1. Research Data Management
Basics and how to apply them

Date: 27 September 2023
Speakers: Dr. Fabian Schmid, Dr. Julian Dederke

DISCLAIMER

Proprietary images in this slide deck that are available under http://andresbucher.ch/cc/library.html are licensed under CC BY 4.0 by Andres Bucher. If not otherwise indicated, other icons and images, including the ETH logo, are the property of the respective holder of rights.

# Can somebody reuse some of your research data if you are unexpectedly absent?

From Adobe Stock: File #471445681, education license

Learning Goals

- You understand theory, concepts and importance of **research data management**.

- You can follow every step in the **life cycle of research data**.

- You know **strategies** and **tools** to implement measures for good research data management (RDM).

- You have a general understanding of the principles of **Open Science** and **FAIR data**.
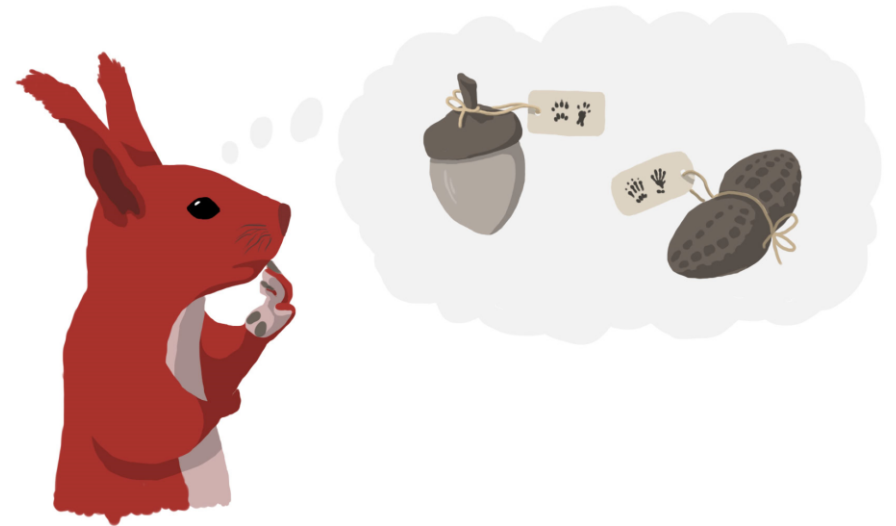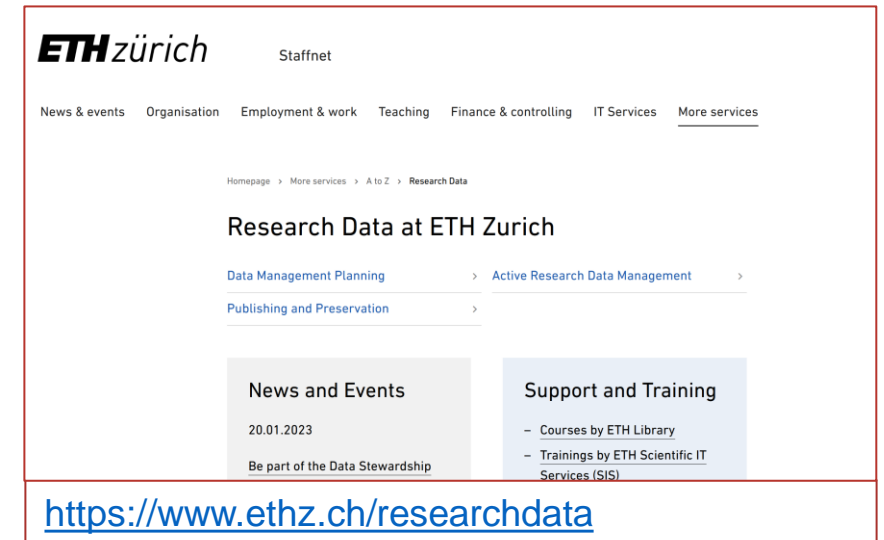
Today's Agenda
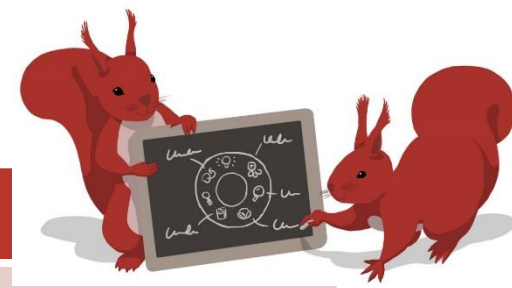
Session	From	To	Content
Welcome and warm-up	09:00	09:10	• Learning goals and survey
Theory about RDM	09:10	09:35	• Problems with reproducibility • RDM basics and Open Science • Miro-Exercise 1: What are YOUR data
The Research Data Life Cycle	09:40	09:50	• Your tasks and benefits
Data documentation I	09:50	10:40	• Importance of metadata • Miro-Exercise 2: Metadata in your discipline
Data documentation II	10:45	11:00	• File naming
BREAK	11:00	11:15	BREAK
Data reuse and preservation	11:15	11:55	• FAIR data • Long-term preservation • Miro-Exercise 3: Data long-term preservation
RDM support at ETH Zurich	12:00	~12:15	• Tools and support at ETHZ • Next workshops and contact points
Q&A	~12:15	12:30	

But first: Let's get to know each other
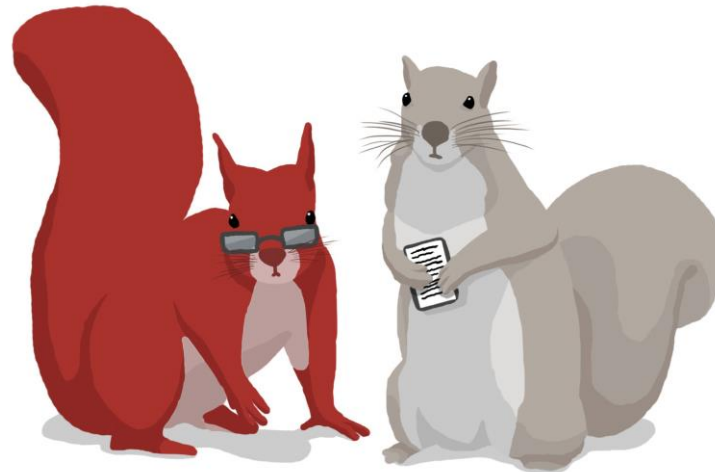
Nice to meet you, we are….

- From the

  - **Research Data Management and Digital Curation Team** at ETH Library, ETH Zurich

- **Sharing a scientific background** ourselves

- Here to **discuss data management** as part of your research

- To **learn** more **about your needs** in the process

- And to **motivate you** to **think critically** about the chances and **limitations of data management and data reuse**

https://www.ethz.ch/researchdata

https://www.library.ethz.ch/rdm

And who are you?

I would like to ask you some questions…

Theory behind Research Data Management (RDM)

What is research data management and why should it concern you?

Open Science: Framework to achieve more openness in science

UNESCO recommendations on Open Science (2021): https://en.unesco.org/science-sustainable-future/open-science/recommendation
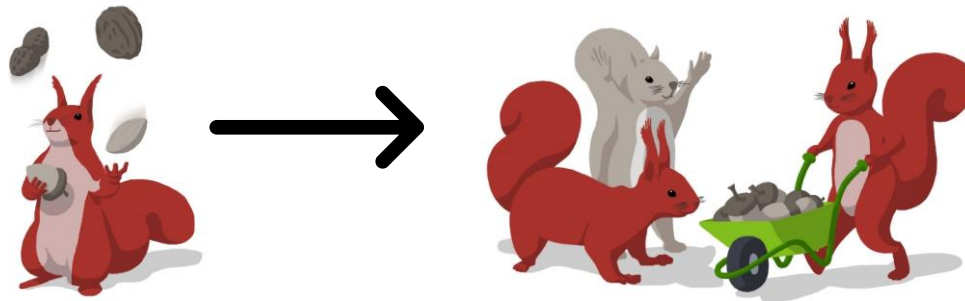
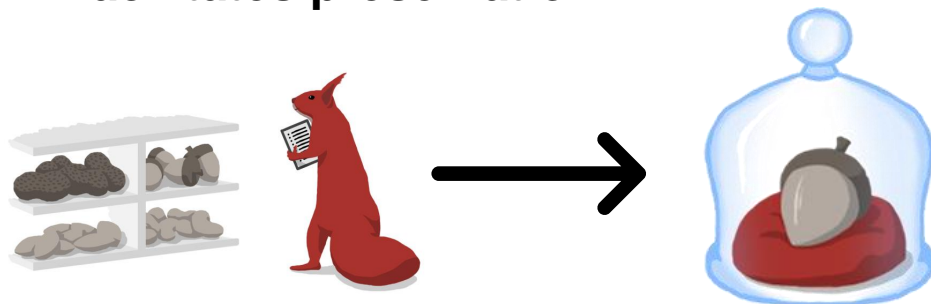# What is proper RDM good for?

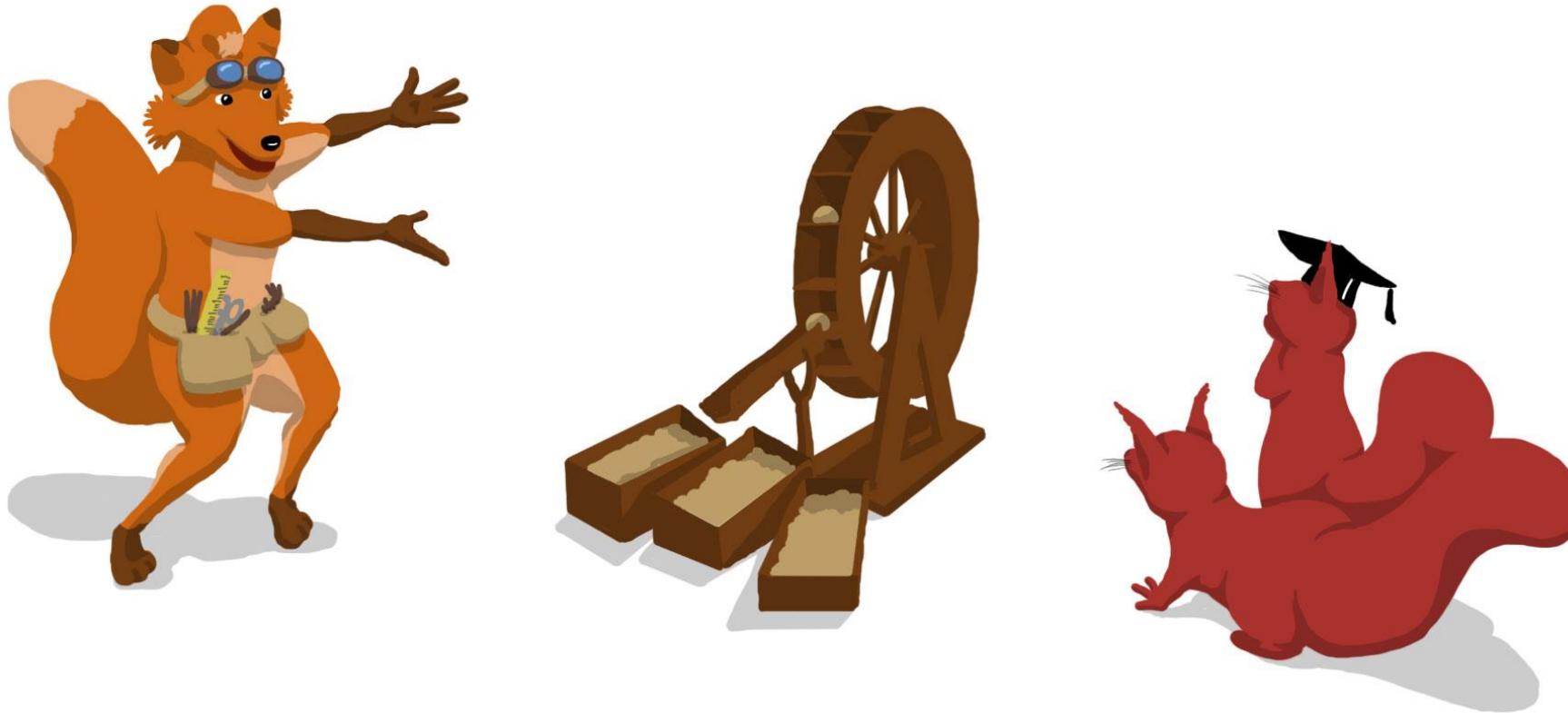**RDM supports your own work**

RDM is a prerequisite for sharing and publishing reusable data

RDM facilitates preservation

Open Science: Dissemination of scientific knowledge

Goal: Increase transparency and reusability of research data and results

Open Science at ETH Zurich: https://ethz.ch/en/research/open-science.html

Reasons for reforming current scientific practices

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

Barriers to reproducing preclinical results included unhelpful author communication, but critics argue that one-time replication attempts don't tell the whole story.

original article in **Nature**, https://doi.org/10.1038/d41586-021-03691-0

Reproducibility in Cancer Biology: Challenges for assessing replicability in preclinical cancer biology

Timothy M Errington, Alexandria Denis, Nicole Perfito, Elizabeth Iorns, Brian A Nosek

Center for Open Science, United States; Science Exchange, United States; University of Virginia, United States

original aricle in **eLife** in Dec 7 2021, https://doi.org/10.7554/eLife.67995

NEWS | 21 June 2022

Many researchers say they'll share data − but don't

Reasons included a lack of informed consent or ethics approval to share; misplaced data; and that others had moved on from the project.

original article in **Nature**, https://www.nature.com/articles/d41586-022-01692-1

Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

Monya Baker

Nature **533**, 452–454 (2016) | Cite this article

Survey sheds light on the 'crisis' rocking research.

original aricle in **Nature**, https://doi.org/10.1038/533452a

Assessing the reproducibility of research results in EU Framework Programmes for Research

Final report

European Commission, Directorate-General for Research and Innovation, _Assessing the reproducibility of research results in EU Framework Programmes for Research : final report_, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/186782

Maybe as a start here: What are data?

"'[R]esearch data' are defined as **factual records** […] used as primary sources for **scientific research**, and that are commonly accepted in the scientific community as necessary to validate research findings"

(OECD 2007, https://doi.org/10.1787/9789264034020-en-fr, p. 13).

The Research Data Life Cycle

The Research Data Lifecycle (DLC)

The cycle (clockwise):
- Define the research question
- Create project proposal and DMP
- Collect, store and document
- Evaluate and select
- Process, analyse and interpret
- Publish and manage access
- Preserve
- Verify and re-use

- **Reuse** relates to
  - primary data
  - processed data
  - analysis scripts / code

- **Open or restricted** access
- Sensitivity / confidentiality

- **Publication** of
  - manuscripts/books/papers
  - data and documentation
  - analysis scripts / code
- **Open Access** options

- **Long-term view** on the life-cycle of your research data

- **Data Management Plan**
  - part of projects/applications
  - planning of DLC steps

- Data **PLUS** Metadata
  - **metadata:** "Data about your data"
  - administrative, descriptive, structural, technical metadata
  - analysis steps, code creation, versioning, traceability of changes

- Decisions what to keep and what to discard

- Formulation of results and conclusions

The Research Data Lifecycle (DLC)

- **Reuse** relates to
  - publication
  - publication
  - access

Special case:
sensitive and confidential data
(e.g., personal health data)
Link: Factsheet: Data Protection
in Research Projects

- **Open or restricted** access
- Sensitivity / confidentiality

- **Publication** of
  - manuscripts/books/papers
  - data and documentation
  - analysis scripts / code
- **Open Access** options

The lifecycle wheel contains: Define the research question · Create project proposal and DMP · Collect, store and document · Evaluate and select · Process, analyse and interpret · Publish and manage access · Preserve · Re-use

- Formulation of results and conclusions
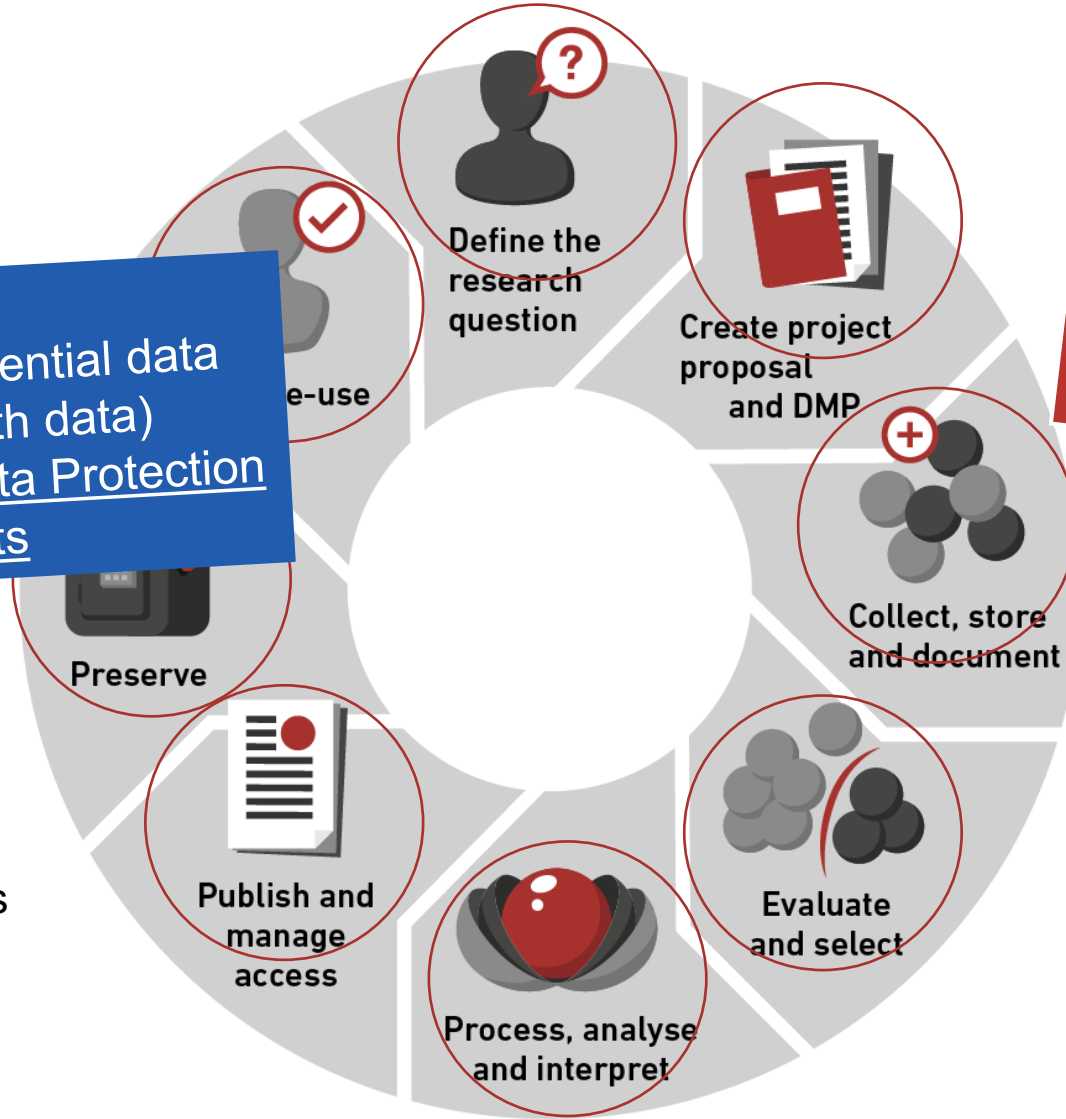
- **Long-term view** on the life-cycle of your research data

More Information on the DMP as well as the requirements of funders and ETH Zurich in the 2nd workshop on 4 October

- Data **PLUS** Metadata
  - **metadata:** "Data about your data"
  - administrative, descriptive, structural, technical metadata analysis steps, code creation, versioning, traceability of changes

- Decisions what to keep and what to discard

Tasks along the Research Data Lifecycle (DLC)

- **Assure reusability:**
  - **choose** a suitable **repository** (FAIR, non-commercial) **and archive**
  - **assure findability** with persistent identifiers (e.g., DOI)
  - **increase visibility**

- **Ensure** security, openness, access rights

- **Choose appropriate outlets** for
  - manuscripts/books/papers
  - data and scripts
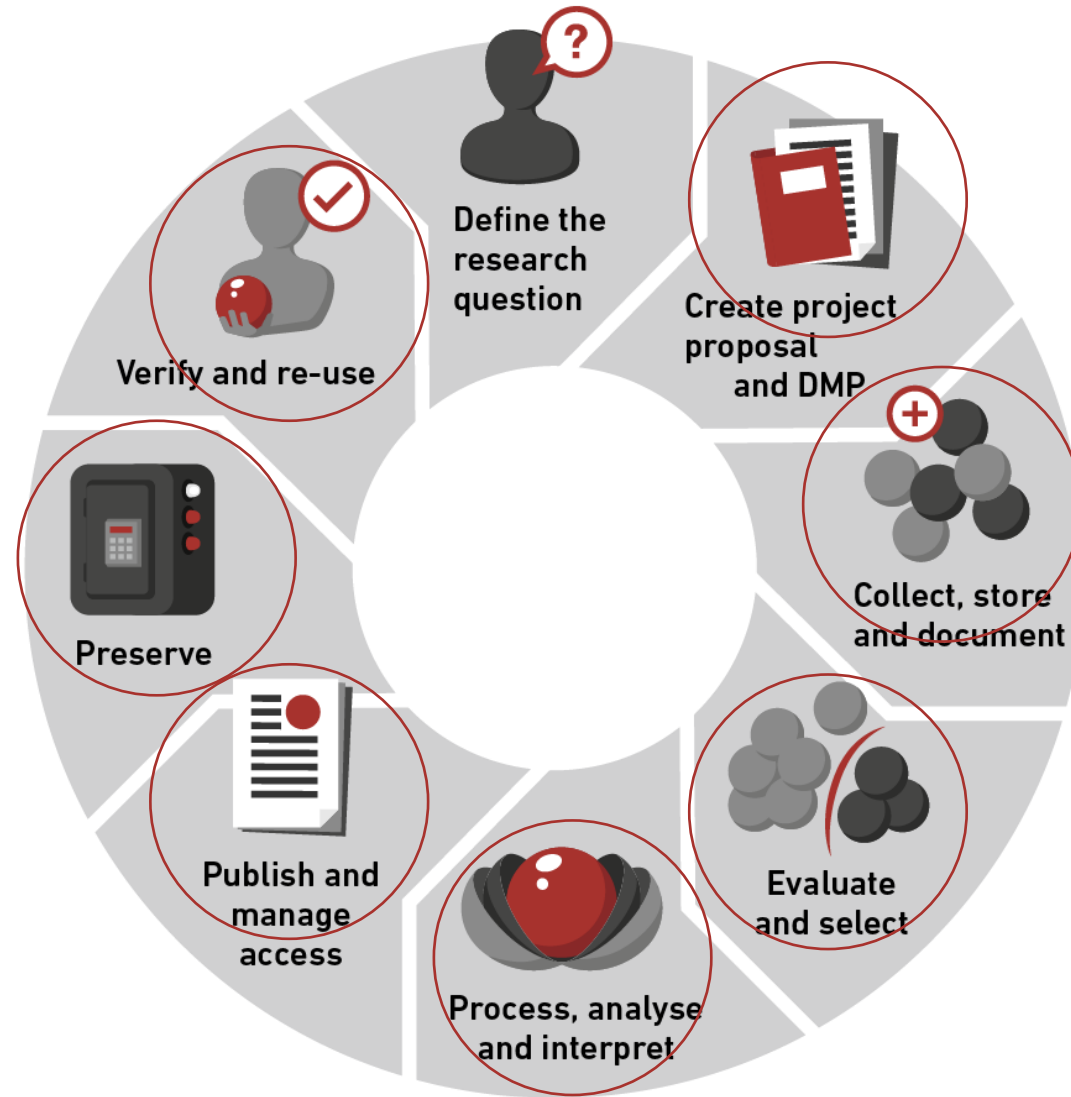
- **Data Management Planning:**
  1. at the start of the project
  2. continuous development and update

- Define which data you **collect where, when, and how**
  - daily data management

- Analysis steps and tools:
  - **choose** appropriate **tools**
  - **process** data
  - **save and control versions** of scripts, code & processed data

- **Aggregate** data; **preparation** for publication

Data documentation: Importance of metadata

Reproducibility

What do these two images
have to do with reproducibility?

Reproducibility: It's all about structure and a story behind it

DATA

SORTED

ARRANGED

PRESENTED VISUALLY

EXPLAINED WITH A STORY

Ravi S. Kudesia
@rskudesia

Brilliant demonstration of the difference between data and the story the data tell. It's one thing to sort, arrange, and visualize data and another thing to make sense of data by putting all the pieces together! 🙏 @OpenAcademics @AcademicChatter (via Mónica Rosales Ascencio)

Homework from doctoral students: Construct and Reproduce

Metadata and their relevance for reproducibility and reuse

Metadata – what are they?

- Any **auxiliary information describing the characteristics** of the actual research data

- **Comments in code** follow good practice in programming

Image: Adobe Stock #96808114

- Metadata to be entered in a form when uploading to a **repository**

  - **Check early** which metadata are required or possible

  - Collect the required information early and **keep it together with the data** you plan to upload

  - Often there **are more metadata required** than repositories demand as obligatory

  - e.g. in a **Readme-file** (parts of the metadata may even be in file / folder names)

Readme-files as crucial documents for data documentation

Image: Adobe Stock #456396154 (Educational licence)

Image: Adobe Stock #509101215 (Educational licence)

Image: Adobe Stock #281541156 (Educational licence)
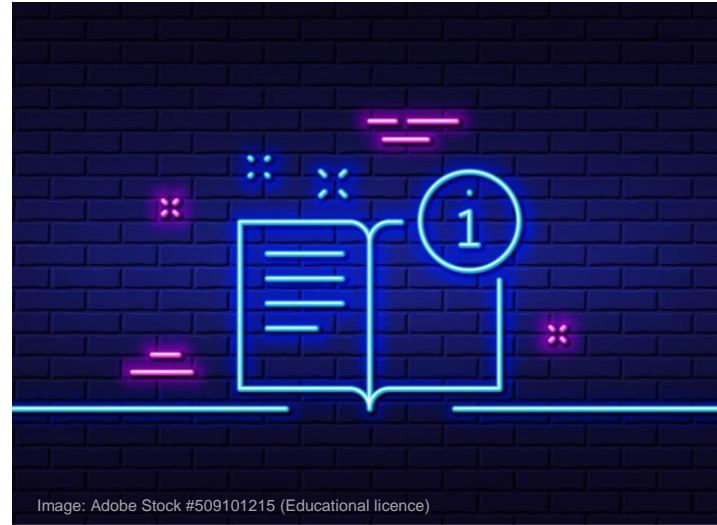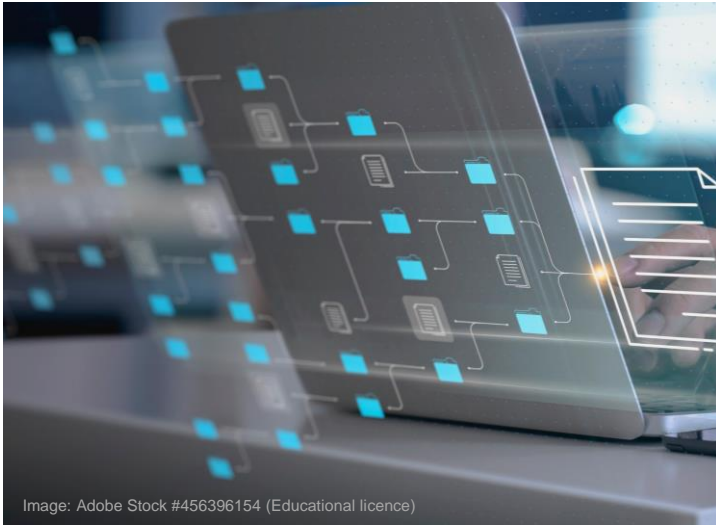
- **Guide for writing README files:** https://unlimited.ethz.ch/x/3CeSCw

- Use a **common structure** and consider creating a **template**

- **Controlled vocabulary**: Define terms and and their intended use

Implementation: community-specific metadata

Scientific metadata improves verification, **validation of research results** and **reusability of research data**

Image: Adobe Stock #96808114

Image: Adobe Stock #174698805

Image: Adobe Stock #465625340

Challenges

- **Diverse types of metadata** for the same data type (e.g., microscopic image vs. astronomy image)
- **Specialised** scientific user communities in a **dynamic scientific landscape**.

Solutions

- **Disciplinary metadata standards**: ideally **unique**, kept **traceable** and **intelligible**
- **Databases** for available disciplinary metadata standards: DCC List of Metadata Standards, RDA Metadata Standards Catalog, https://fairsharing.org/standards/

Implementation: community-specific metadata
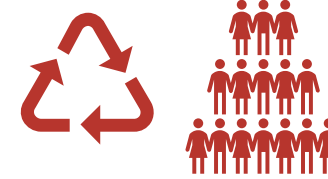
Scientific metadata ensures verification, **validation of research results** and **reusability of research data**

Image: Adobe Stock #96808114
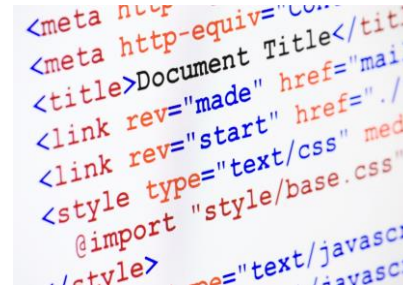
Image: Adobe Stock #174698805

Image: Adobe Stock #465625340

POTENTIAL GOAL: Using a **unique metadata standard** for data types occurring in your scientific discipline (e.g., seven SI base units)
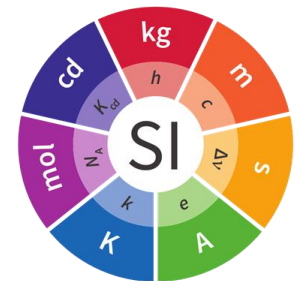
Importance of metadata: Future self

- An important collaborator: your future self

  - E.g., in two years, when you finish your dissertation or prepare a publication

  - E.g., in 10 years, when working at another institution

Data documentation II: File naming

File naming
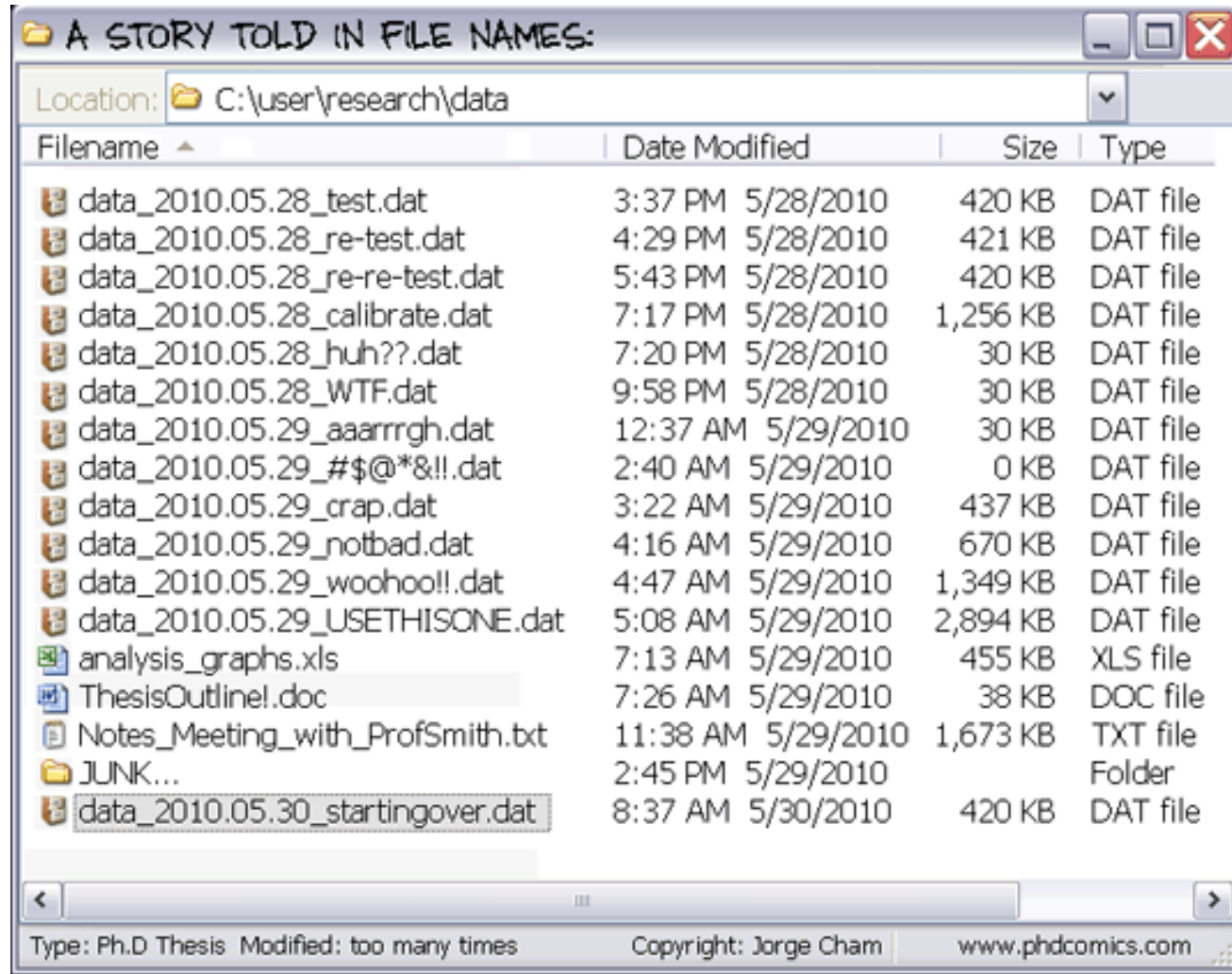
"Piled Higher and Deeper" by
Jorge Cham
www.phdcomics.com

General advice: File naming
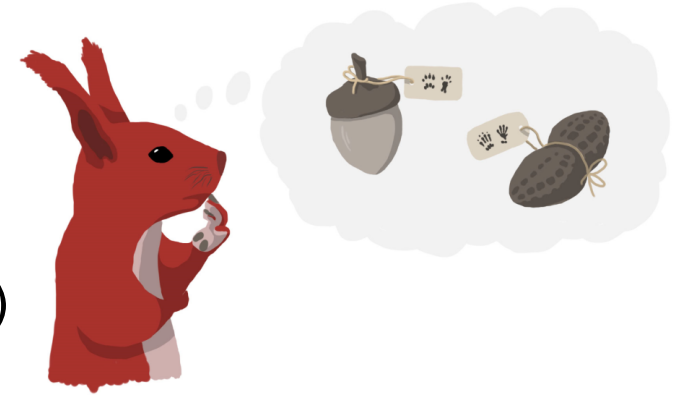
- Use **unique** file names that reflect the **content** (not only for yourself!)

- **Don't rely** solely on the **file extension** regarding file content

- Use **ASCII** characters, **no spaces,** and **points only before file extensions**
(special characters such as ~ ! @ # $ % ^ & * ( ) ` ; < > ? , [ ] { } ' " and | should be avoided)

- Unique file names should be **independent of upper-** and **lower-case** letters (not all OS are case-sensitive)

- Write dates like **YYYY-MM-DD** or **YYYYMMDD** (you will get chronological order of your files, if at the beginning of the file name)

- Use **leading zeros** in a **sequential numbering system** (e.g., file_v01 to file_v10 for a 1 to 10 numbering sequence)

README file for documentation of systematic file naming method

File naming schema (example)

File type	Filename schema	Schema key	Example filename
Microscopic image	[Date]_[microscope]_[signal]_[image number]	Date: Date of image capture in YYYYMMDD format, microscope: name of microscope, signal: source of light or dye used, image number: sequential order of taken images from 1:100	20210623_lsconf28_GFP_001.tiff
…			

File name abbreviations (example)

File name descriptor	Abbreviation's key
Name of microscope	lsconf28: laser scanning confocal microscope located in room 28 of the microscopy facility at…
…	

Reproducibility and reuse in the context of long-term preservation

Reproducibility and reuse as use cases for preservation

Collected data

Preserved data

Reproducible data
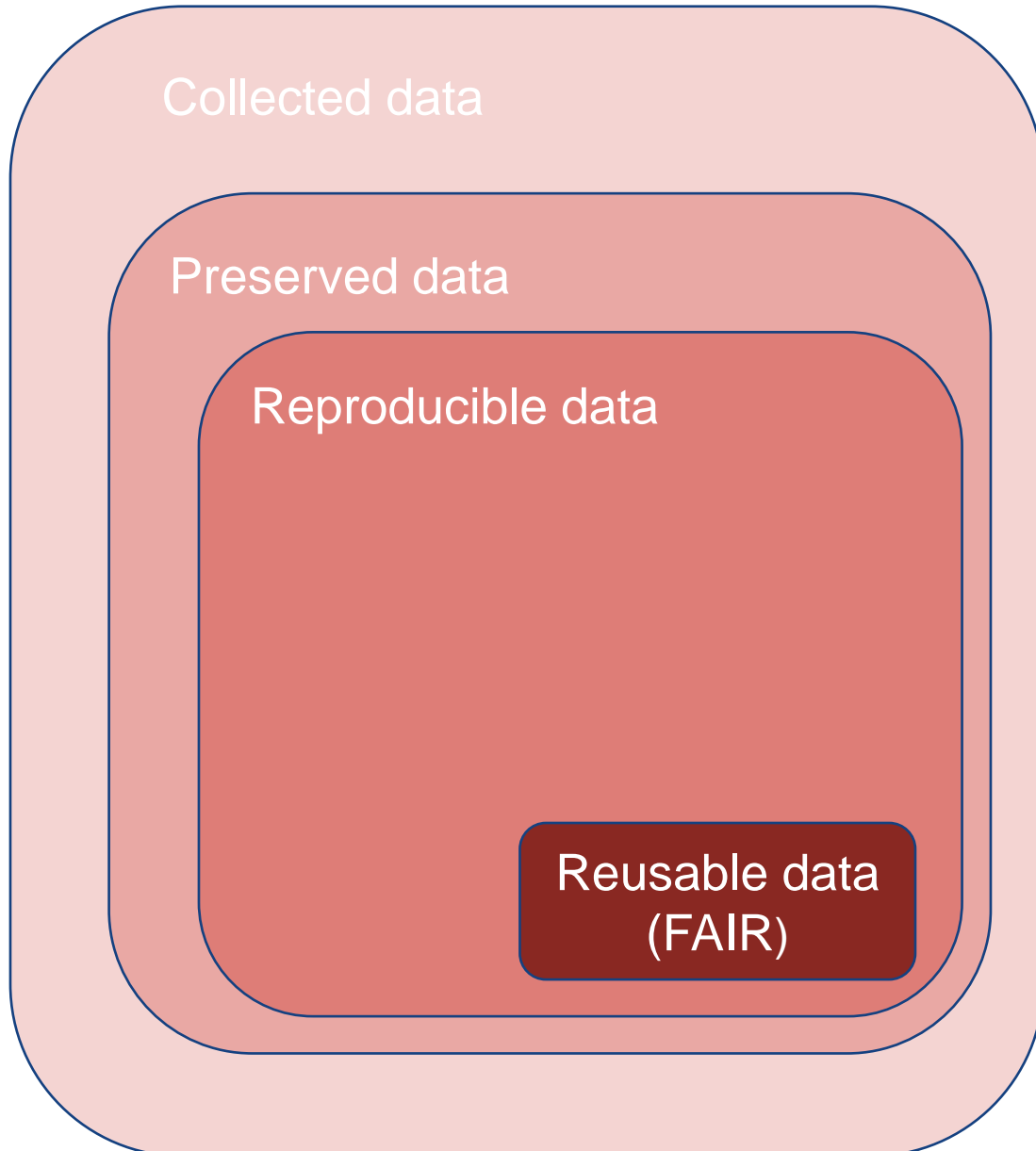
Reusable data
(FAIR)

- **Preservation**
  Includes control of what is where and measures for long-term usability.

- **Reproducible data**
  Research data and **procedural steps** ensuring the **reproduction of research results** are carefully archived and documented

- **Reusable data**
  Data are provided according to the FAIR principles, typically in an online repository with a persistent identifier (e.g., DOI), sufficient metadata including license information.
  They are **F**indable, **A**ccessible, **I**nteroperable, **R**eusable.

The FAIR data principles matrix

https://www.go-fair.org/fair-principles

How do you know if your data is FAIR?

Findable	F1. **(Meta)data** are assigned a globally unique and persistent identifier F2. Data are described with rich metadata F3. **Metadata** clearly and explicitly include the identifier of the data they describe F4. **(Meta)data** are registered or indexed in a searchable resource
Accessible	A1. **(Meta)data** are retrievable by their identifier using a standardised communications protocol A1.1 The protocol is open, free, and universally implementable A1.2 The protocol allows for an authentication and authorisation procedure, where necessary A2. **Metadata** are accessible, even when the data are no longer available
Interoperable	I1. **(Meta)data** use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. **(Meta)data** use vocabularies that follow FAIR principles I3. **(Meta)data** include qualified references to other (meta)data
Reusable	R1. **(Meta)data** are richly described with a plurality of accurate and relevant attributes R1.1. **(Meta)data** are released with a clear and accessible data usage license R1.2. **(Meta)data** are associated with detailed provenance R1.3. **(Meta)data** meet domain-relevant community standards

The FAIR data principles matrix

https://www.go-fair.org/fair-principles

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}
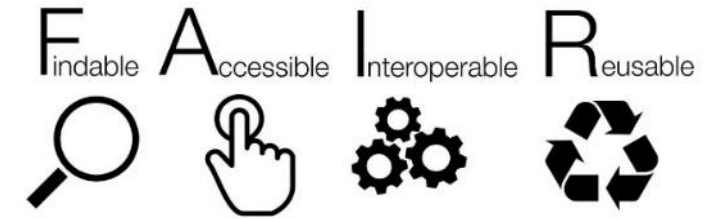
How do you know if your data is FAIR?

What does it mean for YOU?

- To make data FAIR a **lot of metadata** are required

- $\mathbf{F}_{AIR}$: Your research data need a **persistent identifier** (e.g. DOI)

- $_F\mathbf{A}_{IR}$: **Protocols for data transfer** are open (e.g. **https** and **ftp**) to access at least the metadata and to allow **authentification of creator(s)** and to set **user-specific rights**.

- $_{FA}\mathbf{I}_R$: (Meta)data should be **machine-readable** and follow **common standards** (e.g. Dublin-Core)

- $_{FAI}\mathbf{R}$: **Sufficient documentation** which enables exploitation by a **peer**; preferring **open, long-term viable file formats** for your data and metadata

- Many issues for compliance with FAIR data principles concern the **data repository** (e.g. discipline-specific repositories)

Value of data: the stages of data management

active stage

semi-active stage

data in usage

data in retention

Data repository
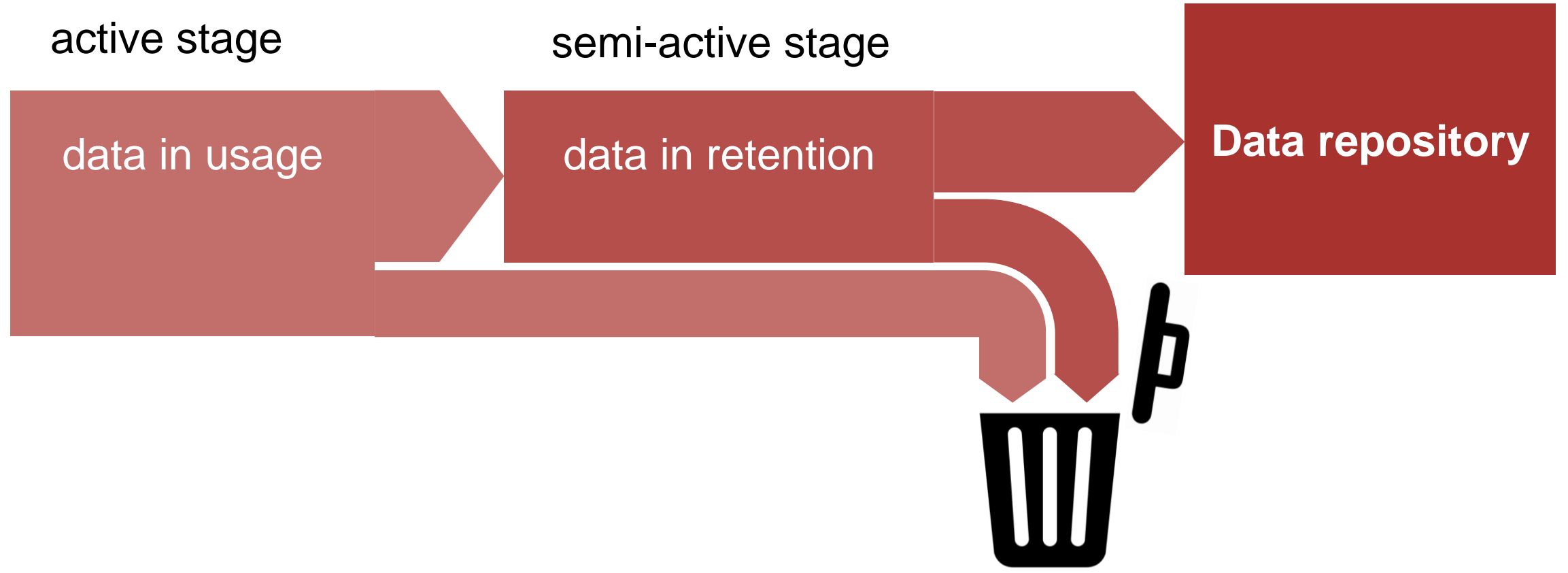
whole image: adapted from ETH Zurich University Archive

Long-term value of research data

Already existing research data

- **Highly reused** (for research, teaching, outreach, product development, commercial application, policy making)

- Represent a **landmark of scientific discovery** (e.g., from nobel prize winner)

Own research data

- **Unique**: not preserved professionally elsewhere

- **Difficult** or **costly** to reproduce

- **Unrepeatable**: observational data (astronomical events, environmental events, animal behaviour)
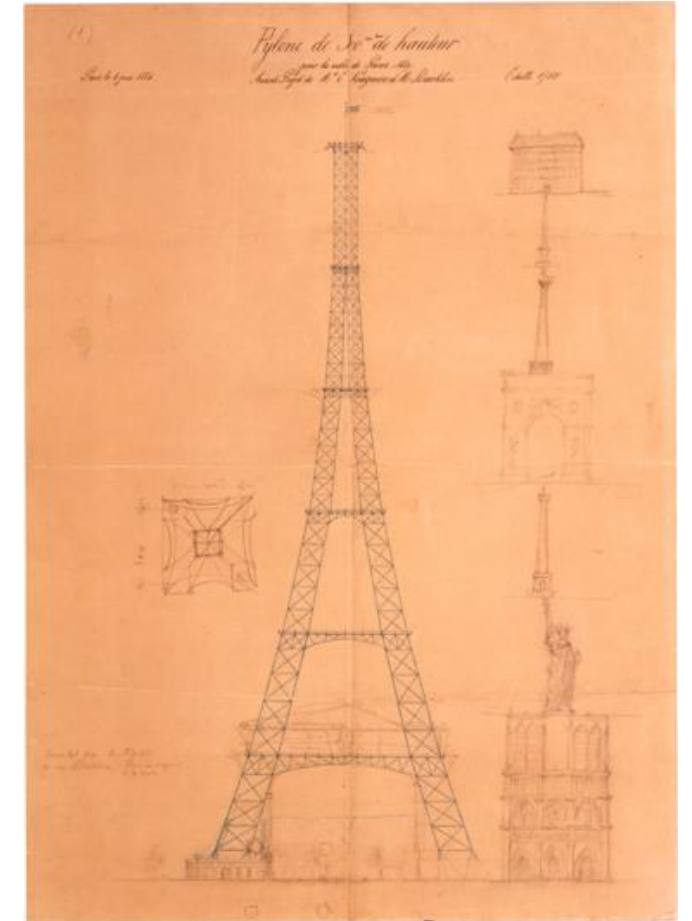
ETH Zurich University Archive, Hs 1092: Köchlin

Criteria for selection of what data to share and keep

Clearing up

- Move personal files to your private storage.

- **Remove redundant files** (e.g., draft versions of publications).

- Check **your DMP** and documentation: are they accurate? Otherwise **update!**

- **Hand over data and documentation** to your supervisor **not last minute**.

- There should be **a written agreement** (template available at legal office) on access and conditions of data reuse after your leaving of the group stipulated ideally **at the start of a new research project**.

Be realistic!

- Estimate your and your community's *current* needs and expectations vs. *future* ones.

- Keeping and preserving vs. sharing and/or publishing: you need to **decide what applies to which data**.

- What **information and additional tools would you or any qualified person need** to reproduce your results in a few years?

File format issues – the 'wishlist' from a preservation point of view

- **Open standards** (non-proprietary)

  - If proprietary, convert or, if not possible, include data viewer

  - When in doubt, **keep original and create a copy** in an open or exchange format

- **Well documented**

- **Widely-used and supported** by many tools

- **Uncompressed** (or at least losslessly compressed)

- **Unencrypted**

- **Don't rely** just on the file extension; check **actual format** before file conversion for instance

- Consider that **data might be used in different operating systems**

File format recommendations (for preservation and reuse)

Data	Recommended file format
Text	in ASCII code
Text documents, PowerPoint	PDF/A1-b, (PDF)
Data from spreadsheets	CSV
Spreadsheets (Data + Formulas)	(CSV), (ODF, OOXML)
Data containers	ZIP, TAR
Analysis code - Matlab (*.m), - Python (*.py), R (*.R) - Excel (*.xlsx)	*.m → HDF *.py / *.R → in UTF-8 format or in ASCII text *.xlsx → suitable to only a limited extent; CSV is better
Images (e.g. PNG, JPG)	Uncompressed TIFF, JPEG2000
Audio files	WAV (*.wav) (uncompressed, pulse-code modulated)

More information: https://documentation.library.ethz.ch/display/DD/File+formats+for+archiving

ETH Research Collection and ETH Data Archive

Repository and Digital preservation solution for ETH Zurich, operated by ETH Library

Researcher

Documents

Access

Peers

Research Collection
- University Bibliography
- Open Access Repository
- Research Data Repository

ETH Data Archive
- Long-term preservation
- Copy of all documents from Research Collection
- Large data archives

ETH Research Collection and ETH Data Archive

Repository and Digital preservation solution for ETH Zurich, operated by ETH Library

Research Collection
- University Bibliography
- Open Access Repository
- Research Data Repository

More Information on the topics of Open Access, Data Publishing and the ETH Research Collection in the 6th workshop on 8 November.

ETH Data Archive
- Long-term preservation
- Copy of all documents from Research Collection
- Large data archives

Long-term storage* services at ETH Zurich

Name	Maximum file size	Long-term preservation**	Descriptive Metadata	Costs	Link / Contact
ETH Data Archive (dark archive) via Research Collection	• **10 GB** per file • 50 GB per entry	**Yes** • via data upload in the ETH Research Collection • 10+ years	**Mandatory** per dataset	**Free** of charge for ETH members	https://library.ethz.ch/en/archiving-and-digitising/archiving/digital-long-term-preservation/eth-data-archive.html
Long term Storage (LTS)	• max. per file: **2 TB** • optimal range per file: 10 – 200 GB	**No** • storage only • storage time: on demand	**No** (Readme-file recommended)	**Free** of charge for ETH members	https://ethz.ch/services/en/it-services/catalogue/storage/lts.html
Libdrive (service by Research Collection)	• e.g., **500GB** per file • Multiple TB per entry	**No** • storage only • Storage time: 10 years	**Mandatory** per dataset	**Free** of charge for ETH members	https://documentation.library.ethz.ch/display/RC/How+to+publish+large+datasets

**Long-term preservation: for well described formats, usability is monitored and preserved by active measures.
  Long-term **storage**: data are kept as they were uploaded (sufficient to comply with ETH guidelines for unpublished data)

A glimpse into the toolbox

Tools and services

ETH Library

Services and support along the DLC at ETH Zurich

Illustration CC-BY, Andres Bucher

- **Preservation and Archiving Services**
  - ETH University Archives
  - ETH Web Archives
  - Preservation planning
  - Metadata preparation

- **ETH Research Collection**

- **Publication services**
  - Open access publishing
  - DOI registration
  - Indexing in Google Scholar

- **Reference management software: licences and courses**

- **DMP guidance, templates, and review:**
  - http://www.snf.ch
    → DMP Guidelines
  - https://documentation.library.ethz.ch/x/WwKsB
    → DMP Instructions

- **Active data management services (storage, wikis, openBIS etc.)**

- **High-performance computing**
- **Research with strictly confidential data**

Define the research question

Create project proposal and DMP

Collect, store and document

Evaluate and select

Process, analyse and interpret

Publish and manage access

Preserve

Verify and re-use

ETH Library

Scientific IT Services

Advice / Training / Infrastructure

ETH library: https://library.ethz.ch

Scientific IT Services: https://sis.id.ethz.ch/

Services and support along the DLC at ETH Zurich
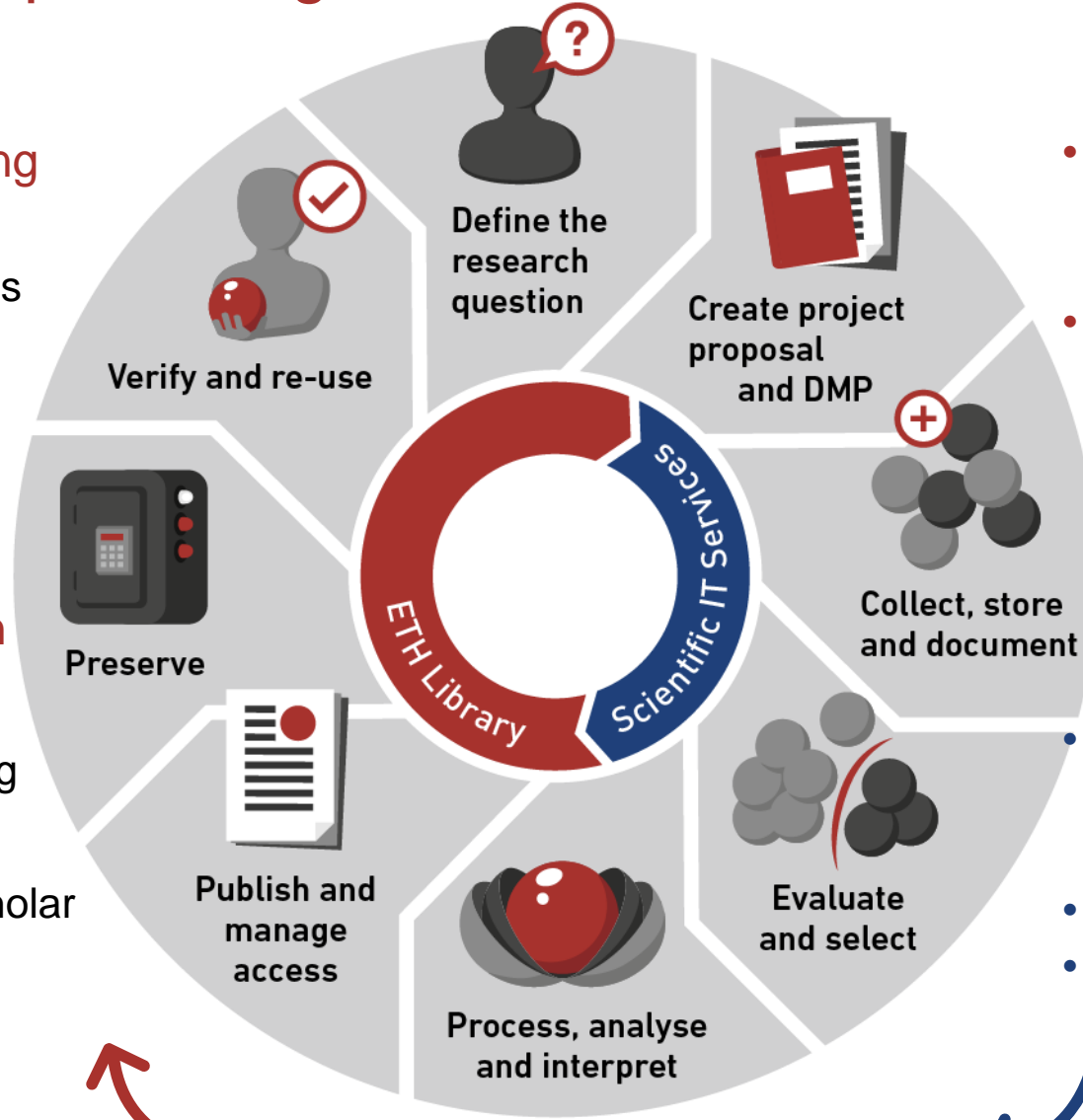
- **Preservation and Archiving Services**
  - ETH University Archives
  - ETH Web Archives
  - Preservation planning
  - Metadata preparation

- ~~ETH Research Collection~~

  More information in the 6th workshop on 8 November

  - DOI registration
  - Indexing in Google Scholar
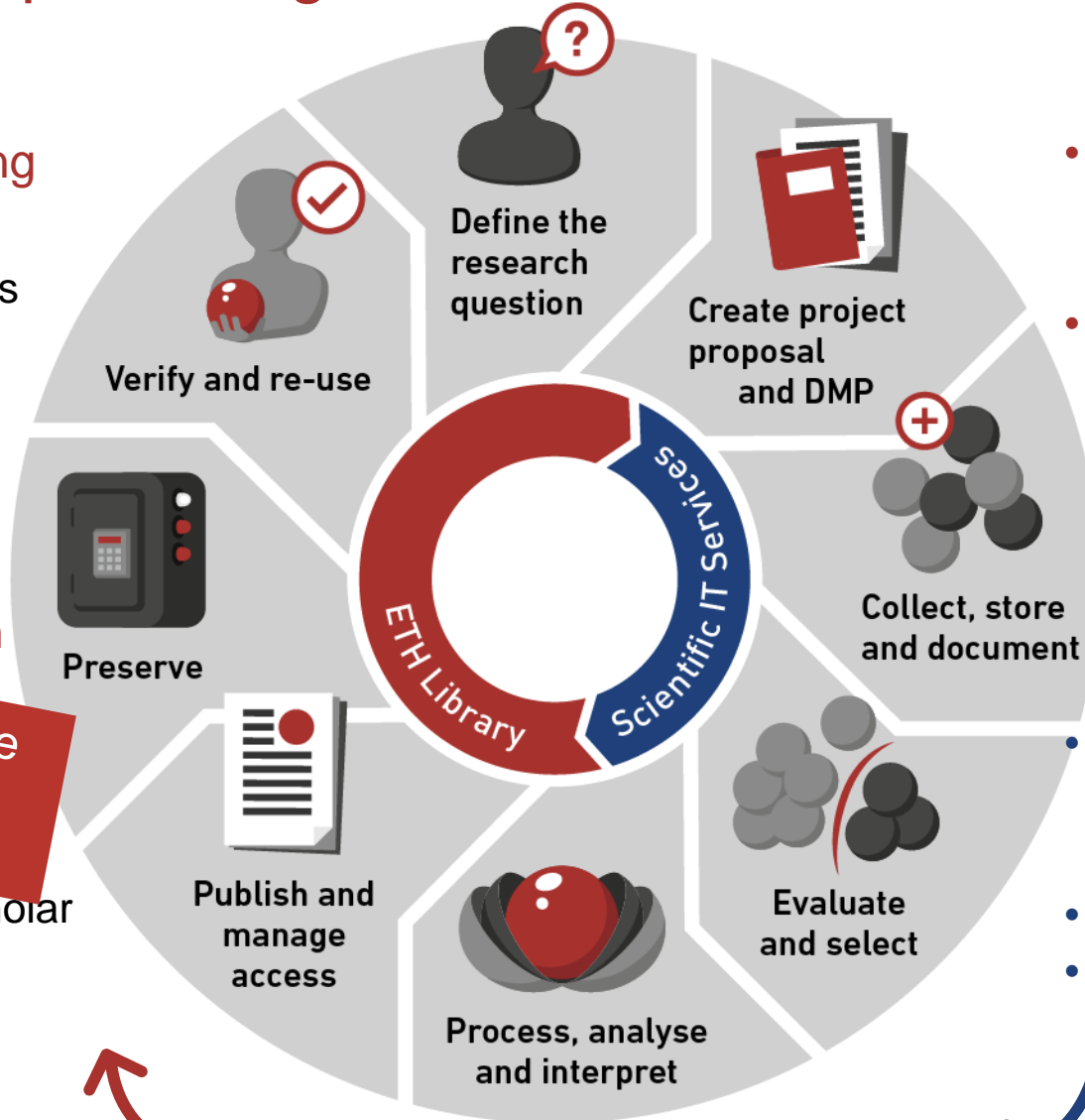
ETH library:
https://library.ethz.ch

Illustration labels: Define the research question · Create project proposal and DMP · Collect, store and document · Evaluate and select · Process, analyse and interpret · Publish and manage access · Preserve · Verify and re-use · ETH Library · Scientific IT Services

Advice / Training / Infrastructure

- **Reference management software: licences and courses**

- **DMP guidance, templates, and review:**
  - http://www.snf.ch
    - → DMP Guidelines
  - https://documentation.library.ethz.ch/x/WwKsB
    - → DMP Instructions

  More information in the next three workshops on 18, 25 October and 2 November

- **High-performance**
- **Research with strictly confidential data**

Scientific IT Services:
https://sis.id.ethz.ch/

Data Storage at ETH Zurich

- **Network Attached Storage (NAS): standard storage solution** at ETH Zurich, used in most research groups, offered by IT Services

- **There are more options for your data (always check with <u>your IT Service Group (ISG) first</u>):**
  - Cost Defined Storage (CDS): For very large storage requirements (>100 TB)
  - Long-term Storage (LTS):
    - Tape storage for data worthy of keeping ("Datenendlager")
    - Data should be compressed and sized between 10 - 200 GB

- **The following are mainly for data sharing and file transfer:**
  - Polybox
  - Microsoft 365 OneDrive

… Cloud-Services must be used with caution

Image: ✓ Free to use. ✓ No attribution required.

Image: Adobe Stock #615830132 (Educational licence)

- ETH Zurich has clear rules on your work-related use of cloud services

- Access rights and the geographical location/jurisdiction of cloud storage services matter:
  e.g. data protection standards in CH/EU vs. USA

ETH Tools: File sharing and collaborative work in the cloud

polybox
- Web client (https://polybox.ethz.ch) and local sync client (to be installed)
- Use polybox as a logical memory stick and store your data also in the ETH IT infrastructure; no standard backup procedure
- Data hosted at ETH Zurich
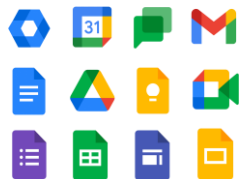- Not suitable for strictly confidential data

Microsoft 365
- https://unlimited.ethz.ch/display/itkb/Microsoft+365
- Data stored in Switzerland or EU
- OneDrive, SharePoint and MS-Teams can be used for confidential data

Google Workspace
- https://unlimited.ethz.ch/display/itkb/Google+Workspace
- Data stored in Europe for the core services
- So far, only suitable for public and internal data (not for confidential data)

Regularly updated list of released external cloud services
https://ethz.ch/staffnet/en/service/information-security/usage-external-cloud-services/list-external-cloud.html

File transferring tools for research data

Globus subscription for ETH Zurich (file transfer)

- More information: https://scicomp.ethz.ch/wiki/Globus_for_fast_file_transfer, https://www.globus.org/

- Can transfer big volumes of data fast, e.g. between institutions, or between your workstation/notebook and the HPC Euler Cluster

- For support and additional features contact globus-support@id.ethz.ch

SWITCHfilesender

- More information: https://www.switch.ch/services/filesender/

- Secure, web-based service for sending files of up to 300 GB

- Data hosted in a SWITCH cloud in Switzerland

- Free of charge for university members

Infomaniak

- More information: https://www.swisstransfer.com/en

- Files are stored in Switzerland

- Sending files of maximum 50 GB is free of charge

Code Management

- Many journals require **code availability** after publication and **during review** (see *Nature 555, 142*)

- **Version control systems**

  - Software tools specialized on **managing and documenting changes** to source code over time

  - Used for managing large code bases

  - They are the standard in professional software dev...

- **Tools**

  - ID-SIS provides hands-on trainings on git for code management (info @ https://sis.id.ethz.ch/consulting)

More information in the 4th and 5th workshop on 25 October and 2 November.

https://renkulab.io/
by Swiss Data Science Center (SDSC)

https://subversion.apache.org
also available at ETH Zurich

GitLab
https://gitlab.ethz.ch
hosted at ETH Zurich

GitHub
https://github.com
in the cloud
SNSF requires justification for use of commercial tools!

Collaborative writing / notes / document management tools (examples)

Cloud based – use with consideration:

Hosted at ETH Zurich:

SharePoint

https://www.ethz.ch/services/en/it-services/catalogue/web-application-hosting/sharepoint.html

Confluence

https://www.ethz.ch/services/en/it-services/catalogue/web-application-hosting/wiki.html

https://teams.microsoft.com/

jupyter

https://jupyter.org/

Overleaf

https://www.overleaf.com

EVERNOTE

https://evernote.com

Simplenote

http://simplenote.com

Authorea

https://www.authorea.com

hypothes.is

https://hypothes.is

OneNote

https://products.office.com/en/onenote/digital-note-taking-app
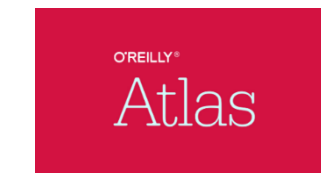
Atlas

https://atlas.oreilly.com

Sometimes, on-site installations are also available

Reference management tools (examples)
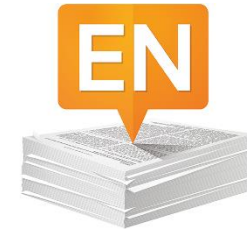
Increasingly cloud-based or synchronised with cloud.

MENDELEY

www.mendeley.com

http://www.jabref.org

EN

endnote.com

zotero
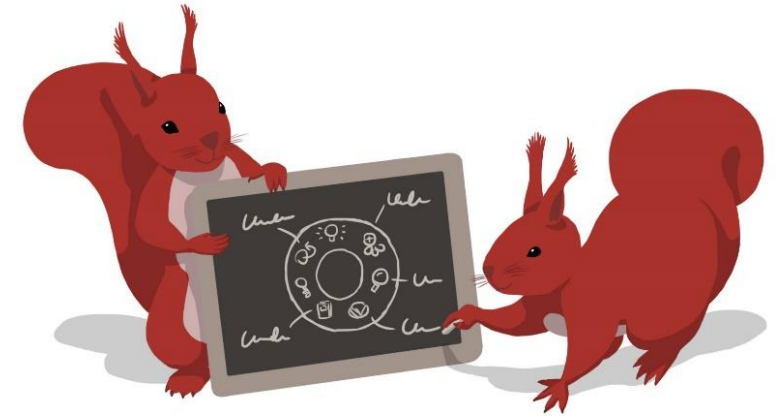
www.zotero.org

BibSonomy

www.bibsonomy.org

https://tagpacker.com
(Web-Bookmarking)

Criteria for choice of tools and some final remarks

☐ **Location** of the service and its servers

☐ Legal regulations on **data protection**

☐ **Sustainability** and **trustworthiness**

☐ **Access rights** for your data

☐ **Licences** and immediate/long-term **costs**

☐ How can you get your **data back**?

▪ There are **several options** available

▪ There is **no "best for all"**, it all depends on your research field, data types and research workflow

▪ RDM is the basis for having **FAIR** published data

▪ RDM requires WORK & TIME, but the time spent on this is an **investment** for the future!

Feeling overwhelmed?

Where to get useful information, help and further training

Additional links and trainings
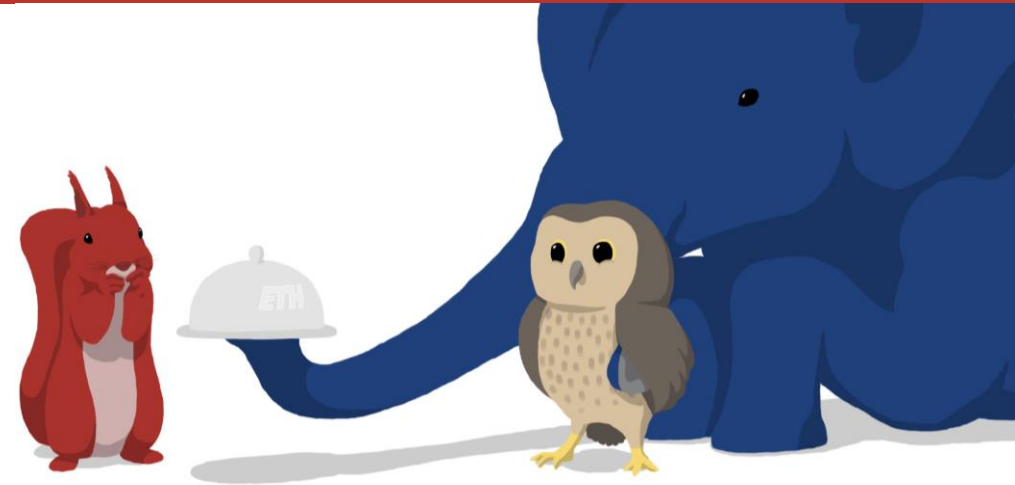
Research Support Services @ ETH Library

Scientific information

Publications
- (E-) Journals
- (E-) Books
- Handbooks

Data
- Data Bases
- Object collections
- Archives

Research

Information retrieval
- Online Access / Loan
- Media/Data purchase on demand

Consulting and training
- Research data management / plans
- Copyright
- Reference Management

Publication and preservation

Publication services
- Open Access publishing
- DOI registration
- Indexing in Google Scholar

Research Collection
- Publication platform
- Open Access
- Research Data

Preservation services
- Metadata preparation
- Preservation planning

**Archiving services**
- ETH University Archives
- ETH Web Archives

What we offer

- **General information** about RDM at ETH Zurich

  - https://library.ethz.ch/rdm

  - http://www.library.ethz.ch/digital-curation

  - https://ethz.ch/services/en/service/a-to-z/research-data.html (together with SIS)

- **Wiki-page** with instructions and downloads

  - DMP: list of guides and templates, guidance for the SNSF DMP, instructions for a DMP at ETH Zurich

  - Data publishing: publishing and preservation steps, workflow for data publication at ETH, suggestions for data availability statements

  - Data formats: file formats for archiving, recommendations for TIFF

  - Suggestions for drafting a data management strategy

- **Consulting** on questions related to RDM
  data-management@library.ethz.ch or researchdata@ethz.ch

- **Individual trainings** on RDM (upon request for five and more people)
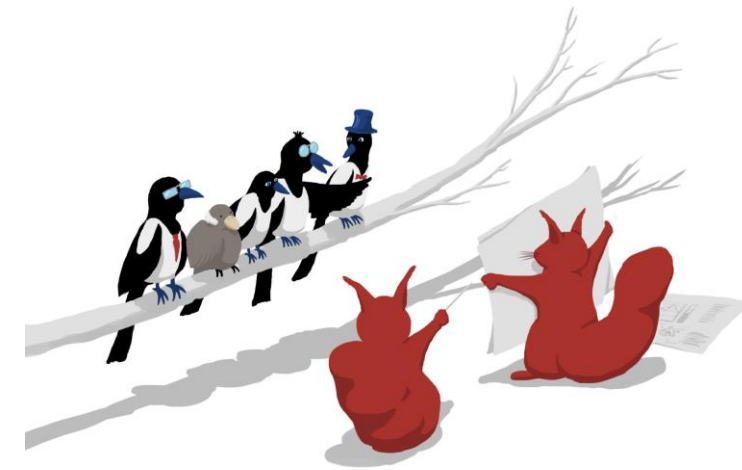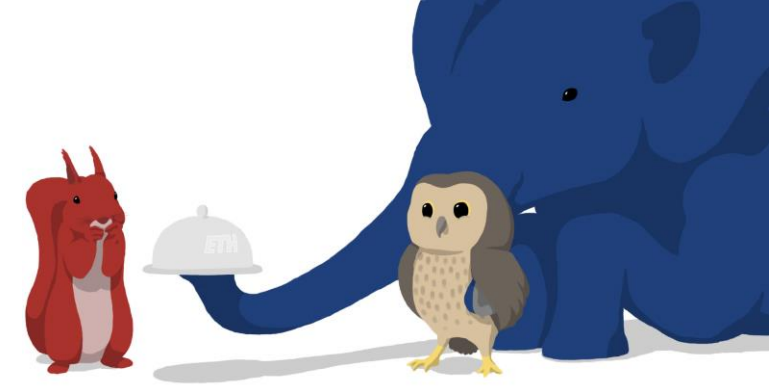
IT services and ETH transfer

IT Services

- Storage provisioning (usually via your IT Support Group)

- ARDM services based on openBIS: https://sis.id.ethz.ch/services/index.html#research-data-management

- SIS bioinformatic co-analysis service for -omics data (contact: sis.helpdesk@ethz.ch)

Learn more in the 4th and 5th workshop on 25 October and 2 November

ETH transfer: https://ethz.ch/en/industry/transfer.html

- Open software disclosure workflow with ETH Data Archive

- Advice on intellectual property, patents, licensing of software etc.

Further training

- **ETH Library**: Training courses on **information research, reference management, data management, scientific writing and open access and more**

- **6th ETH Research Data Management Summer School, 10-14 June 2024,** *registration opens in January 2024*

- **SIS**: Trainings on **coding best practices, Python, Python for NGS, git, openBIS and more**

- Courses offered by the ETH Information Center for Chemistry/Biology/Pharmacy

What messages are you taking home with you?

Why spend time and effort on RDM?

Duties

Meet funders' and institutional requirements (SNSF, EU Horizon)

Keep work in accordance with good scientific practice: transparent and valid

Take part in the discussion with your community, institutions and funders

Benefits

Preserve non replicable data

Raise your impact by citable data

Avoid redundant data creation

Facilitate collaboration

Enable data re-use and sharing

Highlight patterns or connections that might otherwise be missed

Take-home messages

Best start early – but it is never too late

- Keep it as **simple as possible** – but **distrust** it!

- **Separate different categories** of data right from the point when they are created.

- Try **to agree on a structure** which suits most projects in **your group**.

- *By writing* just a one-page document for everyone's reference …

   … agree on simple rules on **what is stored where**.

   … agree on **conventions for file and folder naming**.

- *Just do it!* – This might not be as easy as it sounds…

Take-home messages

- Don't let all this make you nervous – it **must not dominate** your work!

- Talk to **senior** colleagues and your PI

    - What is **common practice** in your discipline?

    - Get into **deeper discussions**

- Appoint a **data steward** in the group/institute/department

- Check what your **local service providers** can offer

- Check for ETH funding opportunities for projects involving Open Science and Open Research Data practices: https://ethz.ch/en/research/open-science.html

We are here to support you!

ETH zürich

Dr. Julian Dederke, Dr. Fabian Schmid
Research Data Management and Digital Curation
julian.dederke@library.ethz.ch / fabian.schmid@library.ethz.ch

ETH Library
Research Support Services
Weinbergstrasse 74
8092 Zurich
http://www.library.ethz.ch/rdm
data-management@library.ethz.ch

ETH Library and Scientific IT Services
researchdata@ethz.ch