rhein-main-universitaeten.de



**RHEIN-MAIN
UNIVERSITÄTEN**

Goethe-Universität Frankfurt am Main
Johannes Gutenberg-Universität Mainz
Technische Universität Darmstadt