

Welcome

Thilo Paul-Stüve, Britta Petersen

Central Research Data Management of Kiel University

Disclaimer: Please note that you are leaving the CAU pages once you open this presentation in your browser. This presentation includes links to other third party websites and services. When you click on these links you will leave this presentation and will be redirected to the respective another websites. These sites are not under our control.

RDM@CAU is not responsible for the content of linked third party websites. Please be aware that the security and privacy policies on these sites may be different than CAU policies, so please read third party privacy and security policies closely.

To see this document as an interactive LiaScript rendered version, click on the following link/badge:

[course](#) on LiaScript

If you need help, feel free to ask us any questions:

fdm@rz.uni-kiel.de



This work is licensed under a [Creative Commons Attribution 4.0 International License](#) with exception of the used material from other copyright holders.

Please be nice!

Some rules for today:

- Please mute your microphone when you do not have the floor
- Please hear each other out and let each other finish
- Draw attention to yourselves when you want to say something
- Please help each other
- Please do not do anything on the side
- Please ask if you have not understood something
- Please contribute actively
- Please allow mistakes → positive culture of mistakes.

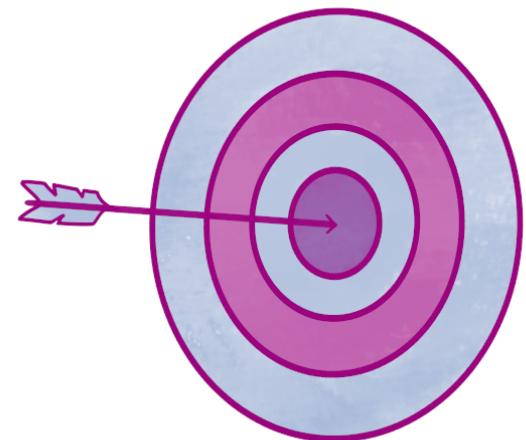


Source: Pixabay

Goals today

At the end of the workshop you...

- have a basic idea of the general concept of RDM and know some important related terms.
- can describe what research data is and know the research data lifecycle.
- can explain the importance of documentation and know what metadata is.
- can identify and assess data formats.
- can distinguish back up and long-term storage.
- can describe the FAIR-principles.
- can describe what a DMP is.
- know the requirements of funders and RDM related CAU contacts.
- had some time to exchange with peers.
- hopefully also had some fun!



Source: Cleo Michelsen

Agenda

Let us have a look at our workload for today:

- Expectations
- Research data and research data management
- Research data lifecycle
- Data organisation
- Documentation & metadata
- Data formats

LUNCH BREAK

- Back up & long-term storage
- FAIR principles and Open X
- Data publication
- Data management plan (DMP)
- RDM related organisations & funder requirements
- RDM @ CAU



Source: Pixabay

Warm up!

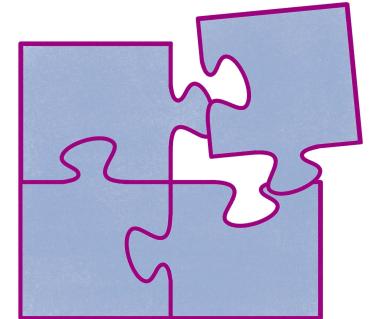
Let us play a game...

Hide your camera (use a sticker or your finger).

I will read statements to you.

Each time you can agree with the statement show yourself and waive.

That 's it !



I like to drink coffee in the morning.

I am working in the field of nature science.

If I have to decide to go to the cinema or to a concert, I decide for the concert.

I am working in a field of the humanities.

I know the FAIR data principles.

I am working in a field of engeneering

I have an ORCID.

I have a pet (or more than one).

I am using open data for my research work.

I am working with personal data.

I am teaching students next to my PhD.

I am writing code for my PhD.

I was sent to this workshop by my supervisor and really don´t know what am I supposed to do here.

Expectations

We would like to know a little bit more about your expectations and your fears regarding this workshop.

Group work:

Collect your ***hopes/wishes*** and your ***fears*** regarding this workshop, using oncoo cards.

<https://www.oncoo.de/0fou>



Review Expectations

Please discuss in your group (1) what are your hopes/wishes for this workshop and (2) what are your fears

Diese Abfrage ist unter folgendem Code erreichbar:

0fou

Bitte notieren Sie sich den Code, um später darauf zugreifen zu können. Die Schülerinnen und Schüler erreichen die Kartenabfrage unter folgender Adresse:

<https://oncoo.de/0fou>



Research data and research data management

Which aspects belong to the topic "research data management?"

Today you are supposed to learn something about research data management. What do you think: Which aspects belong to the topic?

Let's collect together!

Enter words that you associate with the term "research data management".

You may enter as many words as you like: <https://answergarden.ch/3005403>



Which aspects belong to the topic "research data management?"

Type your answer here...

Submit

40 characters remaining



preserve data after
research projects

proprietary vs. open
data/software

versioning policy

research team

cleaning

accessing data

tools for storing data
safely

publications

open standards

saving

analysing

scientific publication

organizing

reserving and sharing

data collection

organising data

storage

archiving

regulations

sharing

long-term linking

data structure

titles

format

findability

format data to be shared

fair

access

acccess

titles

data formats

data visualization

finding

organisation

collecting

clustering

format

publication

saving personal vs result
data

data organization

planning

data format

collection

data formats

data visualization

finding organisation

clustering

format

publication

We use cookies to make this website function properly. [Cookie policy](#)

Got it!

Research Data Management

What is research data management?

‘Research data management is an explicit process covering the creation and stewardship of research materials to enable their use for as long as they retain value.’

[DCC Glossary](#)

‘Research Data Management (RDM) is the methodical handling of the information produced or re-used during the course of academic research.’

[University of Edinburgh Research Data Service](#)

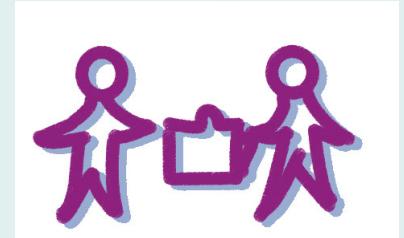
What do you think is the benefit of good research data management?

Think-Pair-Share

You will have some minutes with your partner. Please discuss!

- What might be the benefit of research data management?

Share your thoughts with us when you are back!



The benefits of good research data management are...

- Reproducibility of results
- Findability of the data
- Traceability of research
- Transparency of research
- Quality assurance and quality attribute
- Efficiency in the research process
- Sustainability
- Knowledge generation
- No redundancy
- Structuring, organisation
- Enable collaborations
- ...
- ...

Research Data

Group work:

Find out with what kind of data your group members are working with.

Together, add everything that you think can be research data to your collection.

Please use the following pads to take notes during break out session:

https://zumpad.zum.de/p/2023-01-26_research-data_G1

https://zumpad.zum.de/p/2023-01-26_research-data_G2

https://zumpad.zum.de/p/2023-01-26_research-data_G3

https://zumpad.zum.de/p/2023-01-26_research-data_G4



What are research data?

'In short data means whatever is necessary to validate or reproduce your research findings, or to gain a richer understanding of them.'

[University of Edinburgh Research Data Service](#)

'Any information you use in your research.'

[University of Cambridge PrePARE Project](#)

'All researchers work with data and/or datasets, but what you call data will depend on your discipline. As a humanities scholar you might talk about your primary sources or texts. If your research is in a social science, you may think in terms of survey results, interviews and statistics. You will probably have different terms again for the outputs of your experiments and observations if you are a scientist.'

[Monash University Library Data Management Brochure](#)

Examples for Research Data

- Audio and video recordings
- Diaries
- Geographic information system (GIS) data
- Laboratory and field notebooks
- Model, script and research software code
- Pictures and figures
- Questionnaires and codebooks
- Samples and artifacts
- Sensor data
- Sequence data
- Spectra
- Text and spreadsheet documents
- Text corpora and annotations
- Topography data
- Transcripts



Research data lifecycle

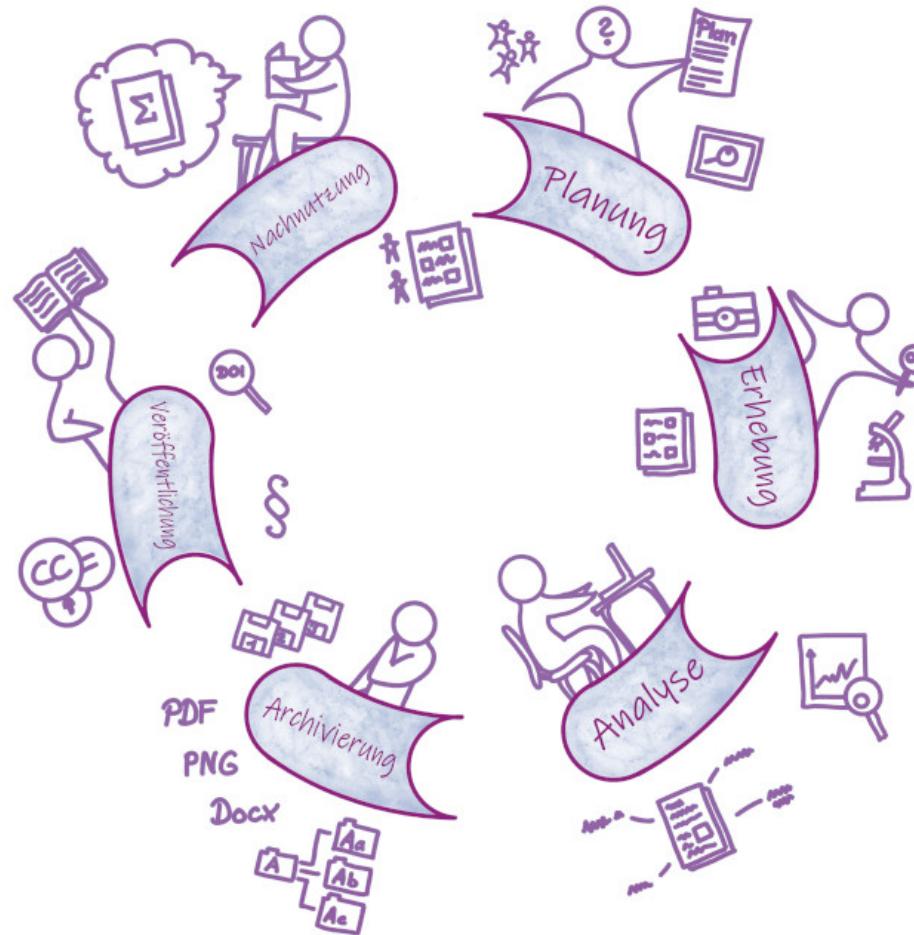


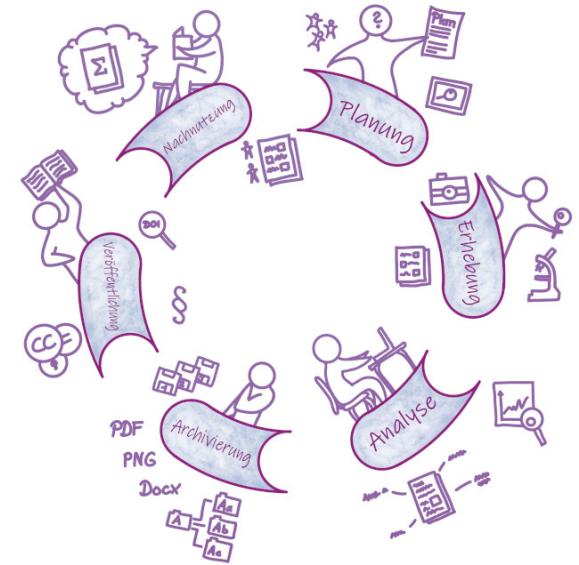
Illustration: Cleo Michelsen, based on UK Data Service

Individual work:

Think about your own PhD project and add keywords to the stations of the research data lifecycle that describe what steps and procedures at each station are relevant to your research data.

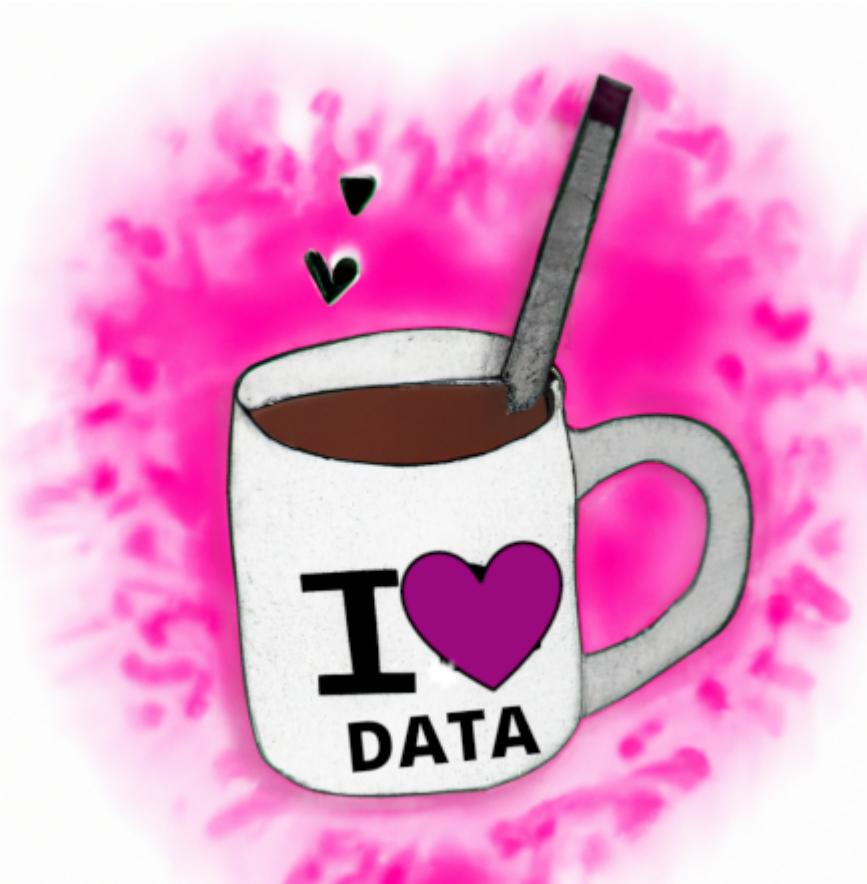
Find your workspace here: <https://miro.com/app/board/uXjVPvwAFbI=/>

- Does this research data lifecycle fit to your research project?
- Are there any deviations?



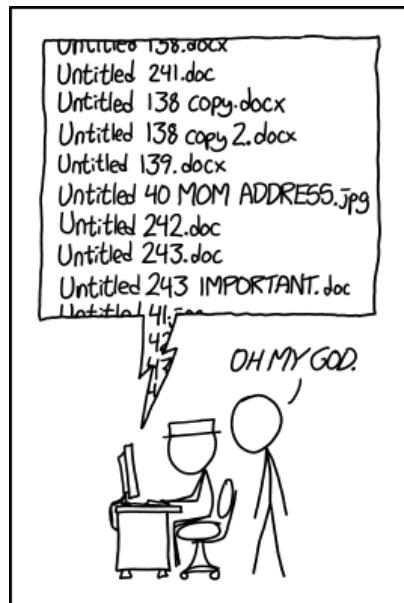
BREAK

Have a break!



Data organisation

It may seem trivial, but structured folder and file naming is a first step in research data management!



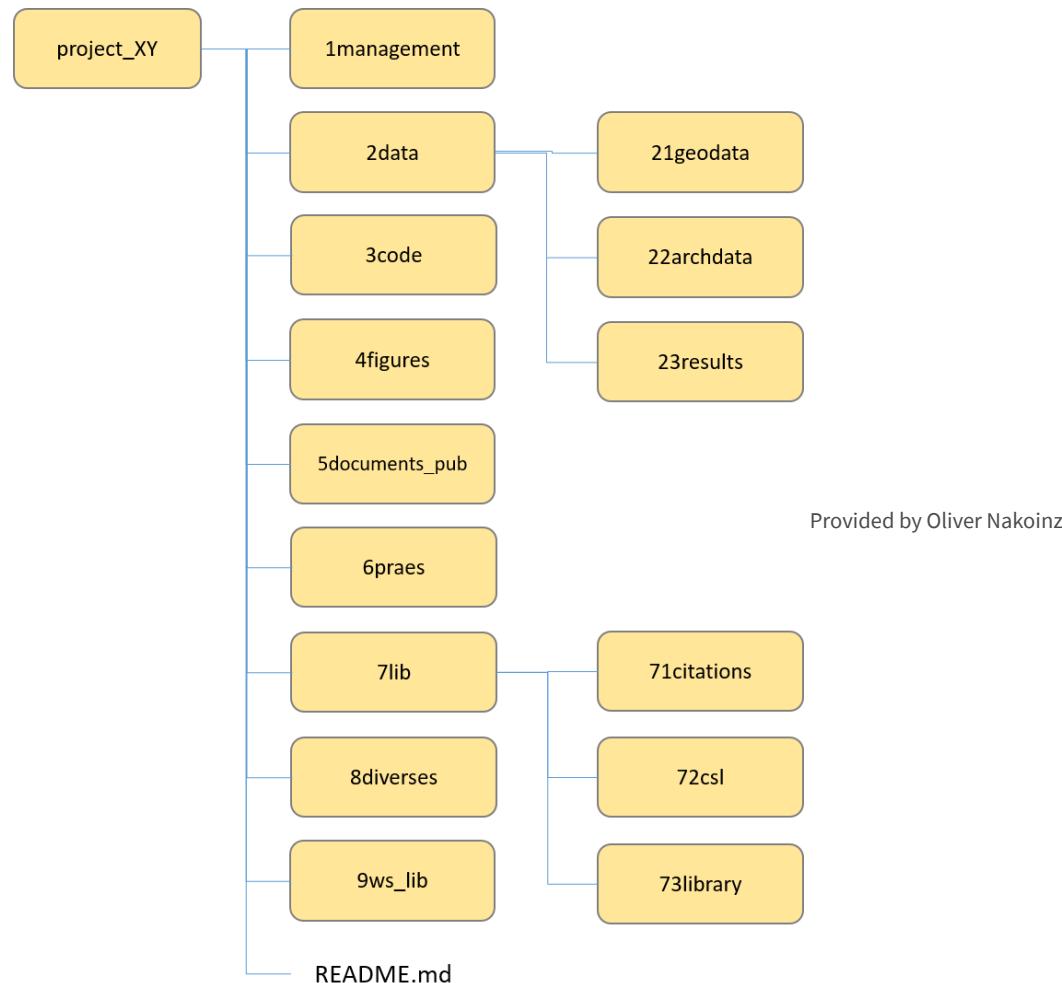
<https://xkcd.com/1459>. Shared under CC-BY-NC License

Organize your Files

- Try to find meaningful names → no “fantasy names”
- Use a uniform scheme and a logical structure
 - Folder structure and file names
 - Hierarchic order with first things first
 - Think about your preferences in *sorting!*
- Follow [ISO 8601](#) for dates and times
 - Date & time, e.g., YYYYMMDDThmmss±hhmm
 - Date, e.g., YYYY-MM-DD
- Versions? Use the [semantic versioning scheme](#) (Major.Minor.Patch), e.g., 2.0.0
- Avoid of blanks and special characters “”
- *Document* your naming conventions and abbreviations used
 - Readme.md

Examples for Naming Conventions

Example for a folder hierarchy



Example for a file name following a naming convention

[Project name]_[Approach]_[Location]_[Person-ID]_[Date].[Format-Suffix]

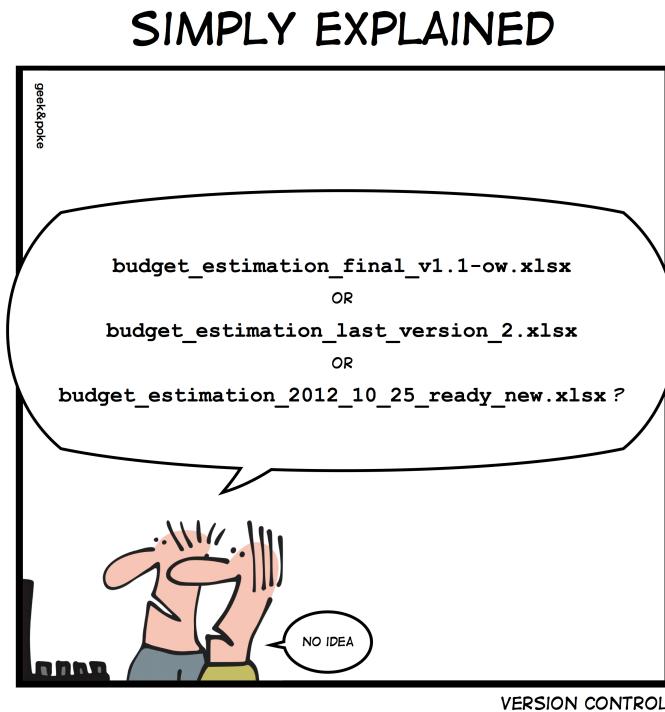
Rebel-Hunting_Interview_DS-1-Orbital-Battle-Station_Organa_1976-05-25.mp4

Which one is a good name?

Which examples are following a good naming convention?

- Olga_170413_probe17k
- Naturepaper-marc-mira-ready!
- Vm4520132Schmidt.pdf
- Kristall_765_spektr_2016-12-03.csv
- 170413_probe17k.olga
- Naturepaper+karl+britta+james&nal
- Olga170413probe17k
- Krst_765_spkt_161203
- Naturepaper+karl+britta+james ready!_revised
- 012_maus_mrna_20200912
- Nature_karl&jan_endendversion
- 28q8QGlHKwrRw.pdf
- Tagung_Digitale_Wissenschaft.pdf
- 647749157.pdf

Version Control



Idea from Jen Simmons and John Albin Wilkins during episode #40 of "Web Ahead" about Git:
<http://5by5.tv/webahead/40>

- Store versions in separate files
 - Semantic versioning (Major.Minor.Patch), e.g.,
 - **0.1.0** (a beta)
 - **1.0.0** (a release version)
 - **1.0.1** (a release with slight corrections)
 - can be included in file names
 - define what you consider to be a "release" or a "slight correction"
 - Document your versioning scheme and constantly document your changes
 - Readme.md
 - Version control table
 - Data dictionary
 - Working in a team? A lot of changes? Use a distributed version control system!
 - Git
 - GitLab, GitHub
-

What a version control table could look like

Versionsnr.	Changes	Date	changed by
1.0	Release	2016-11-2	KL
1.1	Erased spelling mistakes	2016-11-20	KL
1.2	Changed layout	2017-02-20	GN
2.0	Add new chapter (3.1.)	2017-02-20	GN

Example for Version Information within a document:



**Revised Report of the Research Data Alliance Data Versioning
Working Group**

**Principles and best practices in
data versioning for all datasets big and small**

Version 1.1

Jens Klump, Lesley Wyborn, Mingfang Wu, Robert Downs, Ari Asmi, Gerry Ryder, Julia Martin

Version Information		
Release	Date	Description
Version 1.0	2020-01-16	Original submission Klump, J., Wyborn, L., Downs, R., Asmi, A., Wu, M., Ryder, G., & Martin, J. (2020). Principles and best practices in data versioning for all data sets big and small. Version 1.0. <i>Research Data Alliance</i> . DOI: 10.15497/RDA00042 .
Community review: 2020-01-28 until 2020-02-28		
Version 1.1	2020-04-06	Updated with minor changes following community review Recommended Citation: Klump, J., Wyborn, L., Downs, R., Asmi, A., Wu, M., Ryder, G., & Martin, J. (2020). Principles and best practices in data versioning for all data sets big and small. Version 1.1. <i>Research Data Alliance</i> . DOI: 10.15497/RDA00042 .

Have a look here: [Principles and best practices in data versioning for all data sets big and small](#)

Example for a documented versioning scheme:

The screenshot shows a Zenodo record page for a document titled "OstData Publizierte Forschungsdaten versionieren".

Document Preview: A thumbnail image of the document's first page, which features the "OSTDATA" logo and the title "Publizierte Forschungsdaten versionieren". Below the title is a short explanatory text about the versioning scheme.

Metadata:

- Publication date:** February 14, 2022
- DOI:** [10.5281/zenodo.6076538](https://doi.org/10.5281/zenodo.6076538)
- Keyword(s):** Forschungsdatenmanagement, Versionierung, Forschungsdaten
- Subject(s):** Forschungsdaten
- Published in:** Materialien zum Forschungsdatenmanagement in der Ost-, Ostsüd- und Südosteuropaforschung.
- Communities:** OstData – Research Data Service for Central,

Statistics: 202 views, 137 downloads. [See more details...](#)

Indexed in: OpenAIRE

Have a look here: [Zenodo](#)

Develop a naming convention

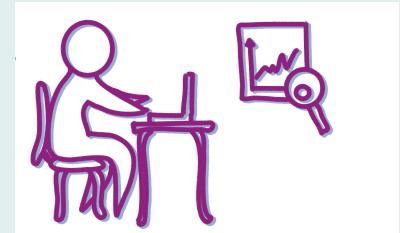
Individual work or group work for people working on the same project

What would be a good folder structure and a good file naming convention for the files related to your PhD project?

Please document

1. to which files your folder structure and your naming convention applies
2. a scratch of a folder structure
3. the descriptive aspects to be used in file names and their order
4. the abbreviations used
5. the versioning scheme, if applicable

Create an own zumpad or use your working space on the miro-Board to take notes.



Data documentation

Group work:

You are working in a research group working on the ecology of forests and take over data from a previous project:
[average_d.xlsx](#)

- Speculate what kind of data it could be.

Discuss and take notes

- Apart from the data itself, what information do you need to be able to work with a dataset?
- What do you notice in regard of data quality?



A good data documentation should include

- Information on the collection of data
 - Methods, units, time periods, locations, technique used, etc.
 - Structure of the data and their mutual relationships
 - Explanation of variables, labels and codes
 - Differences between different data set versions
 - Measures for data cleaning
 - Information on access and terms of use
 - Licensing
 - Ideal world
 - Description of the research undertaking
 - Goals
 - Hypotheses



Digitalbevaring.dk

Short research assignment:

Which guideline of the DFG Guidelines for Good Scientific Practice deals with *documentation*?

Guideline 12: Documentation, „Scientists and scholars shall document all information relevant to the achievement of a research result as comprehensibly as it is necessary and appropriate in the concerned scientific field in order to be able to verify and evaluate the result. [...]“

Deutsche Forschungsgemeinschaft. (2019). Leitlinien zur guten wissenschaftlichen Praxis. Kodex. <http://doi.org/10.5281/zenodo.3923602>, S. 17.

Understanding Data

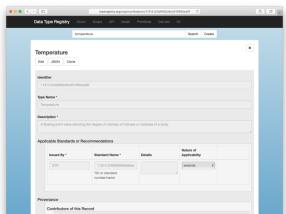
> Date: 364.07{{2-3}}*****



The DOI® System ISO

26324

*****{{3-4}}*****



Temperature in Kelvin 364,07 K

≈ 42,6°C

Data about Data

- Identifier: 10.1594/dwd-weather-data
- Identifier Type: DOI
- Unit: K
- Data Type Identifier: 11314.3/0a9062a9cb51995dea9f
- Date: 2019-07-25T15:00:00Z
- Location: 52.5178687 7.3057642
- Creator: Deutscher Wetterdienst
- ROR: 02nrqs528

Origin, Location and Meaning of Data

- Identifier: 10.1594/dwd-weather-data
- Identifier Type: DOI

*****{{4-5}}*****



- Unit: K
- Data Type Identifier: 11314.3/0a9062a9cb51995dea9f
- Date: 2019-07-25T15:00:00Z
- Location: 52.5178687 7.3057642
- Creator: Deutscher Wetterdienst
- ROR: 02nrqs528

Origin, Location and Meaning of Data

- Identifier: 10.1594/dwd-weather-data
- Identifier Type: DOI

- Unit: K
- Data Type Identifier: 11314.3/0a9062a9cb51995dea9f

- Date: 2019-07-25T15:00:00Z
- Location: 52.5178687 7.3057642
- Creator: Deutscher Wetterdienst
- ROR: 02nrqs528

Origin, Location and Meaning of Data



- Identifier: 10.1594/dwd-weather-data
- Identifier Type: DOI
- Unit: K
- Data Type Identifier: 11314.3/0a9062a9cb51995dea9f

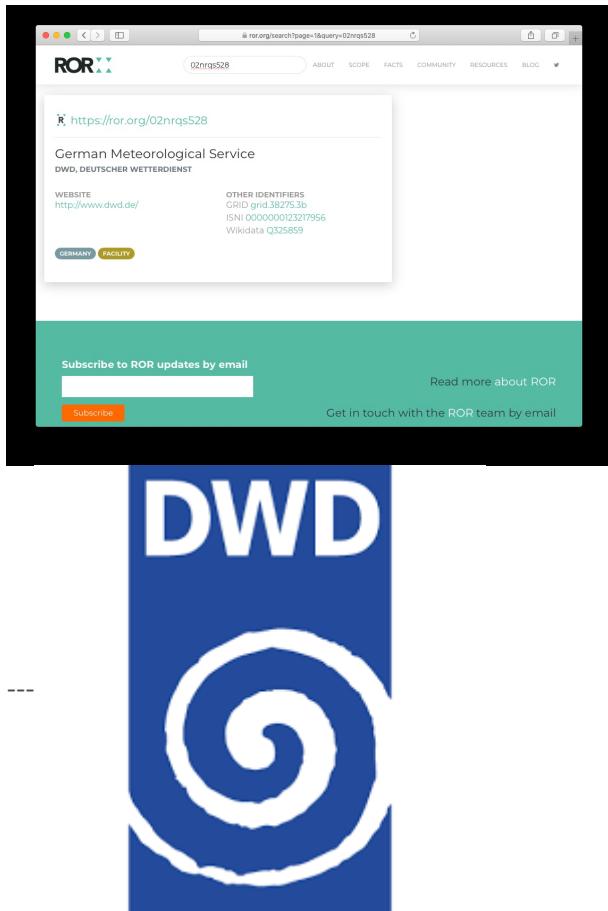
- Date: 2019-07-25T15:00:00Z

- Location: 52.5178687 7.3057642
- Creator: Deutscher Wetterdienst
- ROR: 02nrqs528

Origin, Location and Meaning of Data

- Identifier: 10.1594/dwd-weather-data
- Identifier Type: DOI
- Unit: K
- Data Type Identifier: 11314.3/0a9062a9cb51995dea9f
- Date: 2019-07-25T15:00:00Z

- Location: 52.5178687 7.3057642



- Creator: Deutscher Wetterdienst

- ROR: 02nrqs528

Origin, Location and Meaning of Data

- Identifier: 10.1594/dwd-weather-data
- Identifier Type: DOI
- Unit: K
- Data Type Identifier: 11314.3/0a9062a9cb51995dea9f
- Date: 2019-07-25T15:00:00Z
- Location: 52.5178687 7.3057642

- Creator: Deutscher Wetterdienst
- ROR: 02nrqs528

Origin, Location and Meaning of Data

- Identifier: 10.1594/dwd-weather-data
- Identifier Type: DOI
- Unit: K
- Data Type Identifier: 11314.3/0a9062a9cb51995dea9f
- Date: 2019-07-25T15:00:00Z
- Location: 52.5178687 7.3057642
- Creator: Deutscher Wetterdienst
- ROR: 02nrqs528

- **Description:** Air temperature measurement at the weather station Lingen, Germany, on 29 July 2019 in Kelvin

Traceability

Nordwest-Zeitung

Jobs Immo Auto Kleinanzeigen Trauer Hochzeit Guide Shop Tickets nordbuzz Fußball Werben Kontakt

NWZonline durchsuchen Wetter in Oldenburg

Benutzername Passwort Login ?

PLUS Nachrichten Region Sport Blogs Ratgeber Videos

Fotos Termine Märkte Prospekte ePaper Archiv Abo

NWZONLINE.DE - NACHRICHTEN - PANORAMA - BLAULICHT -
MESSWERT IN LINGEN: 42,6 GRAD - DEUTSCHLANDS HITZEREKORD IM EMSLAND GEBROCHEN 26.07.2019

MESSWERT IN LINGEN

42,6 Grad – Deutschlands Hitzerekord im Emsland gebrochen



Metadata

- Data about data
- Administrative data
 - Information on the management of the data
 - Mostly generic
- Subject data
 - Individual aspects or data sets in more detail
 - Structured with respect to the research discipline
- Generic standards
 - [DataCite Metadata Schema](#)
 - [Dublin Core Metadata Initiative](#)
- Discipline-specific standards
 - [Metadata Standards Directory](#)



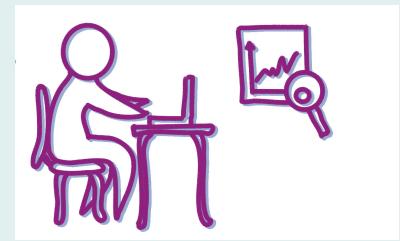
Digitalbevaring.dk

Individual work:

Go to <http://rd-alliance.github.io/metadata-directory/subjects/>

Search for a meta data standard appropriate to your research subject!

There is no standard? What can you do?



File formats

Partner work:

Collect all file formats you are working with and post them to the miro-Board.

Give some explanation on your file formats to your partner.

- What kind of data hides behind the format?
- What kind of software is needed to read it?
- Anything special to consider?

<https://miro.com/app/board/uXjVPvwAFbl=/?moveToWidget=3458764543971666820&cot=14>



Choosing file formats

- Non-Proprietary, unencrypted, uncompressed and commonly used
- Open-standard-compliant, documented and royalty-free

Data Type	Recommended	Trade-off Matter	Not Recommended
Tabular	CSV, TSV, ODS	XLSX, SPSS portable	XLS, SPSS
Textual	TXT, MD, HTML, ODT	DOCX, RTF, PDF/A	DOC, PDF, PS
Presentation	ODP, HTML	PPTX	PPT
video	MP4, MKV, OGG	WEBM	WMV, MOV, QT, Flash
Audio	MP4, FLAC, WAV, OGG	MP3, AIF	
Image	TIFF, PNG	BMP, JPG	PSD, GIF
Vector	SVG		AI
Generic	XML, JSON, RDF		
Container	Bagit, Frictionless, Data Package	ZIP, TAR	

BREAK

Have a break!



Back up & long-term storage

Where do you store your data?

Recommendations for your back up

- At least 3 copies of a file
- On at least 2 different media
- At least one of which is remote
- Test data recovery at the beginning and at regular intervals.

How do you store your (sensitive) data?

Protect your (sensitive) data:

- Hardware (e.g. separate lockable room).
- File encryption
- Password security
- At least two people should have access to your data



Back up vs. long-term storage

Back up	Long-term storage
Automatic backup of all data	Storage of only selected data
All versions	Final version only
to prevent data loss (technical, e.g. defective, or human, e.g. accidentally deleted)	Integrity backup
	Long-term storage
	Searchability

FAIR Data Principles

An important goal of structured research data management is to keep data accessible, reusable and verifiable in the long term and independent of individuals.



Illustration: Patrick Hochstenbach in Engelhardt, Claudia et. al. (2021)

Findable

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Reusable

R1. Meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

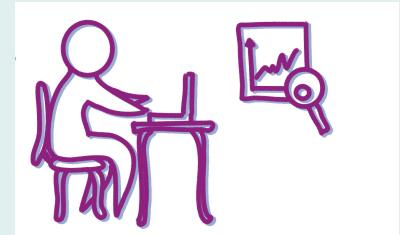
R1.3. (Meta)data meet domain-relevant community standards

How FAIR is your data?

Individual work or group work for people working on the same project

Think about your own project and evaluate the handling of your research data against the FAIR principles.

- Which of the FAIR principles do you already comply with?
- Where do you (or all of us) need to improve?
- What do you need in order to comply with the FAIR principles?



Open X

Open Access

- Distribution of research output online, **freely available** for everyone
 - [Panton Principles](#): **no restrictions** for use und republication
 - In practice: **attribution** to the original author or creator
 - Attribution stacking
- Formerly only open publication of journal articles, papers, etc.
- Part of **Open Science**
 - Like Open Educational Resources, Open Source Research Software, Open Methodology, Citizen Science
- Open Data
 - Publication of data without legal or technical restrictions
 - **Indispensable** for data science
 - Generic **non-proprietary** formats
 - **Metadata**
 - **FAIR?**

Data publication

How to publish and share your data?

Supplement to a peer-reviewed article ("enhanced publication")

- as a supplement to the associated article
- as a data set in a repository with a link to the corresponding article.

Example:

Citation:

Leu, Eva; Brown, Thomas A; Graeve, Martin; Wiktor, Jozef M; Hoppe, Clara Jule Marie; Chierici, Melissa; Fransson, Agneta; Verbiest, Sander; Kvernvik, Ane C; Greenacre, Michael J (2020): Fatty acid composition, ice chemistry and algae biomass during an Arctic sea ice algal bloom on landfast ice near the settlement of Svea, in Van Mijenfjorden, Spitsbergen. PANGAEA, <https://doi.pangaea.de/10.1594/PANGAEA.925015> (dataset in review)

Supplement to:

Leu, Eva; Brown, Thomas A; Graeve, Martin; Wiktor, Jozef M; Hoppe, Clara Jule Marie; Chierici, Melissa; Fransson, Agneta; Verbiest, Sander; Kvernvik, Ane C; Greenacre, Michael J (2020): Spatial and Temporal Variability of Ice Algal Trophic Markers - With Recommendations about Their Application. *Journal of Marine Science and Engineering*, 8(9), 676,  <https://doi.org/10.3390/jmse8090676> 

Independent information object in a research data repository

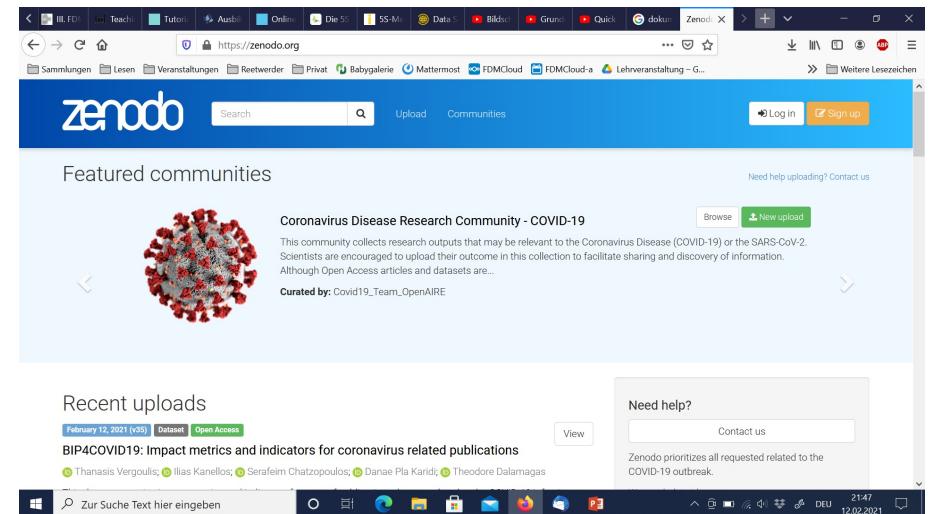
- disziplinspezifische Repositorien, z. B. Datorium, Pangaea
- disziplinübergreifende Repositorien, z. B. ZENODO
- institutionelle Repositorien, z. B. Refubium, edoc

Example:



The screenshot shows the homepage of the PANGAEA Data Publisher. At the top, there's a navigation bar with links like "Sammeln", "Lesen", "Veranstaltungen", "Reetwerder", "Privat", "Babygalerie", "Mattermost", "FDMCloud", "FDMCloud-a", and "Lehrveranstaltung - G...". Below the bar, the PANGAEA logo is displayed with the text "Data Publisher for Earth & Environmental Science". A "Not logged in" link is visible. The main content area features a "Submit Data" button and a "TOPICS" section with categories: CHEMISTRY (199049), LITHOSPHERE (130304), OCEANS (99410), BIOLOGICAL CLASSIFICATION, and ATMOSPHERE (28951). A search bar at the top right says "Search for measurement type, author name, project, taxa,...". The bottom of the page includes a search bar "Zur Suche Text hier eingeben" and a system status bar showing "DEU 12.02.2021 21:49".

Source: <https://www.pangaea.de/> Zugriff 10.02.2021



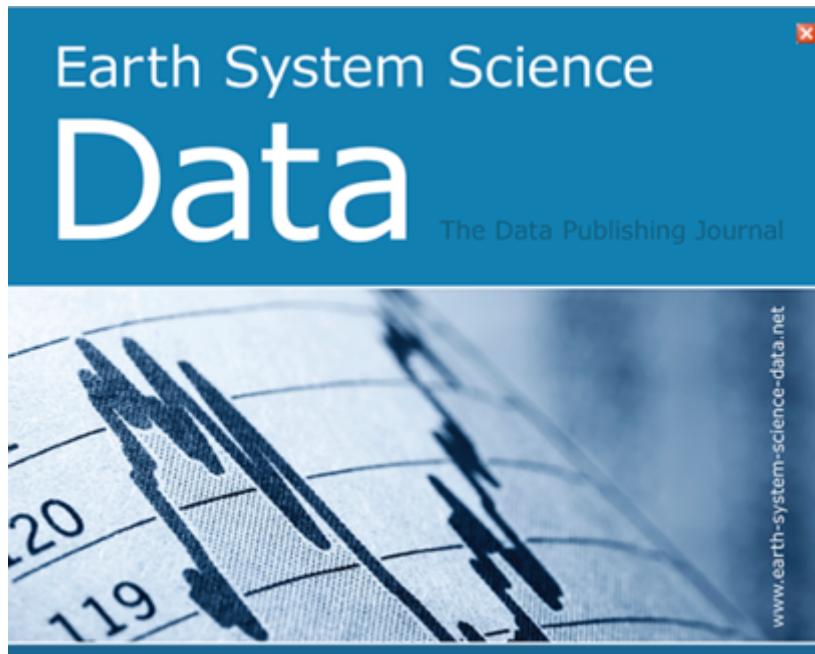
The screenshot shows the homepage of the ZENODO platform. At the top, there's a navigation bar with links like "Sammeln", "Lesen", "Veranstaltungen", "Reetwerder", "Privat", "Babygalerie", "Mattermost", "FDMCloud", "FDMCloud-a", and "Lehrveranstaltung - G...". Below the bar, the ZENODO logo is displayed with "Search", "Upload", and "Communities" buttons. A "Log in" and "Sign up" button are also present. The main content area features a "Featured communities" section with a thumbnail of the COVID-19 virus and the text "Coronavirus Disease Research Community - COVID-19". It includes a "Browse" button, a "New upload" button, and a note about the collection being open access. Below this, there's a "Recent uploads" section with a thumbnail for a dataset titled "BIP4COVID19: Impact metrics and indicators for coronavirus related publications" and a "View" button. A "Need help?" contact form is on the right. The bottom of the page includes a search bar "Zur Suche Text hier eingeben" and a system status bar showing "DEU 12.02.2021 21:47".

Source: <https://zenodo.org/> Zugriff 10.02.2021

Data journals

- publish detailed description of data
- partly peer-reviewed

Example:



Source:

<https://www.earth-system-science-data.net>, Zugriff 10.02.2021



ISSN: 2352-3409

Data in Brief

> Open Access

Editors-in-Chief: Hao-Ran Wang, G

> View Editorial Board

Data in Brief provides a way for resear
other's datasets by publishing data ar

- Thoroughly describe your data, faci
- Make your data, which is often bur
- Increase traffic towards associated..

[Read more](#)



Source: <https://www.journals.elsevier.com/data-in-brief>, Zugriff 10.02.2021

Repositories

What is a repository?

*"A repository (Latin *repositorium*, 'storehouse') is a managed place for storing ordered documents that are accessible to the public or to a restricted group of users. An archive (Latin *archivum*, file cabinet'), on the other hand, manages only historical documents."*

"Digital research data repositories are information infrastructures that store and organize digital research data...as permanently as possible...to ensure the discoverability and accessibility of the data..."

Source: Esther Asef, Katarzyna Biernacka, Elisabeth Böker, Sarah Ann Danker, Juliane Jacob, Janna Neumann, Britta Petersen, Jessica Rex und Ute Trautwein-Brunns (2021): Data Sharing interaktiv vermitteln

How to find a repository

re3data.org

- Collection of repositories
- Worldwide
- Various disciplines
- Researchers, funders, publishers and institutions

re3data.org

Search Browse ▾ Suggest Resources ▾ Contact DataCite

social Search Toggle short help

← Previous 1 2 3 4 5 6 7 ... 28 Next → Sort by ▾

Found 677 result(s)

TESS
Time-sharing Experiments for the Social Sciences
Subject(s) Social Sciences Social and Behavioural Sciences Humanities and Social Sciences
Content type(s) Standard office documents Plain text
Country United States
Time-sharing Experiments for the Social Sciences (TESS) offers researchers the opportunity to capture the internal validity of experiments while also realizing the benefits of working with a large, diverse population of research participants.

Social Science Data Archive - UCLA
SSDA
Subject(s) Social and Behavioural Sciences Humanities and Social Sciences
Content type(s) Raw data Scientific and statistical data formats Structured text Archived data Audiovisual data other
Country United States
The Social Science Data Archives maintains a collection of machine-readable survey, census, and administrative data files and provides access to publicly available data. A portion of the collection is focused on Los Angeles from a demographic, economic, social, or political viewpoint. Los Angeles-specific data may also be extracted from files in the collection that contain data from studies covering demographic areas larger than Los Angeles.

European Social Survey
ESS Data
Subject(s) Social and Behavioural Sciences Education Sciences Empirical Social Research Criminology Public Health, Health Services Research, Social Medicine Geography Humanities and Social Sciences Social Sciences Jurisprudence Medicine Medicine Life Sciences Geosciences (including Geography) Natural Sciences

Source: re3data About. <http://service.re3data.org/about>. Zugriff 10.02.2021

risources.dfg.de

- Offer of the DFG
- Information portal
- Germany-wide
- Research Infrastructures
- For researchers

The screenshot shows the homepage of the RIsources portal. At the top, there is a navigation bar with links to 'DFG Homepage', 'Impressum / Kontakt', 'Hilfe', and 'English'. The DFG logo and the RIsources logo ('The Research Infrastructure Portal funded by DFG') are also present. Below the navigation bar, there is a blue header bar with tabs for 'Leitseite', 'Katalog', 'Suche', and 'Über RIsources'. The main content area is titled 'Katalog' and contains a section titled 'Auswahl' with three columns: 'Wissenschaftsgebiet', 'Kategorie', and 'Bundesland'. The 'Wissenschaftsgebiet' column lists various academic fields with counts: Agrar-, Forstwissenschaften, Gartenbau und Tiermedizin (12), Bauwesen und Architektur (4), Biologie (20), Chemie (16), Elektrotechnik, Informatik und Systemtechnik (4), Geisteswissenschaften (30), Geowissenschaften (einschl. Geographie) (34), and Maschinenbau und Produktionstechnik (0). The 'Kategorie' column lists research categories with counts: Fachinformationsdienste (FID) (0), Forschungsarchive (0), Forschungsbibliotheken (0), Forschungsdaten-Repositorien (90), Forschungsschiffe und Unterwasser-Boote (0), Genomics-, Transcriptomics-, Proteomics-, Metabolomics-Einrichtungen (0), Hochleistungs-Rechner, Rechenzentren (0), In situ Meeres- und Süßwasserobservatorien (0), and Mineralien- und Rohstoffobservatorien (0). The 'Bundesland' column lists German states with counts: Baden-Württemberg (23), Bayern (11), Berlin (13), Brandenburg (1), Bremen (3), Hamburg (3), Hessen (6), Mecklenburg-Vorpommern (1), NRW (0), and Niedersachsen (6).

Source: http://risources.dfg.de/index.html#q=*&sort=RI_SORT_DE%20asc&rows=10&RI_EXT=Y

Zugriff 10.02.2021

Individual work Search re3data.org for a repository that is appropriate for your data.

Let us discuss your search and selection.



Licences

- Licenses regulate conditions of subsequent use of published data.
- Free licenses allow the use, redistribution and modification of copyrighted works
 - are usually available for free use and only need to be linked to
 - Prerequisite is that you are the copyright holder

Selection of the license depends on the type of data:

- e.g. Creative Commons (CC) licenses for articles, monographs, images, etc.
 - Open-Database-License (ODbL) for DB or CC starting with version 4
 - General Public Licence (GNU) for software
- If no license is granted, the stricter copyright applies, as far as applicable to data



CC0 (Public Domain)



CC BY (Namensnennung)



CC BY-SA (Namensnennung - Weitergabe unter gleichen Bedingungen)



CC BY-ND (Namensnennung - Keine Bearbeitung)



CC BY-NC (Namensnennung - Nicht kommerziell)



CC BY-NC-SA (Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen)



CC BY-NC-ND (Namensnennung - Nicht-kommerziell - Keine Bearbeitung)

Data sets: Open Data Commons



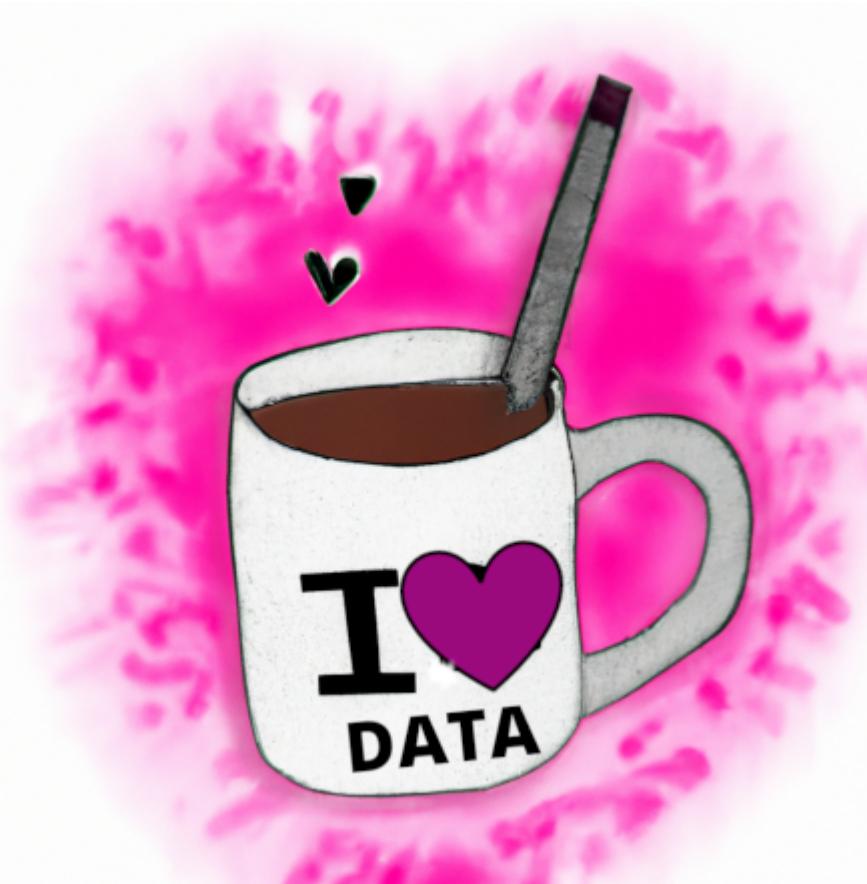
- Open Data Commons Public Domain Dedication and License (PDDL)
- Open Data Commons Attribution License (ODC-By)
- Open Data Commons Open Database License (ODbL)

Take care!

no license \neq free license

BREAK

Have a break!



Data management plan (DMP)

What is a data management plan?

- All information that adequately describes and documents the collection, processing, storage, archiving, and publication of research data in the context of a research project.
- "[...] analysis of the workflow from the generation of the data to their use."¹

[1] J. Ludwig, H. Enke (Hrsg.) Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Verlag Werner Hülsbusch: Glückstadt, 2013.

Components of a DMP

- Administrative information
 - Project name, data originator, other contributors, contact, funding program, etc.
- Project abstract
 - Data set descriptions
 - Data types, formats, scope
 - Metadata and standards information
 - Data sharing
 - Archiving and backup of data
 - Responsibilities
 - Legal and ethical aspects (e.g., licences, GDPR, Nagoya protocol, CARE)
 - Costs

Length can vary from a few paragraphs to several pages!

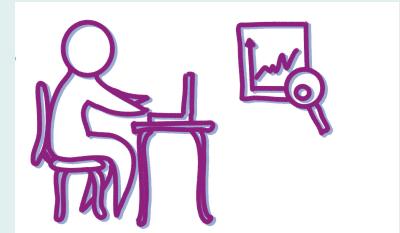
Sketching out a DMP

Individual work

Download the CAU template for data management plans: [CAU_DMP_Template](#)

Have look at the template and try to sketch out a DMP for your research project.

- What information do you already have?
- What information is missing to fill the template?



RDM related organisations & funder requirements

Research Data Alliance

Nationale Forschungsdateninfrastruktur (NFDI)

Deutsche Forschungsgemeinschaft (DFG)

Horizon 2020 & Horizon Europe

Research Data Alliance

- International organisation founded in 2012
 - **Vision:** Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society
 - **Mission:** RDA builds the social and technical bridges that enable open sharing of data
- **Bottom-up** development of practices, infrastructures, tools, technologies, services, approaches, policies, etc.
- **Practitioners** come together in [Birds of a Feather-Groups \(BoF\)](#), [Interest Groups \(IG\)](#) or [Working Groups \(WG\)](#)
- Regional chapters, e.g., [RDA Europe](#) or [RDA Deutschland e.V.](#)
- Strong **influence** on European Commission, BMBF, DFG, ...



Nationale Forschungsdateninfrastruktur (NFDI)

- National research data management initiative in Germany
- Initiated by the [German Council for Scientific Information Infrastructures](#)
- Horizontal linking of existing actors
 - Discipline-specific [NFDI consortia](#) with binding roadmaps
 - Bring into use existing infrastructure
 - Identify and fill gaps
- Interoperability of data and infrastructure
- Use of NFDI will probably get mandatory
- Participation in the work of NFDI consortia possible



Nationale
Forschungsdaten
Infrastruktur

- Code of Conduct: [Guidelines for Safeguarding Good Research Practice](#)
- Guideline 7 – Quality assurance
 - Disclosing of origin of data, organisms, materials and software used
 - Reuse of data is clearly indicated; original sources are cited
 - Description of nature and scope of research data generated
 - Handling of research data in accordance with requirements of relevant subject area
- Guideline 13 – Public access
 - Making available research data where possible and reasonable
 - Making available software programmed by researchers with source code
- FAIR Principles

Horizon 2020 & Horizon Europe

- Framework Programme for Research and Technological Development of the European Commission:
- [Horizon 2020 Online Manual – Data Management](#)
- [Horizon Europe – Programme Guide – Open Science](#)
- [Horizon Europe – Data Management Plan Template](#)
- Open access to research data is applicable by default
 - as open as possible, as closed as necessary
- Make research data findable, accessible, interoperable and re-usable (FAIR)
- DMP should include information on:
 - The handling of research data during & after the end of the project
 - What data will be collected, processed and/or generated
 - Which methodology & standards will be applied
 - Whether data will be shared/made open access and
 - How data will be curated & preserved (including after the end of the project)



RDM @ CAU

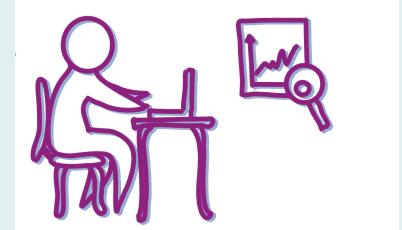
Individual work

You have 5 minutes to perform an internet search!

Are there any guidelines for handling research data at CAU?

Are there specific guidelines for handling research data at your institute?

Who are the contact persons at CAU and/or your institute if you have questions about handling your digital research data?



Kiel University's Recommendations on Handling Research Data

- **Research data** are the **basis and result** of scientific work and are therefore of particular significance. It is important to handle research data responsibly to ensure that **scientific work is traceable, verifiable and reusable**
- Handling research data is **subject to constant change** due to **advances in research**
- Kiel University continuously assists with this ongoing process by implementing appropriate measures

Principle 1

- **Research data** are all data created in the process of research work through observation, collection, experiments, simulation, and further processing for the **purpose of producing and validating research results**
- **Research data management** covers **all areas of data administration**: data acquisition, data preservation, data security and data integrity, as well as the release and publication of data

Principle 2

- Research data management is performed in accordance with current professional standards while respecting the DFG (German Research Foundation) proposals for safeguarding good scientific practice
- Research data management is documented in a data management plan
- **Responsibility** for research data management lies with the head of the particular research activity.

Principle 3

- Research data is made accessible to the public in national or international subject-specific archives under appropriate licensing; if no such archives exist, central university services are used
- **Scientific and legal interests** of the researchers, the protection of the personal details of participants, patients and other individuals affected by the collected data, as well as other **obligations to third parties** – for example, cooperation partners – as well as ethical, legal and economic restrictions **must be observed**
- Kiel University assists researchers through its **central infrastructure facilities**
 - Acquisition and preservation of research data
 - Formulating and maintaining standards for handling research data
 - Appropriate training and advisory service offers created in cooperation with the specialist disciplines
- Kiel University recommends and assists with anchoring the significance of subject-specific, structured research data management in teaching by the particular subject areas

Take-Away Messages

> Practical Take-Away Messages

1. Document your data

- Use documented naming and versioning conventions
- document changes
- think about metadata necessary to understand your data

2. Formats

- Generic and open standard file formats last longer than proprietary file formats
 - Open Document Format (ODF)
 - Comma separated values (CSV)
 - Raw text files (TXT, MD)
 - Data container formats for exchange, archival and publication, e.g., [BagIt](#), [Frictionless Data](#)

3. Storage

- Central infrastructure with backup for storage
 - Desktop and laptop for work on current research data only
 - Systematic file and folder naming and hierarchy
 - Provide *Readme* files
 - Data Management Middleware for handling data and metadata, e.g., [iRODS](#)
 - DFG [Guidelines for Safeguarding Good Research Practice](#) require 10 years of preservation at least!

4. Publication

- Discipline-specific Repositories with specific metadata support
 - [re3data: Registry of Research Repositories](#)
 - National or international initiatives
 - NFDI (work in progress)
 - [European Open Science Cloud Services](#)
 - Institutional Data Repository: opendata@uni-kiel
 - Generic Repositories
 - [Zenodo](#)

5. Licensing

- [Creative Commons](#): data with a necessary creation height; ideally CC0 or CC BY
- [Open Data Commons](#): databases, raw data

Questions

Nearly done!

Time for open questions!



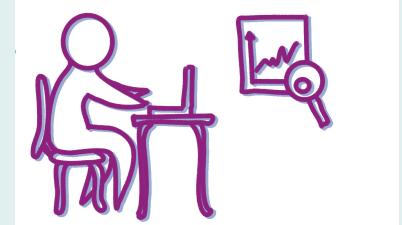
One Minute Paper

Individual work

Please take a piece of paper or create an own pad (e.g. <https://zumpad.zum.de/>).

You have one minute.

Please write down the most important points of our workshop today.



Feedback

Please give us some feedback!

You have a date with some friends tonight.

Your friends remember that you attended a workshop on research data management today and asks: "Well, how was it"?

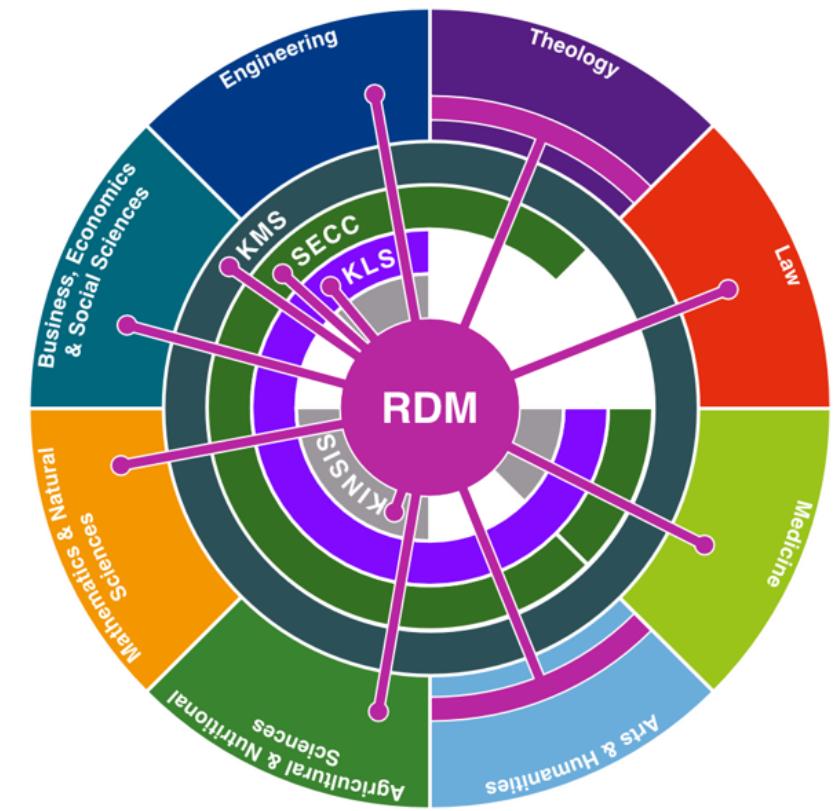
What do you answer?



CAU contacts

RDM contacts at CAU:

<https://www.fdm.uni-kiel.de/de/team>



Thank you!

