2025

Basics of Research Data Management

Lib4RI

Felder, Fabian
Ulrich-Nath, Moushumi,
Bachofner, Anusch
Förster, Christian

Agenda

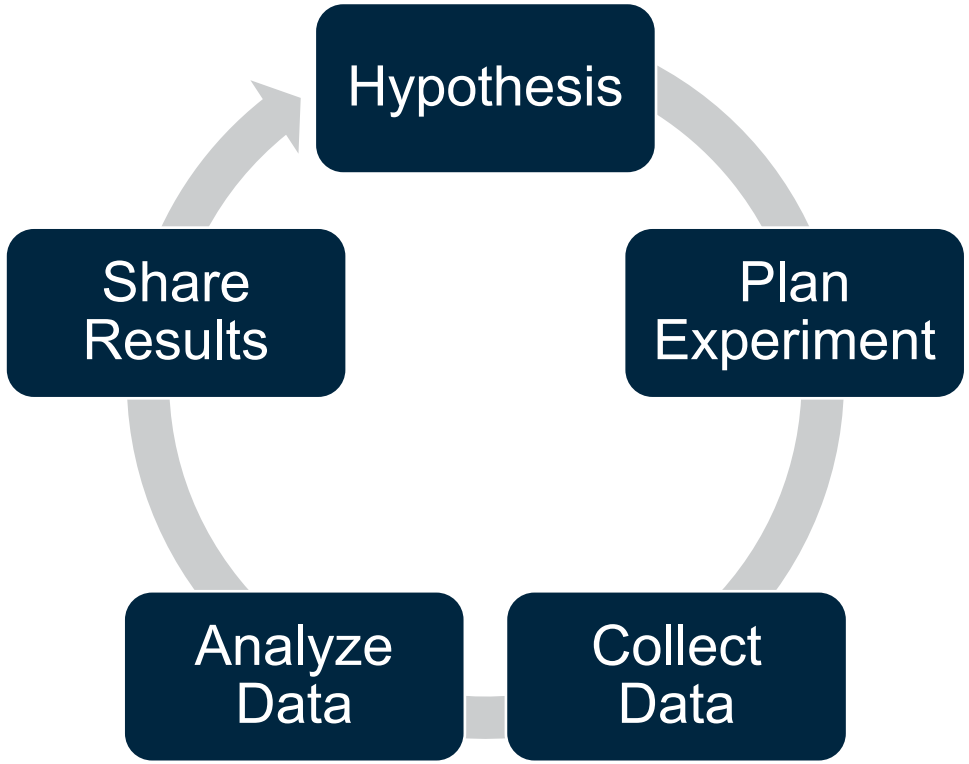
Topic	Time
Introduction	9:00 – 9:20
• Open Science, FAIR, RDM	
Data Collection, Processing and Analysis	9:20 – 10:00
• File folders, naming, versioning, formats	
Pause	10:00 – 10:10
Documentation	10:10 – 10:30
• README & Metadata	
Storage, Preservation, and Sharing	10:30 – 10:50
• Repositories, data availability statements, licensing	
Pause	10:50 – 11:00
RDM Services and Support	
• Eawag	11:00 – 11:30
• Empa	11:30 – 12:00

Lib4RI

Welcome

Open Science, FAIR, and RDM

Scientific Method vs Research Data Management

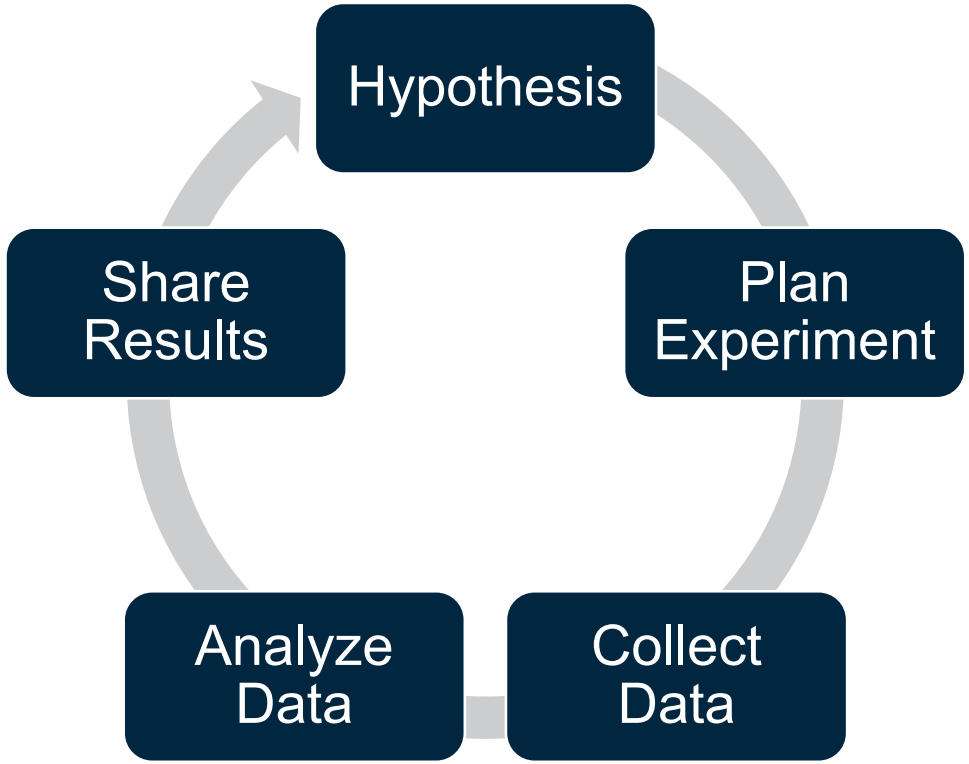


Share Results

Results:

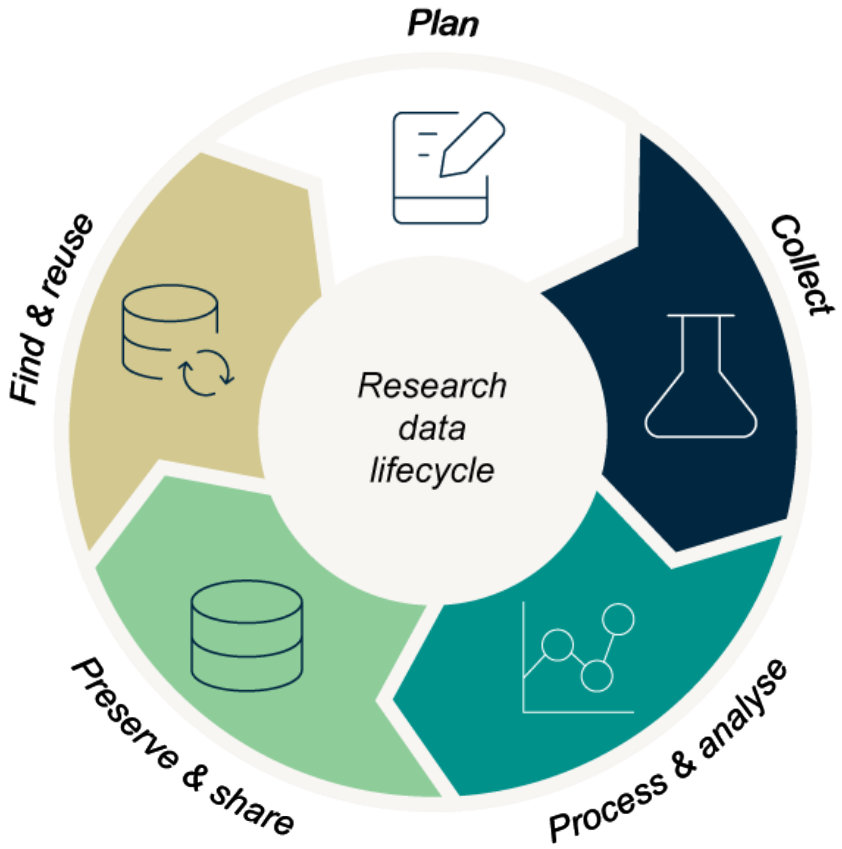
Conclusion:
Amazing results. Requires further research.

Scientific Method vs Research Data Management



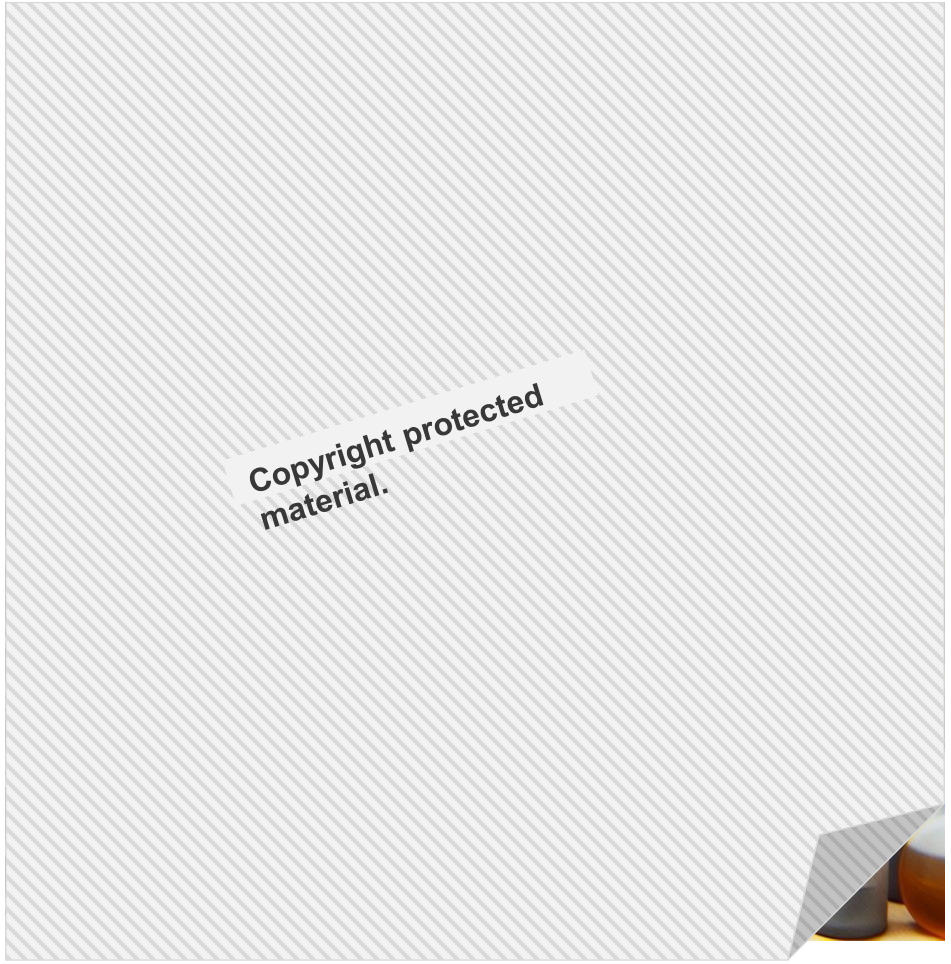
What happens to your data?

Research Data Management



Research data management (RDM) is the process of organizing, storing, preserving, and sharing data that is generated or used in a research project.

Why bother?



What is Open Science?

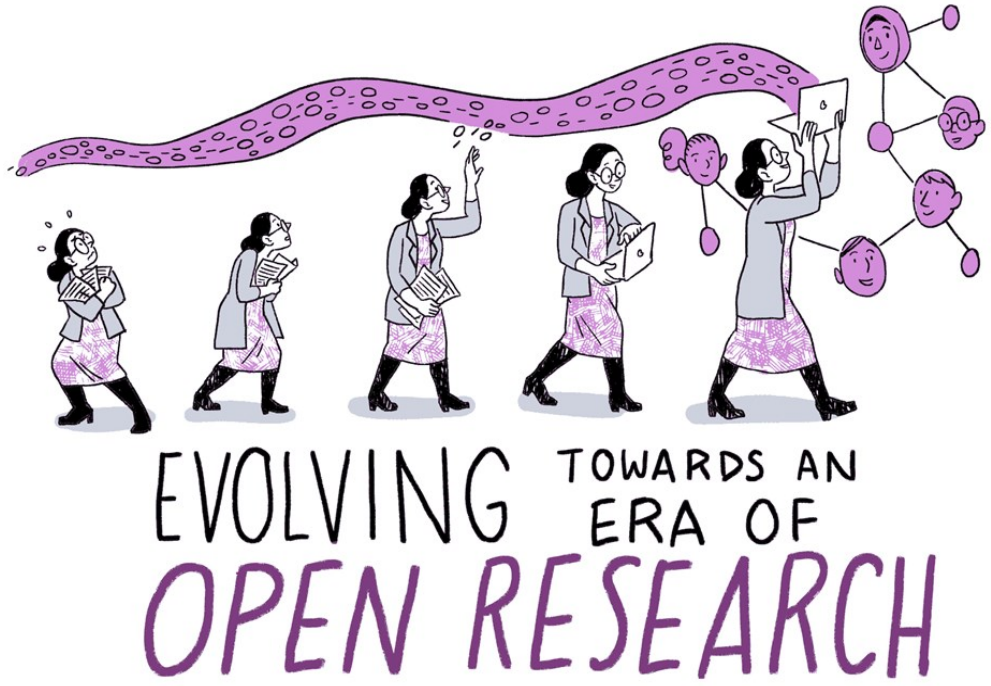
- Q1: What's the first word or phrase that comes to mind when you hear Open Science.

- Q2: I believe Open Science is the future of research.

www.menti.com

Open Science is a culture change.

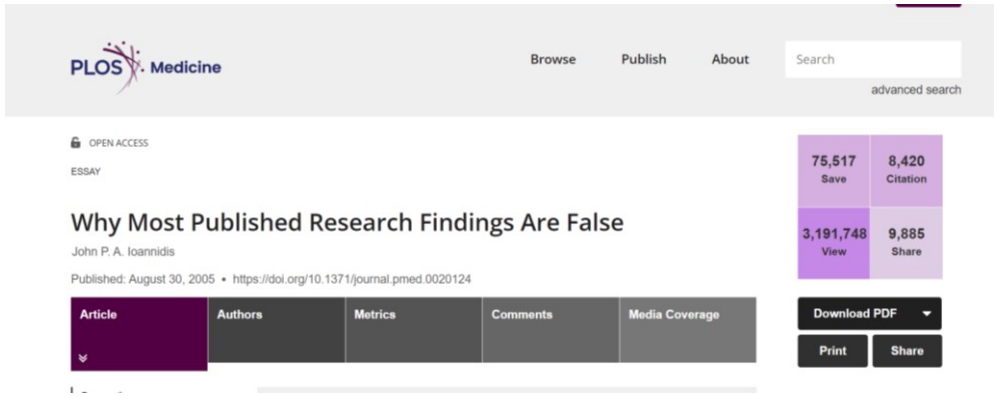
- Accessible
- Transparent
- Collaborative



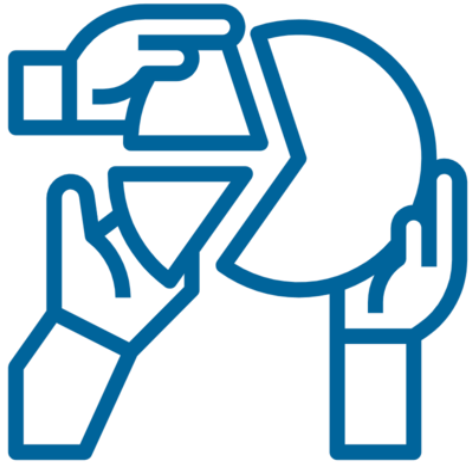
The Turing Way project. CC-BY 4.0. DOI:10.5281/zenodo.3332807.

Open Science is a culture change.

Reproducibility/Replicability Crisis



Global Collaborations



Arm Okay, CC BY 3.0, via Wikimedia Commons

Lib4RI

Open Science is the future.

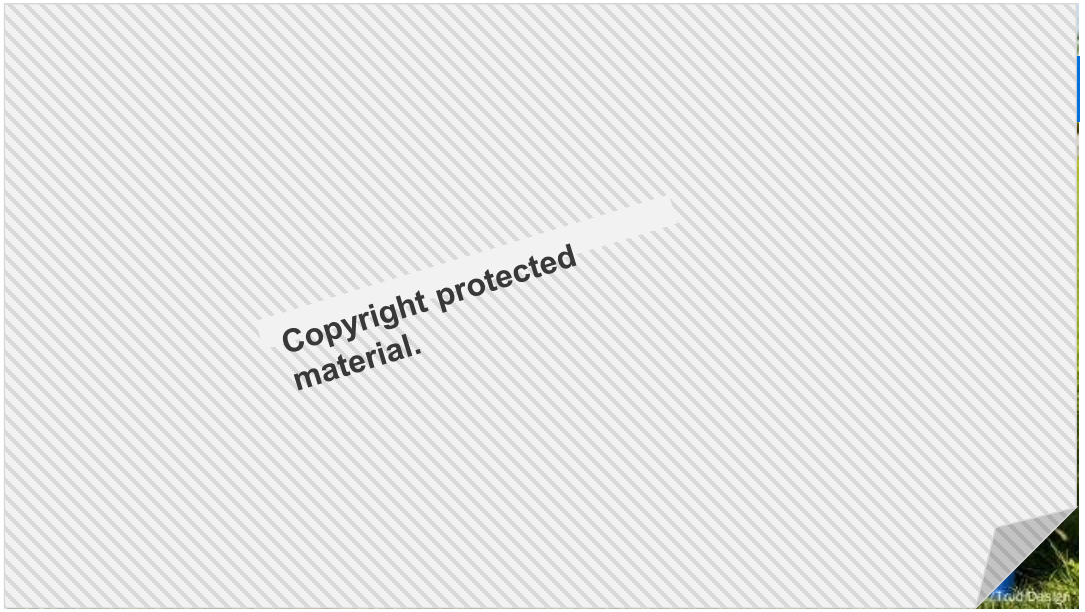
Policy	Incentives	Trainings	Infrastructure
Empa	SNSF	Data Management Campus	Zenodo
Eawag	Horizon		ERIC
PSI			SciCat
WSL			EnviDat

Open Science is not just a data dump.

Cezary p, CC BY-SA 4.0, via Wikimedia Commons

Fotolia/TrudiDesign. Tackling trash – DW – 04/23/2012

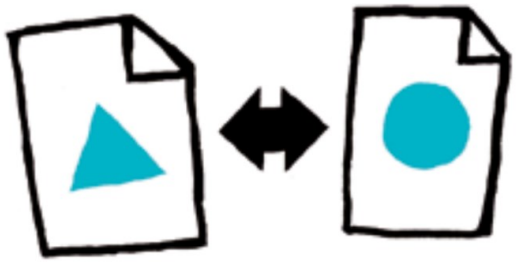
What are some ways you can make your data FAIR?



Fotolia/TrudiDesign. Tackling trash – DW – 04/23/2012

www.menti.com

FAIR are guiding principles for data management.



Metadata – where can people find the data?

Metadata – how can others access your data?

File formats – non-proprietary, standard.

Metadata – provides sufficient context to understand your data.

DOIs

Metadata is accessible.

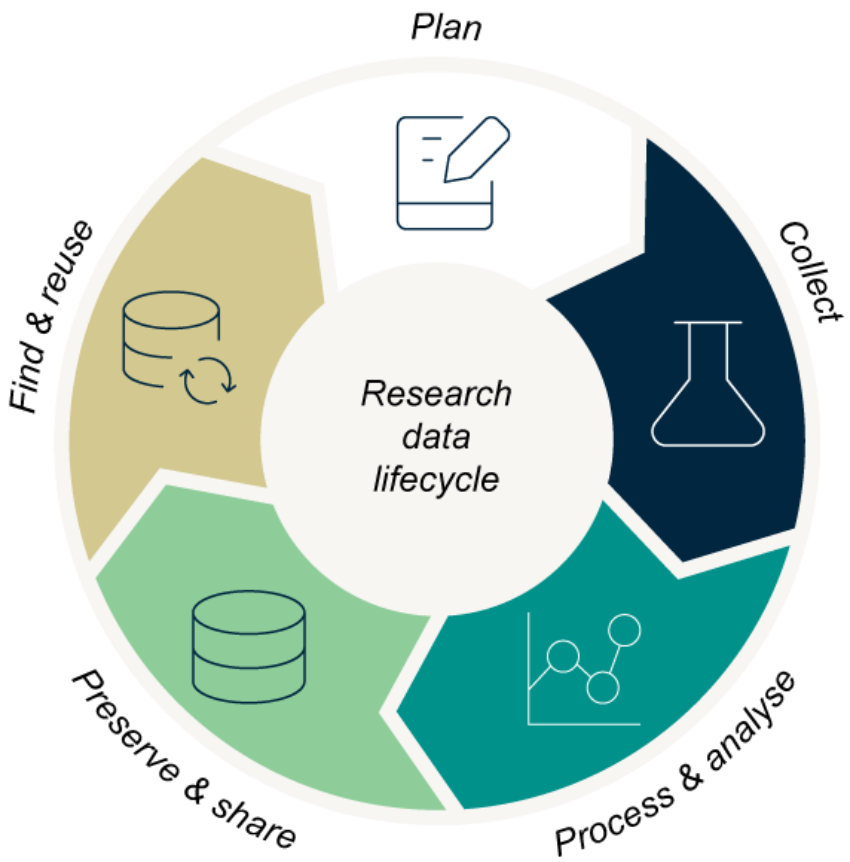
Vocabulary – standard.

Licenses – determines how others can reuse your data.

Data may be restricted.

Studio 4 minutes 34, Studio Lendroit.com, CC BY-SA 4.0, via Wikimedia Commons

Research Data Management



Data Management Plans

How to organize data (activity)
- Folder structures
- File naming
- File formats
- *Versioning*

Documentation (activitiy)
- README & Metadata

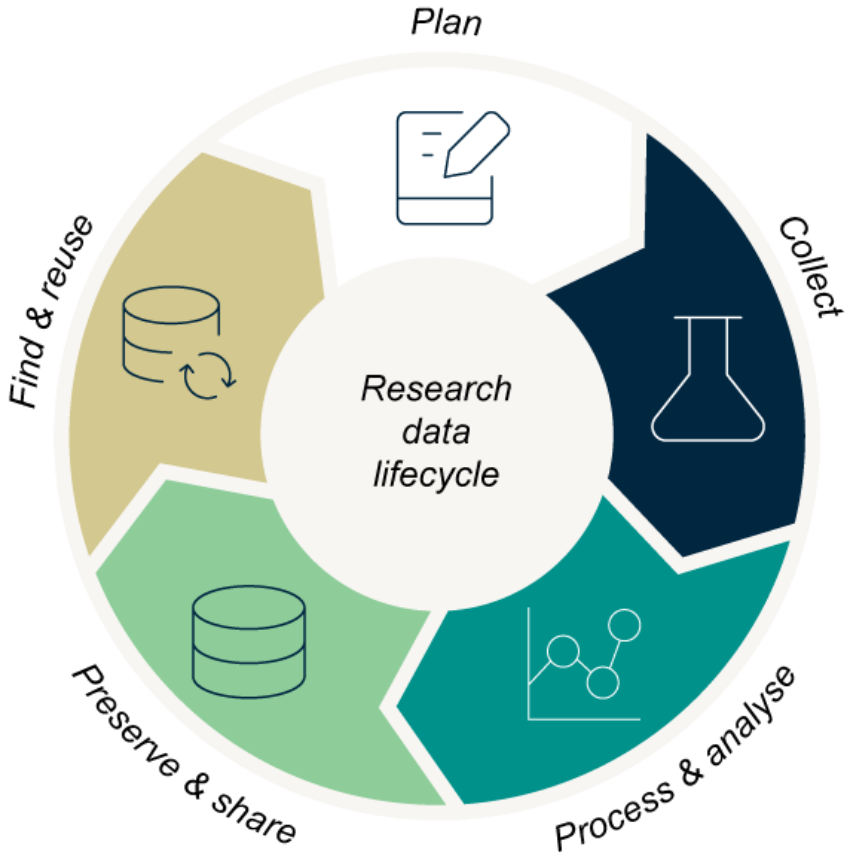
- Storage
- *Repositories*

- Data availability statements
- Licenses

Lib4RI

How to do RDM

Planning


Plan
Collect
Process & analyse
Preserve & share
Find & reuse
Research data lifecycle

Planning – the data management plan (DMP)

What types of data will be collected or generated?

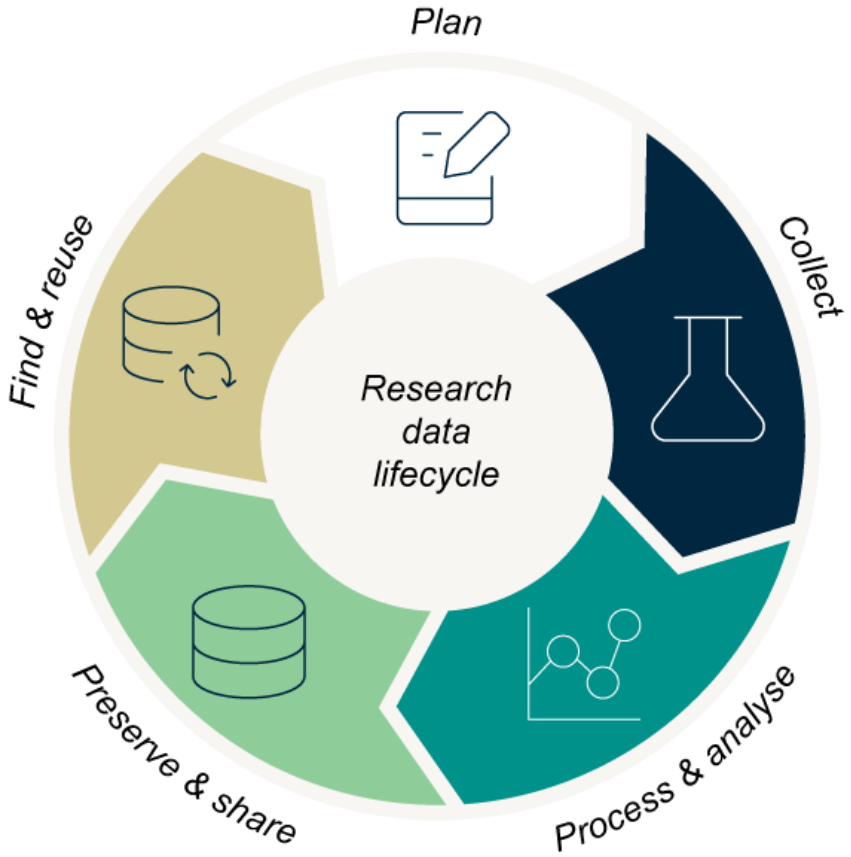
What type documentation will you provide with the data?

How will data be stored during and after the project?

How will your data be shared? How can it be accessed?

Are there any ethical, legal, or security issues to address?

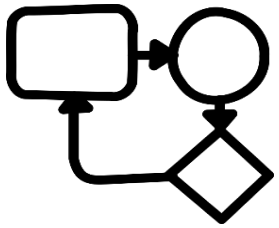
Data Collection, Processing, and Analysis



Data Collection

```
01010100 01101000
01101001 01101110
01101011 00100000
01100100 01101001
01100110 01100110
01100101 01110010
01100101 01101110
01110100 00101110
```

Data
observational, experimental, simulation…



Code
Applications, scripts…

</>

Metadata
Structured information associated with data (and code)
The Who, What, Where, Why & How of data

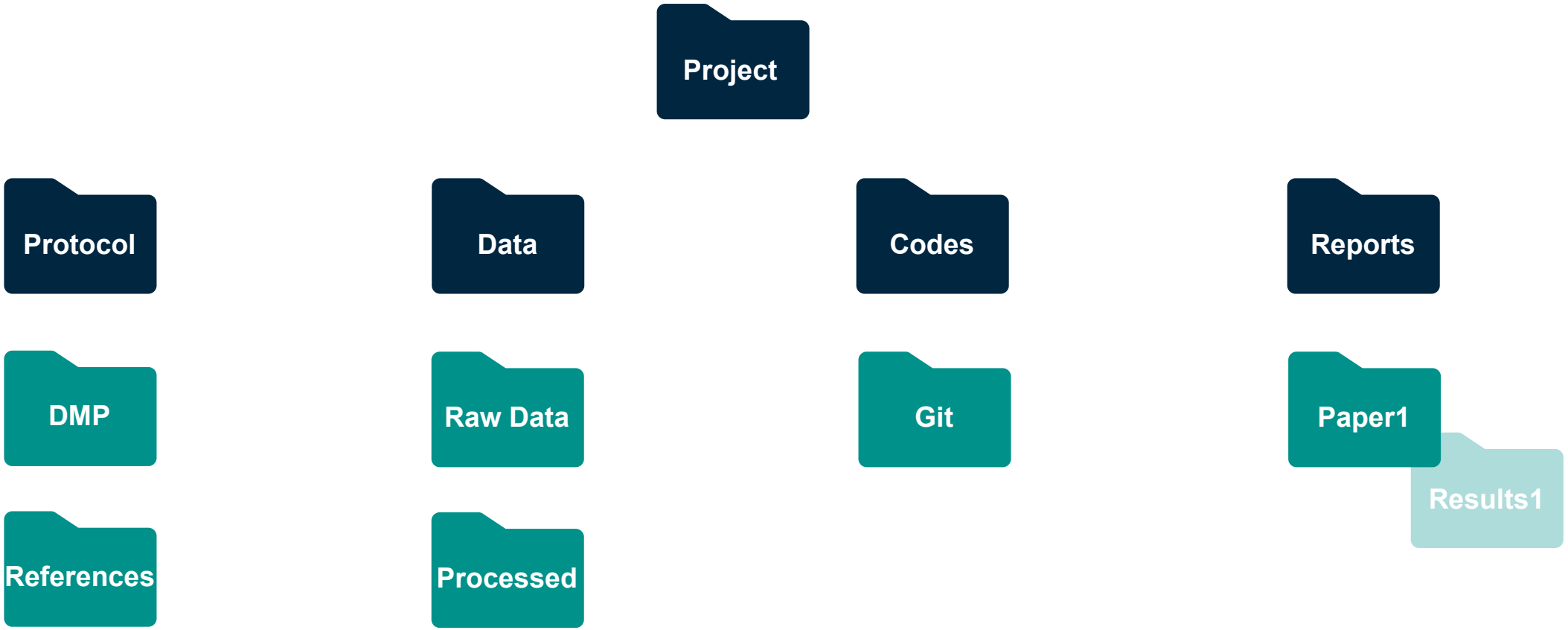
What to do with all this data, code, metadata?

File Organization

File Naming

Versioning

Organize data: File/Folder Structures



File Naming

Be consistent

No special characters or signs @°§°#¬@

Include Date: YYYYMMDD

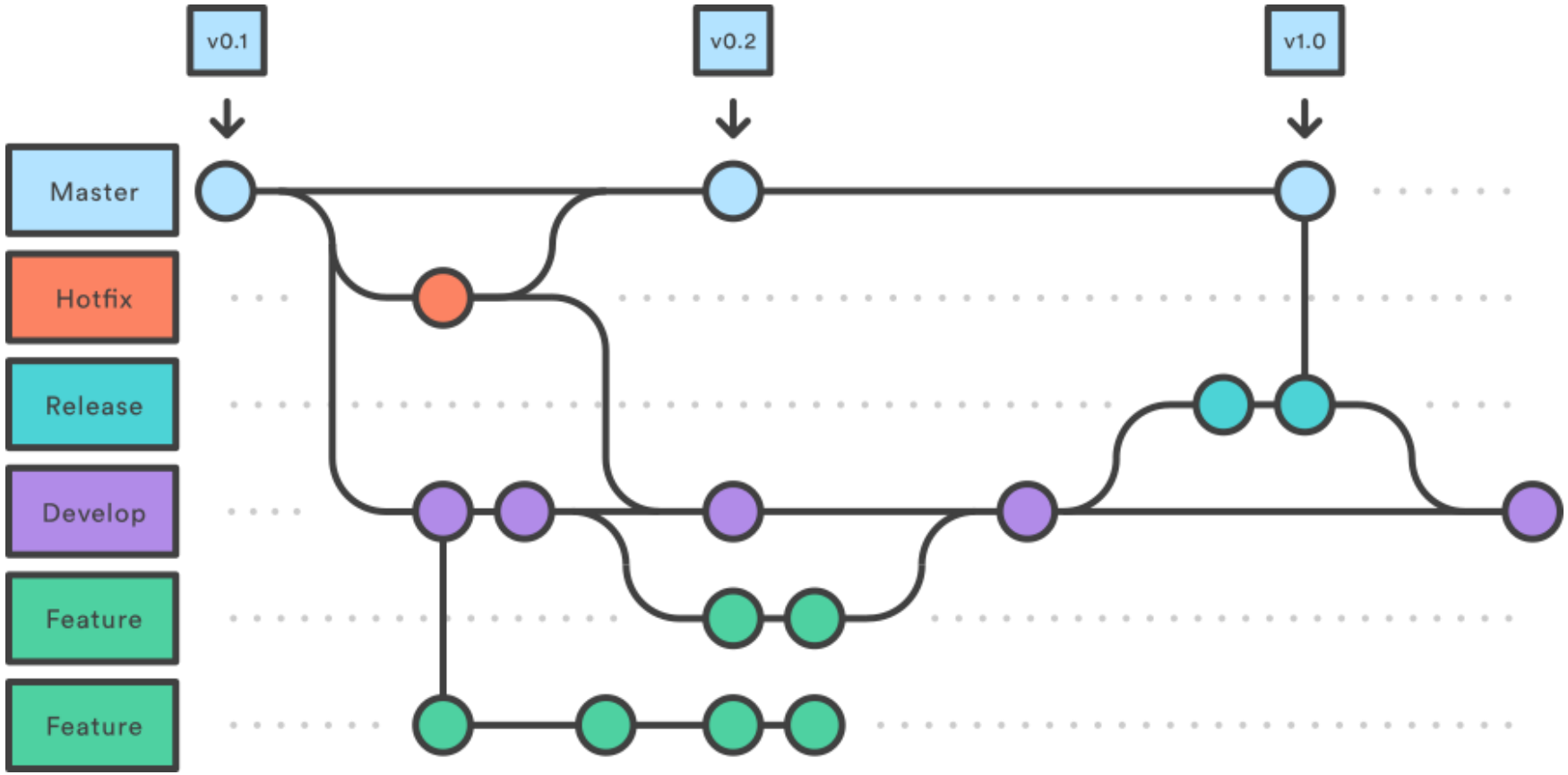
Project Name: short (< 30 characters) and descriptive

Document Type: data, code, results, paper...

Include version: v0.8, v1.2...

Tip: Already started your project?
Use Bulk Rename Utility (Windows), Renamer 6 (Mac), Rename/Thunar Bulk Rename (GNU/Linux).

Software Version Control



Tools for Software Version Control

https://git-scm.com/

- CLI (*Command Line interface)*
- GUIs (*Graphical User Interfaces)*
  https://git-scm.com/downloads/guis

Workstation

Your own server

Gitea

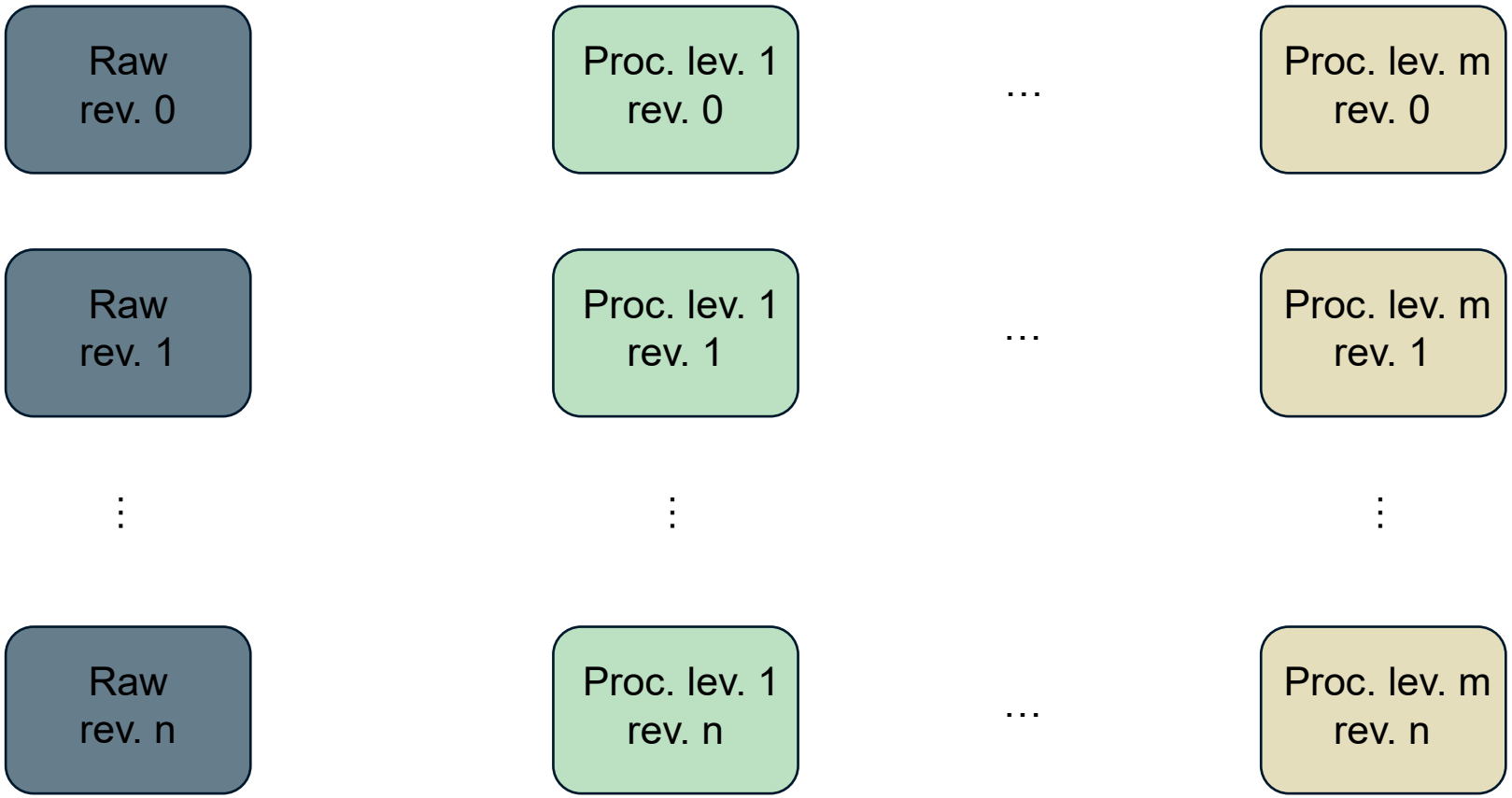
GitLab

Internet

GitLab

Bitbucket

Data Versioning



Tools for Data Versioning

Data Version Control (*https://dvc.org*)

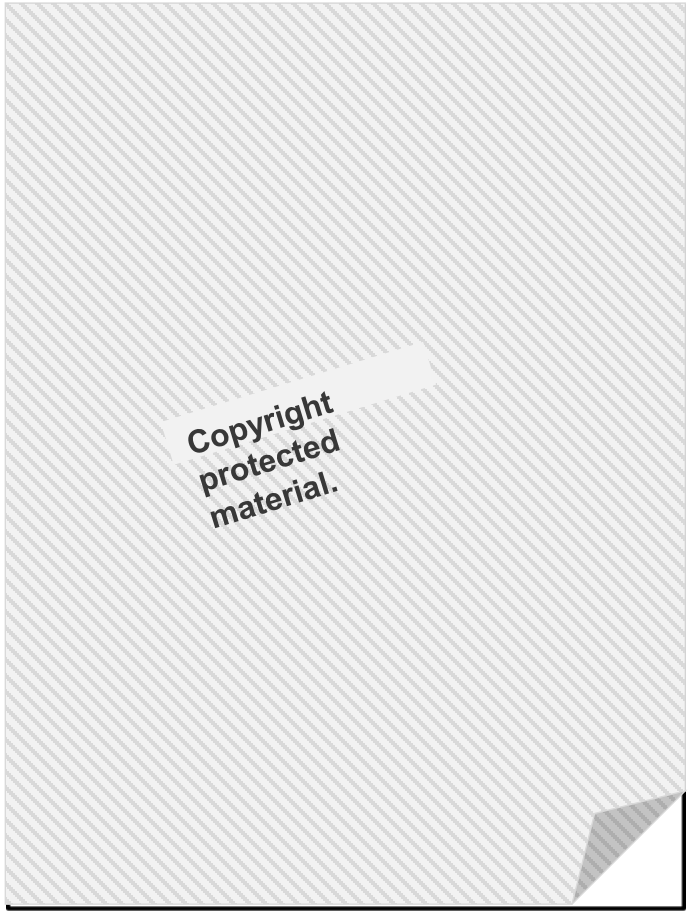
Git Large File Storage (*https://git-lfs.com*)

Lake FS(*https://docs.lakefs.io*)

Tool for data and code organization

Renku (*https://renku.readthedocs.io/en/stable/index.html*)

What to do with all this data, code, metadata?



https://xkcd.com/2116/

Your files are organized, appropriately named, clearly versioned...but what's the point if nobody can open the #*$)@* file??

File Formats

File formats: open and non-proprietary

Data type	Recommended file formats
Text	• PDF/A • Plain Text coded as ACII. UTF-8 or UTF-16 • XML
Spreadsheet	• CSV (NEAD)
Images	• TIFF (uncompressed or lossless compressed) • PNG
Code	• Languages with free environments (e.g. Py or R UTF-8 format of ASCII text)
Audio	• FLAC • Wav

Aim for open and lossless formats
If you are using a proprietary format (ex. MS word extensions xlsx, docx), consider adding an additional format.

Activity: Prepare your data. (15 min, individual or pairs)

Folder Structure

- Does the folder and file organization make sense?
- Separated raw vs processed data?

File Naming

- Is it consistent?
- Does it include:
  - YYYYMMDD
  - Project Name
  - Document Type
  - Version
  - File Extension

Versioning

- Do you have a versioning system?
- Is there a tool you can adopt to help you?

File Formats

- Are your files in non-proprietary, open formats?
  - xlsx → csv
  - docx → pdf
  - images → tiff

Data vs Metadata

https://dataedo.com/cartoon/tag/data-vs-metadata

Documentation: README and Metadata

Feature	README	Metadata
Audience	Humans	Machines (and sometimes also humans)
Format	Free-form text (txt, md)	Structured (XML, JSON)
Scope	How and Why	What and Who
Standards	Flexible	Discipline-specific
Functionality	Help others understand and use data	Facilitate searchability and machine processing

Activity: README (15 min, pairs or trios)

General Information

Sharing/Access Information

Data & File Overview

Methodological Information

Data-specific information for: [filename]

Activity: Data-specific information for: [filename]

Select a dataset/folder/file and fill in the following:

- Number of variables:
- Number of cases/rows:
- Variable List: <list variable name(s), description(s), unit(s) and value labels as appropriate for each>
- Missing data codes: <list code/symbol and definition>
- Specialized formats or other abbreviations used

For access to entire template:
carpentries-incubator.github.io/scientific-metadata/files/AUTHOR_DATASET_ReadmeTemplate.txt

Metadata

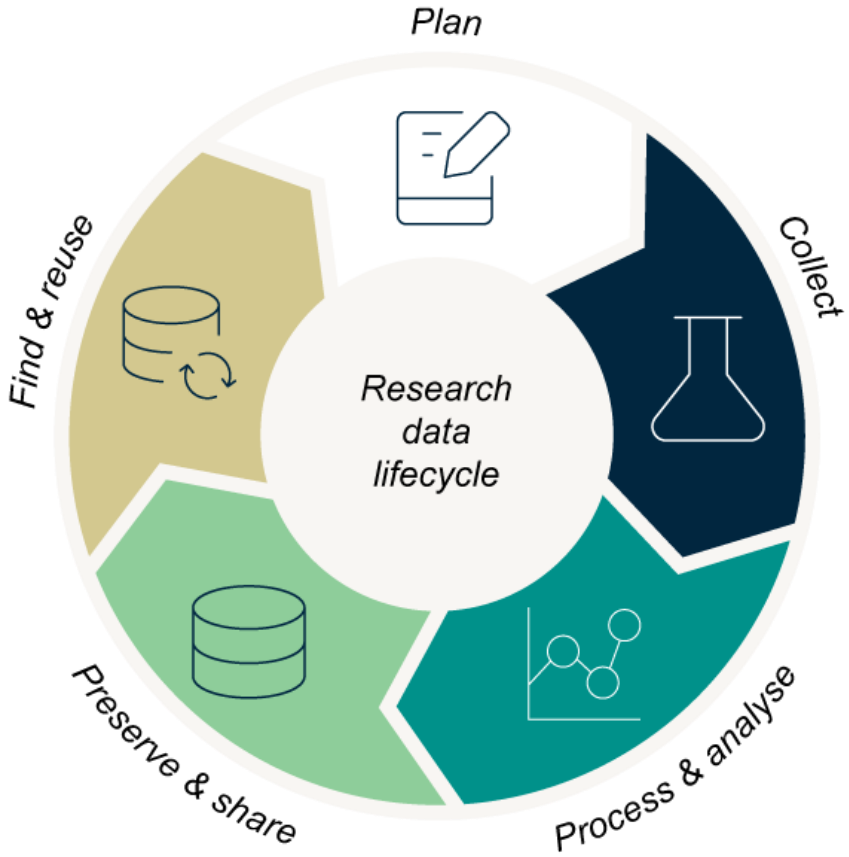
- Definition: Structured data that contains information about other data, but is not the content of the data.

- Metadata is very subject specific. The following directories are helpful:
  - Digital Curation Centre (*https://www.dcc.ac.uk/guidance/standards*)
  - RDA Metadata Standards (*https://rdamsc.bath.ac.uk*)
  - Fairsharing (*https://fairsharing.org*)

- Recommendation: Stick to a list of defined terms (controlled vocabulary) and don't use synonyms to describe the same object (e.g. picture or image)

Activity: Find out which metadata standard is relevant to your field (7 min)

Storage, Preservation, and Sharing

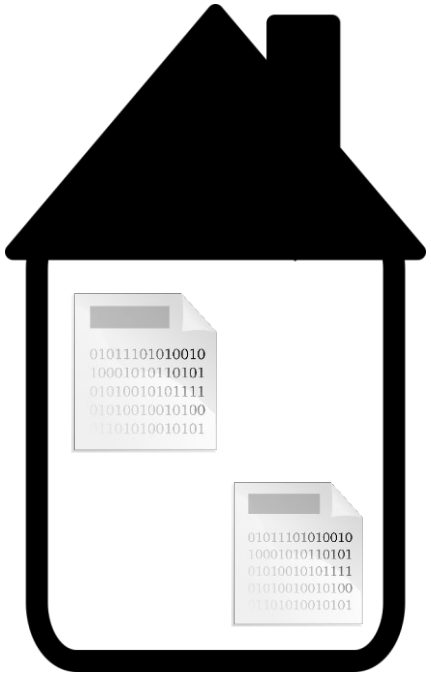


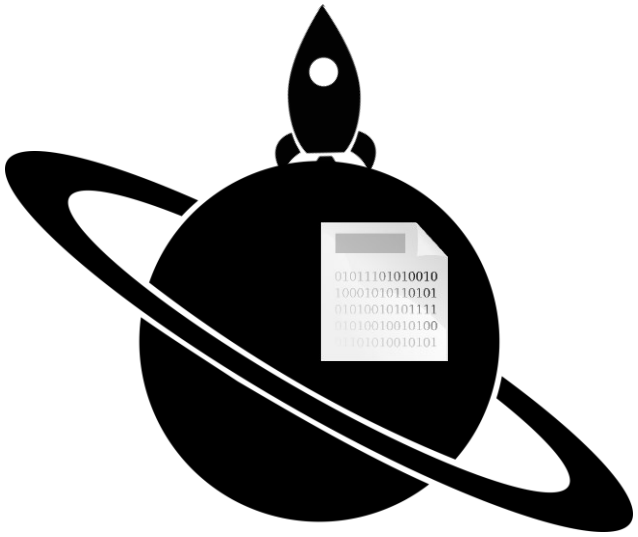
Storage

- Q: Where is your data stored and how is it backed-up?

www.menti.com

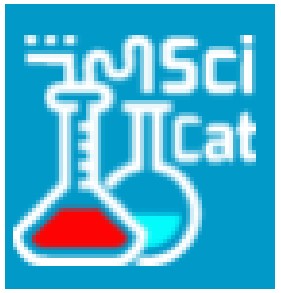
Storage: 3-2-1 backup

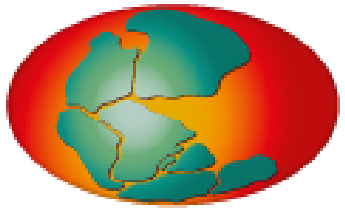




Data repositories: publication vs preservation

ERIC open

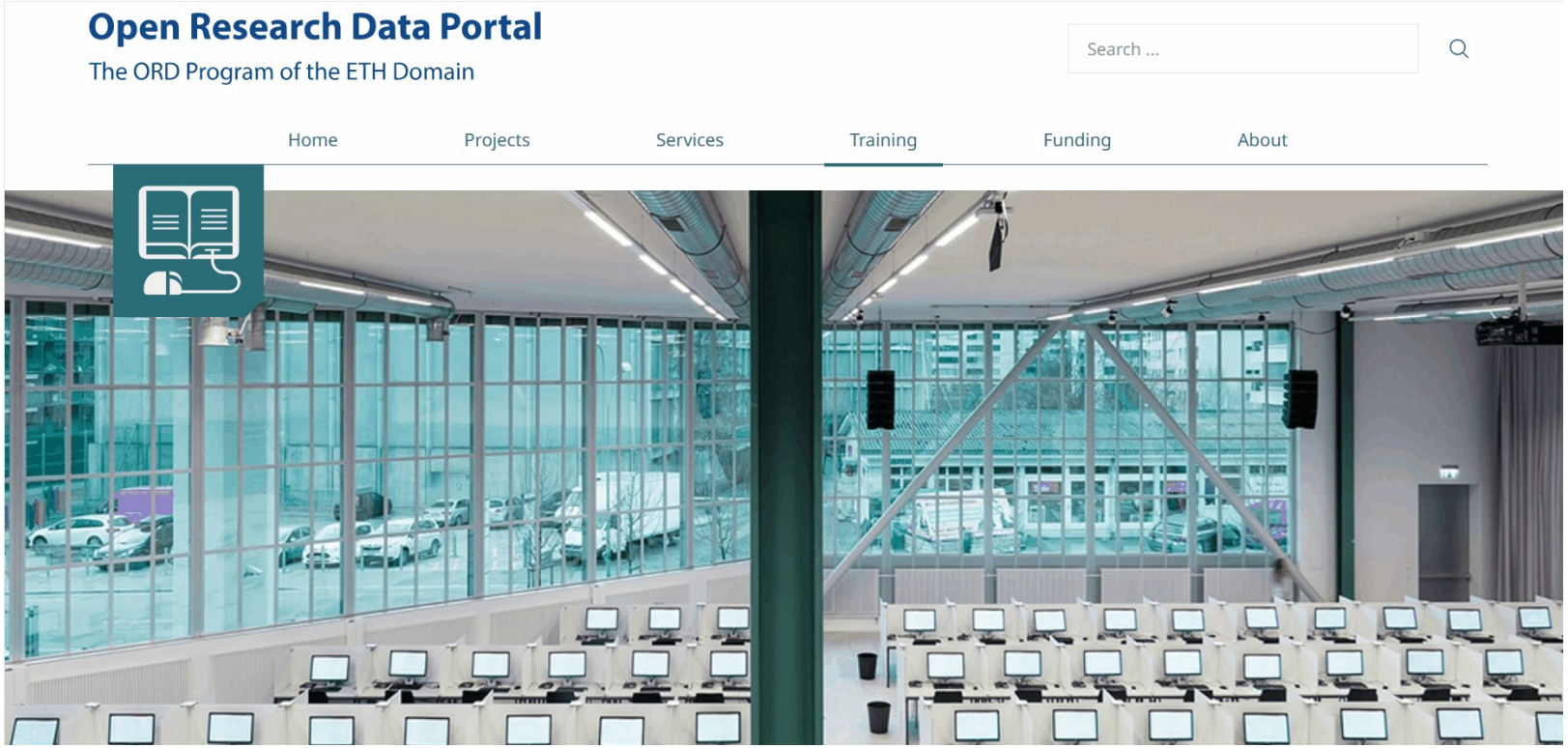




For alternatives: https://www.re3data.org/

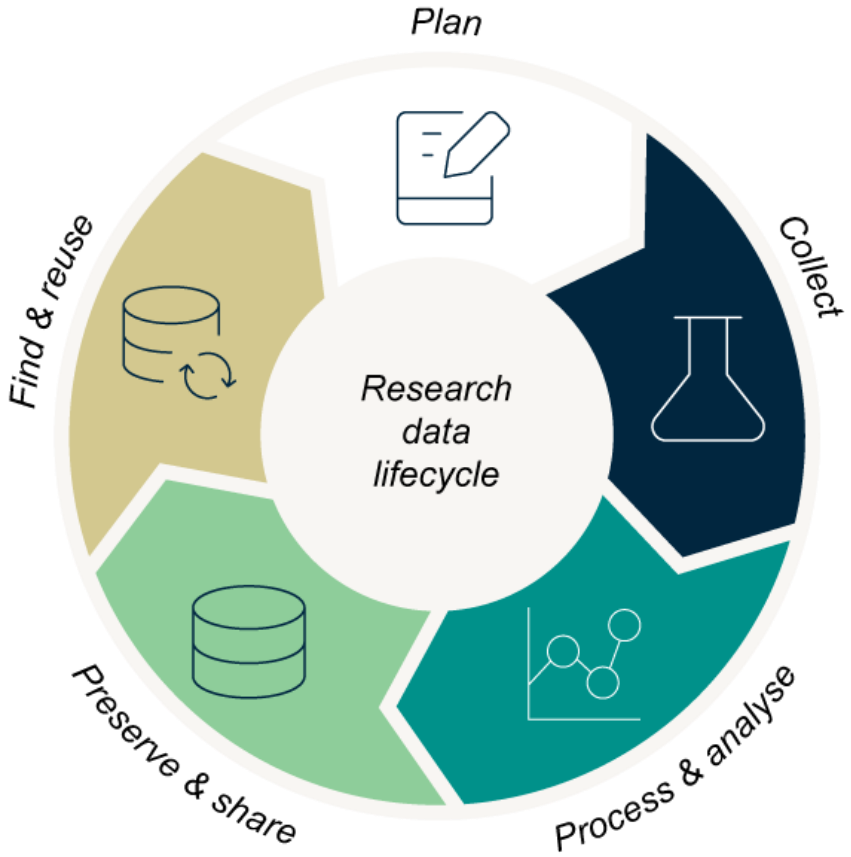
→ DOI

Learn more: Data Management Campus



https://open-research-data-portal.ch/training/

Finding and Reusing



Data availability statements

Where is the data stored (repository name and DOI)?

How can the data be accessed (open, restricted, available upon request)?

How can the data be reused (licensing)?

Very Famous (Open Access) Journal

Example

The data supporting this study's findings is openly available in [Repository Name] at [DOI]. The dataset includes x,y,z and is available under [License].

Licensing

Data

Code

Licensing for Code

Copyleft

- Examples: GPL, LGPL
- Use cases: for projects maintaining open-source is the priority (operating systems, applications, platforms)

Permissive

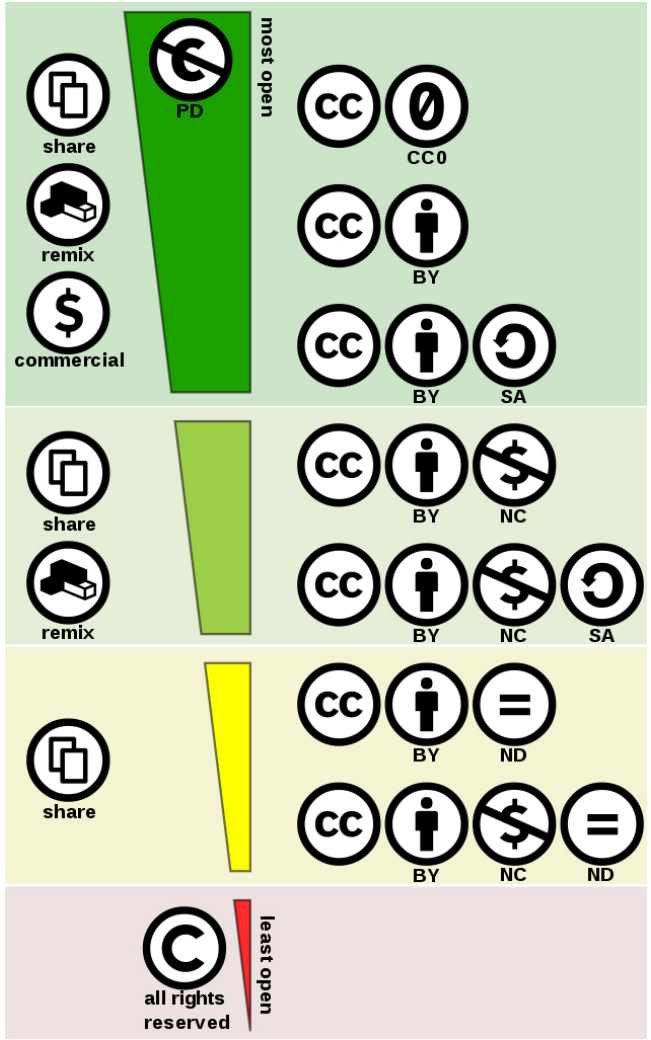
- Examples: MIT, Apache, BSD
- Use cases: for projects which encourage wide-spread adoption and commercial use (libraries, frameworks, tools)

For more information: Licenses – Open Source Initiative

Licensing for Data

creative commons

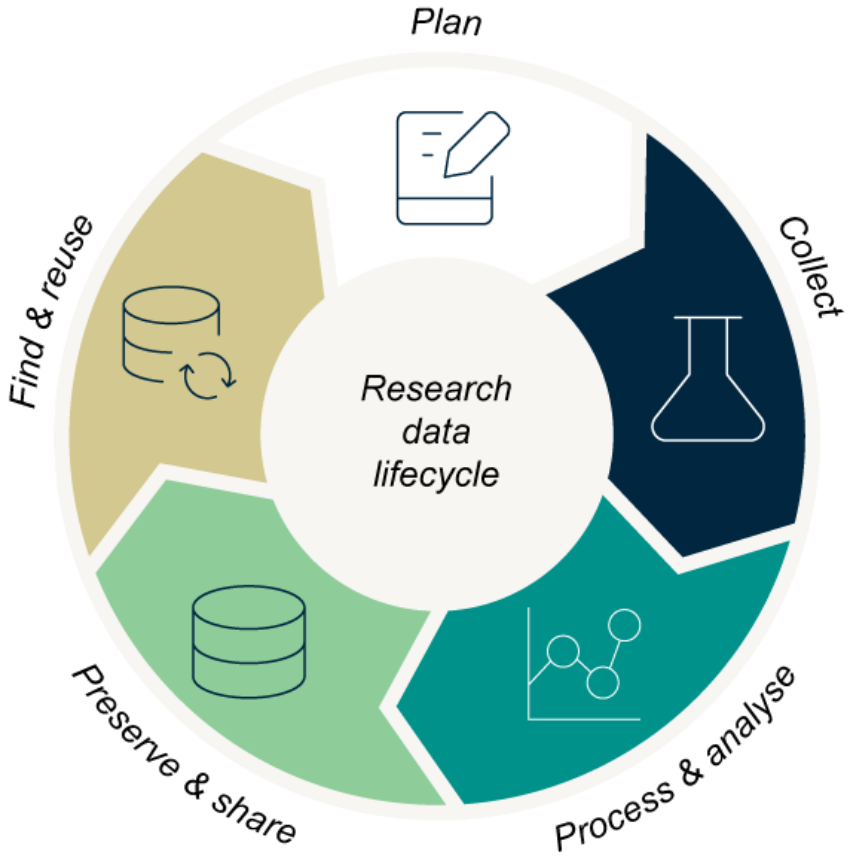
Attribution	Noncommercial	Non- Derivatives	Share alike
Compulsory - Must always **credit** me.	Use it but don't make **money**	Your version must **equal** mine - no changes	If I allow you to change it, **repeat** my CC **licence**



Planning – the data management plan (DMP)

What types of data will be collected or generated?	• File folders, file naming, versioning, file formats
What type documentation will you provide with the data?	• README and Metadata
How will data be stored during and after the project?	• 3-2-1 rule • Repositories
How will your data be shared? How can it be accessed?	• Data availability statements • Licensing

Research Data Management



Lib4RI – Excellent Services for Excellent Research.

fabian.felder@lib4ri.ch
moushumi.ulrich-nath@lib4ri.ch
T: + 41 58 765 57 00