

# Open Science Webinar 26.09.2025

UNIVERSITY OF BERGEN



<https://doi.org/10.5281/zenodo.17206952>

# UiB Library Research Data Team



## Courses and guidance

- Open Science
- Open and FAIR Research Data
- Data Management Planning
- Data Management in the active phase
- Archiving and publishing datasets
- Finding and reusing existing data

## Institutional archive

[DataverseNO: University of Bergen](#)

## More information on our web pages

[Open Access to Research Data](#)

[Data Management Plans](#)

**Contact us:** [research-data@uib.no](mailto:research-data@uib.no)

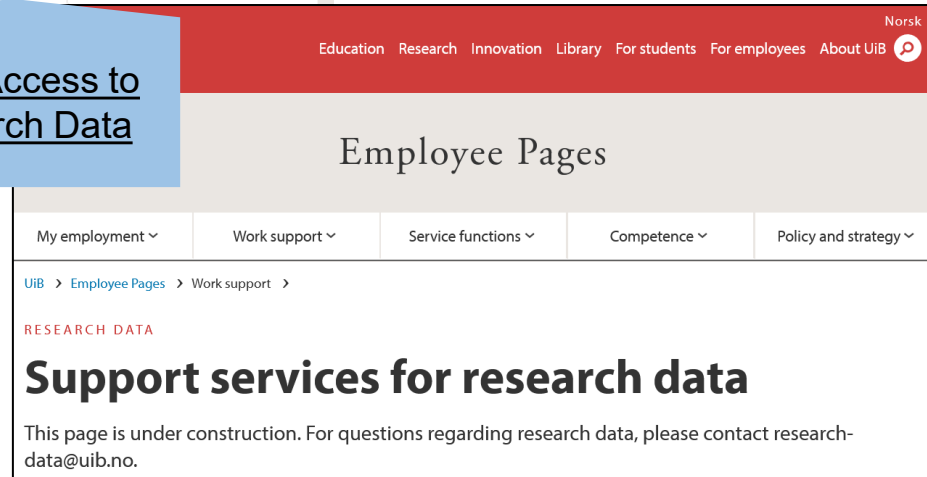
# UiB information resources



<https://www.uib.no/en/ub>

Data  
Management  
Plans

Open Access to  
Research Data



<https://www.uib.no/en/researchdata>

# Agenda

## ? Questions

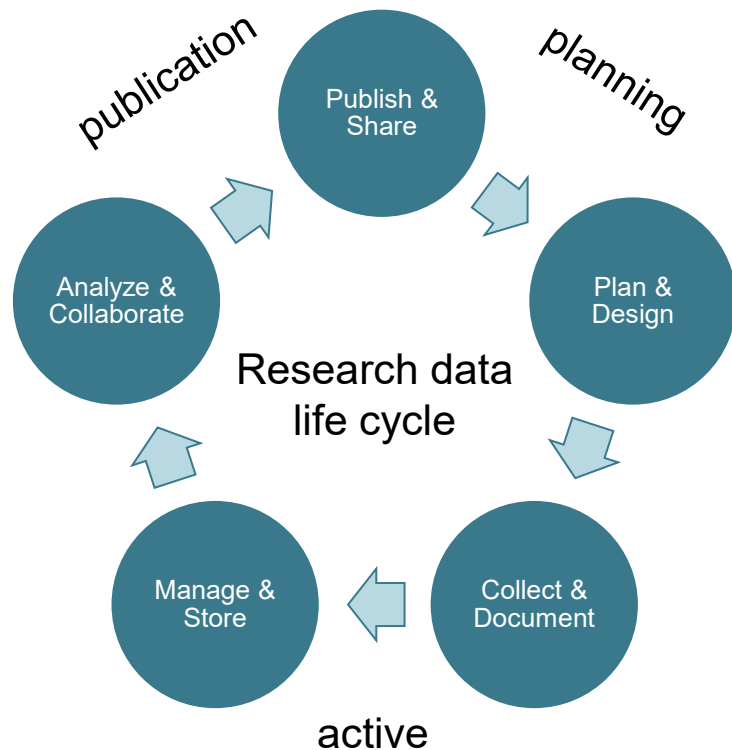
- How can you structure, document and organize your data in a way that is understandable
  - for others
  - for yourself in 3 years

## ! Objectives

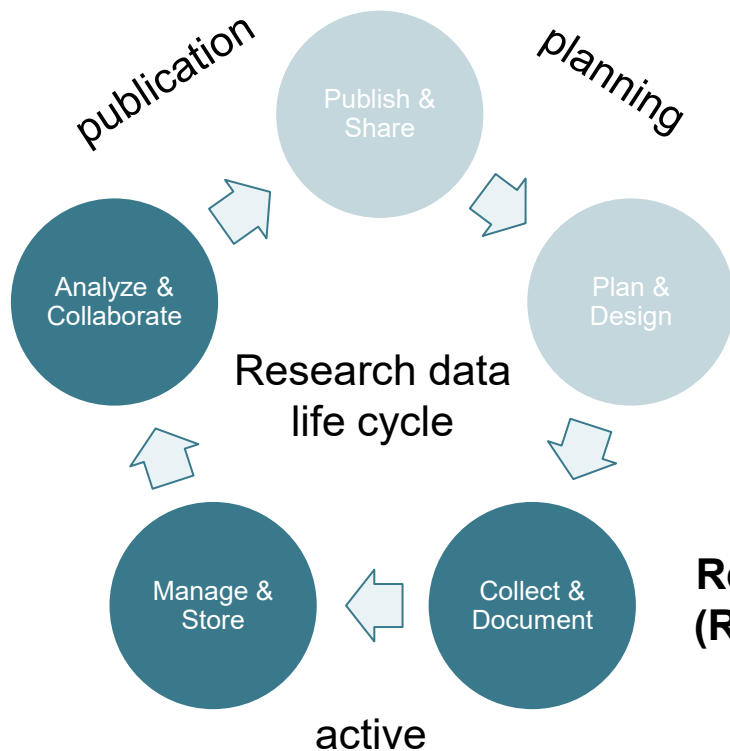
- Be prepared for publishing FAIR research data
- Be prepared to provide details on data provenance
- Make sure you will not loose data



# The research data life cycle

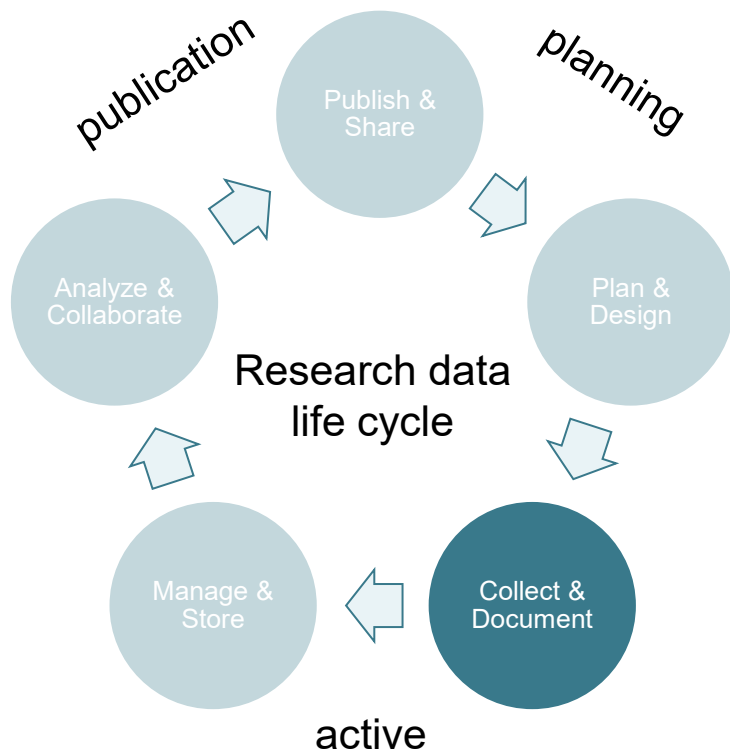


# The research data life cycle



**Research Data Management (RDM) in the active phase of research**

# The research data life cycle



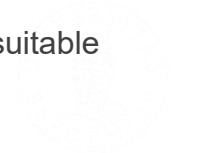
# Keep a log/diary/research journal!

Collect &  
Document

- What was the hypothesis and the chosen approach?
- Document intermediate steps

## ➤ Experimental Sciences: You will likely keep a **lab notebook**

- Archivable document. Formal requirements may vary.
- Follow Good Laboratory (Notebook) Practice.
- Electronic Lab Notes (ELN) can be a convenient alternative but not all solutions are equally suitable (e.g. proof of IPR).





# Non-digital documentation?

Collect &  
Document

If part of your data documentation is non-digital:  
(e.g. traditional lab notebooks)

- Create a digital index/table of content with a short description of each experiment & purpose
- Reference e.g. by running experiment number, date, page, keywords,...
- Copy key information to ReadMe files!



# Keep track of data provenance!

Collect &  
Document

**What was the exact concentration of the reagent you used?**

(you are writing your methods section)

**The thematic analysis based on those interviews from 2020 – which interviewees' statements made up the “identity conflicts” theme?**

(you are replying to reviewer 2)

- You must be able to track the data provenance (data lineage) and trace every single data point back to its origin, including all processing steps!

# Document your data!

Collect &  
Document

*Documentation is a love letter that you  
write to your future self.*

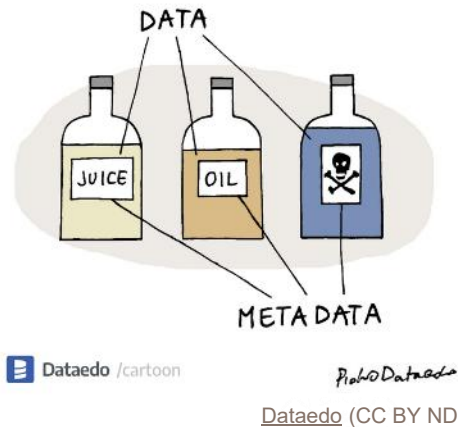
Damian Conway



# Data documentation

Collect &  
Document

- All documentation and metadata (“data about data”) that is needed to understand & reproduce your data:
- Methodology
  - Analytical & procedural information
  - Variable/value/unit definitions
  - Data quality measures
  - Confidentiality, conditions for use
  - File structure



**Remember to also write down steps that seem obvious/trivial to you!**

# Different levels of documentation

Collect &  
Document

## Project or study-level documentation

Title and summary, aims, authors, institutions, funds, methods, license and identifier for datasets, folder structure, file naming conventions, versioning system, relationships between files++

## File-level documentation

Individual records or data points, experimental factors, variables

## Variable/code-level documentation

Units, controlled vocabularies, missing values



# Data documentation: ReadMe files



Collect &  
Document

- Text file (.txt) that describes your data and documents and how you have processed them.
  - Helps you to remember details about your data
  - Ensures other people can read and understand your data

## **A ReadMe file should contain:**

- A short description of data, tables, figures etc.
- Tabular data: definitions of column headings and row labels, data codes, measurement units
- Any data processing steps that may affect result interpretation

```
1 This file was generated on YYYY-MM-DD by NAME.
2 <help text can be deleted before saving>
3
4 GENERAL INFORMATION
5
6 1. Title of dataset: ExpNo
7
8 2. Dataset information <Novel data or secondary data>
9   A. Novel data
10      1. Date of data collection: YYYY-MM-DD
11      2. Geographic location of data collection: <if relevant: latitude, longitude or city/region, State, Country>
12      3. Experimental protocol: <if relevant>
13      4. Codebook: <if relevant>
14      5. Ethics approval: <if relevant, e.g. REK>
15   B. Secondary data:
16      1. Dataset citation
17      2. Access date: YYYY-MM-DD
18
19 3. File list: <data, tables, figures, analysis scripts etc.>
20   A. Readme.txt
21   B. YYYY-MM-DD_ExpNo_Method1.csv
22      1. Column headings, measurement units: <explain all abbreviations>
23   C. ExpNo_Analysis_v1.R
24   D. ExpNo_Analysis_v1_plot1.png
25   E. ExpNo_Analysis_v1_plot2.png
26
27 4. Data processing steps: <explain>
28   A. Experiment executed on YYYY-MM-DD with parameters X.
29   B. Data acquired on YYYY-MM-DD with instrument Y, settings Z.
30   C. Data analyzed with R script ExpNo_Analysis_v1.R.
31      1. Inputs: YYYY-MM-DD_ExpNo_Method1.csv
32      2. Outputs: ExpNo_Analysis_v1_plot1.png, ExpNo_Analysis_v1_plot2.png
```

**Tabular data:** Definitions of column headings/row labels, data codes, measurements units

Any data processing steps that may affect interpretation of the results

# Data documentation: related concepts

- Codebook: Documentation at data(set) level  
[DDI – Create a codebook](#)
- Data dictionary: Explanation of variable names, allowed values  
[OSF Support – How to Make a Data Dictionary](#)

Data documentation resources:

- [The Turing Way - Documentation and Metadata](#)
- [CESSDA Data Management Expert Guide - Documentation and metadata](#)
- [ELIXIR RDMkit - Documentation and metadata](#)





# Different metadata standards

Collect &  
Document

- If you plan to share your data in a discipline-specific data archive, the repository often defines the metadata standard/schema.
- Metadata standards/schemas are structured & machine-readable
- Make yourself familiar with metadata standards in your discipline (if they exist or are applied) & make sure to record all relevant values!

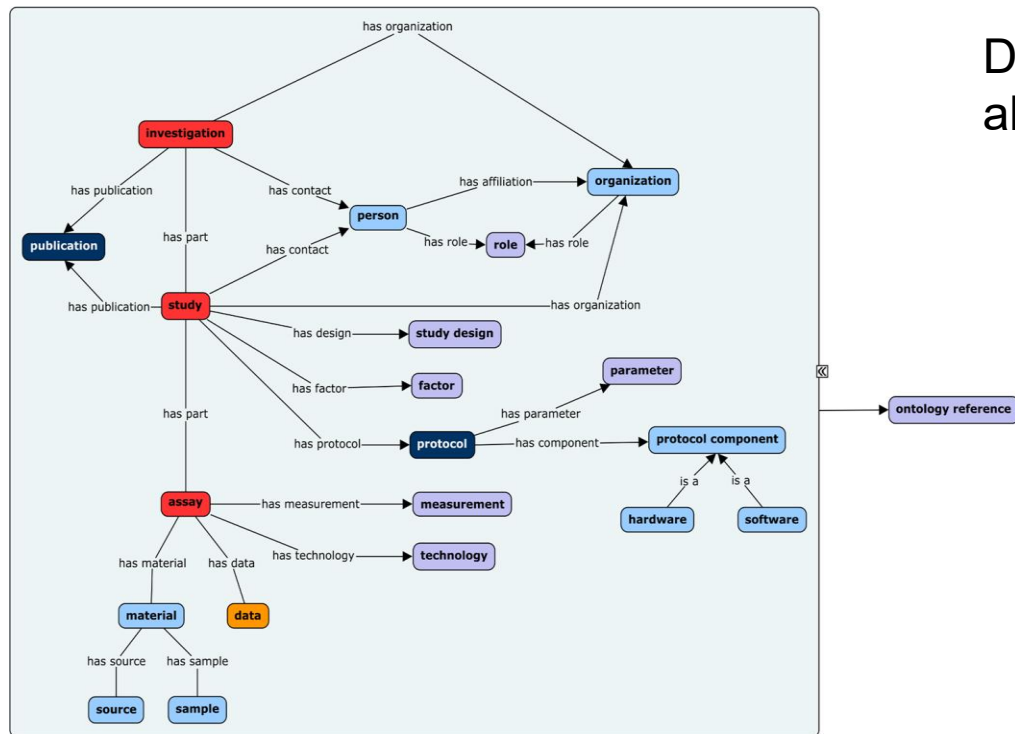
Metadata standard resources:

- [Digital Curation Centre - Metadata standards by discipline](#)
- [FairSharing - Curated list of metadata standards](#)
- [RDA metadata standards catalog](#)



# Metadata schema

Collect &  
Document



Definition for fields and their allowed values:

- Data Types (date, integer, character, url, ...)
- Terms from ontologies
- Terms from controlled vocabularies

# Vocabularies, Ontologies

Collect &  
Document

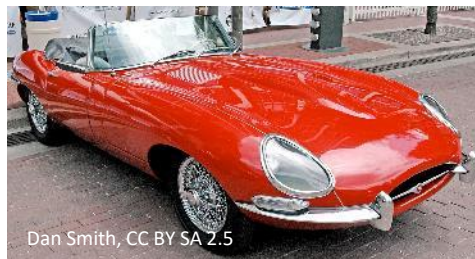
- Ontologies or controlled vocabularies should be applied to describe data in a specific, unambiguous way

«Jaguar»:



*Panthera onca*  
NCBI Taxonomy ID: NCBI:txid9690  
Wikidata ID: jaguar (Q35694)

or



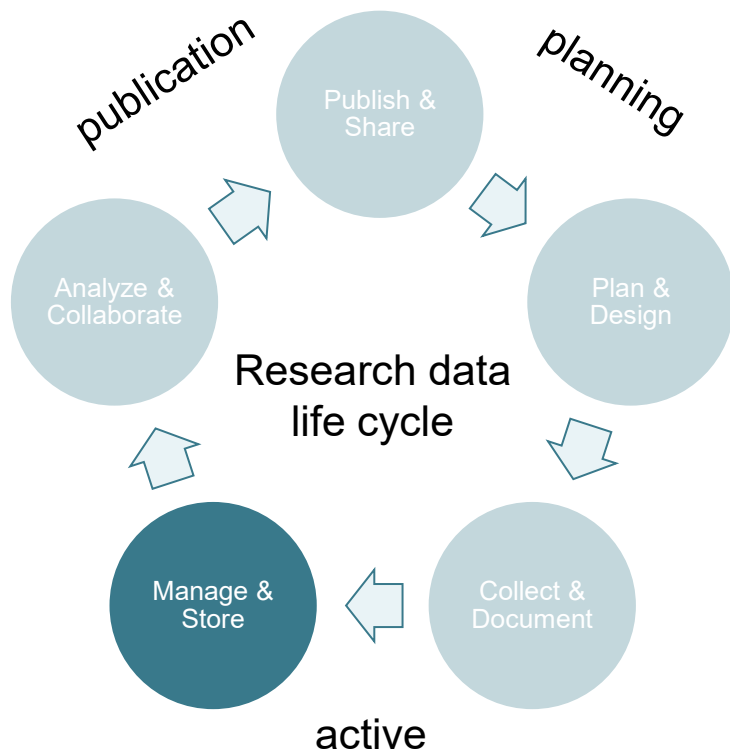
Jaguar E-type (1993)  
Wikidata ID: jaguar cars (Q30055)

?

Ontology and controlled vocabulary resources:

- [Basic Register of Thesauri, Ontologies & Classifications \(BARTOC\)](#)
- [EMBL-EBI Ontology Lookup Service](#)
- [DARIAH Campus: Controlled vocabularies](#)

# The research data life cycle



# Guidelines for your research group/ research environment

Manage &  
Store

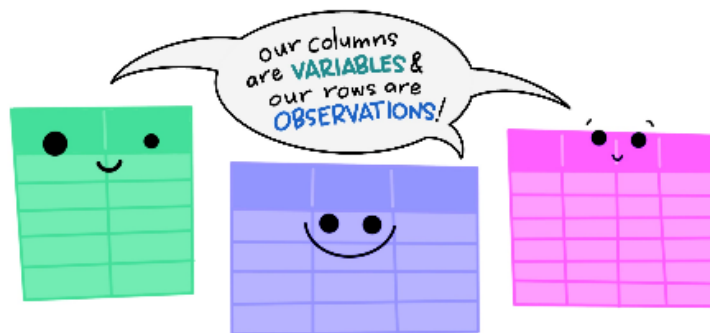
- Did your supervisor suggest you a structure you should follow?
- If you are supervising yourself:  
Consider to provide a set of overarching rules or examples



# Keep your tabular data tidy!

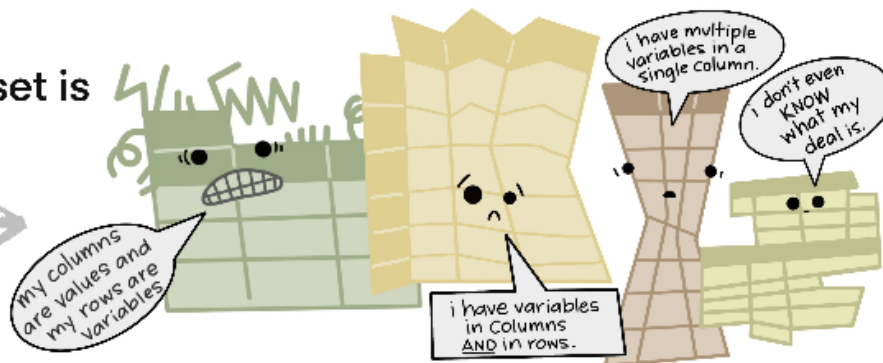
Manage &  
Store

The standard structure of  
tidy data means that  
"tidy datasets are all alike..."



"...but every messy dataset is  
messy in its own way."

—HADLEY WICKHAM



# Keep your tabular data tidy!

Manage &  
Store

## In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation



Tidy spreadsheet resources:

- [Library Carpentry - Formatting data tables in spreadsheets](#)
- [The Turing Way – Data Organisation in Spreadsheets](#)

# Give meaningful filenames!

Manage &  
Store

File naming principles:

- Human readable (meaningful)
- Machine readable (avoid spaces & special characters)
- Play well with default ordering
- From general to specific

**DO**

2025-02-07\_thesis\_intro\_v3.tex

**DON'T**

thesis FINAL.docx





# Organize your data files & folders!

Manage &  
Store

- Organize your data files & folders in a structured manner
- Be consistent!

```
Data
  ReadMe.txt
  |- Method_A
  |- RawData
    |- Exp01
    |- Exp02
    Codebook.txt
  |- Analysis
    |- Exp01
    |- Exp02
```

```
Data
  ReadMe.txt
  |- Exp01
    |- RawData
    |- Analysis
    ReadMe.txt
  |- Exp02
    |- RawData
    |- Analysis
    ReadMe.txt
```

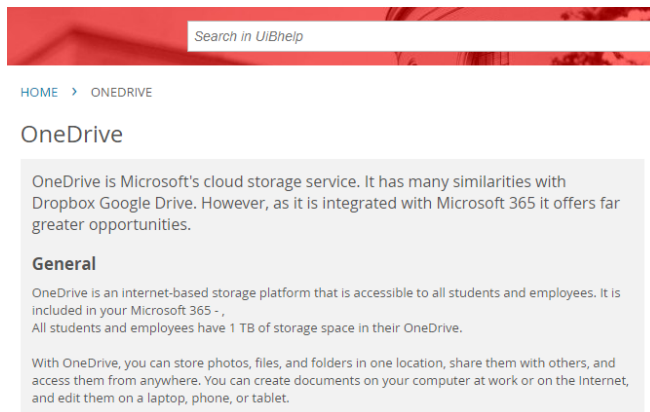
Data organization resources:

- [Turing Way - Data Storage and Organisation](#)
- [CESSDA Data Management Expert Guide - Designing a data file structure](#)
- [ELIXIR RDMkit - Data organisation](#)



# Avoid losing data!

- Use (institutionally) managed storage with automatic backup
- Default@UiB: 1TB on OneDrive
- AVOID: laptops, stand-alone hard drives, USB sticks etc.
- Checksum your data



If you cannot use managed storage

## **3-2-1 rule**

- 3 copies of your data
  - 2 different media
    - 1 off-site copy



# Information security

Manage &  
Store

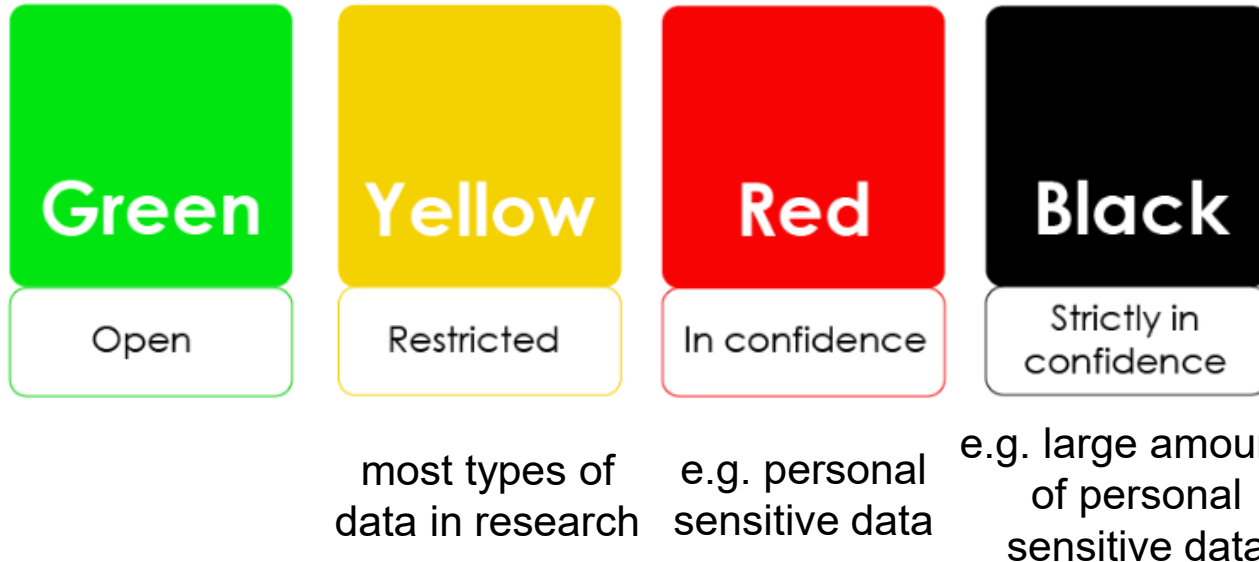
- Make sure to categorize information/data correctly and apply appropriate security measures
- Data can be considered sensitive for many reasons, including Intellectual Property Rights, National Security, or protection of study objects.
- Personal data (data which can be assigned to a person in any kind of way) cannot be shared with a third party, unless specific and explicit consent is secured. Personal sensitive data (GDPR: “special category personal data”) needs special safeguards.



# Information security

Manage &  
Store

- Make sure to categorize information/data correctly and apply appropriate security measures





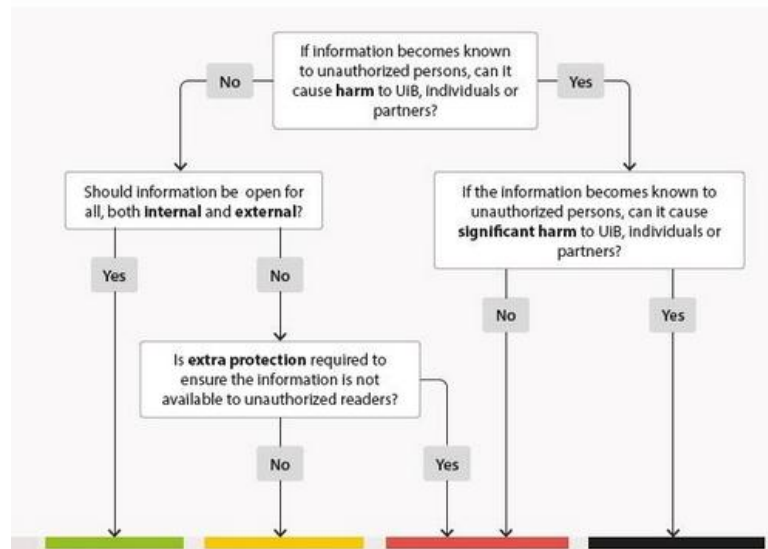
# Information security

Manage &  
Store



## WHAT DATA CAN BE STORED WHERE?

You are responsible for classifying and securing information!



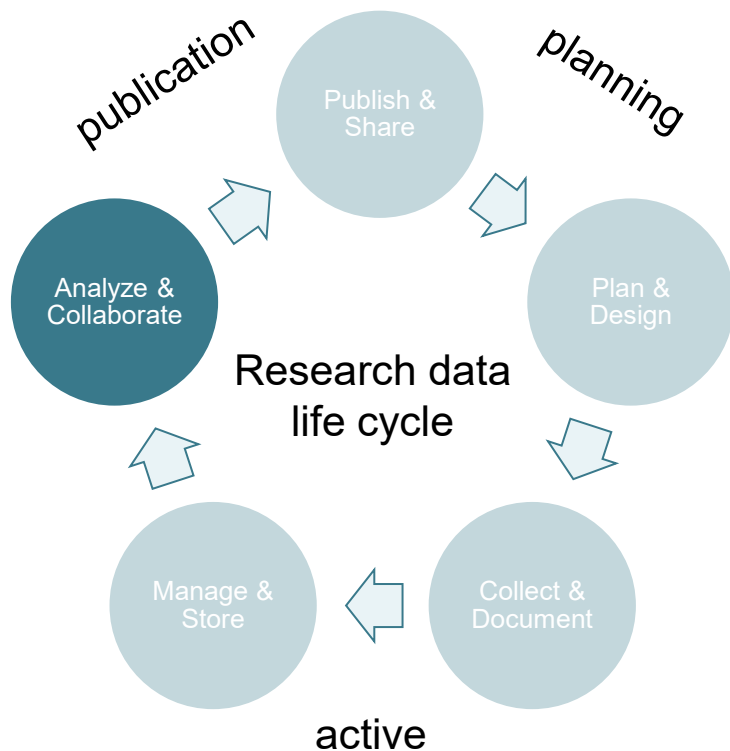
# Store non-proprietary file formats!

Manage &  
Store

- Store data in non-proprietary (open) file formats  
e.g.: .odf instead of .docx, .csv instead of .xlsx, .tif instead of .lsm, .mp4 instead of .mov
  - Store data in uncompressed file formats, if feasible
- 
- Different requirements for short-term and **long-term storage**
  - Long-term: You may lose proprietary software access
  - Long-term: You may change institution
  - Non-proprietary file formats are preferred for data sharing



# The research data life cycle

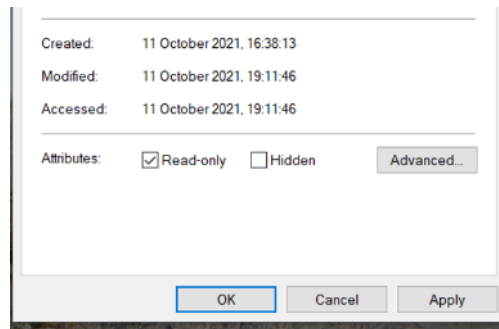


# Stay in control of data analysis!

Analyze &  
Collaborate

Data processing & analysis is part of the data lineage and being transparent

- Raw data should be **read-only**!
- Keep track of data processing steps & file versions!





# Data processing & analysis

Analyze &  
Collaborate

Quantitative data:

- Script-based data cleaning & analysis is easier to reproduce than manual steps.

Qualitative data:

- Code categories should be exclusive, consistent & documented.

Data processing resources:

- [The Carpentries: foundational coding and data science skills to researchers worldwide](#)
- [CESSDA Data Management Expert Guide - Process](#)



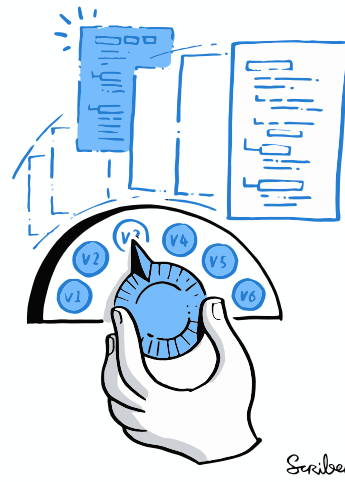
# Version control

Analyze &  
Collaborate

- Helps to understand the history of changes and why something was done in a specific way
- Allows to review changes & restore previous versions
- Makes data processing/analysis transparent
- Is crucial in collaborative projects with simultaneous changes

Version control resources:

- [The Turing Way - Version Control](#)
- [The Turing Way - Version Control for Data](#)
- [Software Carpentry - Version Control with Git](#)
- [RDMkit - How do you manage file versioning?](#)
- UiB GitLab: [git.app.uib.no](https://git.app.uib.no)



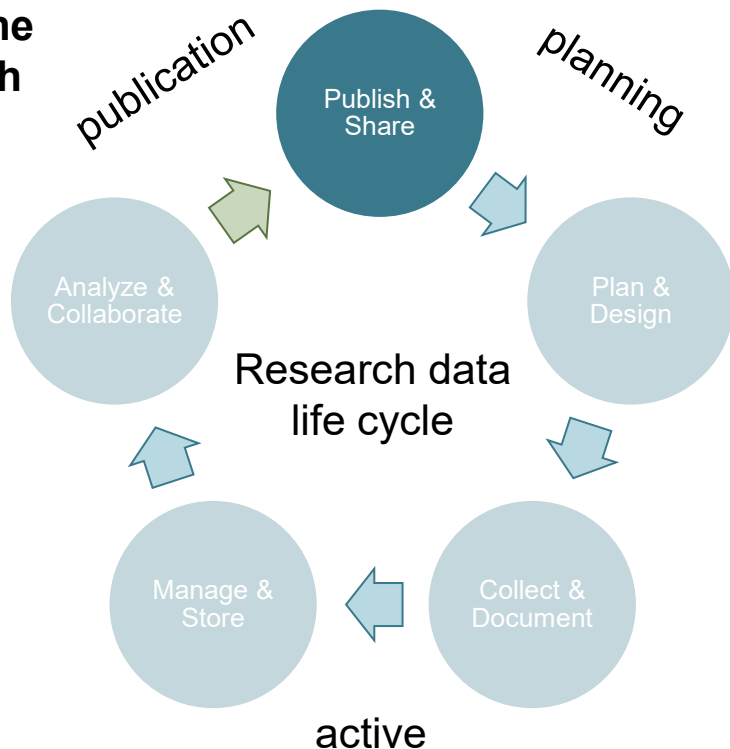
# Collaboration: File transfer

**NB! If there is information security restrictions, they also apply for transferring files!**

- Large files can be shared via <https://filesender.sikt.no/> (FEIDE login)
- Working on collaborative platforms (e.g. OneDrive, Office365, Google Docs, GitHub/GitLab) is more efficient than distributing & merging versions of files (but remember information security)

# The research data life cycle

**Good RDM practice in the active phase of research pays off!**



# The res

## Take-home message

(this will save you pain when publishing your research)

Good RDM practice  
active phase of  
pays off

- Organize your information in a structured way. Develop routines.
- You can never document too much. (your future self will thank you!)
- Be sure about information security and your data backup.
- Be consistent! Stick to the structure/convention you have chosen!



# Further resources



[openscience.no](https://openscience.no)  
[incl. event calendar](#)



[CESSDA Data Management](#)  
[Expert Guide](#)



[The Turing Way - Guide](#)  
[for Reproducible Research](#)



[PhD on Track -](#)  
[Open Science](#)



[ELIXIR RDMkit](#)



[The Carpentries](#)

# Open Science webinars

Autumn 2025

- 29.08.2025 — [Making your research visible using researcher profiles](#)
- 12.09.2025 — [How and where can you archive research data?](#)
- 26.09.2025 — [Stay in control: Research Data Management in active projects](#)
- 17.10.2025 — [How to manage and share sensitive research data](#)
- 31.10.2025 — [How to write a Data management plan \(DMP\)](#)
- 14.11.2025 — [Research Data Management: Humanities, Law, Fine Arts](#)  
(note different time - this webinar starts at **11:15**)
- 28.11.2025 — [Taking advantage of existing research data](#)



<https://www.uib.no/en/ub/calendar>





[research-data@uib.no](mailto:research-data@uib.no)







Except where otherwise noted, this work is licensed under the  
[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

