

Plan your Research Data Management

Elisa Pierfederici, Camilla Hagen Blixhavn & Ivana Malovic
Research Data Management Week
10.11.2025



DSC
Digital Scholarship Centre
at the University of Oslo Library

Questions?
Contact us at research-data@uio.no

The logo for Data Management Week 2025 features a stylized orange folder icon on the left, with three smaller folder icons stacked vertically to its right. The text "Data Management Week" is in orange, and "2025" is in white. Below "2025" is the text "Fall edition" in white.

Data Management Week 2025

Fall edition

November

10th – 14th

Digital Scholarship Center

Monday	<i>Morning:</i> <u>Plan your research data management</u> <i>Afternoon:</i> <u>Version Control using Git and GitHub – Part 1</u>
Tuesday	<i>Morning:</i> <u>Share, archive, and reuse research data</u>
Wednesday	<i>Afternoon:</i> <u>Version Control using Git and GitHub – Part 2</u>
Thursday	<i>All day:</i> <u>Carpentry workshop: Databases and SQL</u>

Agenda

- Research data types
- Research data management
- Data Management Plans

Break

- Data classification and storage
- Data organisation

Break

- Documentation and metadata
- Summary and final questions
- Wrap up and further resources

Research data

What is data?

‘Data’ continues to designate concrete, straightforward, **empirical stuff**

(Dourish 2017).

What is research data?

Research data are **representations** of observations, objects, or other entities used as **evidence** of phenomena for the **purposes of research or scholarship**.

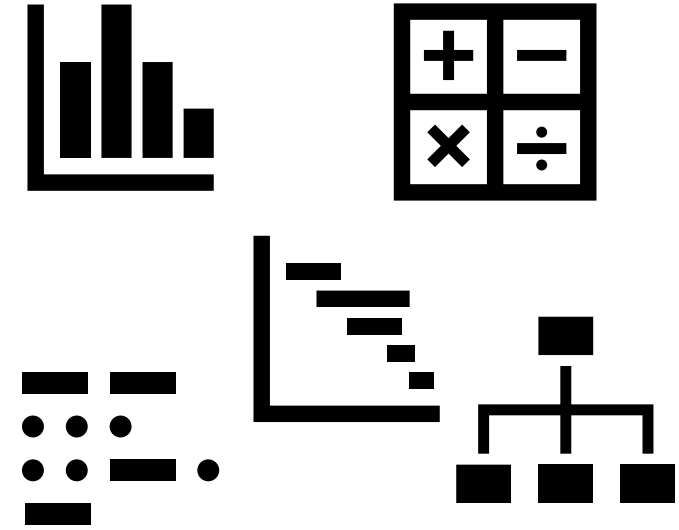
What data are you working with?



Research data types



Qualitative
(descriptive)

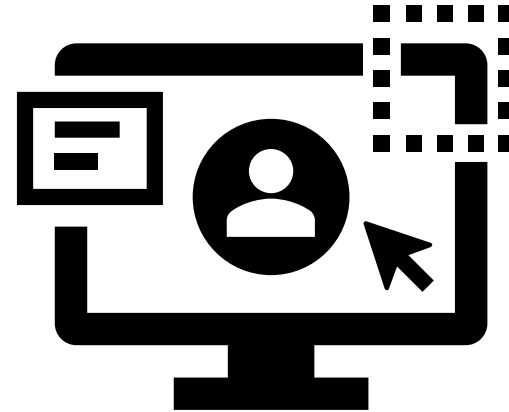


Quantitative
(numerical)

Research data types

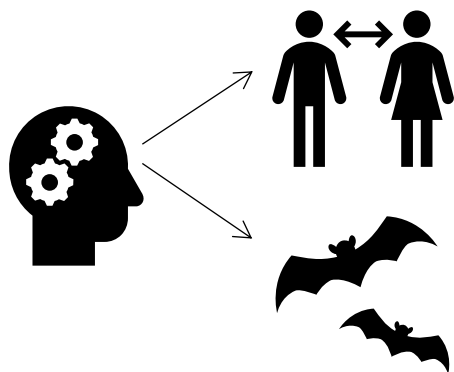


Primary (new)

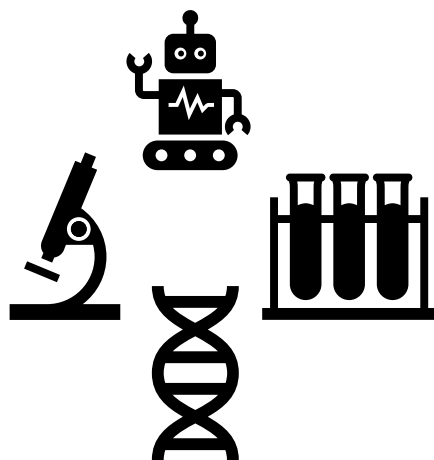


Secondary (existing)

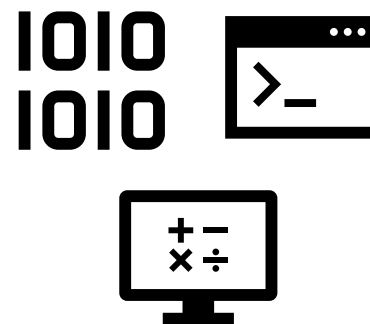
Research data types



Observational



Experimental



Computational



Records

Research data types



Text

documents, transcribed interviews, survey answers, field notes



Numbers

frequencies, occurrences, time of response



Audio

audio recordings of interviews, recordings of sounds, music files



Image and video

photos, images from online content, video recordings

Research data types

- Personal data
 - General (non-sensitive) personal data, e.g. name, address, voice
 - Special categories of personal data (sensitive data), e.g. health information, ethnicity, religion

What is personal data?

Any piece of information about an identified or identifiable physical person.

In other words, any data about living people from which they can be identified.

Direct and indirect identifiers

A direct identifier is information that is sufficient on its own to identify an individual.

e.g. person's name, personal identification number, a facial image, fingerprints

An indirect identifier is information that might fairly easy identify an individual or lead to identification when linked with other available information.

e.g. location data, online identifiers, unusual job title, a position held only by one person at a time, age, education, ethnic background

What constitutes as **general personal data**?

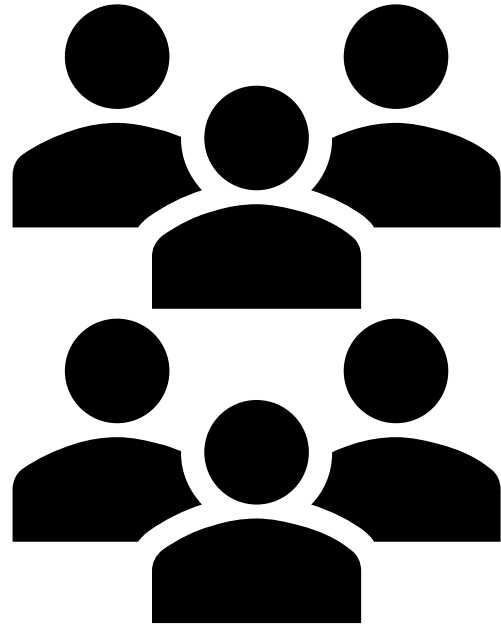
- Name, address, age, phone number, e-mail and Norwegian national identity number
- Video- and audio recordings where individuals are recognizable
- Identifying images of individuals
- Contents of email communication
- Content of case documents, investigations or considerations
- Content of exam papers or grades
- IP addresses (that can identify individuals) and logging of activity in computer systems (where the logs can be associated with individuals)

Special categories of personal data (sensitive data)



- health information and health related conditions
- genetic or biometric information which can be used to identify a physical person
- ethnic or racial origin
- political, philosophical or religious perceptions and beliefs
- sexual orientation or sexual relationships
- trade-union membership

Anonymous data or not?



Anonymization: to irreversibly remove identifying information from the data so that a person cannot be identified based on the data

Pseudonymization: removal or replacement of identifiers with pseudonyms or codes, which are kept separately and protected (note: this is often still perceived as personal data)

Exercise 1: Data types

Think – Pair – Share

Think about what types of data you are or will be collecting in your research projects.

Share and discuss with your neighbor!

03:00

4.0

4th revolution

Cyber physical
systems



3.0

3rd revolution

Electronic and IT
systems, automation



2.0

2nd revolution

Mass production and
electricity

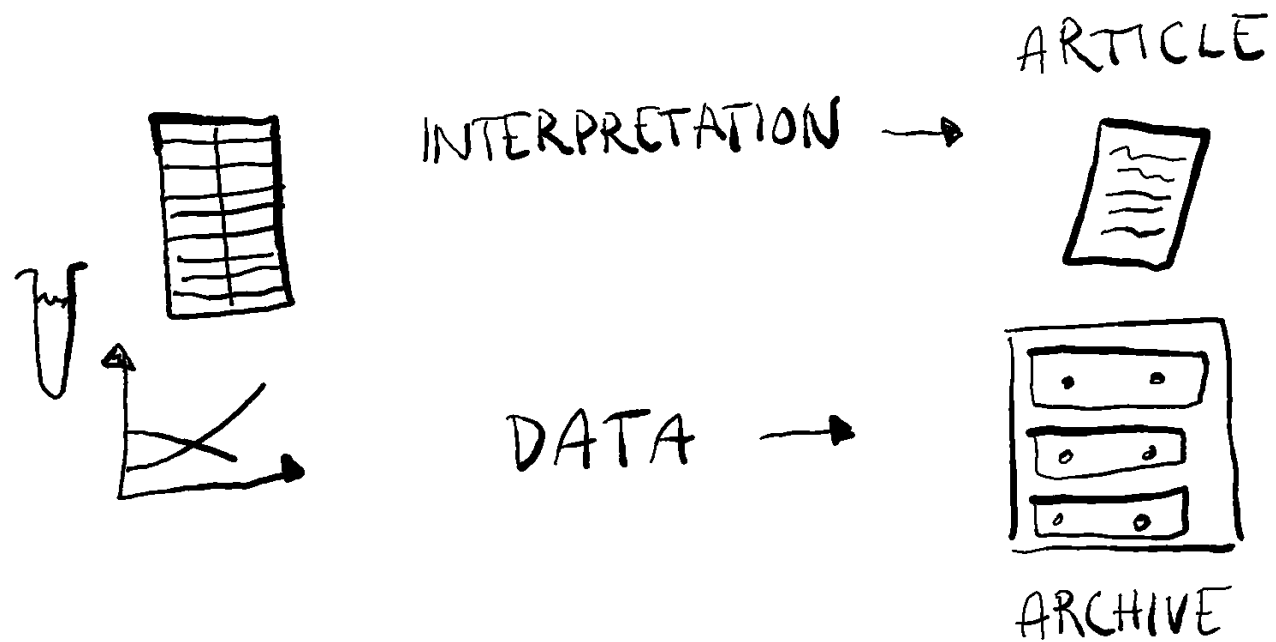


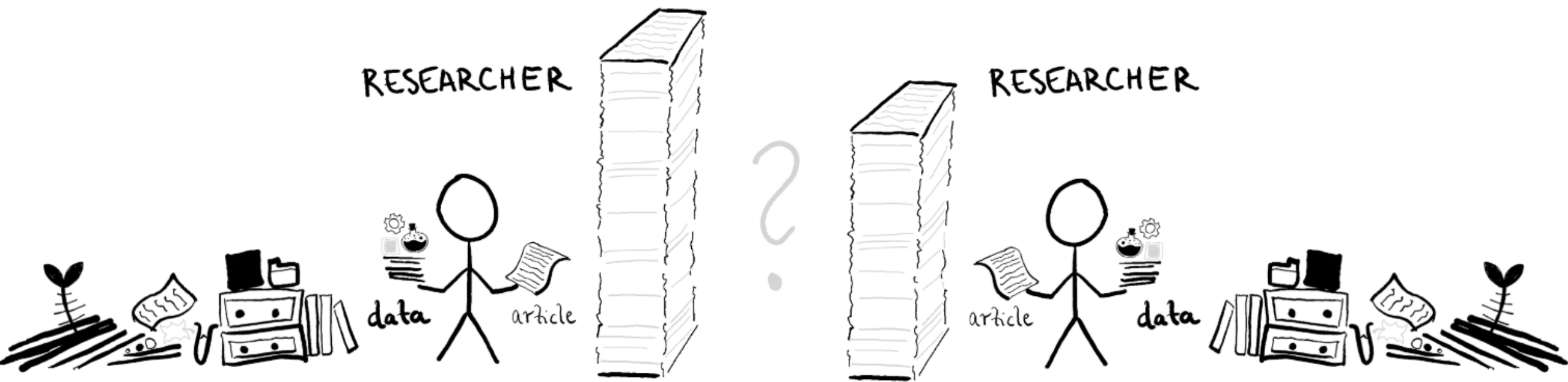
1.0

1st revolution

Mechanization, steam
and water power







Open Science



open
access



open
peer review



open
source



open
educational
resources



open
methodology



open
data

FAIR principles

Findable



Accessible



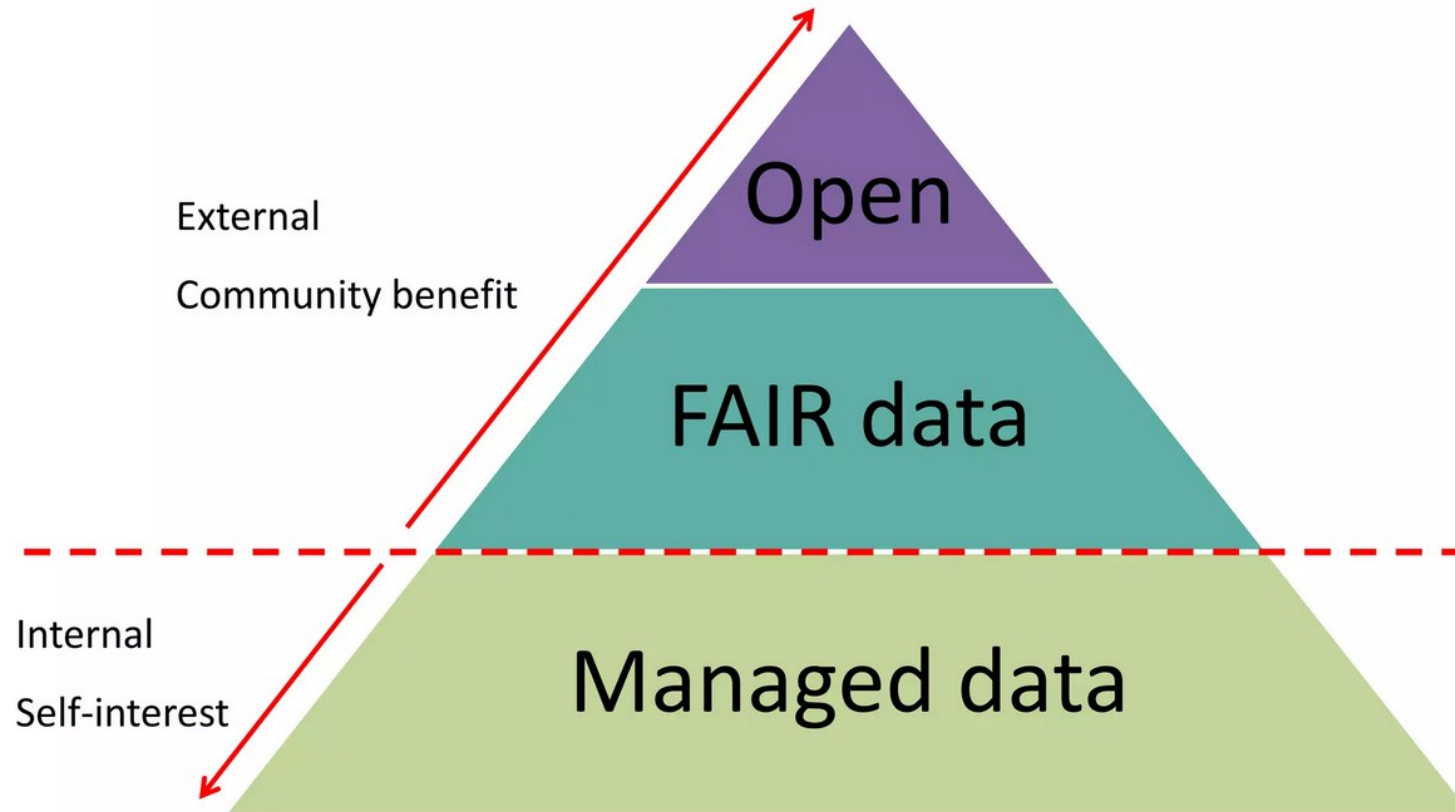
Interoperable

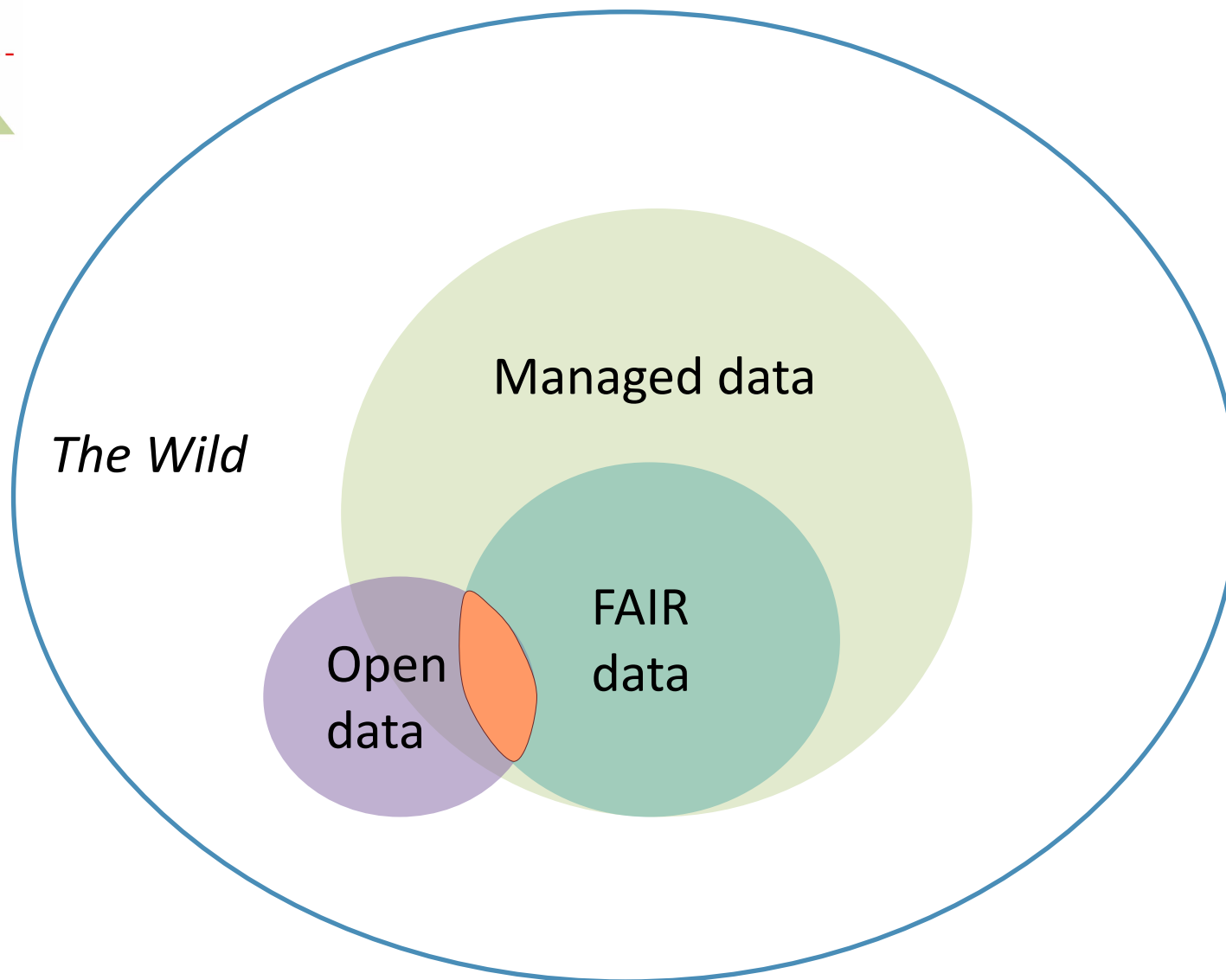
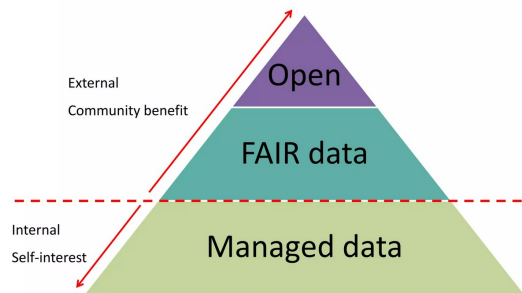


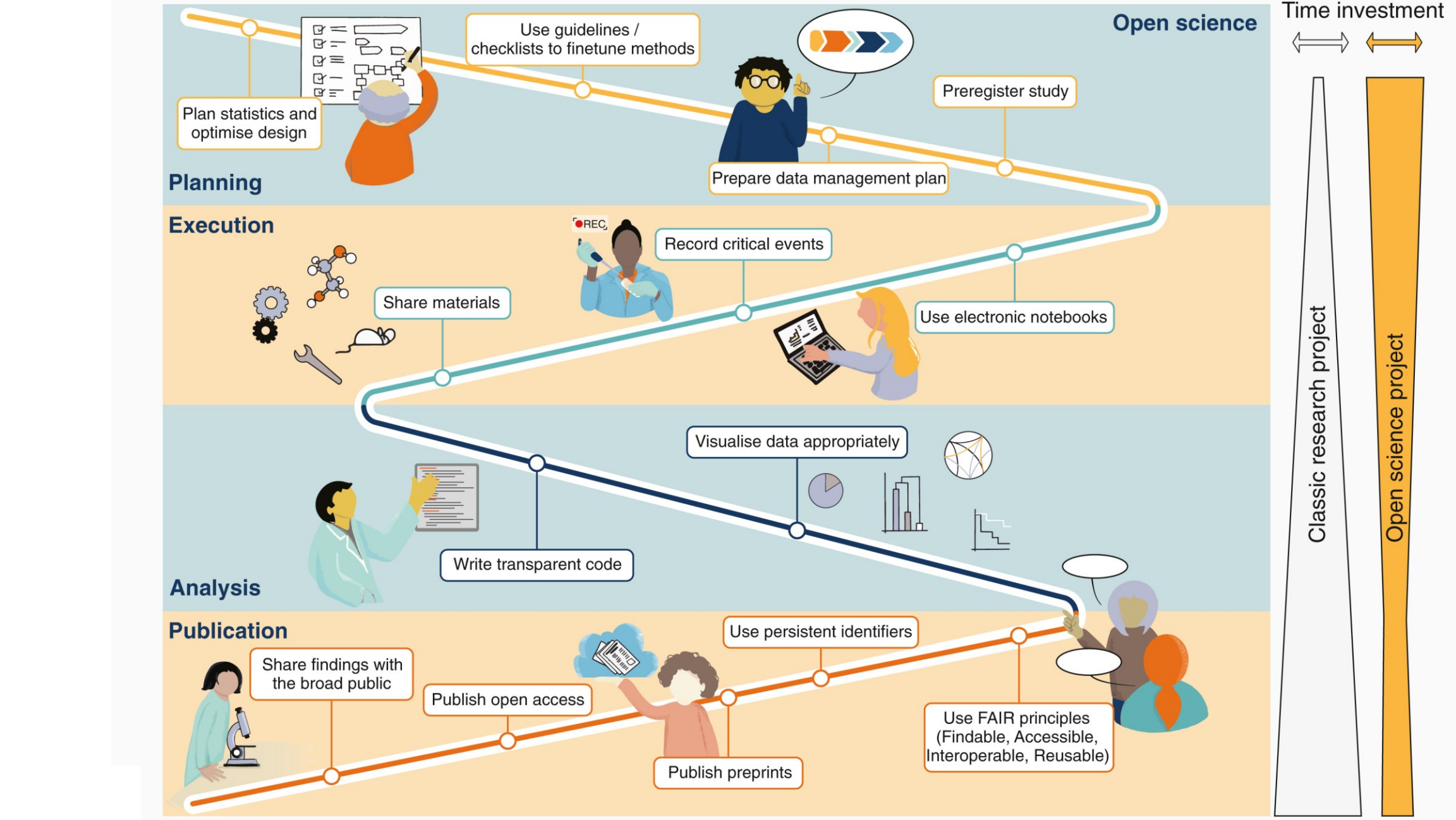
Reusable



As **open** as possible,
as **closed** as necessary







Research Data Management

A DATA MANGMENT TIMELINE

WHEN YOU SHOULD
PLAN OUT YOUR DATA
MANAGEMENT
BUT DON'T

THE FIRST
TEMP FOLDER
IS MADE

THE TEMP FOLDER GETS
A TEMP SUBFOLDER

THERE ARE NOW 30
TEMP FOLDERS

WHEN DATA MANAGEMENT
NORMALLY STARTS

ERRANTSCIENCE.COM

What is research data management (RDM)?



- **Organized** so you can find and reuse them easily
- **Secure and compliant** with funder, ethical, and legal requirements
- **Sharable and citable** to increase visibility and impact
- **Preserved** for long-term value and reproducibility



Protection
Organization
Documentation (and metadata)
Classification
Short-term storage
Licensing
Sharing
Long-term preservation
Dissemination (publishing)
Reusing

The Data Management Plan

Living document that accompanies the research project



- Specifies the **types of data** in the project
- Describes how you **plan to manage data** (organize, document, classify, store)
- **Technical needs**
- Conveys **whether and how the data can be shared**
- **Agreement** between project members
- A **tool** to keep overview over data

Research Data Policy

Research data at the University of Oslo shall:

- be made **openly available** for further usage
- be made available at an **early stage**
- have a **data management plan**
- have **metadata** and be **documented**
- must be securely **archived**
- have **licenses** for access, reuse and redistribution
- made **freely available**



Research Data Policy

All scientists and students must develop and document clear procedures for gathering, storing, archiving, using, reusing, accessing and storage or destruction of research data in connection with their research.



Tools to create a Data Management Plan



Interactive tools and web forms

- [FAIR Wizard Norway](#) - pilot in collaboration with Sikt
- [Data Stewardship Wizard](#) (ELIXIR) - life science data but not only
- [Sikt \(formerly NSD\)](#) - single and easy to use template but with limited functionalities
- [DMP-online](#) - which provides templates following ERC and Horizon Europe recommendation

Text-based templates

- UiO's template (in [Norwegian](#) and [English](#)) - simple Word-based template with links to UiO resources
- [Horizon Europe template](#) - recommended for use with Horizon Europe projects

FAIR Wizard Norway

<https://norway.fair-wizard.com/>



2025-10-22: A revised version of the Knowledge Model, which is underlying the DMP questionnaires, has been released. The 'Update available' button enables a dialogue where the changes either can be reviewed per question or all changes accepted. [Read more about the new version.](#)

Welcome to **FAIR Wizard Norway**, a pilot instance to create data management plans. This pilot is coordinated by Sikt, and this instance is piloted by University of Bergen, The Norwegian University of Science and Technology (NTNU), the University of Oslo, and UiT The Arctic University of Norway.

The pilot is accompanied by the information resources available under plan.research-data.no/.

Not employed at NTNU, UiB, UiO or UiT? See other [Norwegian DSW/ FAIR Wizard instances](#).

Log In

 Email

 Password

[Forgot your password?](#)

[Log In](#)

Or connect with

 Feide

Login with FEIDE is enabled for users at NTNU, UiB, UiO and UiT and is the preferred login option. If you are experiencing issues, please contact your [local research data support](#).

FAIR Wizard Norway

Select phase



Initial DMP

In many research projects, the finished data management plan is one of the early deliverables. Questions that should be addressed in those early months of the project are highlighted as desirable when this phase is selected. Regular updates should be conducted as the project develops.



Midway DMP

Regular updates of the DMP should be conducted as the project develops. Questions that should be (re-)addressed while revising the DMP project are highlighted as desirable when this phase is selected.



Complete DMP (concluding the project)

When a project rounds up, this means that all data has found its final spots. In this phase, questions that document the final state of affairs will be labeled as desirable.

FAIR Wizard Norway

FAIR Wizard Nor...

DATA MANAGEMENT PLANNER

Dashboard

Knowledge Models

Document Templates

Projects

List

Files

Documents

Importers

Settings

Start your DMP

DMP supporting information

Get Help

Elisa Pierfederici

Admin

TEST-MaDaGo

Questionnaire

Preview

Documents

Settings

View

Import replies

Current phase

Initial DMP

Chapters

I. About the Data Manage... 1

II. About the research project ✓

III. Legal and ethical aspects 5

IV. Storing and protecting d... 6

V. Reusing and collecting d... 7

I. About the Data Management Plan

Remember to select the right project phase before filling out/updating the DMP: [User guidance](#)

This chapter collects administrative information about the DMP itself.

[Supporting information for this chapter](#)

✓ I.1 DMP title

RCN/Science Europe

A title following the scheme "Data Management Plan for: Project identificator" is recommended.

Example:

- If your research project's name is "A really fantastic project" and you use the acronym AMAZING, the DMP title should be "Data Management Plan for AMAZING - A really fantastic project, 2020-2024"

☒ Desirable: Initial DMP

☒ External links: [Supporting information: DMP title](#)

Data Management Plan for MaDaGo - Magic Data Golf, 2025 to 2999

Clear answer

Share

TODOs 9

Comments 12

Version history

← Previous

1/8

Next →

☐ View resolved comments

Comments 1

Editor notes

Assigned to you

Anonymous user

11. 2. 2025, 12:56

✓

@Elisa another test

Reply...

Create a new comment...

Research-data.no

Support package for Norwegian higher education



plan.research-data.no

is a knowledge resource to support Data Management planning at Norwegian research institutions.

For researchers writing a DMP



[DMP questionnaire](#)

Start writing a DMP using the tool Data Stewardship Wizard/ FAIR Wizard. You can select between different DMP questionnaires to adapt for project-specific needs.



[Accompanying information resources](#)

Information pages for researchers filling out a DMP, supplementing the DMP questionnaire in Data Stewardship Wizard/ FAIR Wizard.



[Find local and discipline-specific experts](#)

Find the local DMP experts to support you and get inspired by discipline-rooted resources.

About data management planning



[Purpose of a Data Management Plan](#)

Why should you write a Data Management Plan (DMP)?



[Guiding questions pre-start/ pre-award](#)

Identify data management needs and challenges in ahead of starting on a project, particularly beneficial when applying for funding.



[Glossary](#)

Terms used in the context of data management plans and research data management explained.

For supporting staff



[DMP Background knowledge](#)

The knowledge base for support personnel includes the annotation of the Science Europe DMP guidance in the Norwegian context.



[Conceptual DMP reflections](#)

The project work involved conceptual reflections on DMPs which we believe can be of value to other national and international DMP projects.

Writing a Data Management Plan

10 Simple Rules

1. Determine research sponsor **requirements**
2. Identify **data** to be collected
3. Define how data will be **organized**
4. Explain how data will be **documented**
5. Describe how **quality** will be maintained
6. Develop a strong **storage** and **preservation** strategy
7. Define project's data **policies**
8. Describe how data will be **disseminated**
9. Assign **roles** and **responsibilities**
10. Prepare a realistic **budget**

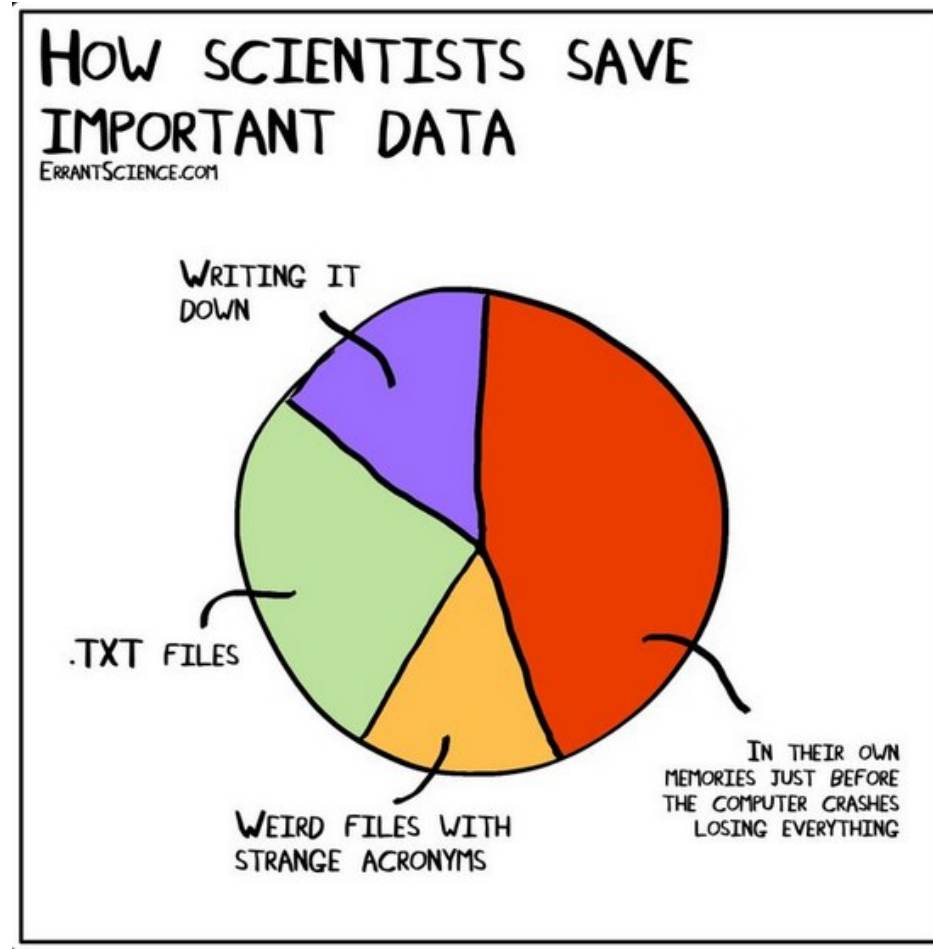
Exercise 2: DMP

Write and discuss

Identify which elements are relevant for your project and note any considerations or actions you'll need to take.

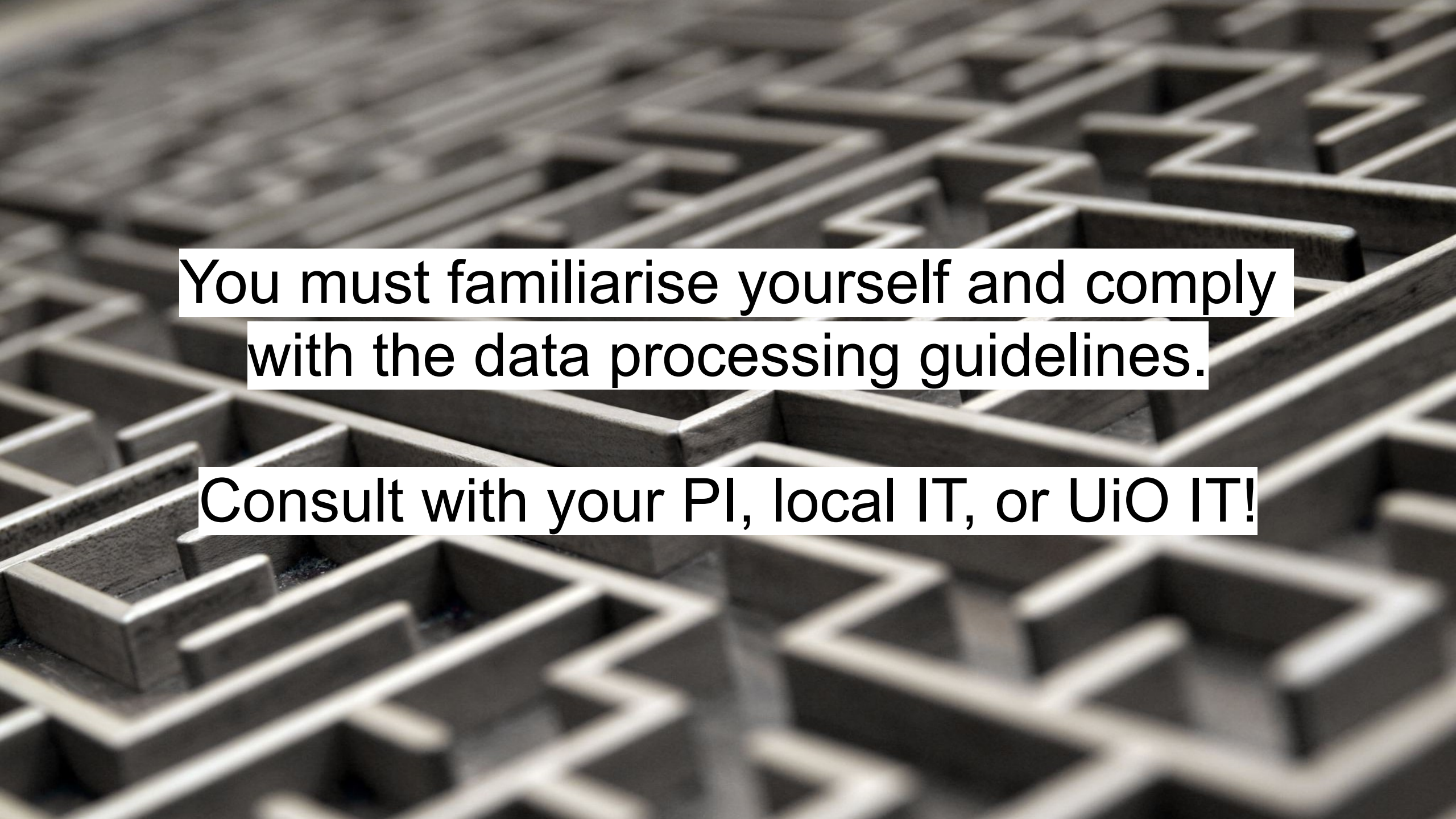
07:00

Break



By Matthew Partridge (@ErrantScience, Aug 19 2021), CC-BY-NC

Data classification and storage selection



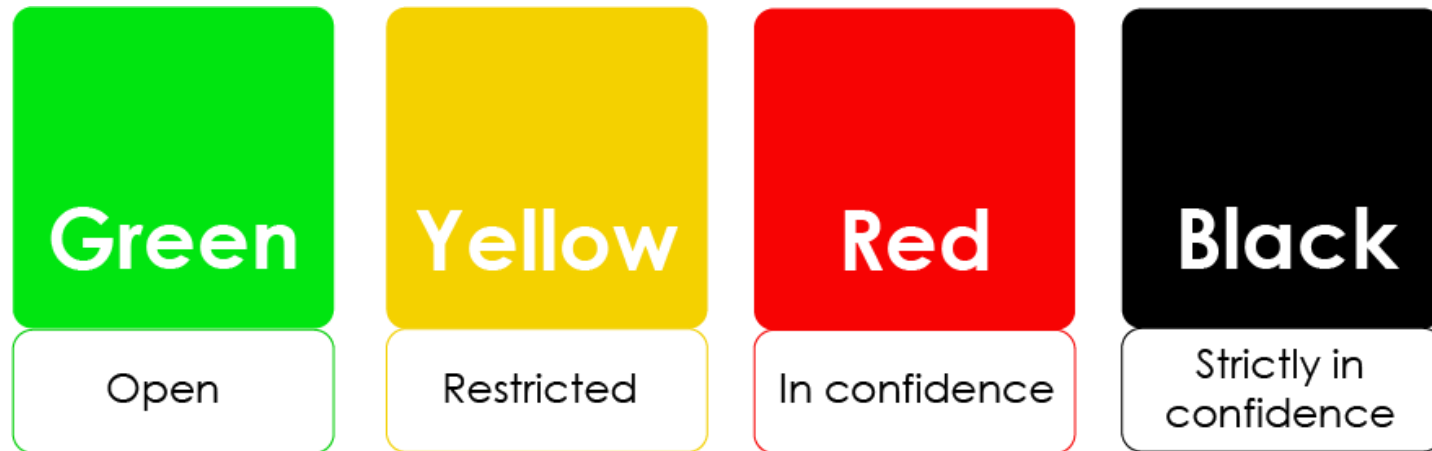
You must familiarise yourself and comply
with the data processing guidelines.

Consult with your PI, local IT, or UiO IT!

Classifying your data

Data classification at UiO is based on how harmful for the university, its partners, public interests, or individuals it would be if the data become exposed to third party

Data Classification Guide

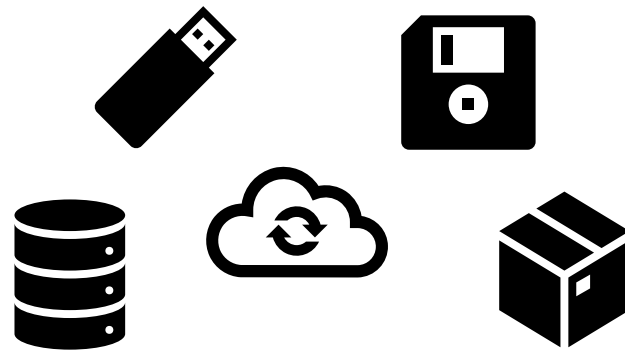


<https://www.uio.no/english/services/it/security/isis/data-classes.html>

Storing your data

Based on the classification of your data, you can choose correct, secure and efficient way to **process**, **store** and **handle** them

Data Storage Guide



<https://www.uio.no/english/services/it/security/lisis/storage-guide.html>

Data Storage Guide

Storage on physical devices

–

	Open (green)	Restricted (yellow)	In confidence (red)	Strictly in confidence (black)
Privately owned computer	Yes	<u>11</u>	No	No
UiO-owned computer, not UiO managed	Yes	<u>11</u>	No	No
UiO owned and managed computer	Yes, <u>4</u>	Yes, <u>4</u>	Yes, <u>1</u> , <u>3</u> , <u>4</u> , <u>15</u>	No
UiO managed cell phone	Yes	Yes	Yes, <u>3</u> , <u>4</u>	No
Memory stick / external hard drive	Yes	No	No	No
Memory stick / external hard drive – encrypted	Yes	Yes	<u>1</u> , <u>3</u>	No

E-mail

+

Storage services

+

Administrative services

+

Content management systems (CMS)

+

Other services

+

Sharing your data

Based on the classification of your data, you can choose correct, secure and efficient way to **share** them

Data Sharing Guide



<https://www.uio.no/english/services/it/security/Isis/data-sharing-guide.html>

Data sharing guide

	Open (green)	Restricted (yellow)	Confidential (red)	Strictly confidential (black)
Intern e-post	Yes	Yes	See below	No
Ekstern e-post	Yes	See below	No	No
"Del"-knapp i Office 365	Yes	Yes	No	No
Teams, OneDrive, SharePoint, Mattermost	Yes	Yes	No	No
FileSender	Yes	Yes	No	No
FileSender + 7-Zip-kryptering	Yes	Yes	Yes	No
Zoom	Yes	Yes	Yes	No
EduCloud Share	Yes	Yes	Yes	No

Green data

Open or freely available to the general public

- This class is to be used if the university or its partners are **not subject to any harm** if the information is exposed to third parties.



Green data

Open or freely available to the general public

- Examples are:

- Webpage presenting a department or a class, published openly on the internet
- Course material which is openly published, but marked with a certain license and/or copyright
- Research data that does not need any protection (the researcher is responsible for this assessment)
- Teaching material that does not need any protection (the teacher is responsible for this assessment)



Green data

Open or freely available to the general public

- Recommended storage solutions:
 - **Anywhere you want**, but ideally **not only** on your private laptop (very vulnerable!)



Yellow data

Restricted information not open for everyone

- This class is to be used if the university or its partners may be subject to **limited harm** if the information is exposed to third parties.
- The information needs a certain protection, and may be accessible to people both within and outside the university, provided that the access is limited and controlled per user.



Yellow data

Restricted information not open for everyone

- Examples may be:
 - Certain work documents
 - Information which is to be kept from the public
 - Many types of personal data
 - Grades
 - Work by students
 - Examination answers
 - Unpublished research data and corresponding works



Yellow data

Restricted information not open for everyone

- Recommended storage solutions:
 - **Not** your private laptop, unless you comply with the [guidelines for using private computers](#)
 - Encrypted memory stick or external hard drive
 - Encrypted laptop maintained by UiO
 - UiO cloud solutions
 - UiO Storage Hotel («lagringshotell»)
 - UiO personal researcher storage
 - UiO shared storage for your unit
 - UiO Teams; UiO Zoom; UiO Nettskjema; UiO Canvas, UiO Vortex, ...



Red data

In confidence information which the university is obliged to protect by law, agreements and other regulations

- This class is used if it **could cause harm** to the university, its partners, public interests, or individuals if the information is exposed to third parties.



Red data

In confidence information which the university is obliged to protect by law, agreements and other regulations

- Examples may be:
 - [certain types of sensitive personal data](#)
 - personnel files
 - certain information about for example protection and safety of buildings and IT systems
 - information about a person's health



Red data

In confidence information which the university is obliged to protect by law, agreements and other regulations

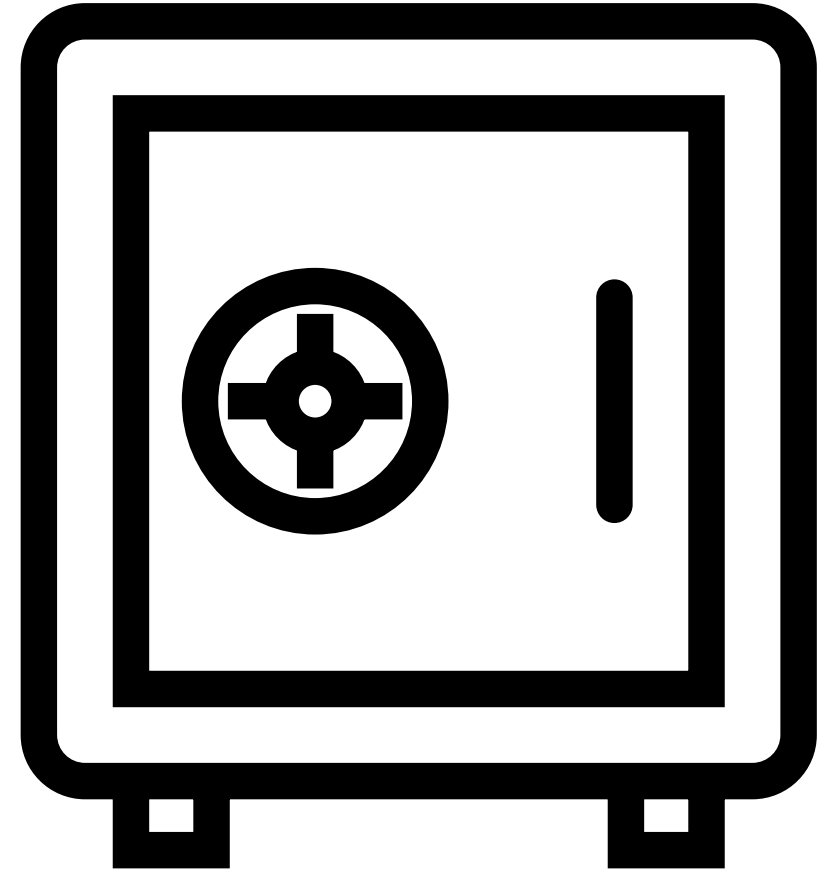
- Recommended storage solutions:
 - Laptop with a fully encrypted disk, an encrypted memory stick or an encrypted external hard drive.
 - Services for sensitive data – TSD
 - [Educloud Research](#)*
 - In Vortex, on your shared research network drive or in the UiO Storage Hotel, if you make an individual risk assessment.



Black data

Strictly in confidence category encompasses the same type of information as «In confidence (red)», but where special circumstances makes it necessary to protect the information even more.

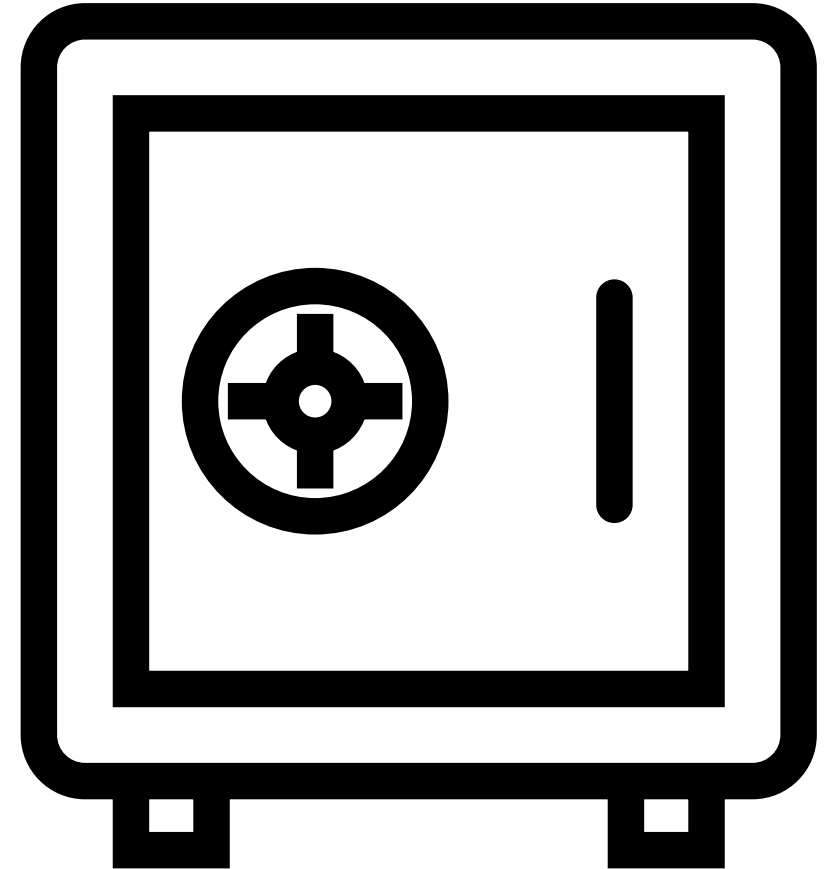
- This class is used if it **could cause considerable harm** to the university, its partners, public interest, or individuals, if the information is exposed to third parties.



Black data

Strictly in confidence category encompasses the same type of information as «In confidence (red)», but where special circumstances makes it necessary to protect the information even more.

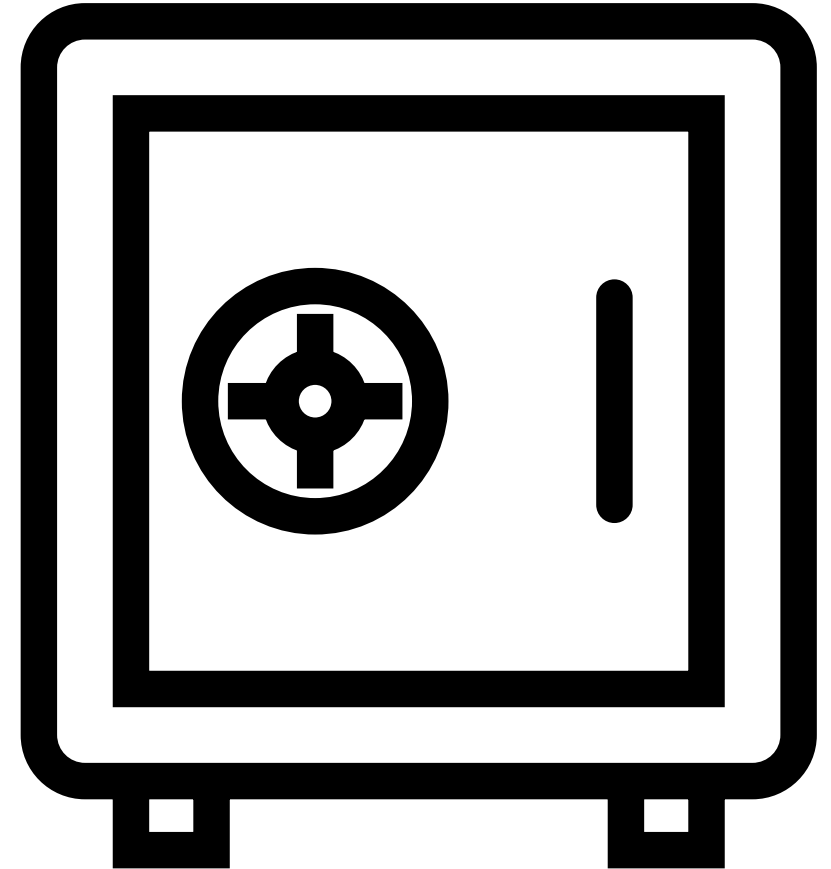
- Examples may be:
 - Large amounts of sensitive personal data
 - Large amounts of data about people's health
 - Research data and datasets of huge economic value



Black data

Strictly in confidence category encompasses the same type of information as «In confidence (red)», but where special circumstances makes it necessary to protect the information even more.

- Placement of data and information in this category should be done in cooperation with the lawyers at USIT and the IT security manager.
- Recommended storage solutions:
 - UiO TSD Services for sensitive data



Classifying your data

Always put the information in a sufficiently safe class.

If you are not sure whether to your information is red or yellow, choose red.



Exercise 3: Data classes

Menti

What are correct data classes in these cases?



08:00

Overview of all cloud services evaluated or under consideration at UiO

- https://www.uio.no/tjenester/it/ny-tjeneste/vurderte_skytjenester.html

(in Norwegian)

- Services with the UiO agreement:
<https://www.uio.no/english/services/it/all/>



Backup



- Use safe storage solutions, where data/files are backed up
- Always keep a copy of your raw data material safely hidden away (like a folder on storage hotel that you never edit)
- Saving files only on your own computer is **not** safe!
- A memory stick or an external hard drive is **not** safe!
- How much extra work would take to recreate all your data or your work?

<https://www.uio.no/english/services/it/store-collaborate/backup/>

Organising the research project

Why file organization?



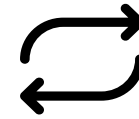
**Saves time and
reduces errors**



**Prevents data
loss**



**Smooth collaboration
& project handover**



**Support reproducible
research & FAIR
requirements**



**Increases the value
and reuse
potential of data**

You will thank you!



SPRINGER NATURE

Used under a CC-BY 4.0 license. DOI: [Scientific Data et al. \(2017\)](#).

You will thank you!

You can understand your own project months or years later

You avoid searching through disorganised folders

You can reuse past work without re-doing it

You handle deadlines and revisions with less stress

Others will thank you!

Your Collaborators

They can quickly find the correct files

They do not need to repeatedly ask for information

They onboard easily into your project

Others will thank you!

Your Collaborators

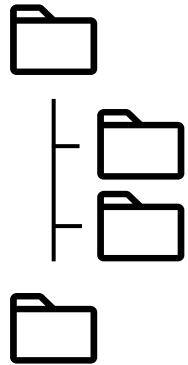
They can quickly find the correct files
They do not need to repeatedly ask for information
They onboard easily into your project

The Research Community

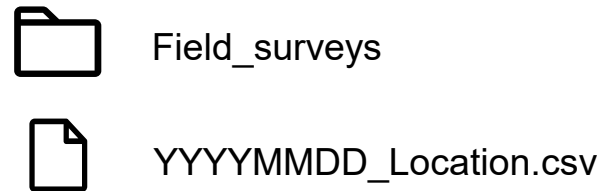
Your work becomes easier to review, reproduce, and build upon
You increase research impact and trustworthiness

What does file organization involve?

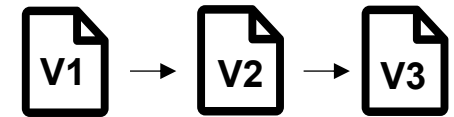
Folder structure or hierarchy



File Naming Convention



Version Control Strategy



Designing a folder structure

SvalbardEx271020

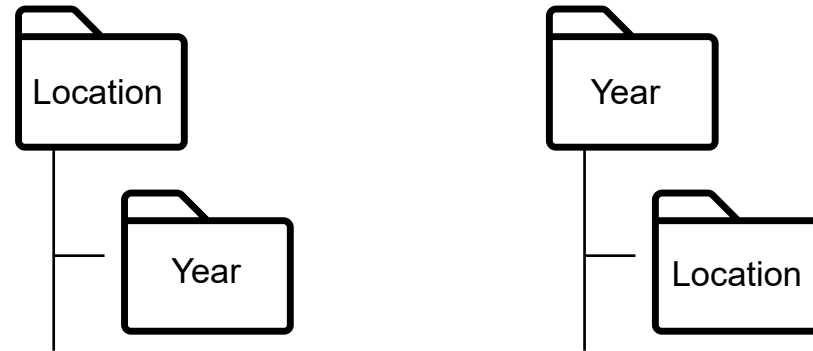
File Home Share View

< > < > > SvalbardEx271020

Name	Date modified	Type	Size
P-10-01	26.10.2020 16:29	File folder	
SEM 2012	26.10.2020 16:25	File folder	
P-10-03-01.tif	11.06.2012 11:17	TIF File	1 880 KB
P-10-03-02b.tif	11.06.2012 11:25	TIF File	1 880 KB
P-10-03-02c.tif	11.06.2012 11:26	TIF File	1 880 KB
P-10-03-05.tif	11.06.2012 11:38	TIF File	1 880 KB
P-10-03-79b.tif	11.06.2012 12:59	TIF File	1 880 KB
P-10-03-92b.tif	11.06.2012 13:26	TIF File	1 880 KB
P-10-03-109.tif	11.06.2012 13:49	TIF File	1 880 KB
P-10-03-119b.tif	11.06.2012 14:02	TIF File	1 880 KB
P-10-03-193b.tif	11.06.2012 15:45	TIF File	1 880 KB
P-10-03-222b.tif	11.06.2012 16:29	TIF File	1 880 KB
P-10-03-222c.tif	11.06.2012 16:30	TIF File	1 880 KB
P-10-03-226b.tif	12.06.2012 09:33	TIF File	1 880 KB
P-10-03-226c.tif	12.06.2012 09:33	TIF File	1 880 KB
Gasser_2014.pdf	27.01.2015 13:11	Adobe Acrobat D...	5 127 KB
Gernigon_and_Bronner_2012.pdf	30.01.2015 15:07	Adobe Acrobat D...	3 445 KB
Gernigon_et_al_2014.pdf	26.01.2015 16:02	Adobe Acrobat D...	54 447 KB
Glorstad_Clark-2010.pdf	15.09.2014 15:06	Adobe Acrobat D...	10 269 KB
Gronlie et al. 1980 Seismic inversion of Bj...	09.10.2012 13:28	Adobe Acrobat D...	643 KB
Gudlaugsson_et_al_1998.pdf	17.10.2014 12:50	Adobe Acrobat D...	3 411 KB
Harland_and_Gayer_1972.pdf	17.10.2014 11:47	Adobe Acrobat D...	1 474 KB
Høy_and_Lundchien_2011_NBarenstSea...	17.09.2014 14:10	Adobe Acrobat D...	13 600 KB
Isaksen 1996 Organic geocem Bjornoya.p...	09.10.2012 13:24	Adobe Acrobat D...	1 439 KB
Klausen et al 2015 Triassic Snadd in Baren...	30.03.2015 13:15	Adobe Acrobat D...	14 911 KB
Klausen_et_al_2014_Triassic_Snadd_Fm.pdf	08.09.2014 13:20	Adobe Acrobat D...	5 069 KB
Klitzke_et_al_Barents_Sea_Region_2014-pr...	21.01.2015 15:22	Adobe Acrobat D...	2 951 KB
Gabrielsen et al. 1990_Structural_E_BS_NP...	11.03.2013 15:45	Adobe Acrobat D...	9 099 KB
Gac_et_al_2013_ultra_deep_EBB.pdf	20.01.2015 12:15	Adobe Acrobat D...	1 008 KB
Abstract.doc	15.09.2014 16:04	Microsoft Word 9...	234 KB
Copy of NuAge_090214-120115-run1.xls	22.01.2015 13:06	Microsoft Excel 97...	1 390 KB
CPP_Svalbard_Bjornoya_Copy of CUMUL...	17.12.2014 15:35	Microsoft Excel 97...	2 429 KB
2013 Wintermeeting poster-1_Edina_Final...	07.01.2013 15:12	Adobe Acrobat D...	130 KB
ICPMS data	26.10.2020 16:31	File folder	
Franz Josef presentation.pptx	09.01.2013 12:06	Microsoft PowerP...	11 334 KB

- Raw & processed data mixed
- 100+ files in one folder
- No naming convention
- PDF literature mixed with TIFF data
- No README

Organize before you start

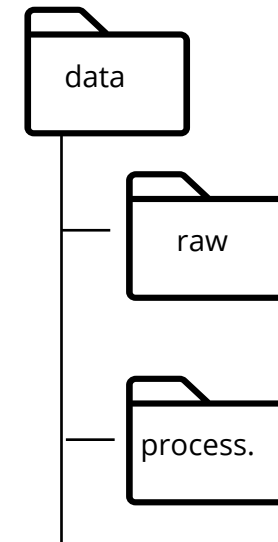


How would you and your colleagues typically looking for files?

Create a clear and efficient folder hierarchy!

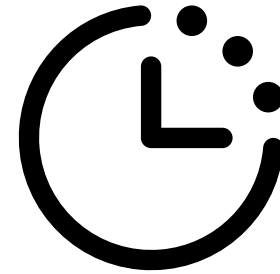
Best practices

- Avoid overly broad folders that accumulate mixed content
- Avoid a too deep structure
- Keep active and completed work separate
- Protect raw data by storing it separately

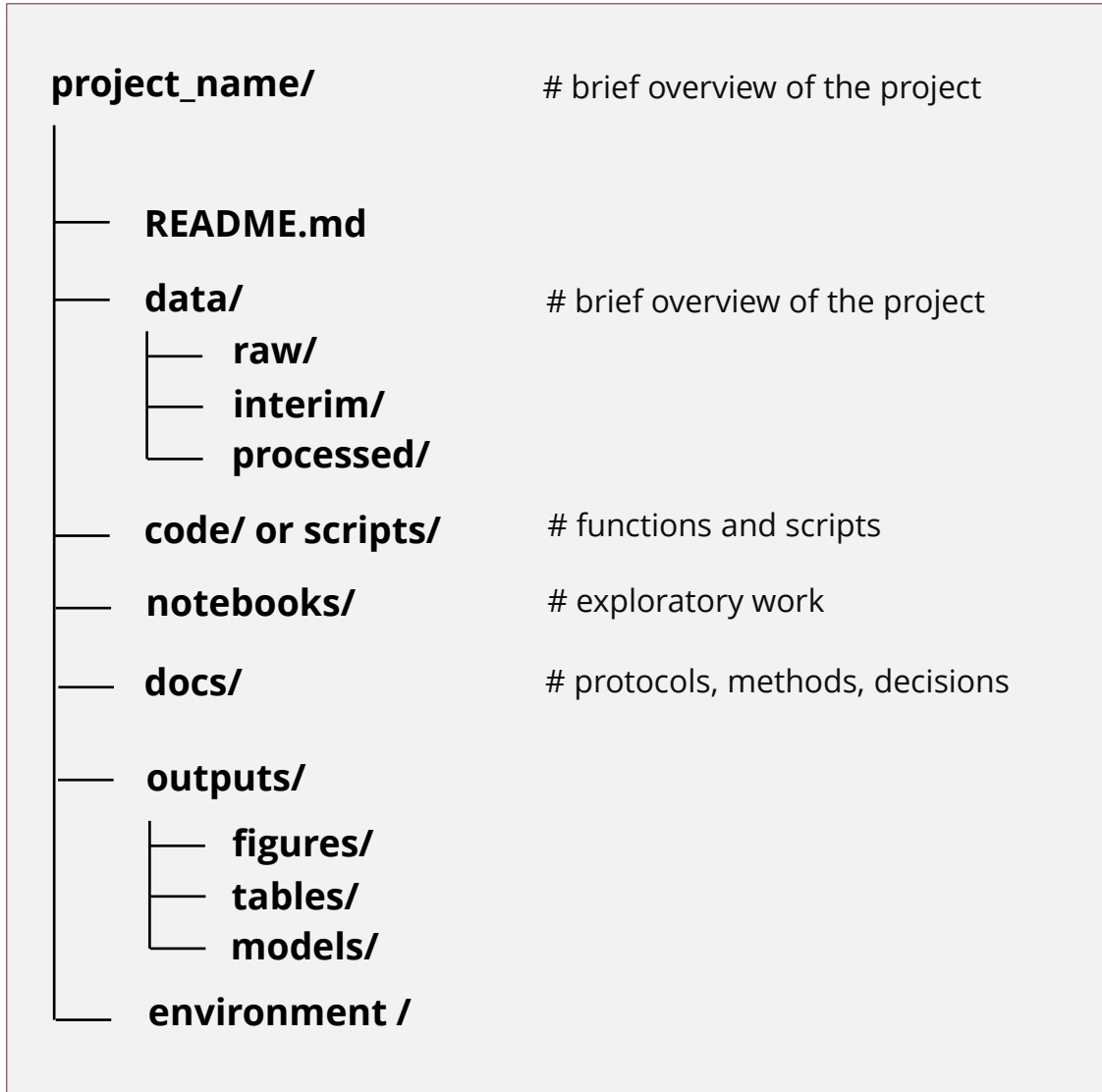


Best practices

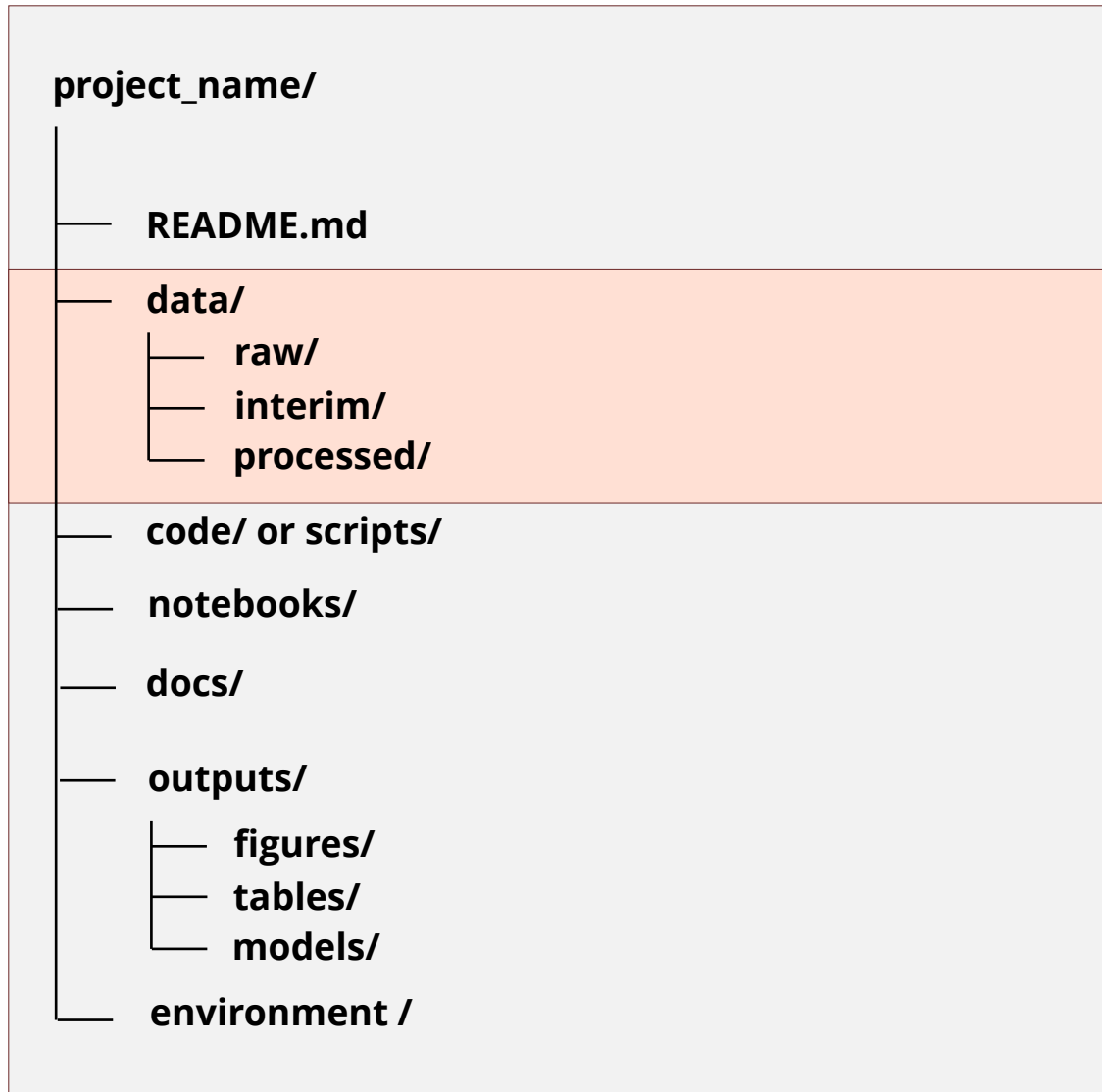
- Avoid overly broad folders that accumulate mixed content
- Avoid a too deep structure
- Keep active and completed work separate
- Protect raw data by storing it separately
- Schedule regular times to tidy and reorganize!



Research Project Folders Layout



Research Project Folders Layout



raw/

- Exactly what you collected or received.
- Never modify these files.
- Allows always going back to the source.

interim/

- Temporary working data.
- Cleaning, merging, harmonising.
- Useful for debugging and replicating steps

processed/

- What you analyse.
- Final clean tables, extracted variables.
- Ready for statistics, modelling, visualisation.

Take-home

- **Organize before you start**
- **Never modify raw data**
- **Script all processing steps**
- **Use clear, systematic folder names**
- **Organise data by its stage (raw → interim → processed).**
- **Separate data, code and results**

File Naming Convention

Core Principles

1

Consistency

Research team must agree upon and consistently use the naming system

2

Logical Ordering

Arrange elements from the most general to the most specific.

3

Brevity & Description

Keep names relatively short (< 32 chr.) but long enough that they still make sense

4

Machine-Readability

Avoid characters that may be problematic for different software or operating systems

Machine-Readability

- **DO** use only alphanumeric characters - letters (A–Z, a–z) numbers (0–9) -, hyphens (-), or underscores (_) as separators.
- **DON'T** use spaces or special characters like ~ ! @ # \$ % ^ & * () ; < > ? , [] { } ' " | \ / : * ?.
- **DON'T** use periods, except for the one before the file extension (e.g., .csv, .pdf).

Elements to include

Survey	20240212_SSprojA_survey_raw_v001.csv
Interview	20230315_UrbanYouth_interview01_transcript_v001.txt
Microscopy	20230518_CellTrack_fluorescence_40x_field12_v001.tif
Field Bio	20230722_BirdSurvey_acoustic_siteA_v001.wav

Elements to include

Survey	20240212_SSprojA_survey_raw_v001.csv
Interview	20230315_UrbanYouth_interview01_transcript_v001.txt
Microscopy	20230518_CellTrack_fluorescence_40x_field12_v001.tif
Field Bio	20230722_BirdSurvey_acoustic_siteA_v001.wav

Record the **date in YYYYMMDD** format to maintain correct chronological sorting across files.

Elements to include

Survey	20240212_SSprojA_survey_raw_v001.csv
Interview	20230315_UrbanYouth_interview01_transcript_v001.txt
Microscopy	20230518_CellTrack_fluorescence_40x_field12_v001.tif
Field Bio	20230722_BirdSurvey_acoustic_siteA_v001.wav

Short **project or experiment identifier** to ensure the file can be traced back to its source.

Essential Elements to Include

Survey	20240212_SSprojA_survey_raw_v001.csv
Interview	20230315_UrbanYouth_interview01_transcript_v001.txt
Microscopy	20230518_CellTrack_fluorescence_40x_field12_v001.tif
Field Bio	20230722_BirdSurvey_acoustic_siteA_v001.wav

Describe the **file's content or conditions briefly**, such as the data type, instrument used, or processing state (e.g., raw_data, processed, microscope).

Elements to include

Survey	20240212_SSprojA_survey_raw_v001.csv
Interview	20230315_UrbanYouth_interview01_transcript_v001.txt
Microscopy	20230518_CellTrack_fluorescence_40x_field12_v001.tif
Field Bio	20230722_BirdSurvey_acoustic_siteA_v001.wav

Use **consistent version numbers** (e.g., v001, v002, v010 instead of v1, v2, v10) for sequential sorting.

Standardised (open) file formats

Recognized as having best guarantees in terms of usability, accessibility and sustainability in the future

Data Type	Recommended	Avoid
Text	.txt, .pdf/A	.docx
Spreadsheet	.csv, .tsv	.xlsx
Audio	.wav, .flac	.m4a, .ogg, .wma
Images	.tiff, .png	.bmp, .gif, .psd
Video	.mp4	.mov, .avi
Statistics	.R, .Rdata, .csv	.sav, .dta, .por

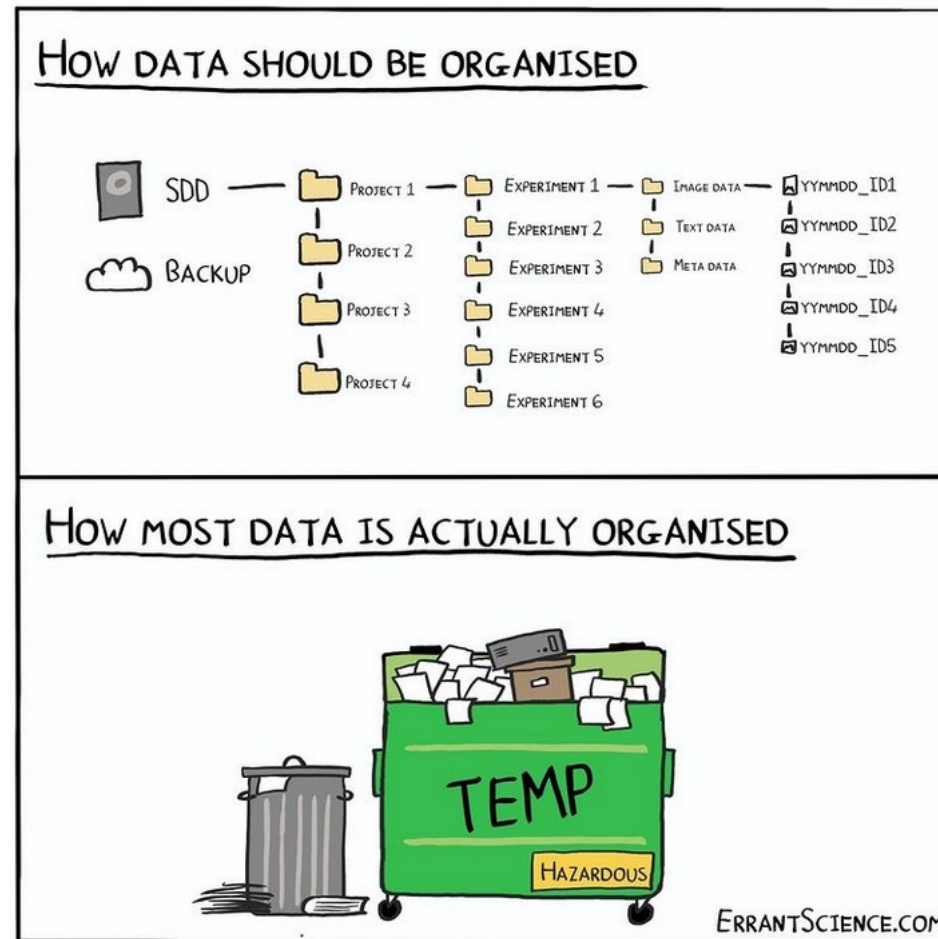
Open formats reduce dependency on specific software vendors and increase the chance that your files remain readable for decades.

Standardised (open) file formats

Trusted data archives often have guidelines and recommendations for this:

DataverseNO: <https://site.uit.no/dataverseno/deposit/prepare/#preferred-file-formats>
Data Archiving and Networked Services (DANS): <https://dans.knaw.nl/en/file-formats/>

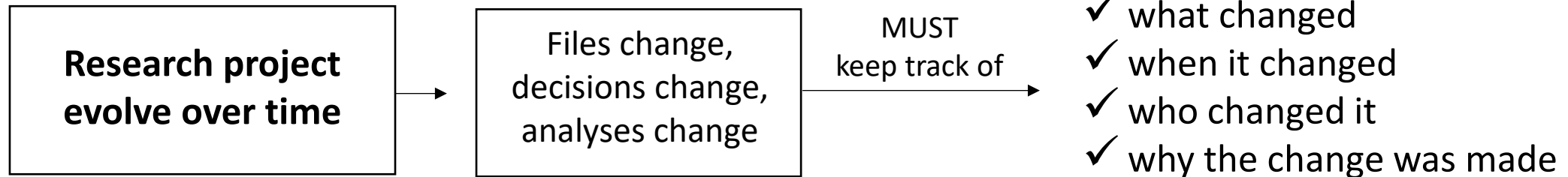
Break



By Matthew Partridge (@errantscience@mas.to Jul 05, 2021), CC-BY-NC

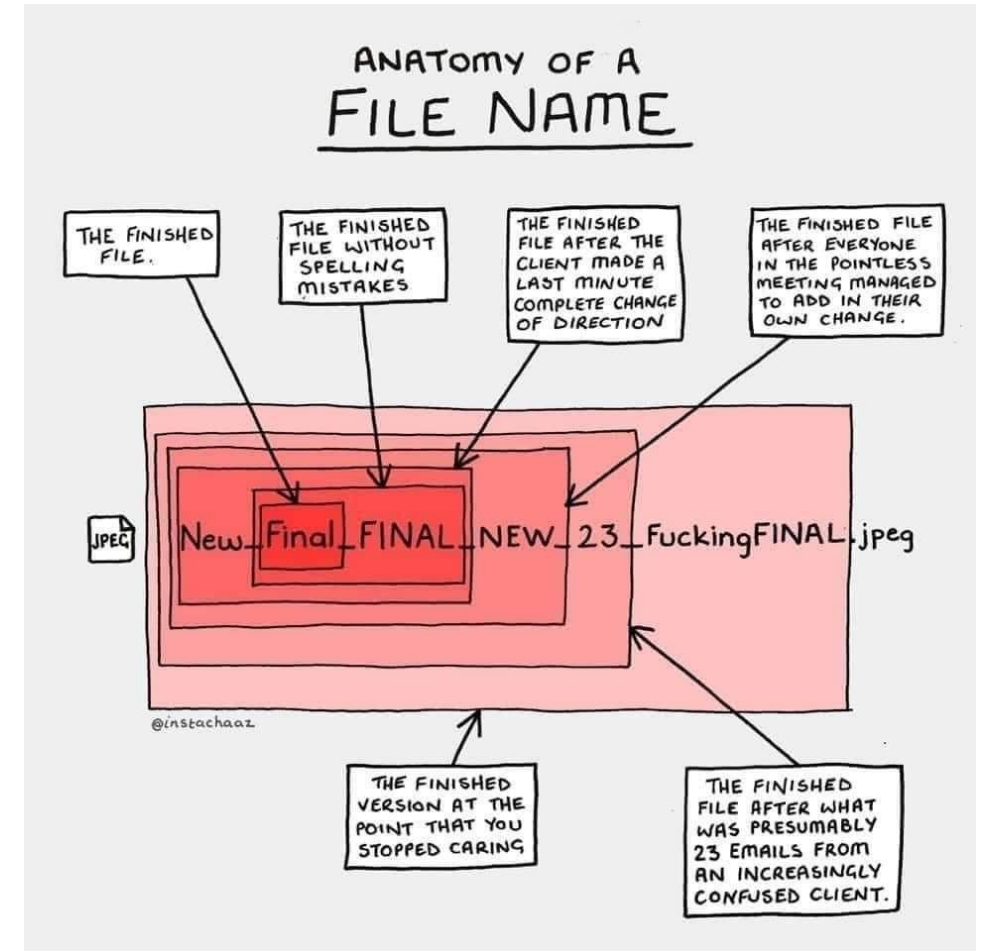
Version control

Why version control matter



What is versioning?

- Systematic way to save new copies of data, documents, or code whenever changes are made
- Help track progress
- Allows you to retrieve earlier versions.



Source: <https://www.workingtype.com.au/blog/2023/10/23/the-science-of-filenames>

Manual versioning

You rename the file to indicate the version.

DataFileName_v1_0 original document

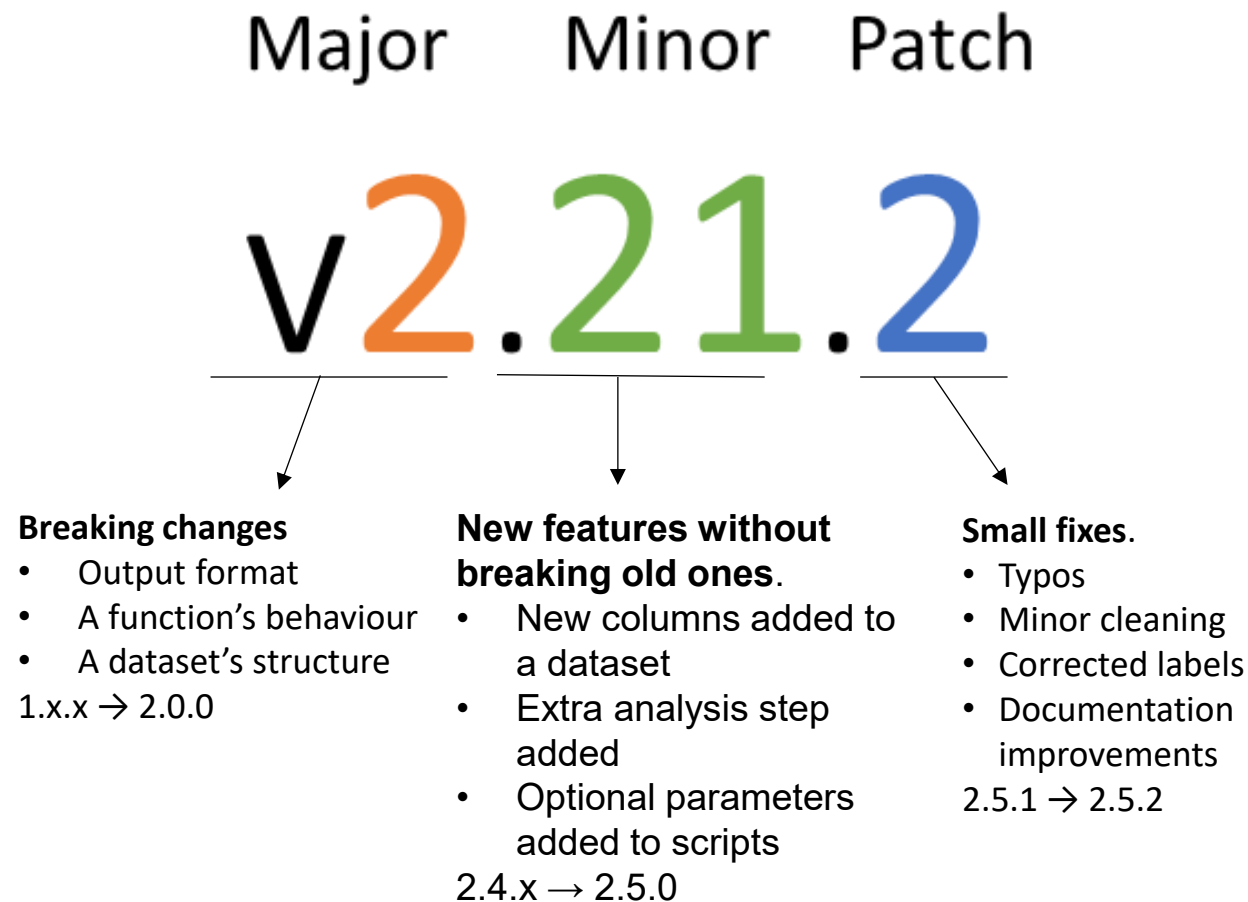
DataFileName_v1_1 original doc. with minor revisions

DataFileName_v2_0 doc. with substantial revisions

- Word documents, Excel files, PDFs, Figures and images, Audio/video, Anything NOT compatible with Git (binary formats)
- Very simple, works for any file type, easy to understand without training, no tools required
- Easy to make mistakes, no record of *who changed what*, no “diff” history, can clutter folders when many versions accumulate, hard to collaborate at scale

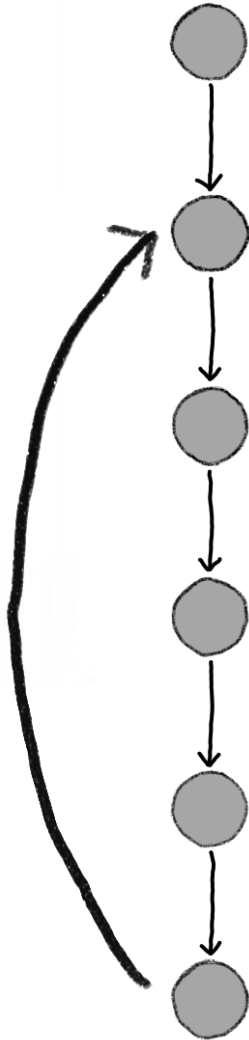
Semantic versioning

Used in computational research. Analysis scripts (R, Python, Stata), computational pipelines, shared analysis notebooks, published datasets (Zenodo, OSF, Figshare)



Automated Version Control

Software that tracks changes to files automatically.



The key idea

- Records **every** change
- Stores the **entire** history
- Allows you to **restore previous versions**
- Multiple people **work on the same files**

Useful when

- many people edit the same files
- your analysis changes over time
- you use code
- you want reproducibility
- you need to restore earlier results
- you want transparency for publication or reviewers

Git: Popular Version Control System

Free and open-source version control system.
The tool that tracks changes.

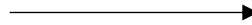


- ✓ records every change
- ✓ remembers who changed what
- ✓ lets you go back in time
- ✓ keeps a full history of your work
- ✓ allows multiple people to work without overwriting each other

Github

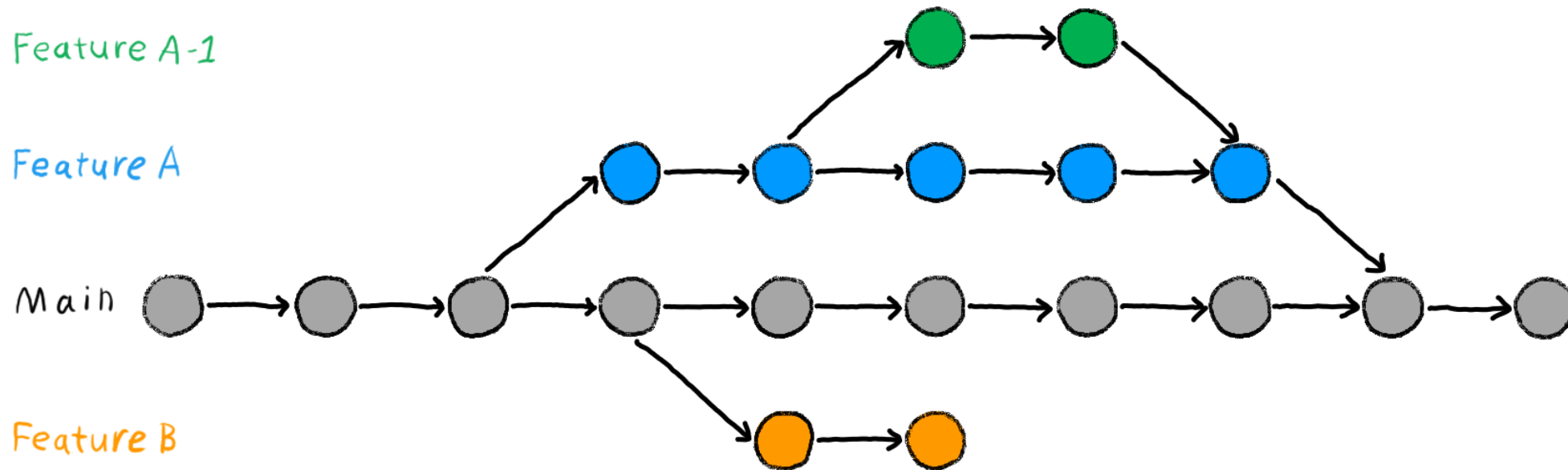
The online library that stores and shares these notebooks.

You *can* use Git without GitHub. You *cannot* use GitHub without Git.



- ✓ stores a copy of your Git project online
- ✓ helps you share your work with others
- ✓ keeps backups
- ✓ enables collaboration, reviews, comments, and transparency
- ✓ lets you publish code with a DOI
- ✓ can make your project public or private

Collaboration using Git and GitHub



Version Control with Git

Hands-on workshop on Version Control using Git and GitHub.

Time and place: Nov. 10, 2025 1:00 PM – Nov. 12, 2025 4:00 PM, UiO, Blindern, Georg Sverdrups Hus, DSC Oasen

[Add to calendar](#)



dScience logo and Data Management Week 2025 Fall edition

```
project_social_science/
├─ participant01
│   ├─ Interview notes FINAL.docx
│   ├─ audio recording 1.m4a
│   ├─ transcript_v3_final.docx
│   └─ survey-results(edited).xlsx
│
├─ participant02
│   ├─ raw data
│   │   ├─ interview_audio.m4a
│   │   └─ transcript.docx
│   └─ Notes from interview 2.docx
│
├─ participant03
│   ├─ audio2_final.mp3
│   ├─ TRANSCRIPT_clean.docx
│   ├─ survey.xls
│   └─ Interview notes.docx
│
└─ extra_stuff_mixed/
    ├─ random_notes.docx
    └─ survey_final_reallyFINAL.xlsx
```

- All messy participant folders sit inside **one project folder**
- Irregular names
- Mixed file types
- No raw/processed separation
- Extra unstructured folder (“extra_stuff_mixed”)

```

project_social_science/
├── participant01
│   ├── Interview notes FINAL.docx
│   ├── audio recording 1.m4a
│   ├── transcript_v3_final.docx
│   └── survey-results(edited).xlsx
│
├── participant02
│   ├── raw data
│   │   ├── interview_audio.m4a
│   │   └── transcript.docx
│   └── Notes from interview 2.docx
│
├── participant03
│   ├── audio2_final.mp3
│   ├── TRANSCRIPT_clean.docx
│   ├── survey.xls
│   └── Interview notes.docx
│
└── extra_stuff_mixed/
    ├── random_notes.docx
    └── survey_final_reallyFINAL.xlsx

```

```

project_social_science/
├── participants/
│   ├── p01/
│   │   ├── raw/
│   │   │   ├── interview_audio.m4a
│   │   │   ├── transcript_raw.docx
│   │   │   └── survey_raw.xlsx
│   │   └── processed/
│   │       ├── transcript_cleaned.docx
│   │       └── survey_clean.csv
│   └── p02/
│       ├── raw/
│       │   ├── interview_audio.m4a
│       │   ├── transcript_raw.docx
│       │   └── survey_raw.xlsx
│       └── processed/
│           ├── transcript_cleaned.docx
│           └── survey_clean.csv
│
├── docs/
│   ├── interview_protocol.md
│   └── codebook.md
│
└── metadata/
    ├── participants_overview.csv
    └── data_dictionary.md

```

Exercise 4: Folder structure

Write and discuss

**Think about the ideal folder structure for
your research project.**

Then discuss in pairs or as a smaller group.

04:00

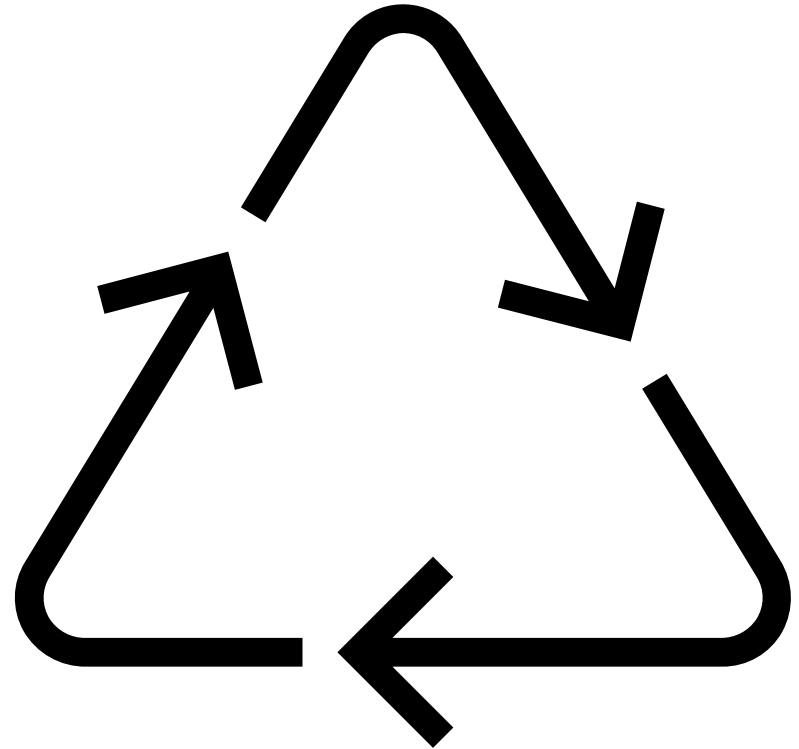
Data documentation

Documentation

- Systematically documented research data is key to making the data publishable, discoverable, citable, and reusable (and FAIR)
- Clear and detailed documentation **improve the overall data quality**

Why documentation?

- Helps others (and «future you») to understand your project and reuse your data:
 - all you need to know about your data
 - all your collaborators need to know about your data
 - all you need to know about your collaborators data
 - all anyone else who wants to work with your data needs to know about it
- You must have documentation when you archive your data



What documentation?

- The **project-level documentation** explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc.
- **Data-level or object-level documentation** provides information at the level of individual objects such as images or variables in a database/table or transcripts, etc.

Examples of documentation

- Field journal
- Lab journals and experimental protocols
- Scripts for analysis
- Questionnaires, codebooks, data dictionaries
- Software syntax and output files
- Methodology reports
- Geolocation, topography, orientation (e.g. when collecting a sample)
- Instrument settings and calibration



Data-level documentation: **Codebook**

- Explains all variables and their codes in the dataset
- It typically contains:
 - variable names, variable labels, variable codes, variable formats, missing data (in quantitative research)
 - codes, code definitions, examples of what to include with a given code (in qualitative research)
- Can be also called **Data Dictionary**

Data-level documentation: **Codebook**

Example

Sex: biological sex self-reported by the participant (0 – Female, 1 – Male)

Gender: self-reported gender (0 – Female, 1 – Male, 2 – Other)

Age: age in months at the date of test (automatically calculated in the system as the difference between the date of test and the date of birth)

	A	B	C	D	E	F	G
1	subjid	sex	gender	recruitment	condition	group	age
2	10001	1		2	9999	9999	272
3	10002	1	1	1	2	2	187
4	10003	1		1	9999	9999	291
5	10004	1	1	1	1	1	300
6	10005	1	1	1	1	1	288
7	10006	1		6	0	9999	254
8	10007	1	1	4	1	2	233
9	10008	0	2	9999	0	0	266
10	10009	1	1	2	1	1	195
11	10010	1	2	2	1	1	207
12	10011	1	1	1	1	2	234
13	10012	1		2	1	2	231
14	10013	1	1	0	1	2	257
15	10014	0	2	0	0	0	309
16	10015	0	2	1	2	2	254
17	10016	0	2	5	0	9999	161
18	10017	1		6	0	9999	
19	10018	1	1	1	1	2	314
20	10019	1	1	5	0	0	188
21	10020	1		0	1	9999	
22	10021	1	1	4	9999	9999	
23	10022	1		1	1	2	294
24	10023	1	1	0	9999	9999	
25	10024	1	1	1	1	1	291
26	10025	1	1	2	1	1	214

Controlled Vocabularies

- Standardized and organized arrangements of preferred terms and phrases used to index and, subsequently, retrieve data and other content
- Provide a consistent way to describe data and their variables
- Common types include term lists, authority files, and thesauri

Controlled Vocabularies

	A	B	C	D	E	F	G	
1	participant number	age	lifestyle	religion	sex	salary group	occupation group	
2	1	45	1	0	1	A	B	
3	2	54	2	1	1	A	D	
4	3	36	1	0	2	B	A	
5	4	47	1	NA	2	C	C	
6	5	30	2	0	0	B	A	
7	6	32	3	0	1	D	B	
8	7	52	3	0	1	A	NA	
9	8	34	1	1	2	A	A	
10								

	A	B	C	D	E	F	G	
1	ID	yob	habits	beliefs	gender	income	profession	
2	101	1971	2	1	1	A	1	
3	102	1983	4	0	2	B	2	
4	103	1977	3	0	1	C	4	
5	104	1974	1	2		A	2	
6	105	1987	1	1	2	A	1	
7	106	1994	3	0	1		1	
8	107	1990	2	2	2	C	2	
9	108	1988	1	3	1	B	3	
10	109	1982	4	1	0	C	4	
11								

Exercise 5: Data variables

Think - Share

How could this table be improved?

cohort-01.xlsx

43	y	m	
12	y	female	
18	m	Male	
9	m	F	
78	y	f	
32	y	male	



cohort-01.xlsx

age	age_unit	sex	
43	year	male	
12	year	female	
18	month	male	
9	month	female	
78	year	female	

cohort-01.xlsx

age_y	age_m	sex	
43	0	m	
12	0	m	
1	6	m	
0	9	f	
78	0	f	

cohort-01.xlsx

age_year	sex	
43.00	male	
12.00	female	
1.50	male	
0.75	female	
78.00	female	

03:00

How to create documentation?

- **More information** is better than less
- **Structured information** is better than unstructured information
- **Re-use templates** for good documentation
- It's become a convention to **create multiple README-files**, both for project-level documentation and for data-level documentation

How to create documentation?

README.txt-files:

- **Simple text files** that provide information about any elements (or any workflows or processes these elements have been used for) in the same directory
- **The first files to open**
- **Provide a map** for exploring your files
- Create **one** README.txt file **per folder**
- Content should **help navigation** through digital files and folders
- A **project-level** README.txt should give the **general project information** and a very coarse overview of file contents and locations
- A **data- (object-) level** README.txt would be **more specific** as to what each file contains

Documentation vs. metadata?

"While data **documentation** is meant to be read and understood by **humans**, **metadata** (which are sometimes a part of the documentation) are primarily meant to be processed by **machines**."



Protection
Organization
Documentation (and metadata)
Classification
Short-term storage
Licensing
Sharing
Long-term preservation
Dissemination (publishing)
Reusing

The logo for Data Management Week 2025 features a stylized orange folder icon on the left, with a bracket connecting it to the text 'Data Management Week 2025'. The text 'Data' and 'Management' are in orange, while 'Week' and '2025' are in white. Below '2025' is the text 'Fall edition' in white.

Data Management Week 2025

Fall edition

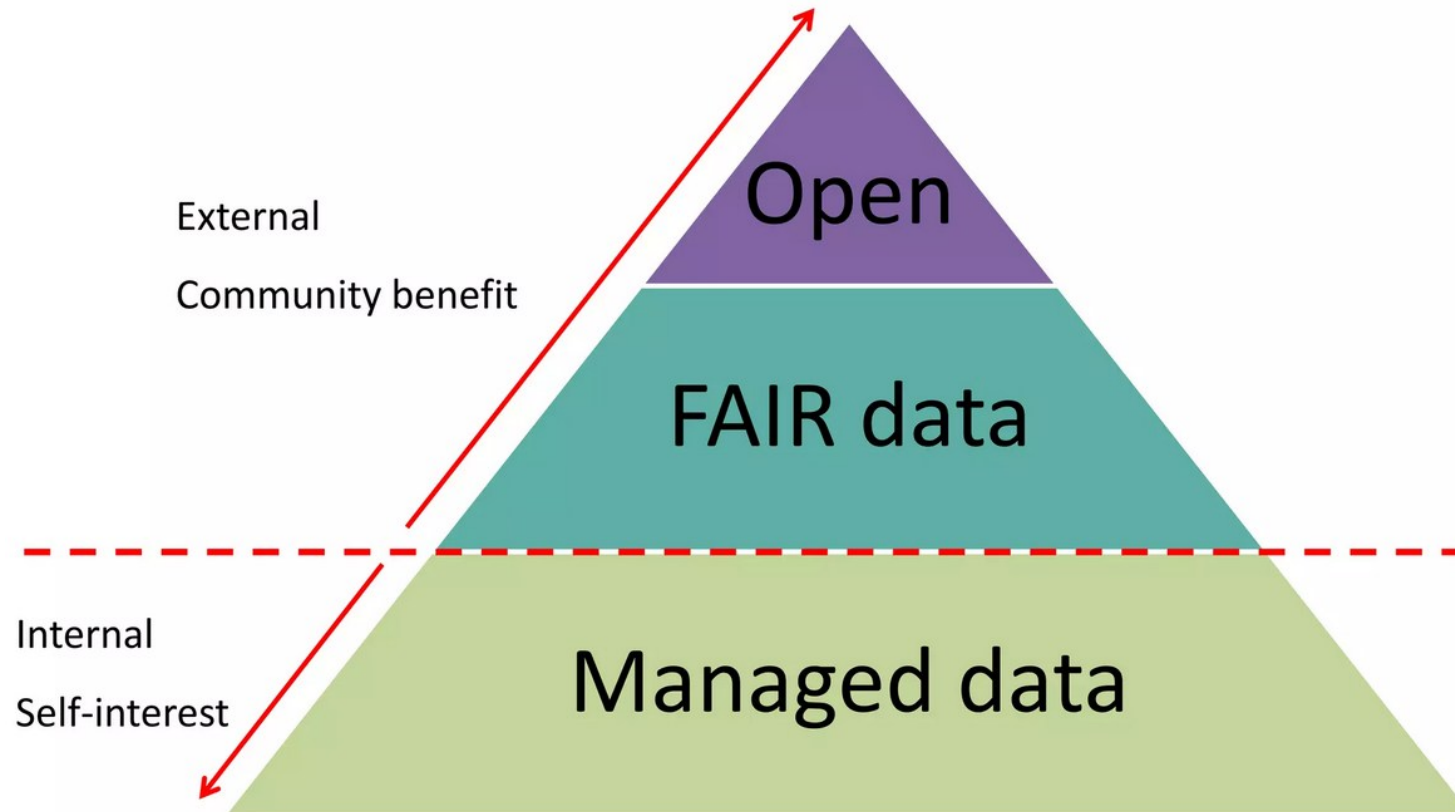
November

10th – 14th

Digital Scholarship Center

Monday	<i>Morning:</i> <u>Plan your research data management</u> <i>Afternoon:</i> <u>Version Control using Git and GitHub – Part 1</u>
Tuesday	<i>Morning:</i> <u>Share, archive, and reuse research data</u>
Wednesday	<i>Afternoon:</i> <u>Version Control using Git and GitHub – Part 2</u>
Thursday	<i>All day:</i> <u>Carpentry workshop: Databases and SQL</u>

As **open** as possible,
as **closed** as necessary



Need help with research data?

**Send an email to
research-data@uio.no**

More resources?

For employees

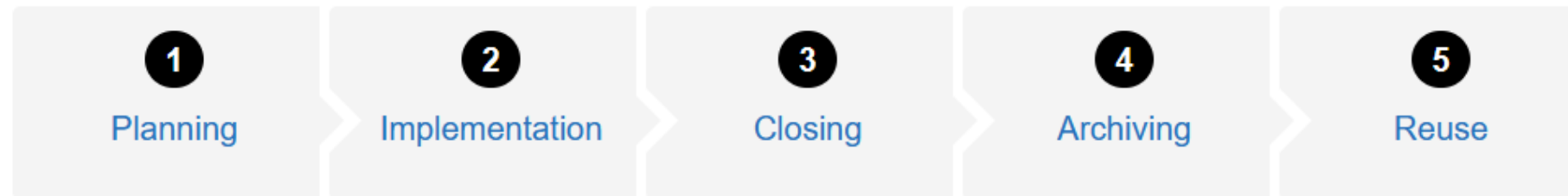
[≡ All content](#)[◀ Research support](#)

Research data

[Norwegian](#)

Research data management includes documentation, organising, licensing, sharing, archiving and reuse.

Data throughout the research project



[← Libraries and centres](#)

Digital Scholarship Centre (DSC)

The Digital Scholarship Centre (DSC) gathers and highlights digital resources from the University Library and UiO that can enhance your research activities. On this page, you will find an overview of methods and tools. We also offer specialized courses and workshops for PhD candidates, instructors, researchers, and other employees at UiO and OUS.

Find resources and courses about:

Open research and research data	+
Science visualization	+
Digital tools and methods	+
Literature search for systematic reviews	+

Open and reproducible research

Learn about how to make your research more open and reproducible and get involved in initiatives and communities that are interested in sharing and improving research at UiO.

More and more researchers and students across disciplines are implementing open research practices, preregistering their hypotheses, methods, and analysis plans and sharing research materials, data and analysis scripts. Digital Scholarship Center can help you learn about and implement these practices in your own research as well as advise on the policies and requirements from funders.

Open Science Lunch



Every last Thursday of the month we meet at noon to discuss topics related to open research.

ReproducibiliTea@UiO



Join us for a Journal Club where we read and discuss papers on open research and meta-science.

Norwegian Reproducibility Network



Join a broader community that aims to promote and enable rigorous, robust and transparent research practices in Norway

Courses and workshops



Click here for the list of upcoming and previous courses and workshops on open and reproducible research at UiO.



Open Science Lunch

Each last Thursday of the month at 12:00 we invite you to join us for a lunch seminar to hear about how to make your research more open. We will discuss research transparency and visibility, open publishing, data sharing, and more!

Time and place: Nov. 27, 2025 12:00 PM – 1:00 PM, [Zoom](#)

Reforming Research Assessment: from Principles to Practice

What does reform in research assessment look like in practice? Join us in this Open Science Lunch for three different perspectives on implementing reforms in research assessment.

ReproducibiliTea

Journal Club

**JOIN IN AND DISCUSS WITH FELLOW
STUDENTS AND RESEARCHERS**

**OPEN RESEARCH, REPRODUCIBILITY
and RESEARCH IMPROVEMENT**



Join us

Everyone is welcome to join us - whether you are an enthusiast of open and reproducible research, a skeptic, or a cautious explorer. Currently, all meetings are hybrid with the possibility of joining on-site at Blindern or via Zoom. Grab a cup of tea (coffee?) and join us!

Subscribe to our mailing list





Carpentry@UiO

Carpentry@UiO is a community of people who are passionate about learning, teaching, and sharing best practices and digital skills for making the research process more reproducible and effective. If you want to get involved, or join one of our workshops, check us out!



The Unix Shell

Shell speeds up repetitive and tedious processes. It is also essential skills needed to use high-performance computing (HPC) resources.



Version Control with Git

Git helps you to keep track of what you've done, for a better collaboration and for yourself in future. In the workshop we use GitHub as well.



Programming in Python

Python is now widely used in scientific computing with various powerful packages. Carpentry@UiO runs workshops for participants with no programming experience ("Plotting and Programming in Python" lesson) and for participants at intermediate level ("Programming with Python" lesson, episodes 10-12).



R for Reproducible Scientific Analysis

R is commonly used for statistical analysis, but it is also a powerful programming language. Workshops on R focuses on teaching best practices for scientific computing: breaking down analyses into modular units, task automation, and encapsulation. Workshops on R may use lessons from Data Carpentry instead.



Using Databases and SQL

Databases include powerful tools for search and analysis, and can handle large, complex data sets. The lesson will show how to use a database to explore research data by using SQL.

Give us your feedback!



<https://nettskjema.no/a/565173>



Follow us!



@DigitalScholarshipCentre
@SenterForDigitalForskerstøtte

DSC NEWS

Senter for digitalforskerstøtte
Digital Scholarship Centre



[https://sympa.uio.no/ub.uio.no/
subscribe/dsc-news/subscribe](https://sympa.uio.no/ub.uio.no/subscribe/dsc-news/subscribe)

Thank you!

Elisa Pierfederici, Camilla Hagen Blixhavn and Ivana Malovic

Digital Scholarship Center, University of Oslo Library

10.11.2025

Questions? Contact us at research-data@uio.no