# Planning for research data management
– *the basics* –

Assist. Prof. Dr. Sebastian Dahle
University of Ljubljana

University *of Ljubljana*
Biotechnical Faculty

Why?

**Professional toolset**

What is data? (Metadata?)
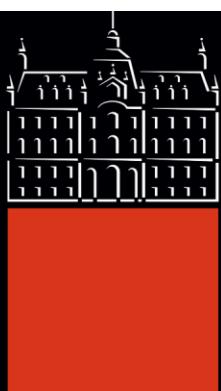
Simply put: **Data is information**

The tangible forms this 'material' may take are e.g. "facts, observations, interviews, recordings, measurements, experiments, simulations, and software; numerical, descriptive and visual; raw, cleaned up and processed"

van Berchum & Grootveld (2017)
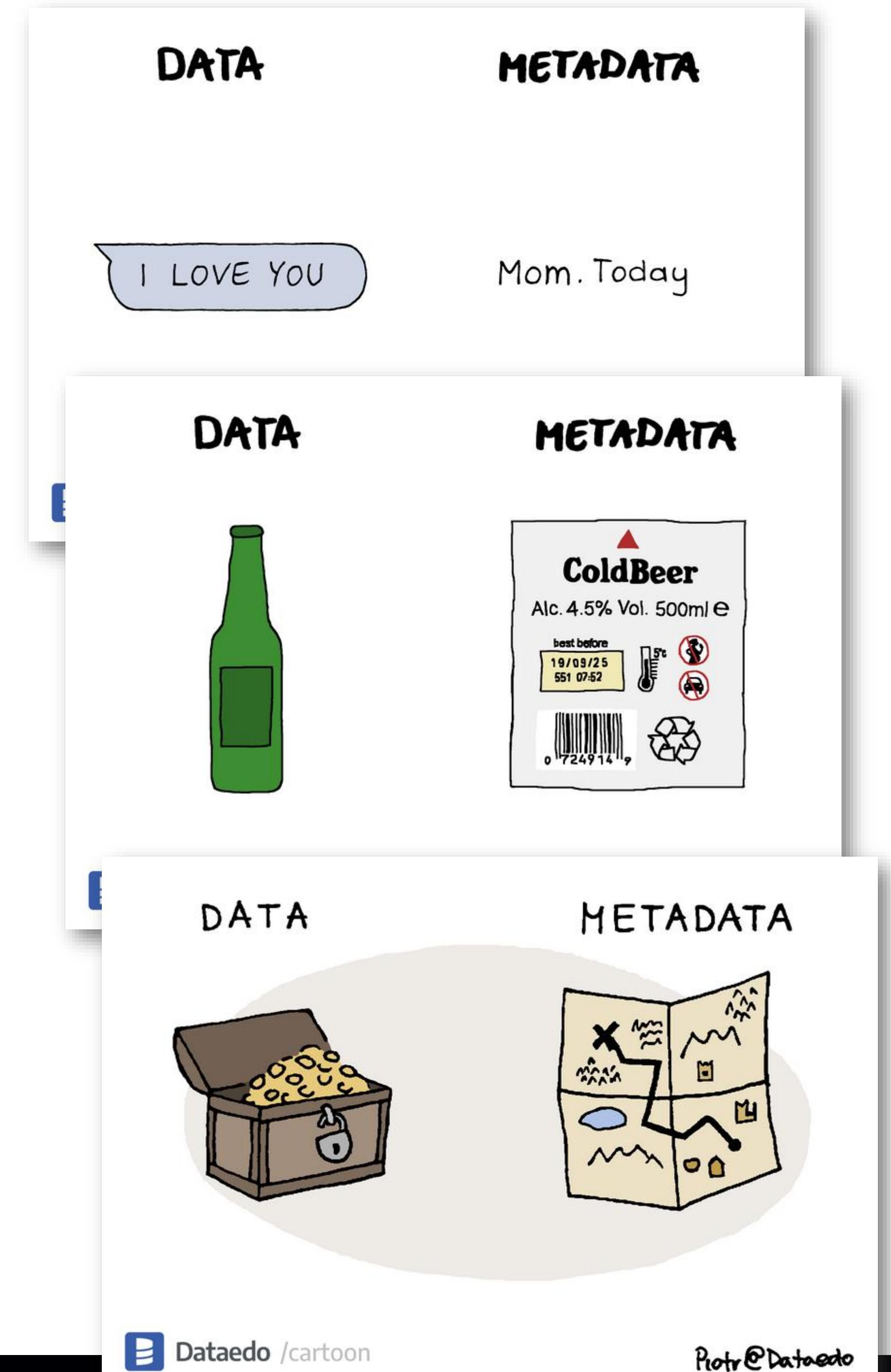
Sanchez, C.: DATA MINDFULNESS

→ Next webinar, Thu, Nov 10th, 2022

University *of Ljubljana*
Biotechnical Faculty
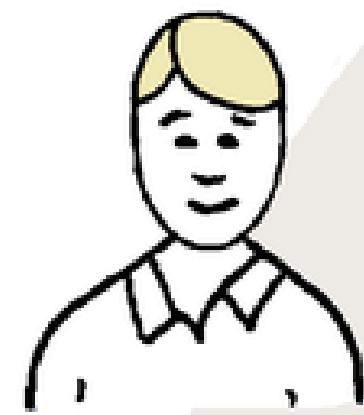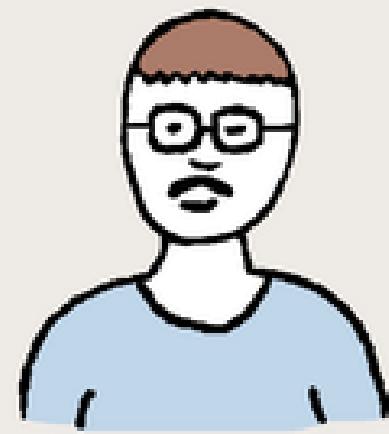
LEVELS OF DATA LITERACY
IN THE ORGANIZATION
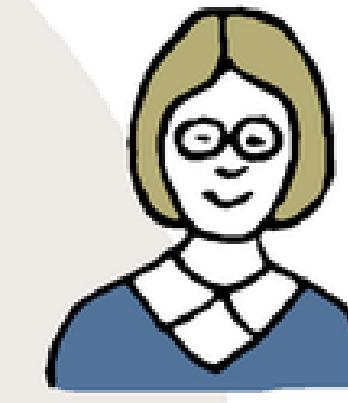
DATA ARCHITECT
MUST KNOW EVERYTHING ABOUT THE DATA
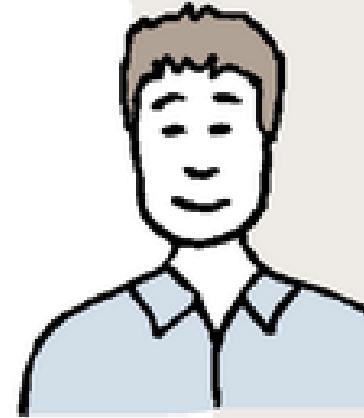
DBA
MUST KNOW WHAT DATA IS STORED IN DATABASES

SUBJECT MATTER EXPERT
MUST KNOW ALL ABOUT DATA IN HIS DOMAIN
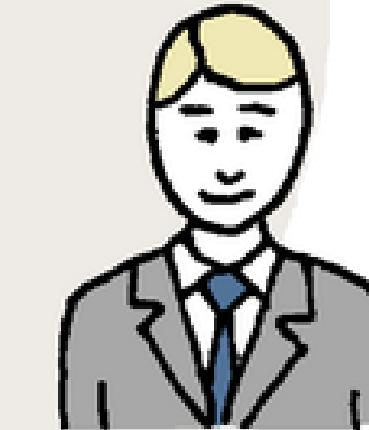
DATA STEWARD
MUST KNOW WHERE TO LOOK FOR DATA

DATA GOVERNANCE DIRECTOR
MUST KNOW WHAT DATABASES EXIST IN THE COMPANY

CHIEF DATA OFFICER
MUST KNOW THE NUMBER TO THE DATA GOVERNANCE DIRECTOR

Dataedo /cartoon

Ploth@Dataedo

https://dataedo.com/

University *of Ljubljana*
Biotechnical Faculty
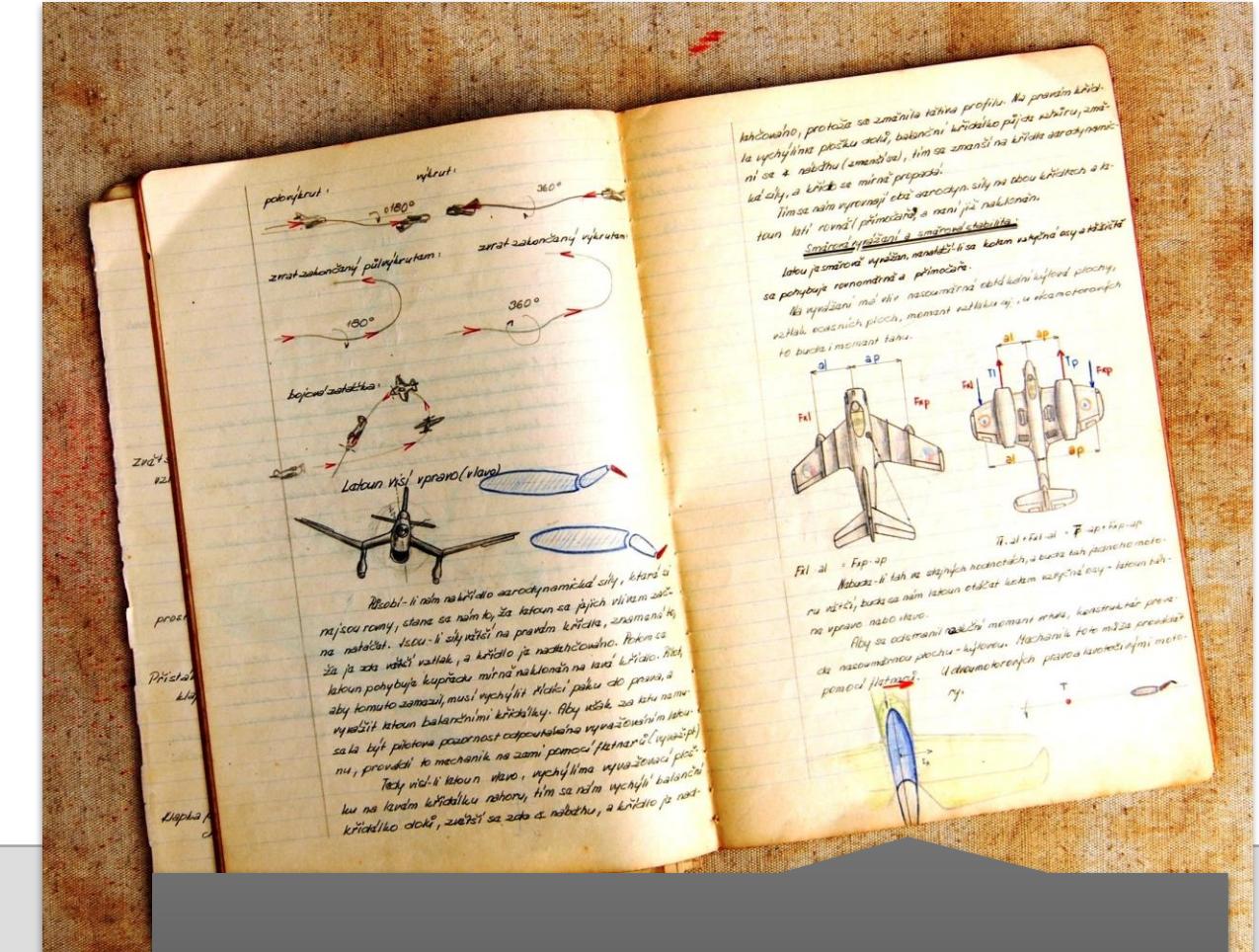
Simple approach

Measure

Analyze

Publish

Versioning

IPR

Proprietary format

File naming

Archiving

Labbook notes
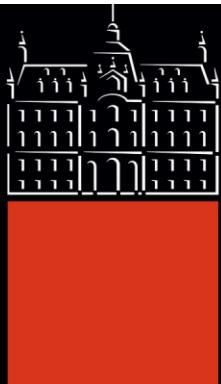
Directory structure

Open Data
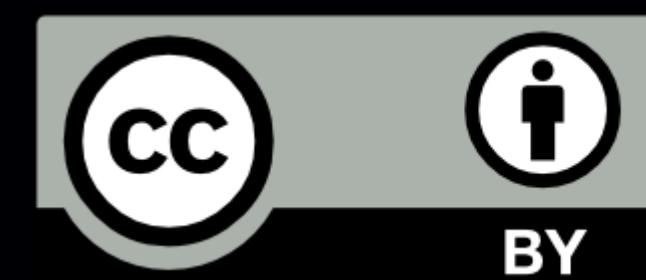
Secure storage

Documentation

FAIR data
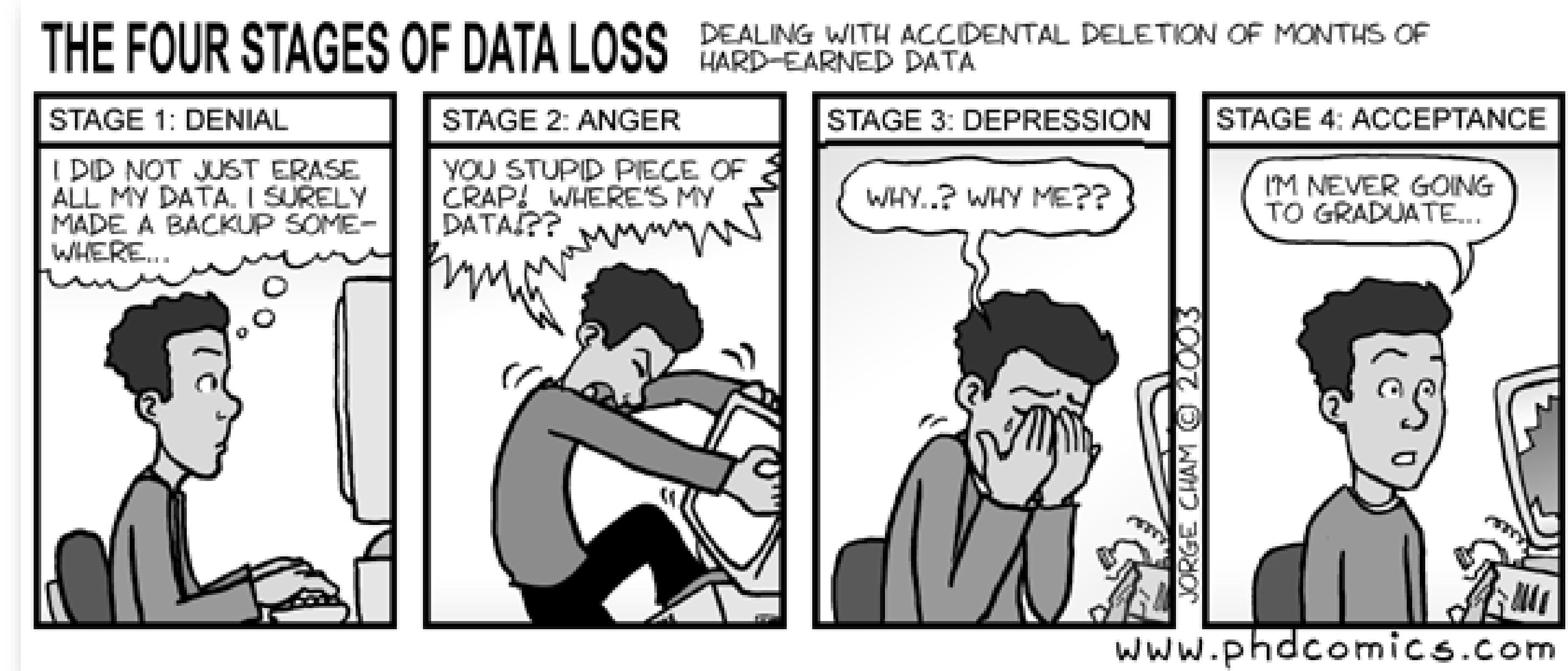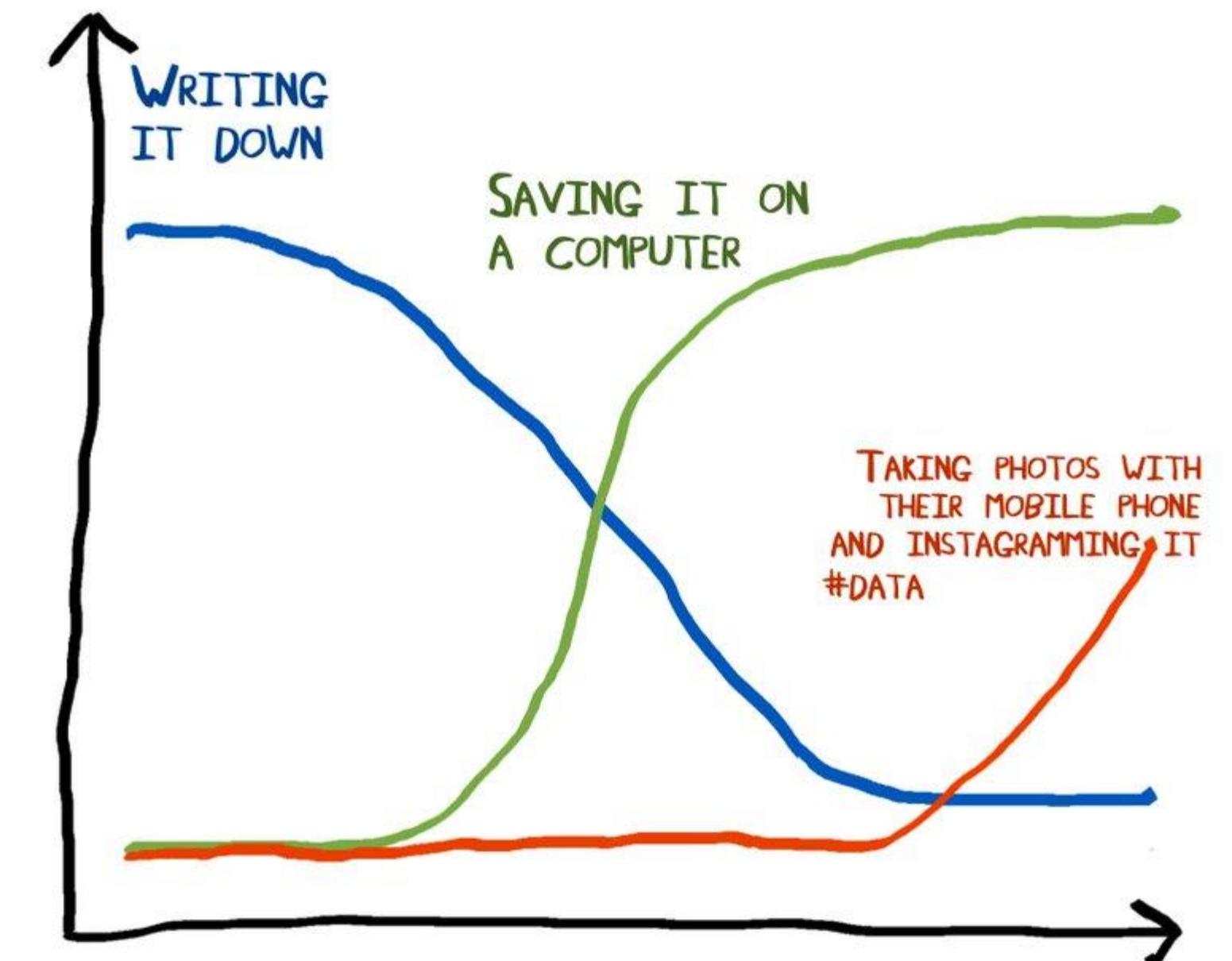
Secure storage

- **RAID**

- **Backups**

- **Multiple locations**

- **3-2-1 rule**

- **Automation!**

A USB drive is lost
EVERY 29 SECONDS
at the Engineering Library.

Save your data. Remember your USB drive.

A TIMELINE OF RESEARCHERS RECORDING DATA
ErrantScience.com

WRITING IT DOWN

SAVING IT ON A COMPUTER

TAKING PHOTOS WITH THEIR MOBILE PHONE AND INSTAGRAMMING IT #DATA

THE FOUR STAGES OF DATA LOSS — DEALING WITH ACCIDENTAL DELETION OF MONTHS OF HARD-EARNED DATA

STAGE 1: DENIAL
I DID NOT JUST ERASE ALL MY DATA. I SURELY MADE A BACKUP SOMEWHERE...

STAGE 2: ANGER
YOU STUPID PIECE OF CRAP! WHERE'S MY DATA!??

STAGE 3: DEPRESSION
WHY..? WHY ME??

STAGE 4: ACCEPTANCE
I'M NEVER GOING TO GRADUATE...

JORGE CHAM © 2003
www.phdcomics.com
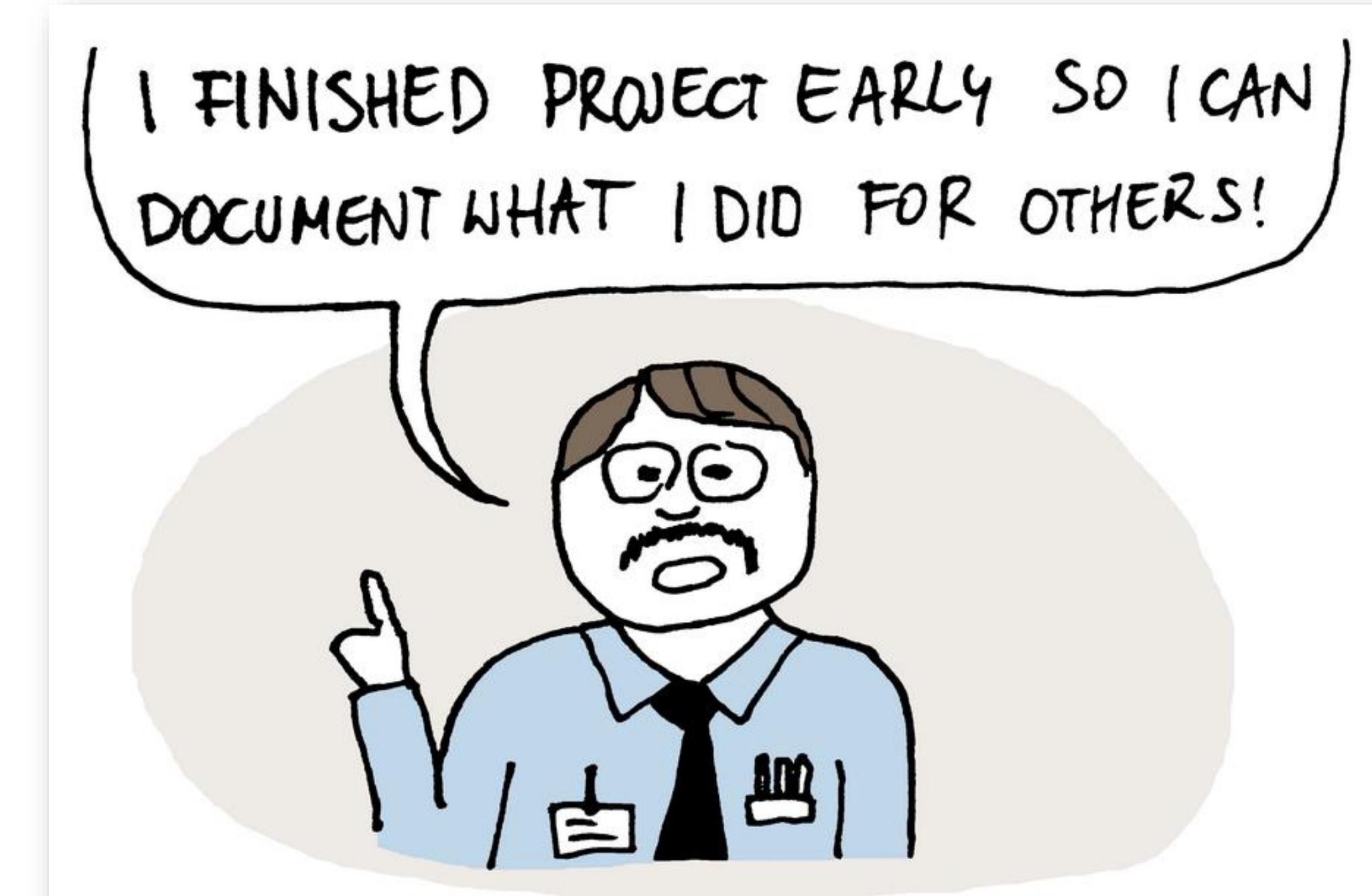
Documentation

➢ **Be intuitive.**

➢ **Be explicite.**

➢ **Rather too much than too little detail.**

➢ **Link data & metadata.**

Also think about:

➢ **File naming**

➢ **Directory structure**

Good documentation saves time and resources!
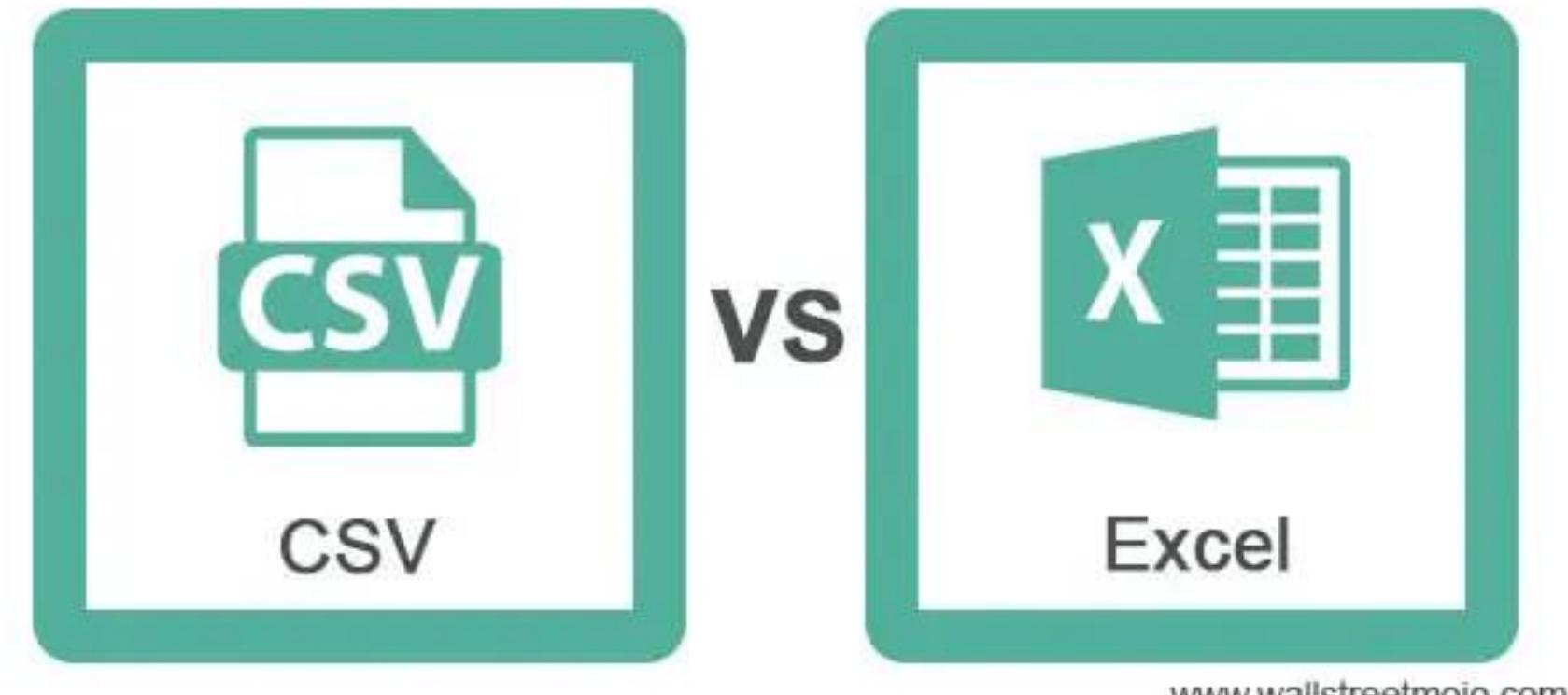
University *of Ljubljana*
Biotechnical Faculty
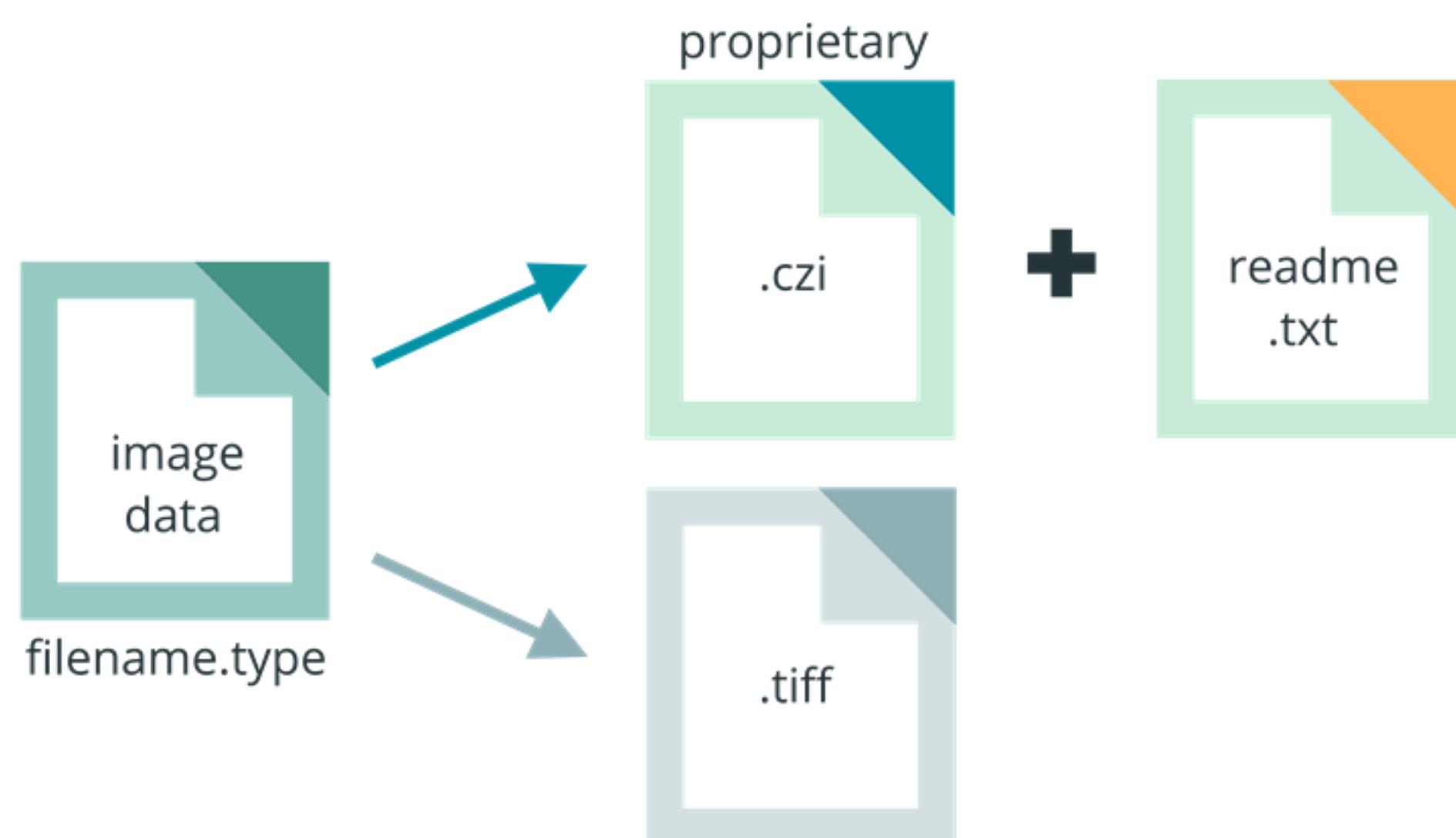
Data formats

- **Limited use of proprietory formats**

- **Lossy vs. uncompressed formats**
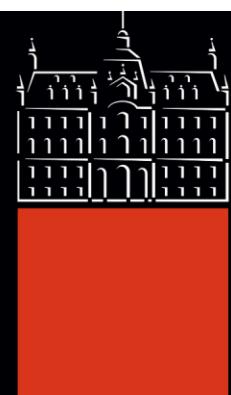
See also:

https://www.openaire.eu/data-formats-preservation-guide
https://www.howtofair.dk/how-to-fair/file-formats/
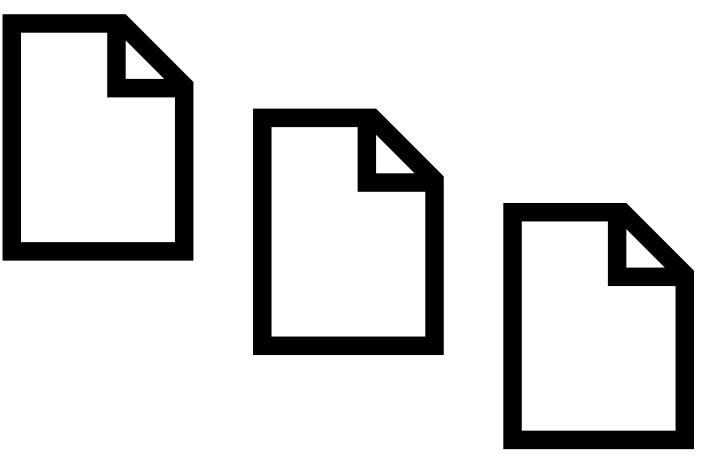
University *of Ljubljana*
Biotechnical Faculty

Versioning
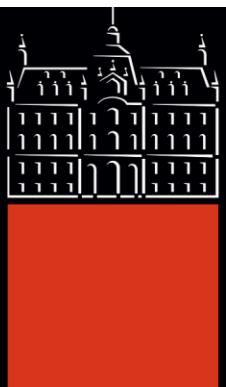
- **Keeping a backlog of changes**
- **Avoid tags such as „final"**

Most common:

- **Date tags (e.g. YYYY-MM-DD)**
- **Semantic versioning**

4.2.1
MAJOR *Minor* patch

"FINAL".doc

FINAL.doc!

FINAL_rev.2.doc

FINAL_rev.6.COMMENTS.doc

FINAL_rev.8.comments5.CORRECTIONS.doc

FINAL_rev.18.comments7.corrections9.MORE.30.doc

FINAL_rev.22.comments49.corrections.10.#@$%WHYDID ICOMETOGRADSCHOOL????.doc

WWW.PHDCOMICS.COM
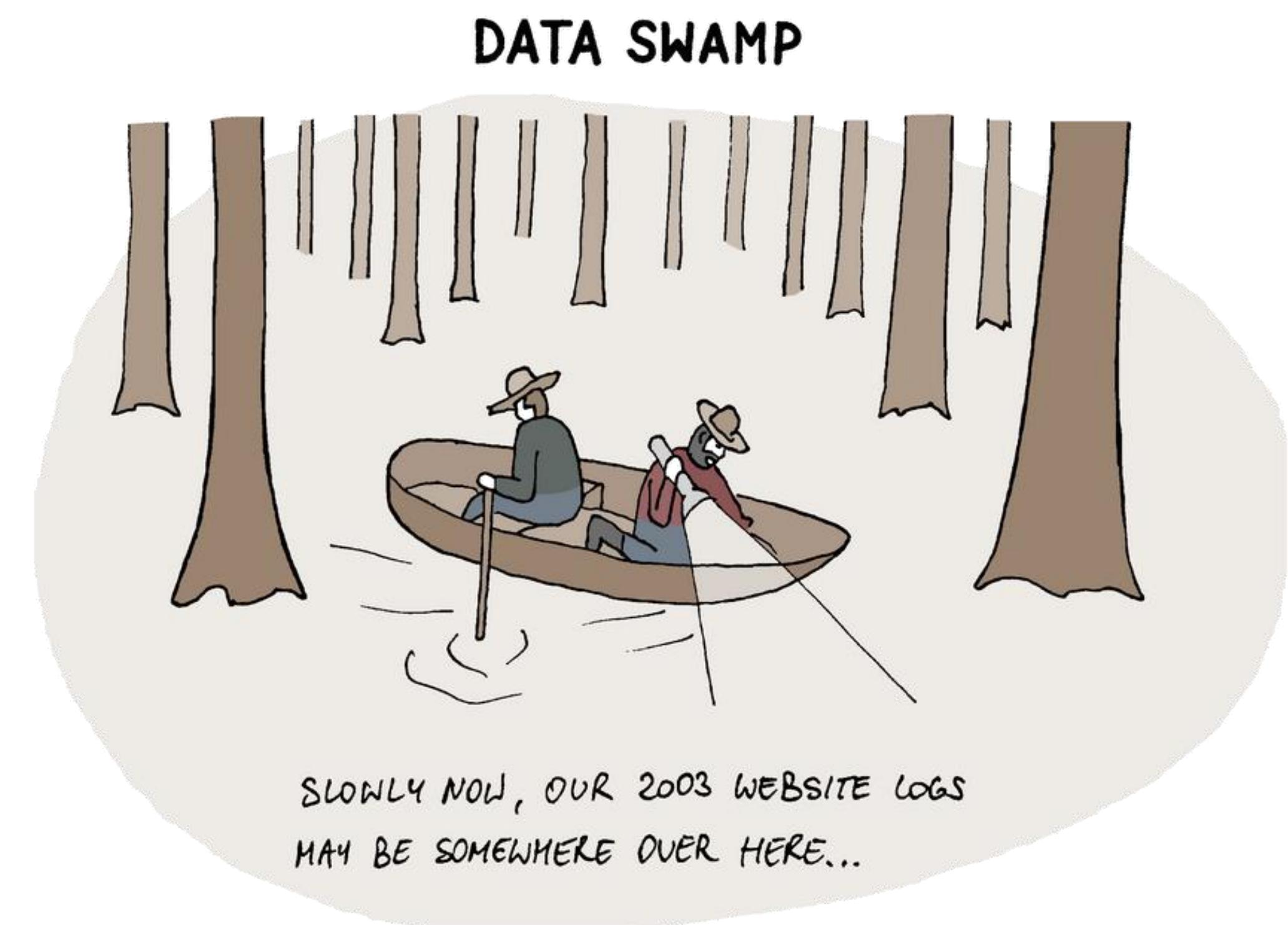
Archiving (& deletion)

DATA SWAMP

SLOWLY NOW, OUR 2003 WEBSITE LOGS MAY BE SOMEWHERE OVER HERE...

Dataedo /cartoon

Piotr@Dataedo

Check your:
- ➢ **Institutional policy**
- ➢ **Funder's policy**
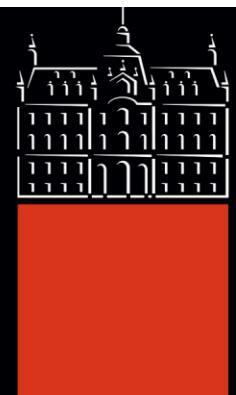
Important considerations:
- ➢ **What data is worth archiving?**
- ➢ **Which data is required to be archived?**
- ➢ **For how long does the data need to be archived?**
  - ➢ many EU funders require min. 10 years
  - ➢ many repositories guarantee min. 20 years

Publishing data & Open data
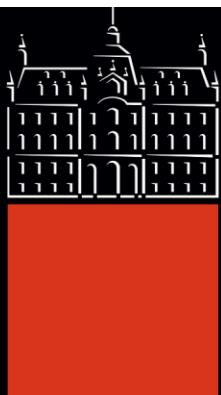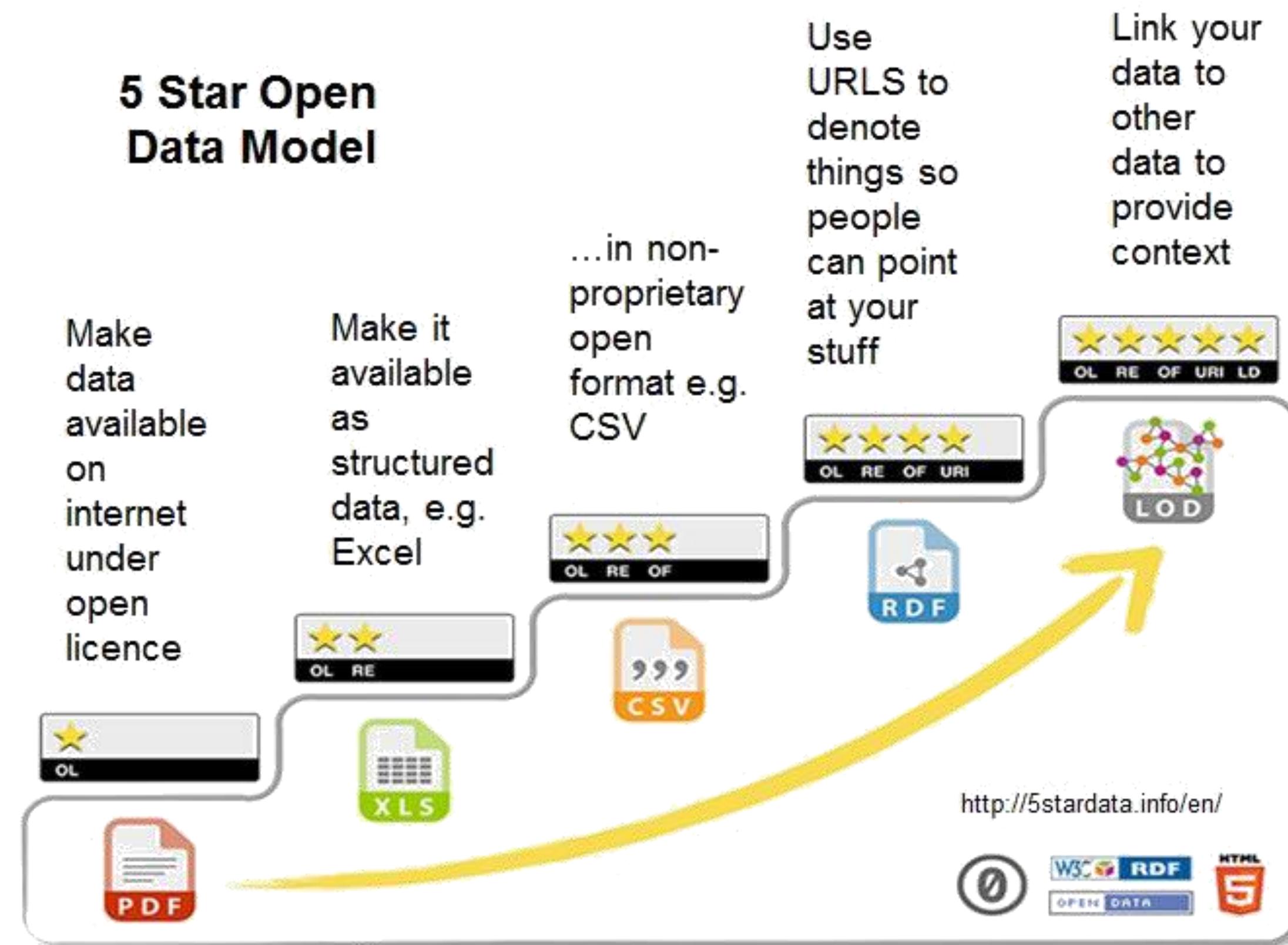
❖ **Reproducibility crisis**
❖ **Data mining & AI**
❖ **Unintended reuse**
❖ **Accountability**

➢ **Curation & metadata**
➢ **Repository selection**
➢ **Licensing**
➢ **Persistent ID**
➢ **FAIRness**
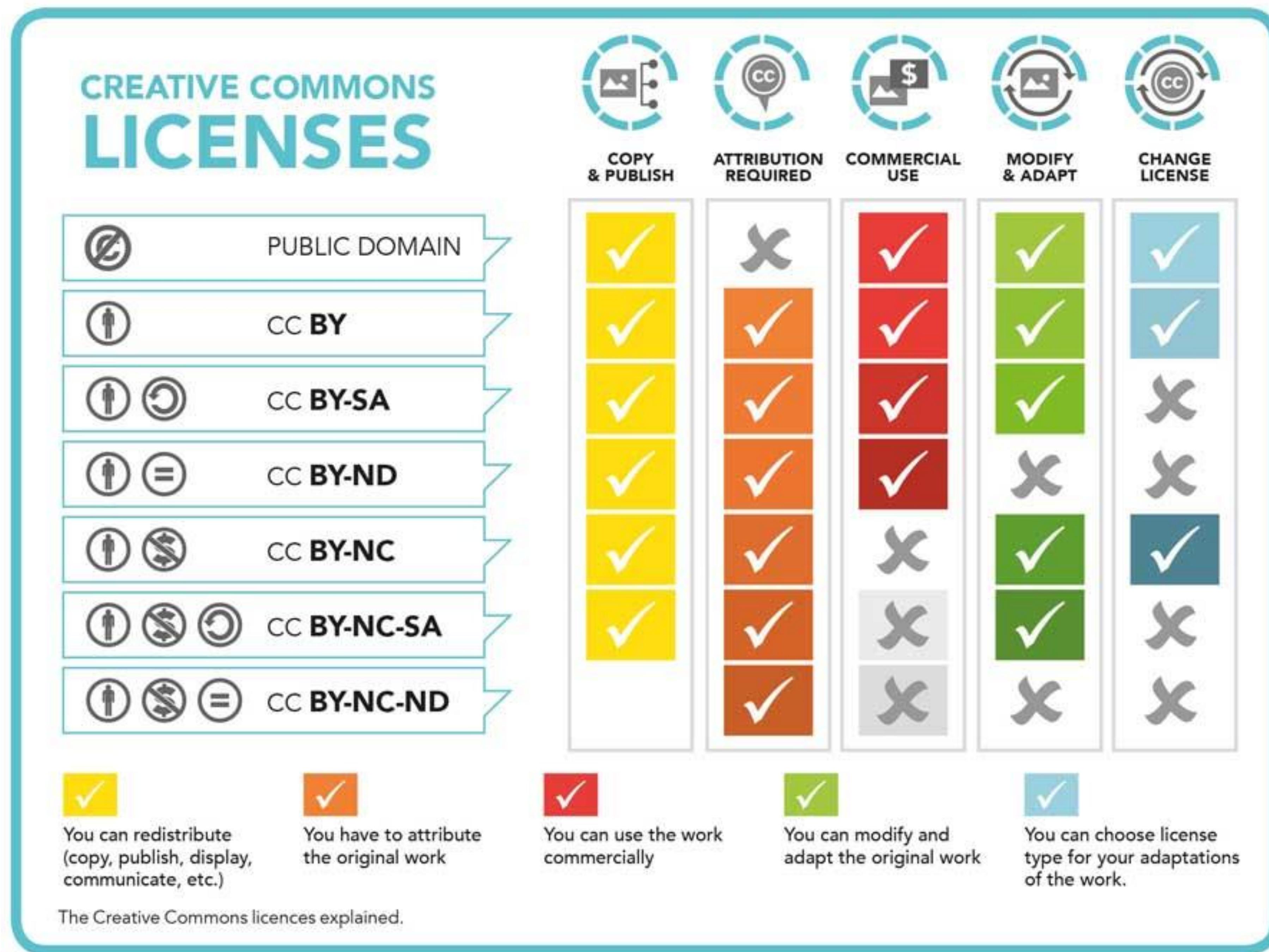
Intellectual Property Rights – Licenses

CREATIVE COMMONS **LICENSES**

	COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
PUBLIC DOMAIN	✓	✗	✓	✓	✓
CC **BY**	✓	✓	✓	✓	✓
CC **BY-SA**	✓	✓	✓	✓	✗
CC **BY-ND**	✓	✓	✓	✗	✗
CC **BY-NC**	✓	✓	✗	✓	✓
CC **BY-NC-SA**	✓	✓	✗	✓	✗
CC **BY-NC-ND**		✓	✗	✗	✗

✓ You can redistribute (copy, publish, display, communicate, etc.)

✓ You have to attribute the original work

✓ You can use the work commercially

✓ You can modify and adapt the original work

✓ You can choose license type for your adaptations of the work.

The Creative Commons licences explained.

Persistent Identifiers (PIDs)

A persistent identifier is a
long-lasting reference to a digital resource.

Examples:

- **Permalinks**
- **DOI system (ISO 26324)**
- **ORCID**
- **ISBN**

ORCID

ISBN 978-3-16-148410-0

doi®

FAIR data

https://www.go-fair.org/fair-principles/
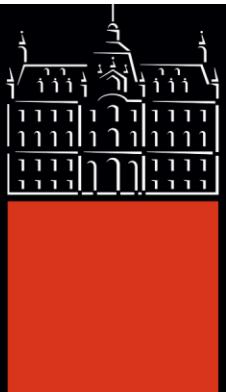
! *FAIR data ≠ Open Data ≠ machine-readable data*

Thu, Nov 10th 2022 → next EUTOPIA webinar

University *of Ljubljana*
*Biotechnical* Faculty

GDPR & sensitive data

Examples of sensitive data:
- Personal data
- Biometric data
- Genetic data
- Health-related data
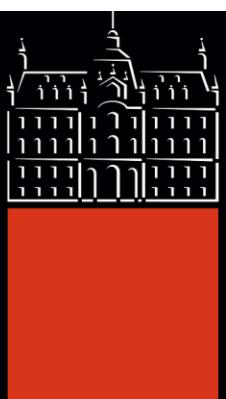
Special care is needed to ensure
- Explicit consent for purpose is obtained
- Data is anonymized / pseudonymized
- Storage periods and safeguards

Always confer with Data Protection Officer!
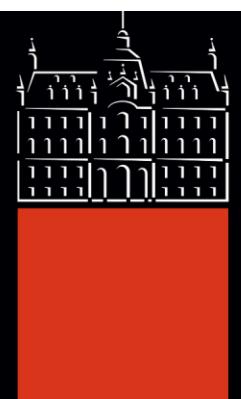
Costs

- ➤ **Data collection**

- ➤ **Documentation**

- ➤ **Storage**

- ➤ **Access & security**

- ➤ **Preservation**

- ➤ **Availability & reuse**

University *of Ljubljana*
Biotechnical Faculty

Research data lifecycle

Source: https://datamanagement.hms.harvard.edu/plan-design/biomedical-data-lifecycle

Data Management Plans

➢ **Help to set up efficient practices**

➢ **Avoid double work and data loss**

➢ **Allow to estimate efforts and costs**

➢ **Help revise dated practices**
    or adapt to new circumstances

➢ **Required by many funders & in some PhD programs**

Tools for Data Management Planning

https://dmponline.dcc.ac.uk/

https://argos.openaire.eu/splash/

https://dmponline.be/
https://dmptool.org/

Repositories

https://www.re3data.org/

https://zenodo.org/

https://github.com/

Some more training resources

CESSDA training:
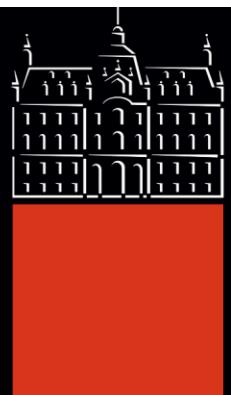
https://www.cessda.eu/DMGuide

https://opensciencemooc.eu/

Science Europe

https://scienceeurope.org/our-priorities/research-data/research-data-management/
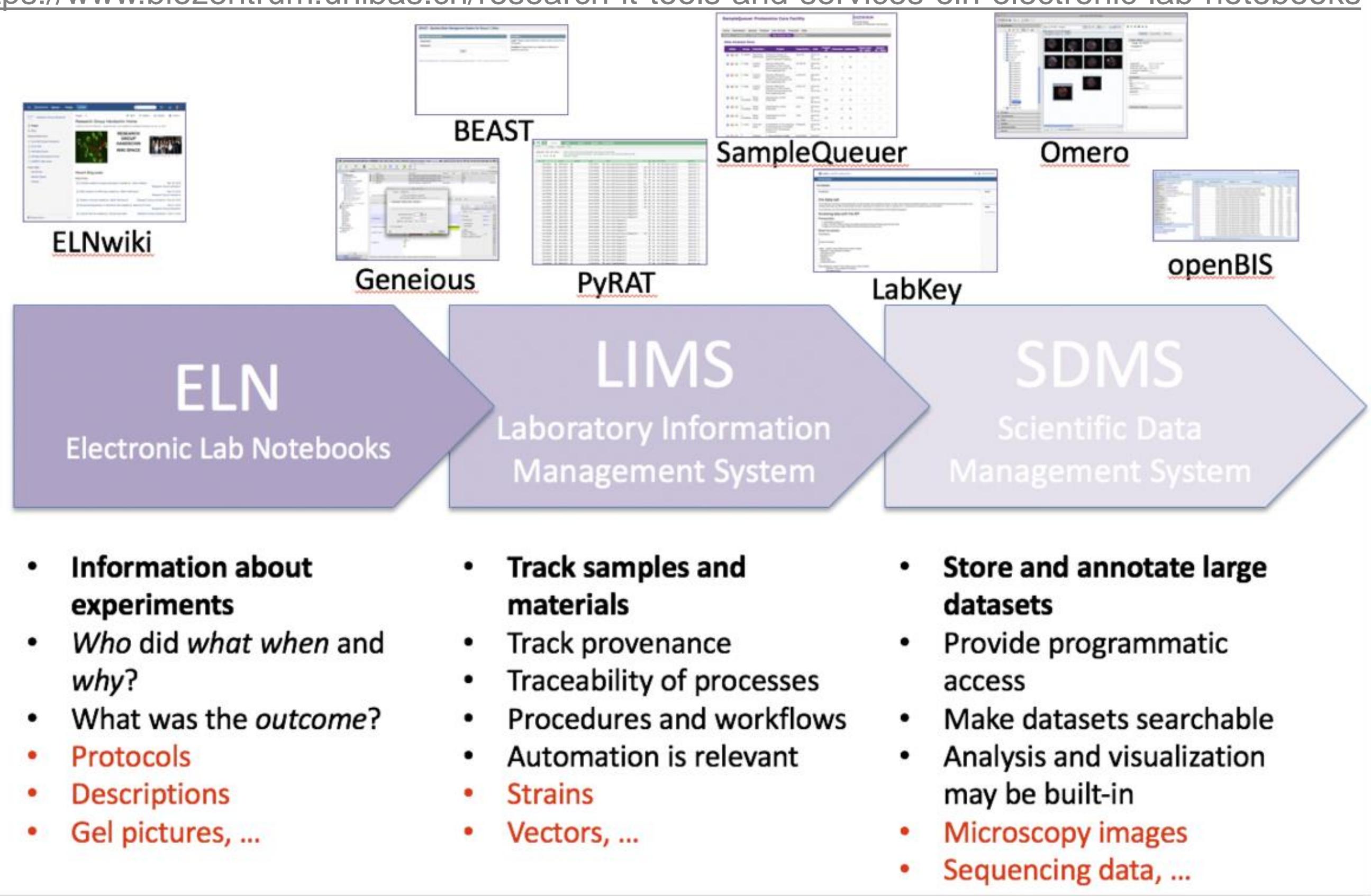
Electronic Lab Notebooks
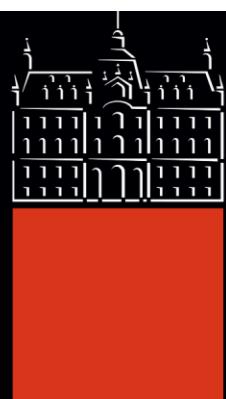
Potential to simplify processes from protocols to documentation and publication.

How to pick an electronic laboratory notebook?
https://doi.org/10.1038/d41586-018-05895-3
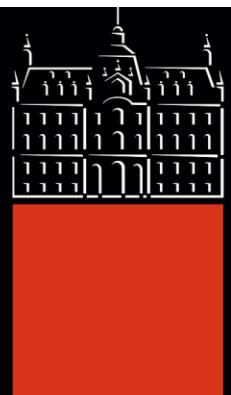
University *of Ljubljana*
Biotechnical Faculty

Where to turn with specific questions?

❑ **Data stewards** *(?)*

❑ **Data protection officer** *(?)*

❑ **Ask your local librarian!!**