KICK-START RESEARCH DATA MANAGEMENT

Open Science Week 2024

23.10.2024 I INES SCHMAHL

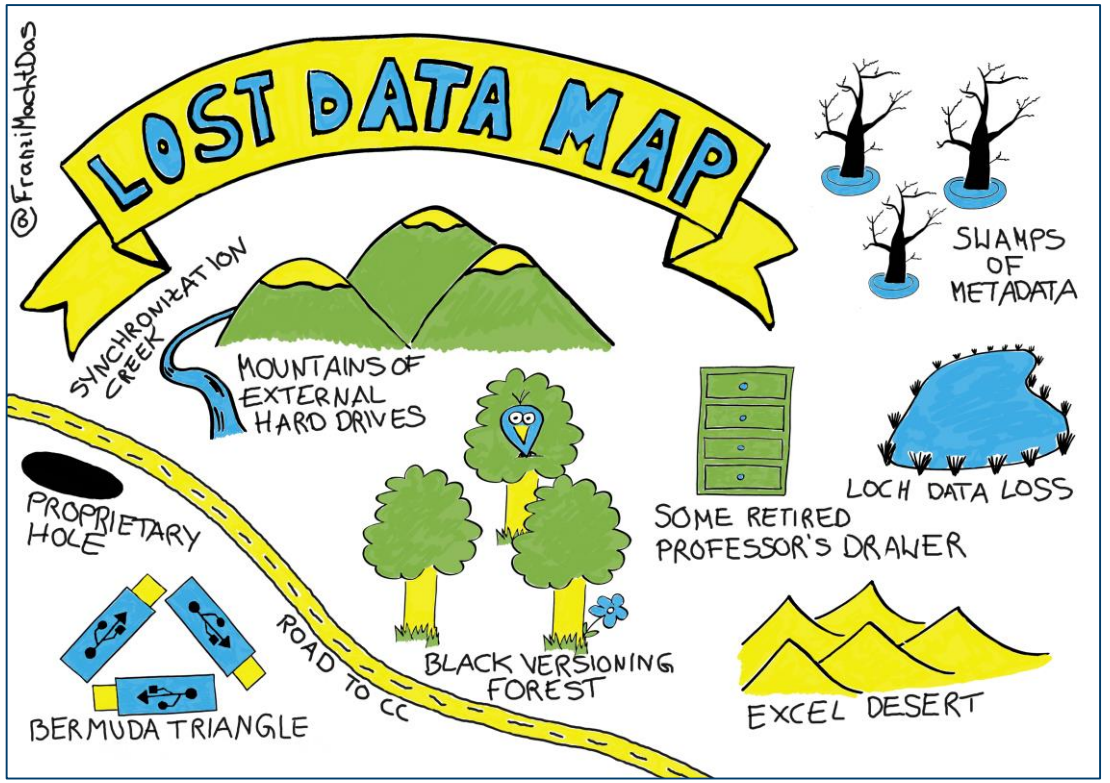
Member of the Helmholtz Association

JÜLICH Forschungszentrum

WHY RESEARCH DATA MANAGEMENT

Today

- More and more (digital) data is produced
- Data is stored on different devices
- Variety of software and file formats



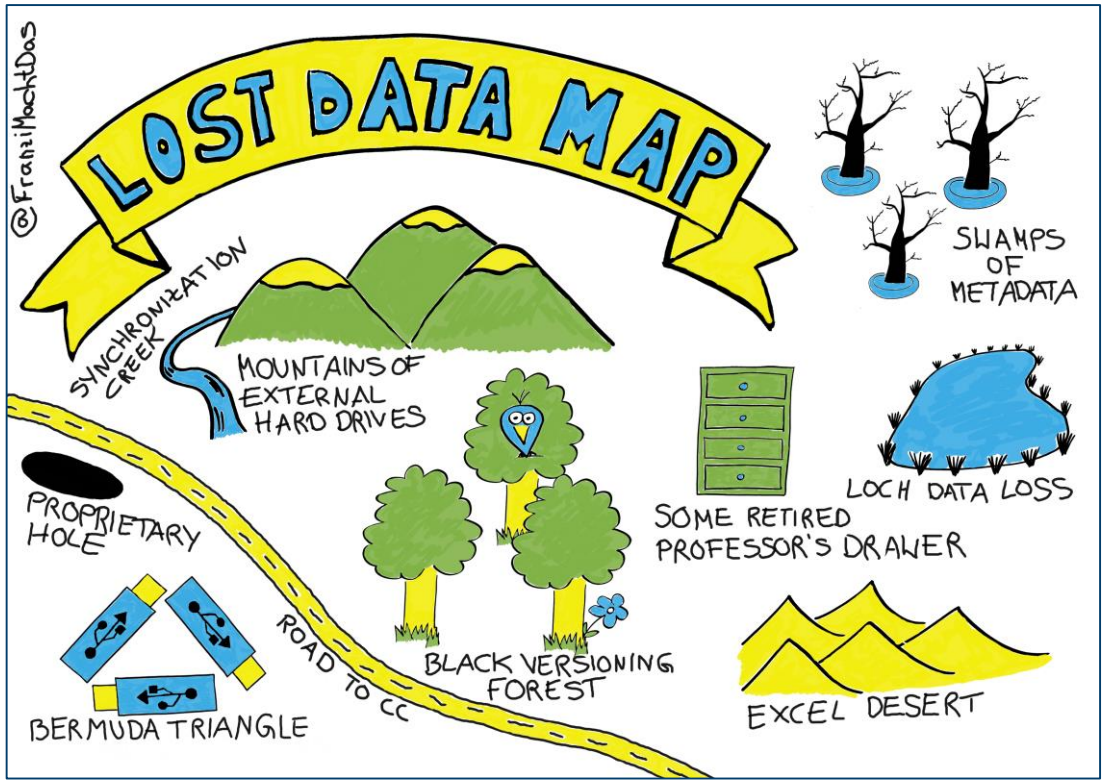
Mau, F. (2019). Sketchnote: Lost Data Map.
Zenodo. https://doi.org/10.5281/zenodo.4388672
published under a CC-BY 4.0 license

WHY RESEARCH DATA MANAGEMENT

Today

- More and more (digital) data is produced
- Data is stored on different devices
- Variety of software and file formats

But we are not taught how to organize data.



Mau, F. (2019). Sketchnote: Lost Data Map.
Zenodo. https://doi.org/10.5281/zenodo.4388672
published under a CC-BY 4.0 license

LEARNING OBJECTIVE MATRIX

Overview

What do we need to know?

1. Basics and concepts
2. Working with data
3. Documentation and metadata
4. Long-term preservation, publication and reuse of data
5. Legal and ethical aspects
6. Infrastructure for support

Grundlagen und übergreifende Konzepte	Allgemeine Grundsätze und Konzepte des Forschungsdaten-managements
	Forschungsdaten-Policies
	Datenmanagementpläne (DMPs)
	FAIR-Prinzipien
	Open X (Open Data, Open Source, Open Science, etc.)
Arbeiten mit Daten	Ordnung und Struktur, Versionierung
	Daten, Datentypen, Datenformate
	Datenspeicherung und Backup
	Datensicherheit
	Datenqualität
	Tools
	(Forschungs-)Software & Coding
Dokumentation und Metadaten	Datendokumentation
	Metadaten und Metadatenstandards
	Persistente Identifikatoren
	Ontologien und kontrollierte Vokabulare
Langzeitarchivierung, Publikation, Nachnutzung	Langzeitarchivierung von Daten
	Publikationswege für Daten
	Repositorien
	Daten nachnutzen
Recht und Ethik	Allgemeine rechtliche Aspekte
	Datenschutz und personenbezogene Daten
	Informierte Einwilligung
	Anonymisierung und Pseudonymisierung
	Ethische Aspekte und
Unterstützungsstrukturen	Rollen im Datenmanagement / Data Stewardship
	Relevante Infrastrukturen
	Didaktik
	Beratung

Petersen, B., Engelhardt, C., Hörner, T., Jacob, J., Kvetnaya, T., Mühlichen, A., Schranzhofer, H., Schulz, S., Slowig, B., Trautwein-Bruns, U., Voigt, A., & Wiljes, C. (2022). Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards (Version 1). Zenodo. https://doi.org/10.5281/zenodo.7034478

BASICS AND CONCEPTS

Overview

- Data life cycle
- Data management plan
- FAIR Principles
- Data policies

BASICS AND CONCEPTS

Data life cycle

Data life cycle presents the workflow of research data during a research project.

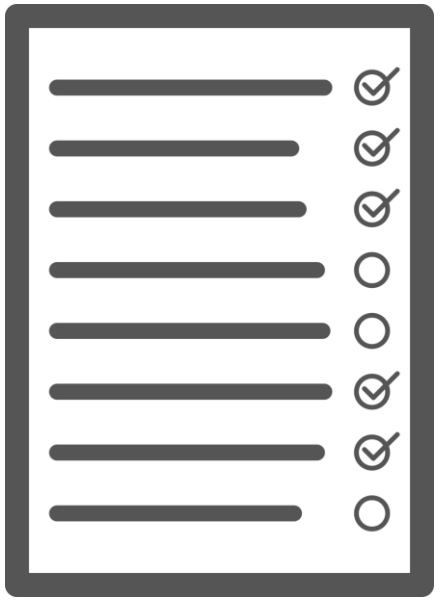
Image source by ELIXIR Research Data Management Kit under CC-BY license downloaded from https://github.com/elixir-europe/rdmkit/blob/master/images/data_life_cycle.png

BASICS AND CONCEPTS

Data management plan

- Project description
- Description of the datasets
- Backup and preservation strategy
- Sharing strategy
- Resources and responsibilities

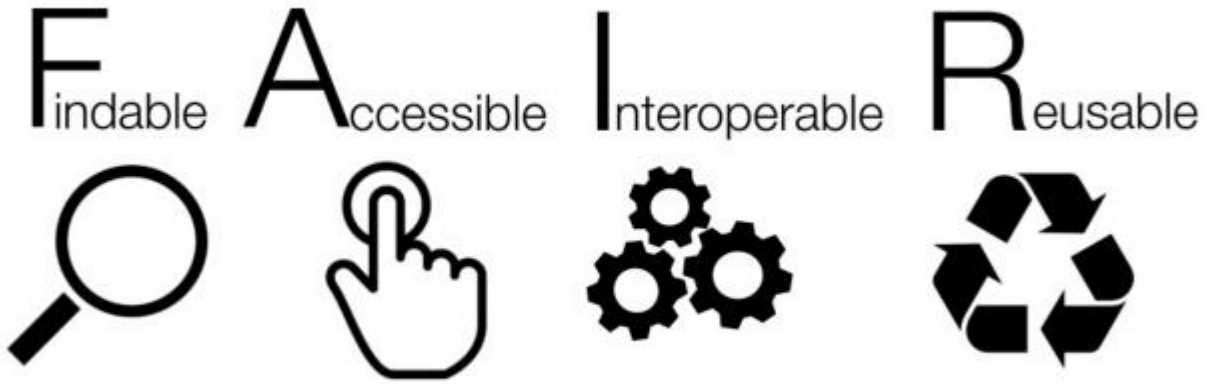
More information on the rdm portal: Data management plans



BASICS AND CONCEPTS

FAIR Principles

FAIR data should be



Published 2016 in the paper **“The FAIR Guiding Principles for scientific data management and stewardship” (**https://doi.org/10.1038/sdata.2016.18)

More information on the rdm portal: FAIR Principles

BASICS AND CONCEPTS

Data policies

Jülich Research Data Handling Guidelines

(published 2019 and currently revised)

Content

- Jülich's commitment to permanent and sustainable support of research data management
- Responsibilities and roles like implementing data managers
- Scientists' duties like to ensure that research data is reusable

More information on the rdm portal: Jülich Research Data Handling Guidelines

WORKING WITH DATA

Best practices for folder and file structure

- Clear folder structure

- Consistent file naming with meaningful names, e.g. for dates YYYYMMDD or YYMMDD format

- Keep track of your versioning, e.g. use tools like Git (open-source software for distributed version control)

More information on the rdm portal: Personal data management

WORKING WITH DATA

Subject-specific best practices

- NFDI (German National Research Data Infrastructure) is an important stakeholder in Germany.

- Mission: Today research data is often stored decentralized. The goal is to make the data accessible for the entire German science system according to the FAIR principles.

- It is organised in community consortia.

More information on the rdm portal: NFDI consortia

DOCUMENTATION AND METADATA

Always give provenance to your data!

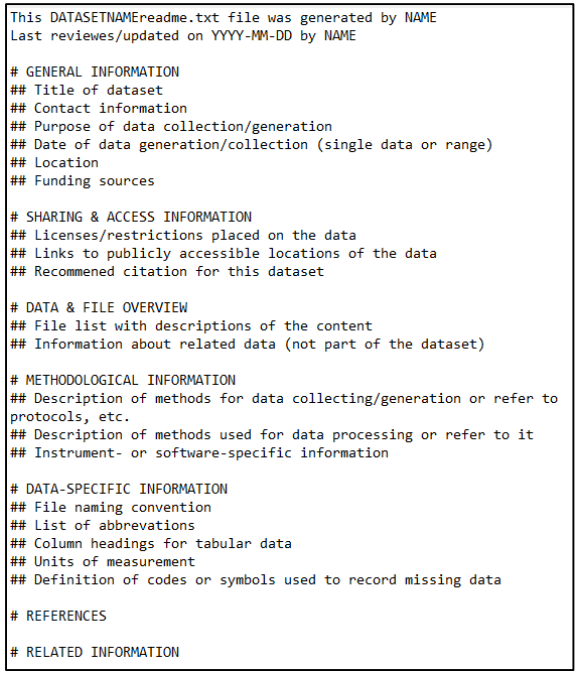
Why and how was the data produced, where, when and by whom?

Tip

Provide a README file to your dataset.

Link template

https://cornell.app.box.com/v/ReadmeTemplate



```
This DATASETNAMEreadme.txt file was generated by NAME
Last reviewes/updated on YYYY-MM-DD by NAME

# GENERAL INFORMATION
## Title of dataset
## Contact information
## Purpose of data collection/generation
## Date of data generation/collection (single data or range)
## Location
## Funding sources

# SHARING & ACCESS INFORMATION
## Licenses/restrictions placed on the data
## Links to publicly accessible locations of the data
## Recommened citation for this dataset

# DATA & FILE OVERVIEW
## File list with descriptions of the content
## Information about related data (not part of the dataset)

# METHODOLOGICAL INFORMATION
## Description of methods for data collecting/generation or refer to
protocols, etc.
## Description of methods used for data processing or refer to it
## Instrument- or software-specific information

# DATA-SPECIFIC INFORMATION
## File naming convention
## List of abbrevations
## Column headings for tabular data
## Units of measurement
## Definition of codes or symbols used to record missing data

# REFERENCES

# RELATED INFORMATION
```

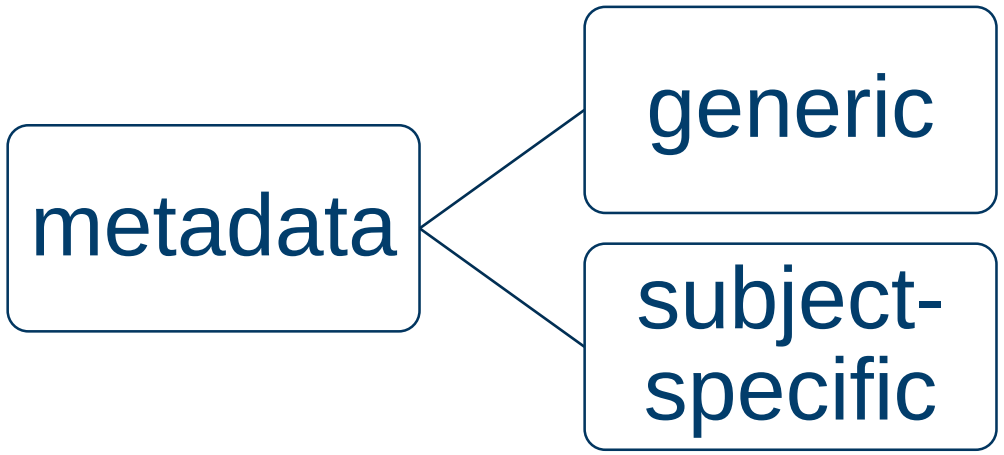
More information on the rdm portal: Personal data management

DOCUMENTATION AND METADATA

Always give metadata to your data!

Metadata are data about data.

What could be proper metadata for a dataset?

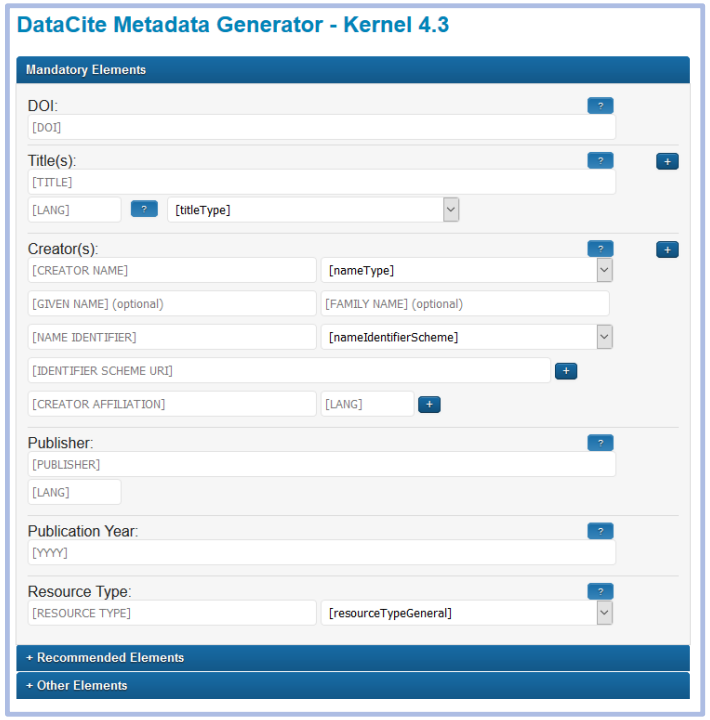


More information on the rdm portal: Metadata

DOCUMENTATION AND METADATA

Always give metadata to your data!

DataCite as a example for a generic schema



https://dhvlab.gwi.uni-muenchen.de/datacite-generator/

LONG-TERM PRESERVATION, PUBLICATION AND REUSE OF DATA

LONG-TERM PRESERVATION, PUBLICATION AND REUSE OF DATA

Find subject-specific data repositories

- Online register for data repositories re3data
- Check out the websites of the NFDI consortia

More information on the rdm portal: NFDI consortia

LONG-TERM PRESERVATION, PUBLICATION AND REUSE OF DATA

Jülich DATA

- Institutional data repository for storing and publishing data at FZ Jülich managed by Central Library
- Free for staff
- Each institute has its own institute collection

More information on the rdm portal: Jülich DATA

LONG-TERM PRESERVATION, PUBLICATION AND REUSE OF DATA

Step 1: Data curation/selection

Data underlying your publication	Additional data
• Raw data Related information • Metadata • Documentation files	<u>Guiding Questions</u> ➢ Is the data reusable? Further: • What is the effort of a reproduction? • Can the data be used for reanalysis (e.g., using new methods)? • How good is the data quality? • Are there subject-specific criteria, e.g. coverage of the subject?

LONG-TERM PRESERVATION, PUBLICATION AND REUSE OF DATA

Step 2: Data preparation

- Check completeness
- Check correctness
- Migrate (if necessary) data into open file formats like .csv for spreadsheets or PDF for text

More information on the rdm portal: Personal data management

LONG-TERM PRESERVATION, PUBLICATION AND REUSE OF DATA

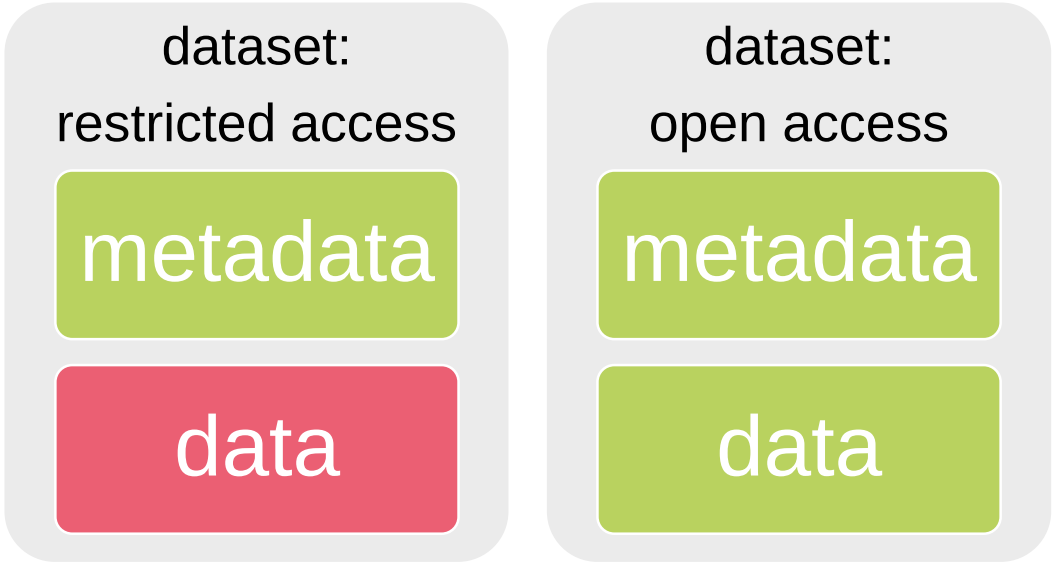
Step 3: Data publication

Publication models

Supplement to a publication	Data repository	Data journal
• Established • Disadvantage: Not so convenient for large amounts of data	1. Subject-specific repository (search via online register re3data) 2. Jülich DATA 3. Advantage: You can grant data with an own license.	• Usually only documentation to the data • List of data journals on forschungsdaten.org

LONG-TERM PRESERVATION, PUBLICATION AND REUSE OF DATA

Access options in a data repository



LEGAL AND ETHICAL ASPECTS

Copyright in Germany

Information is free. → But the work is copyright protected.

In science:

Data, content and results are free. → The form and the representation is copyright protected, e.g. the paper.

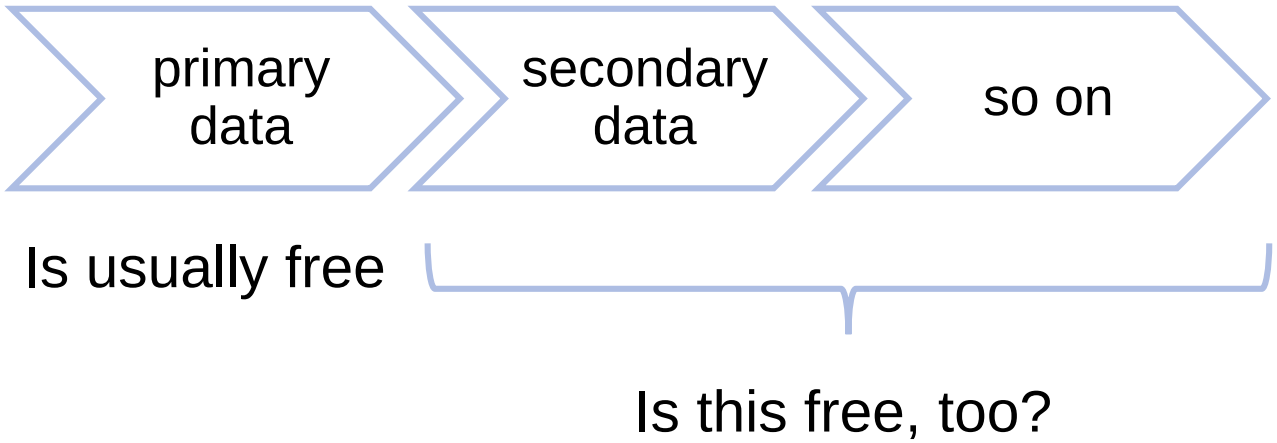
Reason:

Enable free exchange of information in society.

More information on the rdm portal: Legal Aspects

LEGAL AND ETHICAL ASPECTS

Legal aspects for research data



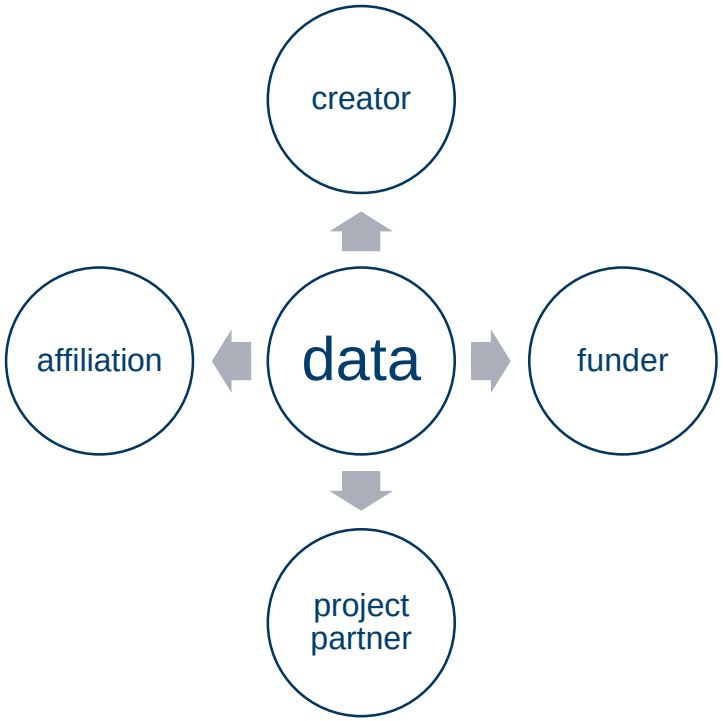
When does copyright apply?

To avoid legal uncertainties → Grant your dataset with a license.

LEGAL AND ETHICAL ASPECTS

Involved parties

- Who is the data creator or collector?
- Are there specifications from the research funder?
- Who has the right of use like employer or project partners?



LEGAL AND ETHICAL ASPECTS

Licenses for research data – Creative Commons

Recommended for research data: CC0 or CC-BY		
	CC-BY	CC0
		
Free to:	• Adapt • Share	• Adapt • Share
Terms:	Give appropriate credit like name of the creator	None (But note: Good Scientific Practice demands to cite your sources.)

INFRASTRUCTURE FOR SUPPORT

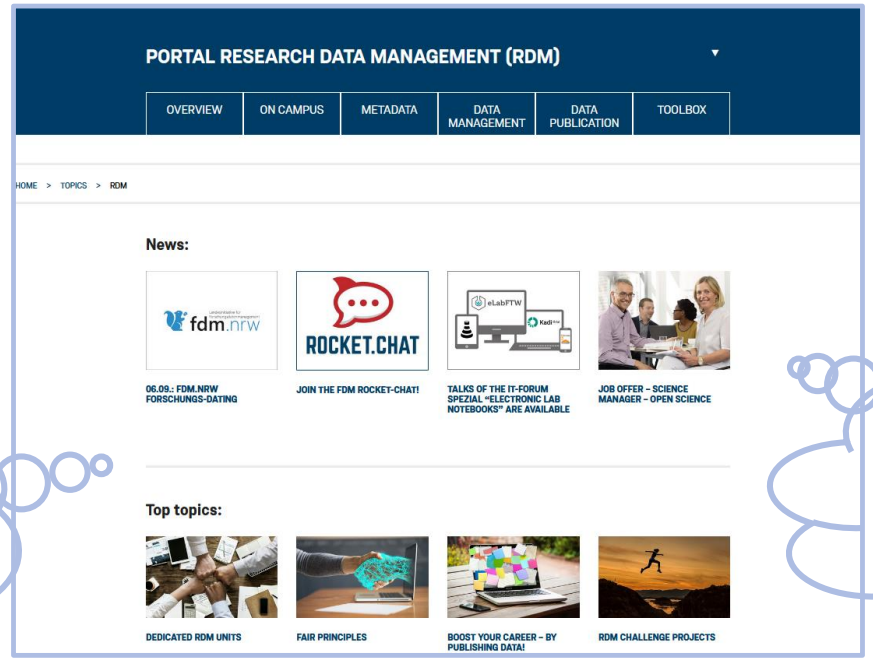
RDM Portal

Where can I find information?

I‘m interested in trainings.

I have a question. Who can I ask?

I‘m looking for tools?



https://intranet.fz-juelich.de/en/topics/rdm

JÜLICH Forschungszentrum

TEAM RESEARCH DATA OF THE CENTRAL LIBRARY

Contact


RESEARCH DATA MANAGEMENT

Research Data Management

The Central Library (ZB) at Forschungszentrum Jülich offers practical support relating to research data management (RDM).

Research data are a foundation of scientific work. The existing diversity of such data matches the diversity of different scientific disciplines, interests, and research procedures. Due to the fast-paced digital transformation and the increasing abundance of data, the handling and usage of research data is changing dramatically. In addition, large research funding bodies increasingly require thorough and transparent research data management that is publicly available, without exceptions. One of the main reasons is that research data are created to a large extent by means of public funding.

Research data include measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections, and samples that are created, developed, or analysed as part of scientific work. They may also include methodological test procedures such as questionnaires, software, models, and simulations, since these can also represent central results of scientific research. Research data occur in every scientific discipline, and – depending on the stage in the life cycle of the data – in various media types and formats, as well as aggregation and quality levels.

Guidelines for Handling Research Data

The Guidelines for Handling Research Data at Forschungszentrum Jülich entered into force on 1 May 2019.

Guidelines for Handling Research Data (PDF, 103 KB)

ZB flyer about research data management at Forschungszentrum Jülich

The flyer was presented for the first time on 25 October 2019 at the information event on how to publish research data and software as part of International Open Access Week. It is available in German and English in an A2 format from the Central Library and is also available to download.

Flyer – Forschungsdaten managen (PDF, 1 MB)

Flyer – Managing research data (PDF, 1 MB)

Get in touch with us:

forschungsdaten@fz-juelich.de

Website (intranet):

Research Data Management

RDM Portal **(intranet)**

RocketChat channel:

#fdm

THANK YOU

forschungsdaten@fz-juelich.de