Data Management, Open Access and the ETH Research Collection

Presentation

Author(s):
Ziehmer, Malin Michelle; Rees, Rainer

Publication date:
2019-05-22

Permanent link:
https://doi.org/10.3929/ethz-b-000343462

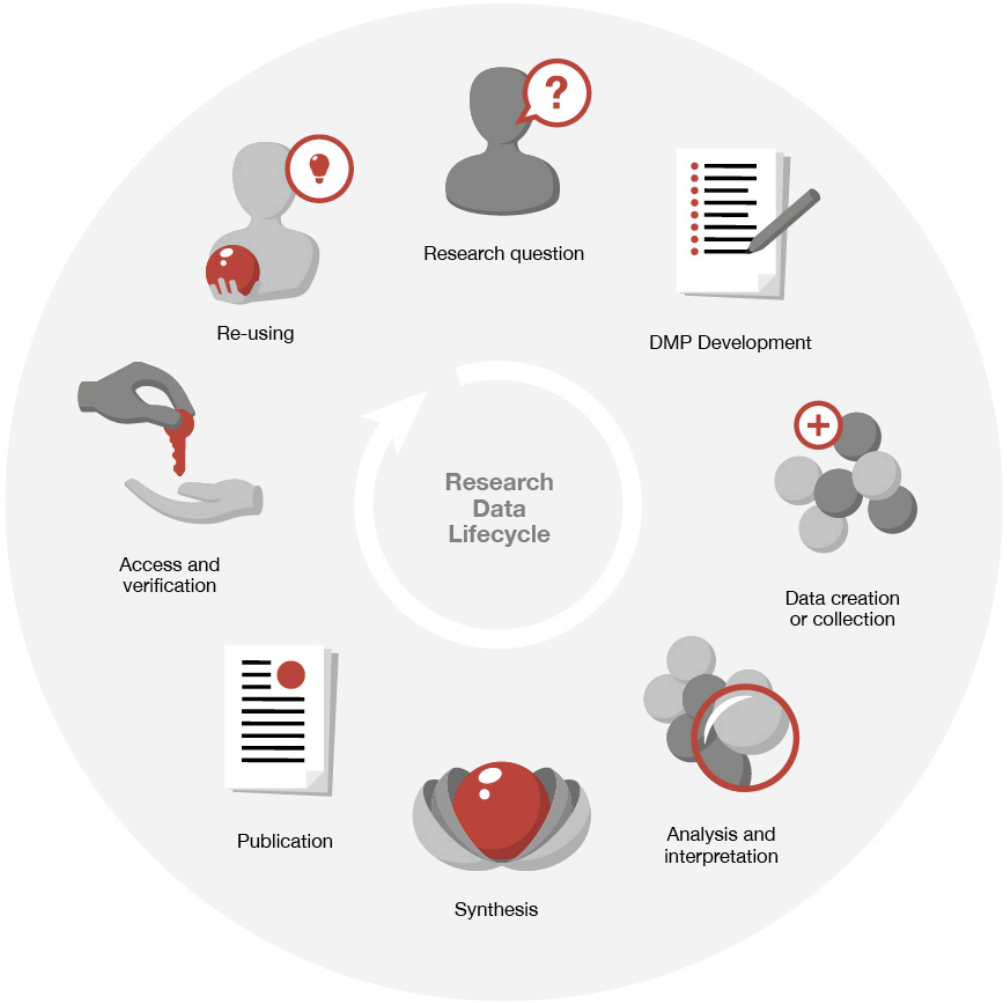
ETH zürich

Research Data Management – A (pretty) short overview

Dr. Malin Ziehmer

What is data management and why does it concern you?

The data life cycle

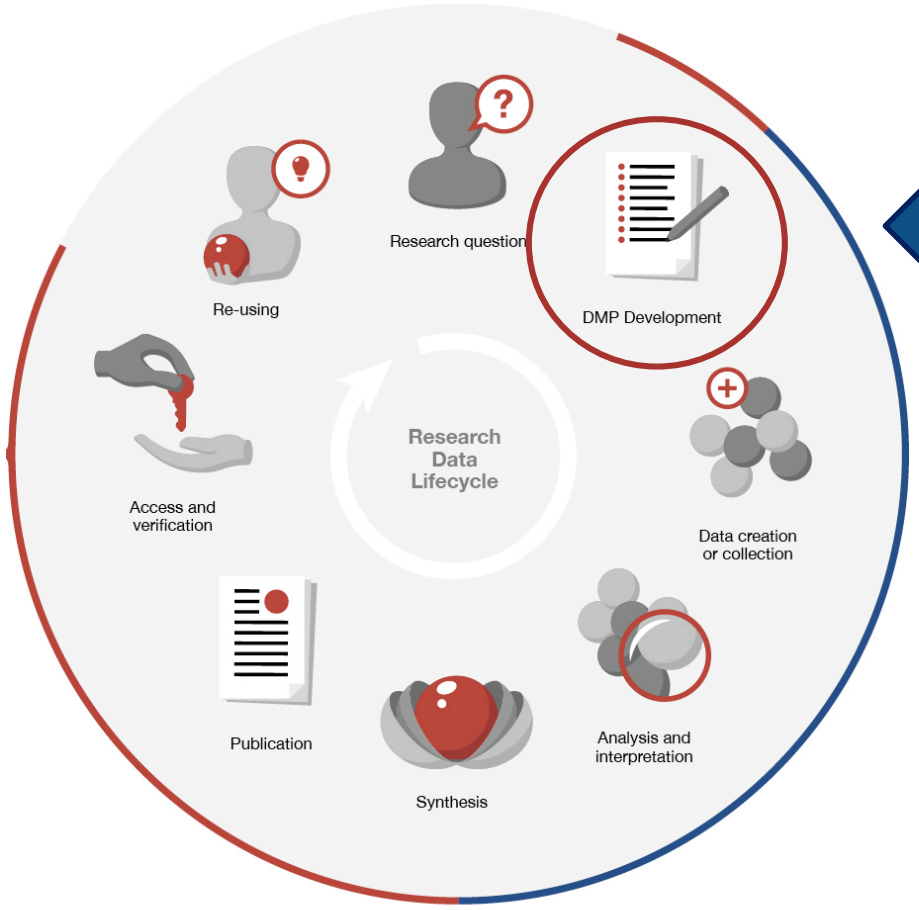


The data life cycle

Publication and preservation: annotate, share, publish, preserve data at the end of the project/publication

Research Support Services

- Publication services
- Preservation services
- Training and consulting



Active data management: annotate, store, backup data while it is produced

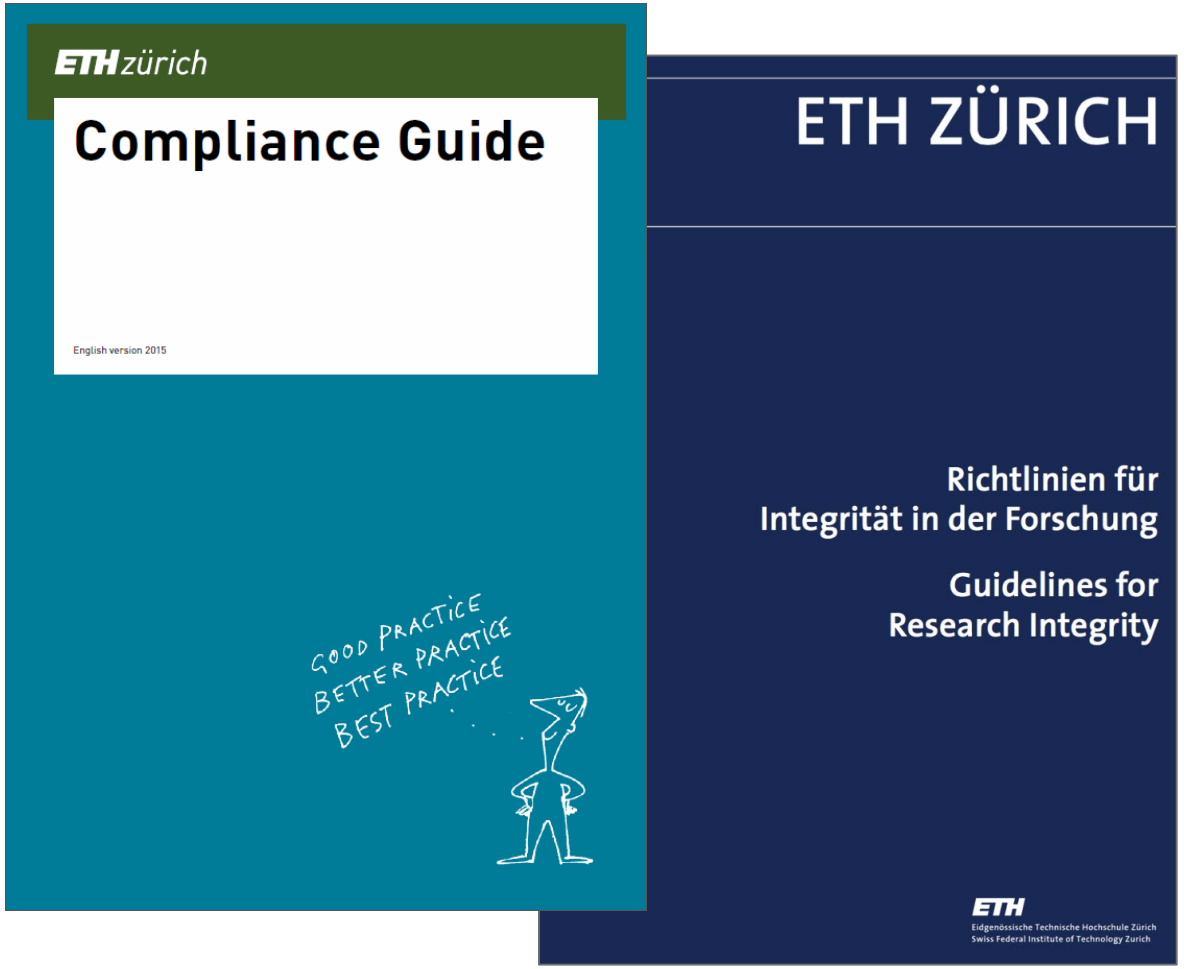
Scientific IT Services

- Active research data management services
- Soft- and hardware
- Training and consulting

Regulations, intellectual property, privacy, and access rights

An overview

What you (should have) received…



…at the beginning of

- your studies
- your PhD
- your employment

at ETH Zurich

… and what they state…

https://rechtssammlung.sp.ethz.ch/Dokumente/133en.pdf

https://doi.org/10.3929/ethz-b-000179298

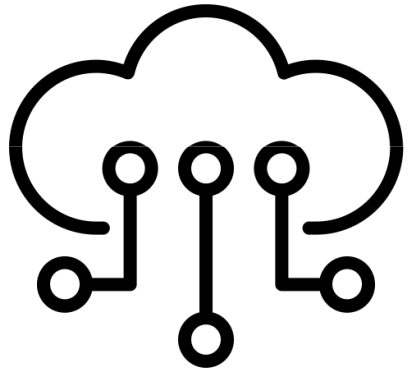
- […] all ETH members […] are required to integrate the general conditions and internal directives into the work process.
- In the research context, **the project manager plays an active role in guiding and monitoring junior scientists**. In particular, he or she is responsible for making sure that everyone involved in the project is aware of the research integrity guidelines.
- Junior scientists are given **appropriate guidance**.
- **Primary data is carefully archived**.

- At the ETH Zurich research is founded on intellectual honesty. Researchers […] are committed to scientific integrity and truthfulness in research and peer review.
- For research data, see Art. 11, in particular.

Roles and responsibilities in RDM

- Project **Members**:
  - **adhere to the principles of good scientific practice** and the guidelines for Research Integrity at ETH.
  - All steps of treatment of primary data must be **documented in a form appropriate to the discipline** and results must be **reproducible**.
- Project **Manager**:
  - responsible for **data management** (data collection, storage, data access, compliance with data protection requirements, **retention for the period prescribed by the discipline** ...).
  - **Ensures that all** research project participants **are aware of the guidelines**.
  - Determines together with the professor, **which departed project members should retain access** to the primary data or materials.

Cloud Computing@ ETH Zurich
Rules and Regulations



© Symbolon from Noun Project

- The removal of sensitive data from ETH Zurich (e.g. research data subject to contractual confidentiality with third parties, important ETH Zurich business data such as financial data, personal employee or student data, reports) is not permitted. **ETH** Zurich **must retain access** to and control over such data **at all times**.

- The **use of cloud** and social media **services** (e.g. Facebook, Google, Dropbox) in research, for exchange with researchers at other universities, or in teaching for exchange with students (lecture folders, etc.) **is permitted as long as no sensitive** ETH Zurich **data are affected and no third party rights**, in particular privacy or intellectual property rights, **are infringed**.

Links:

https://www.ethz.ch/content/dam/ethz/associates/services/Service/IT-Services/files/broschueren/rechtliches/de/Merkblatt_Cloud_Computing_MA.pdf

https://itsecurity.ethz.ch/leaflet_example_cloud_EN.pdf

Privacy

© Hea Poh Lin from Noun Project

- People-related data need to be preserved according to **Swiss data protection law**

  Federal Act on Research involving Human Beings (https://www.admin.ch/opc/en/classified-compilation/20061313/index.html)

  Federal Act on Data Protection (https://www.admin.ch/opc/en/classified-compilation/19920153/index.html)

  Swiss Criminal Code (https://www.admin.ch/opc/en/classified-compilation/19370083/index.html)

- Appropriate **anonymization** might be required
- The **deletion** of individual datasets must be possible at all times
- The study subjects need to sign a **declaration of consent**
- More information: ETH Zürich **Ethikkommission** (German): https://www.ethz.ch/services/de/organisation/gremien-gruppen-kommissionen/ethikkommission.html

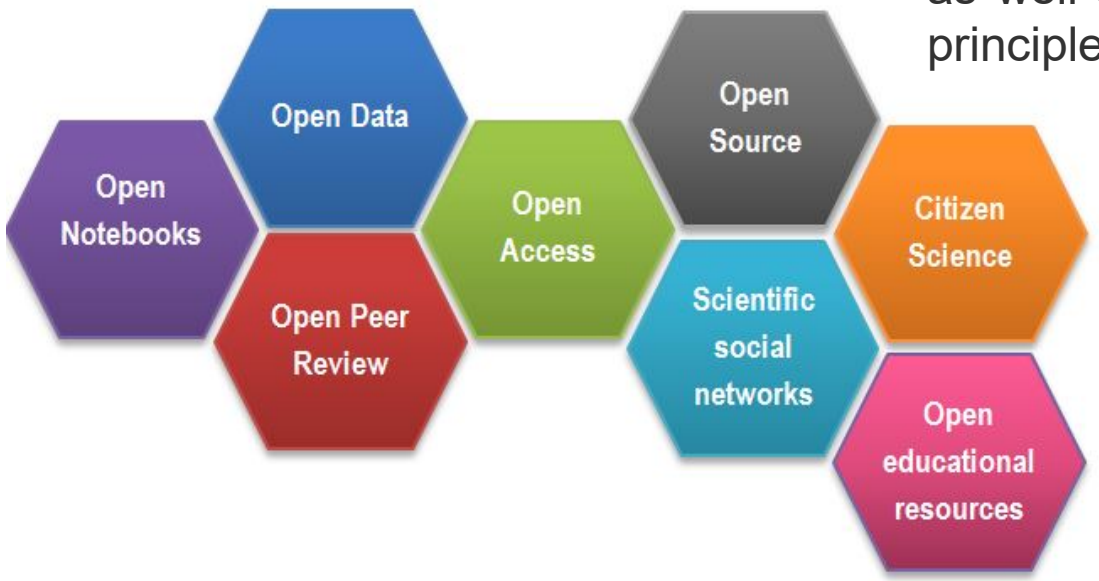
Intellectual Property Rights: What you need to consider

For publications and for data!

- **Respect the rights of others**
  - Third parties
  - Individuals you work with
- In case of doubt: **seek permission** even when a CC-licence is assigned
- Note that according to ETH law, **ETH reserves most immaterial rights** in works by its employees. When in doubt, contact ETH transfer early (www.transfer.ethz.ch)
- Make sure you **keep sufficient rights**
  - E.g. for Open Access Publishing (green path)
  - E.g. with respect to patent applications: ETH transfer early (www.transfer.ethz.ch)

Short excursion: Open Science

Short Excursion: **Open Science (OS)**



"Open Science facets as a beehive" (29.1.2019) by fosteropenscience.eu

"Open Science is the **umbrella term** for all efforts aimed at **achieving more openness in science** and the necessary **paradigm shift**, e.g. **open access to publications and open data** as well as the transition towards research funding based on DORA principles."

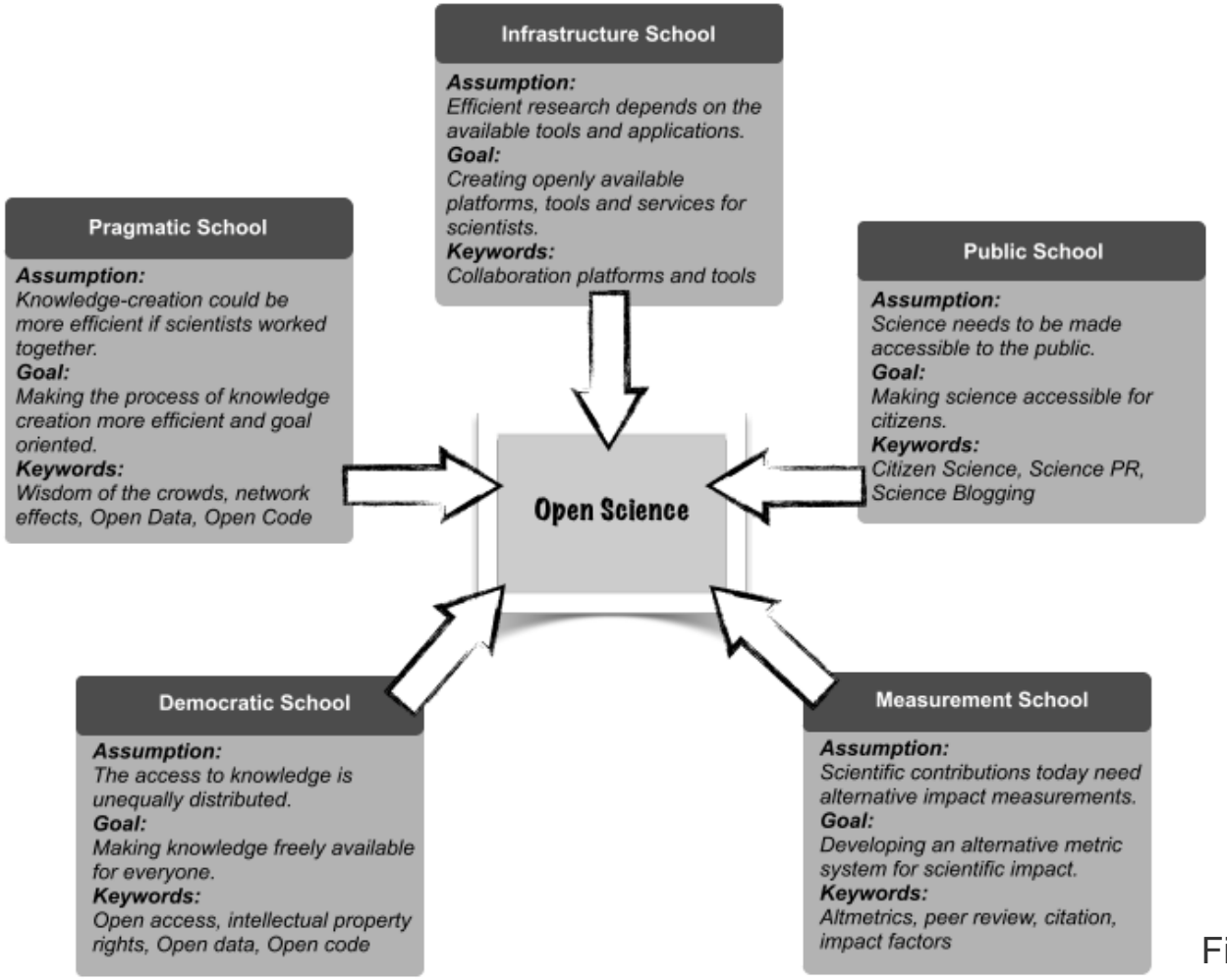
(SNSF, www.snf.ch/openscience)

In short:

The dissemination of scientific knowledge, that is as wide as possible, free of charges, to all users, and accessible online.

(OS MOOC, TU Delft)

Short Excursion: Open Science (OS) Schools of Thought

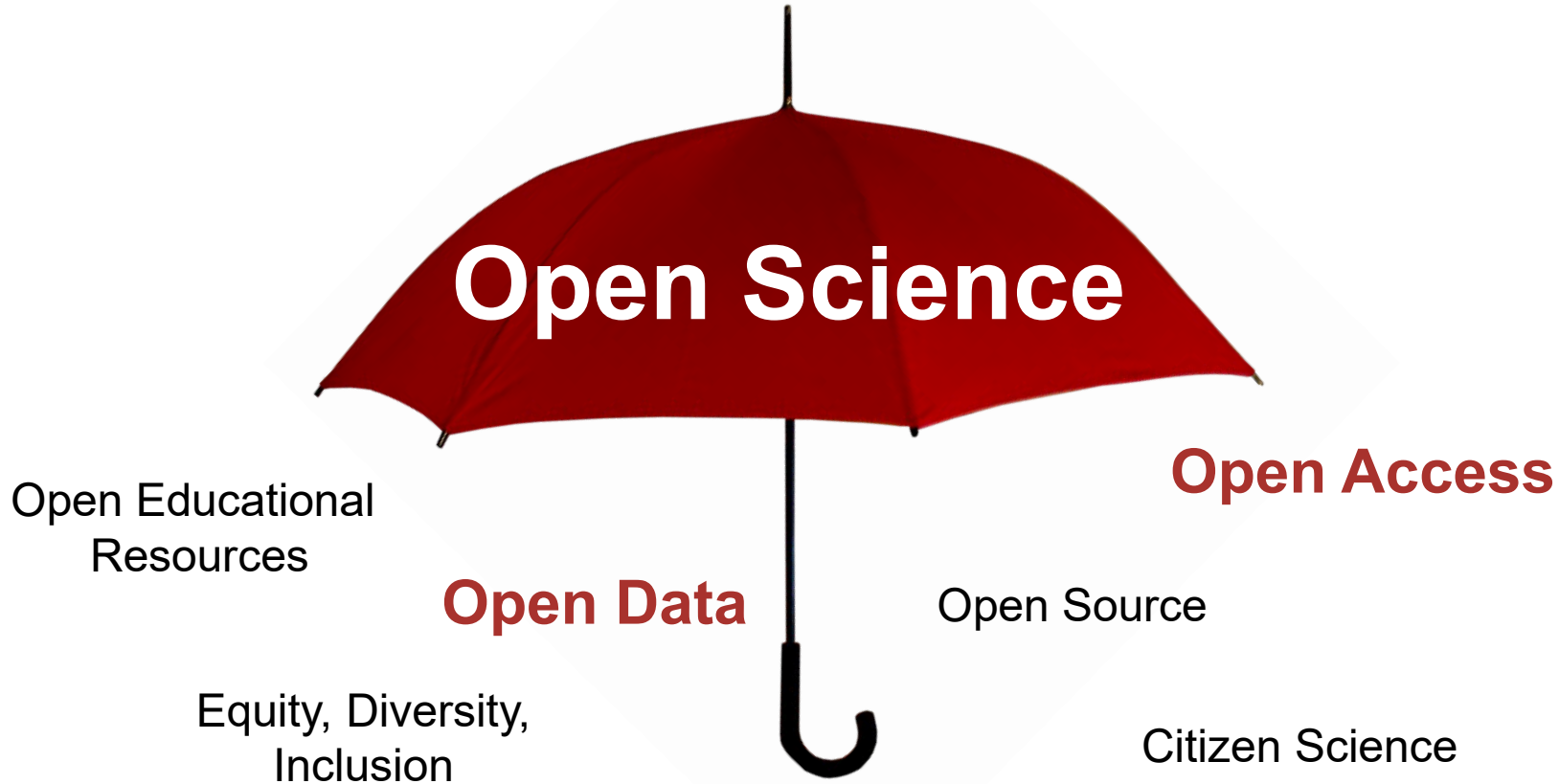


Goals:

- Creating **openly available platforms**, **tools** and **services** for scientists
- Making science **accessible for citizens**
- Developing an **alternative metric system** for scientific impact
- Making **knowledge freely available** for everyone
- Making the process of **knowledge creation** more **efficient** and **goal-oriented**

Five Open Science schools of thought (Fecher and Friesike, 2014)

Short Excursion OS: Components of Open Science



Open Access and Open Data

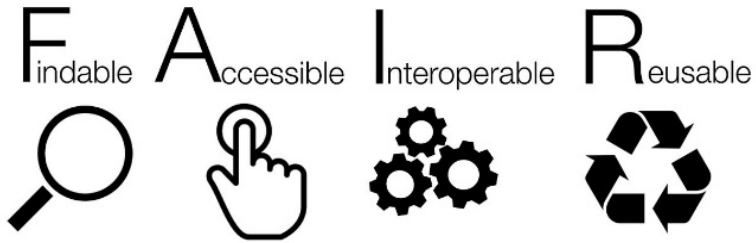
Ideally, an **open access publication** is for anyone to

- read
- download
- copy
- distribute
- print
- search for
- search within
- use in education or another way

with few legal limitations.

Open Data

"Research data should be freely accessible to everyone – for scientists as well as for the general public."
(*SNSF, Open Research Data Policy*)



Data that is **freely** available to **everyone** to use and republish as they wish, **without restrictions** from copyright, patents or other mechanisms of control.

(OS MOOC, TU Delft)

SNSF policy on Open Research Data

Goal of the SNSF:

Research data should be freely accessible to everyone – for scientists as well as for the general public.

Article 47 of the Funding Regulations
(1 Jan 2016, http://www.snf.ch/SiteCollectionDocuments/allg_reglement_16_e.pdf):
"[…] the data collected with the aid of an SNSF grant must also be made available to other researchers for further research and integrated into recognised scientific data pools […]"

→ **A data management plan is just one of the tools to reach this goal**

Data Management Planning

What? Why? How?

What is a Data Management Plan (DMP)?

A brief plan written at the start of a project and updated during its course **to define:**

- **What data** will be collected or created?
- **How** will the **data be documented and described**?
- **Where** will the data be **stored**?
- **Who will be responsible** for data security and backup?
- Which data will be **shared and/or preserved**?
- **How will the data be shared** and with whom?

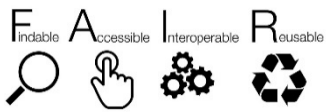
DMPs are e.g. demanded by:

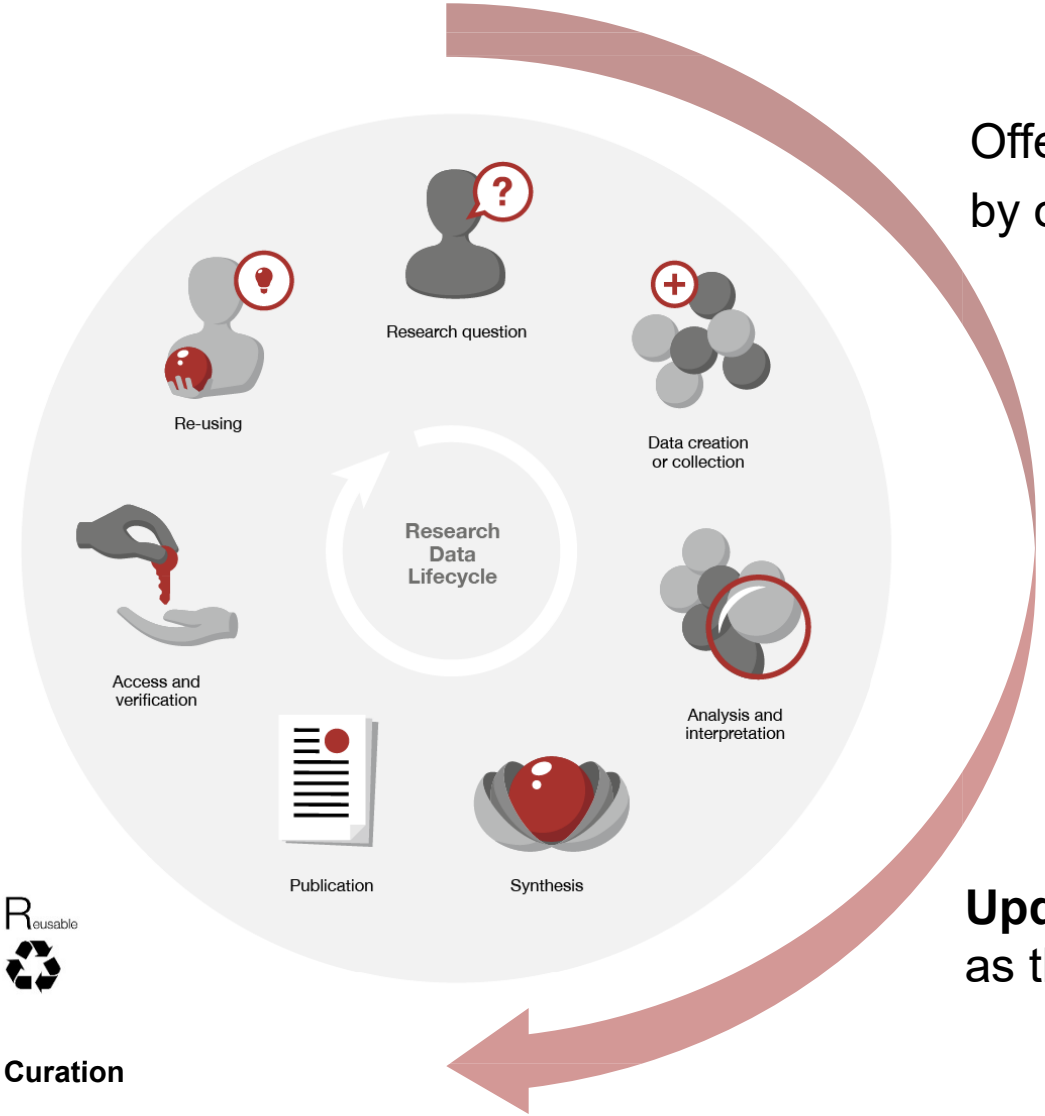
SNSF since October 2017
http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx

Horizon2020 EU funding programme
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Aims of the DMP according to SNSF

- **Planning and documenting the life cycle** of data
- In the ideal case, you only need to **document** your current practice / **best practice** in your field
- **Making data FAIR**:
  - Findable
  - Accessible
  - Interoperable
  - Re-usable





Offering a **long-term perspective** by outlining how the data will be:

- Generated
- Collected
- Documented
- Shared / Published
- Preserved

Updating the plan
as the project progresses

The FAIR data principles matrix

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

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

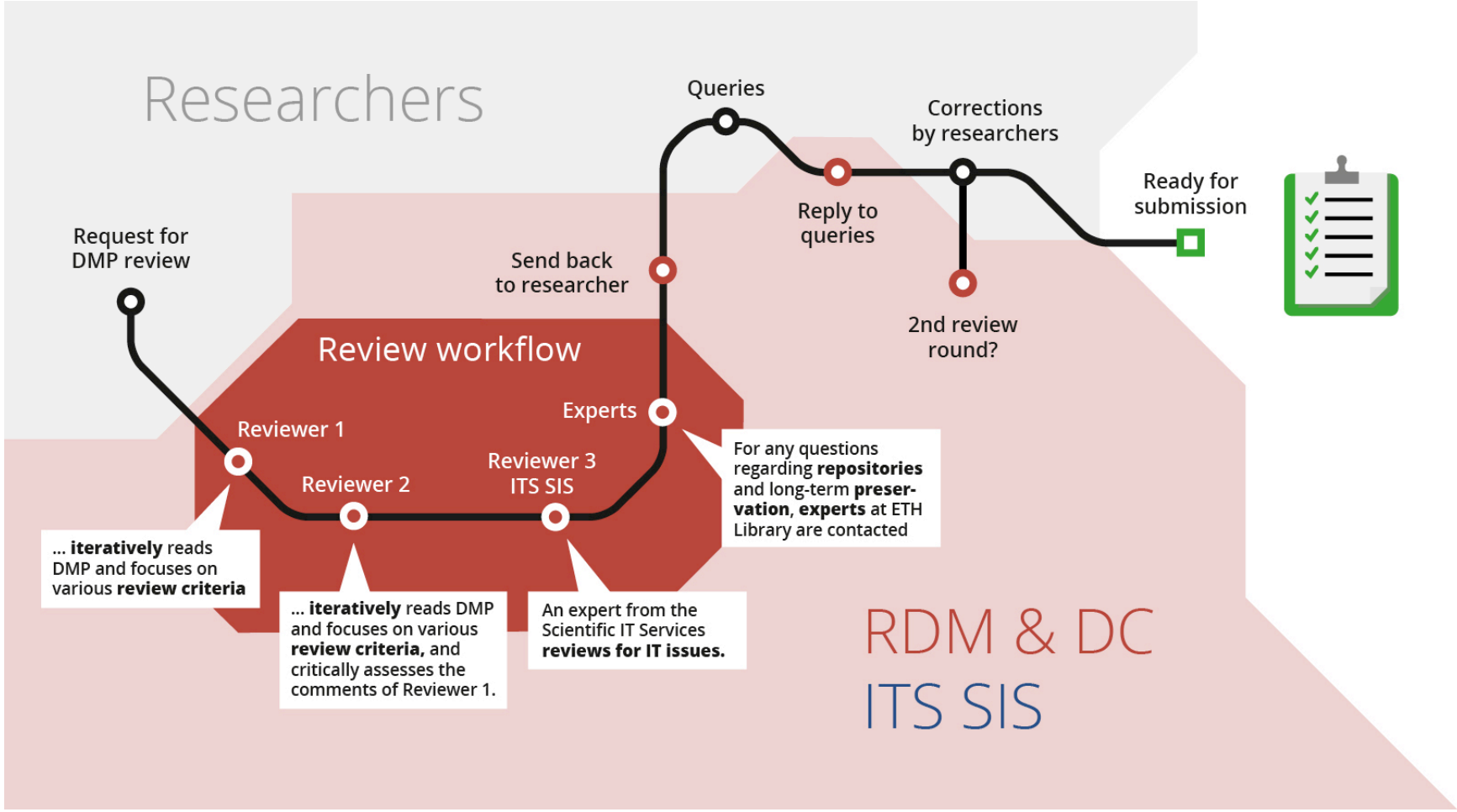
Information to support you

- Collection of **SNSF information** on Open Research Data including **FAQ**: http://www.snf.ch/en/theSNSF/research-policies/open_research_data/

- SNSF's explanation of the **DMP expected content**: http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_en.pdf

- **Guidance for ETH researchers** on filling out SNSF Data Management Plans: https://documentation.library.ethz.ch/display/DD/Guidance+for+ETH+researchers+on+filling+out+SNSF+Data+Management+Plans
  - Includes:
    explanations per question, examples from DMPs, contacts and links specific for ETH Zurich

We review your DMP – contact researchdata@ethz.ch for more



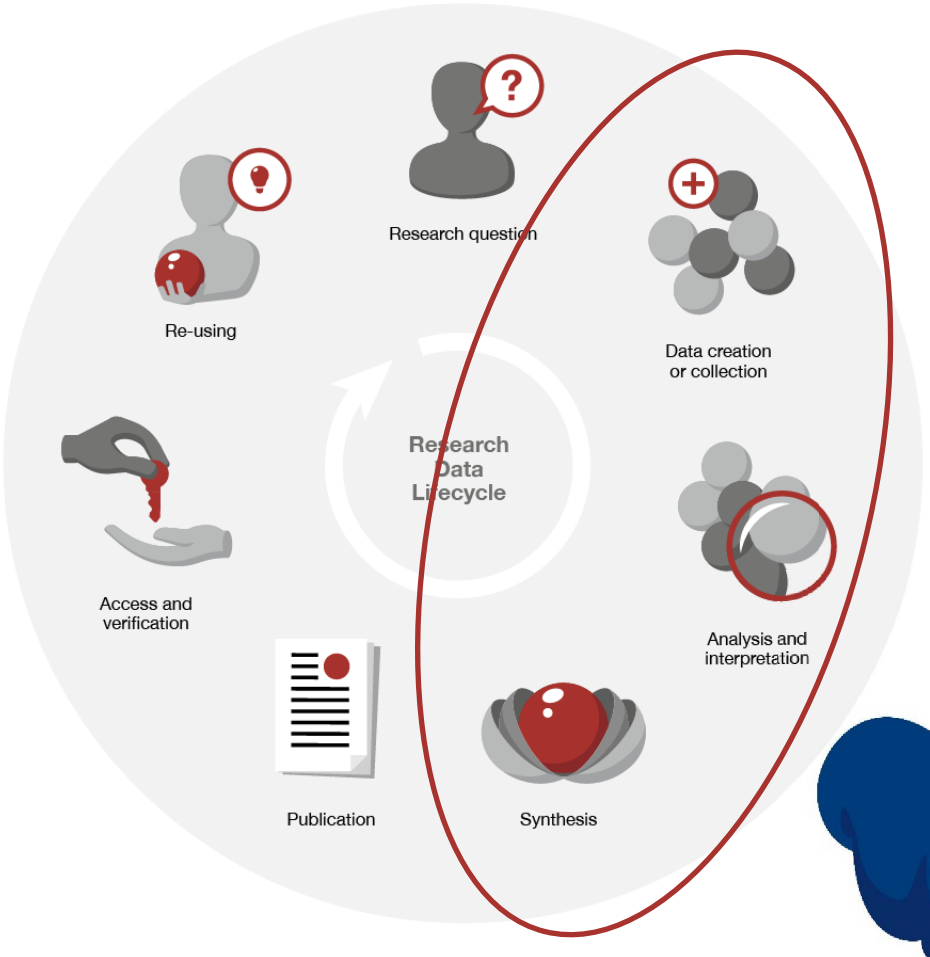
Submission and Assessment of the DMP

- A proposal can only be submitted if a DMP was created
- A DMP for SNSF **must be created online in mySNF**
- You cannot upload a DMP created outside of mySNF
  – except in Lead Agency process

https://www.mysnf.ch

- The DMP is assessed by SNSF staff **for plausibility and compliance** with its Open Research Data policy
- It is **not sent to external reviewers**
- Applicants **can be assigned «tasks»** for enhancing their DMP as part of the funding decision

Research Group Policy



- **Self-critical questions:**
  - What must data look like to **enable us to re-use** it with scientific conviction and **trust** into its quality and correctness?
  - Is this true for our own data? What is missing?
- **Tasks for group leaders (not required but recommended):**
  - Agree on **binding rules**
  - Define **data management responsible (DMR)** within the group
  - **Discuss** and **document rules** (in writing) with DMR

Data-aware workflow?

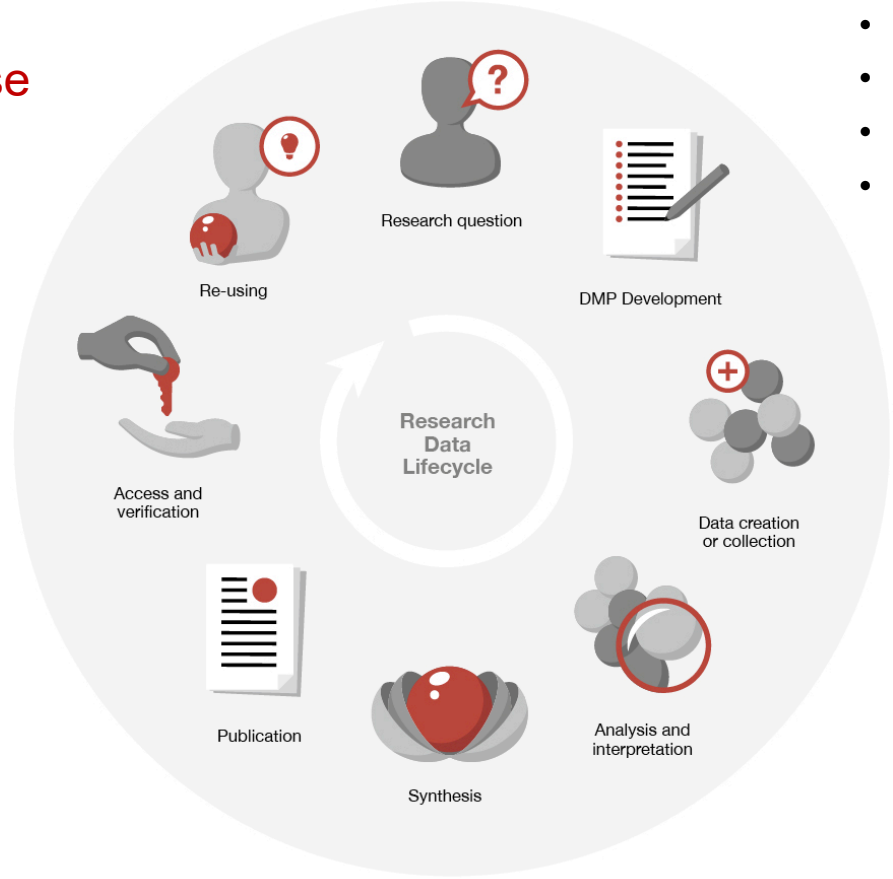
From initial idea, pre-evaluation, development to publishing

Long-term preservation and reuse

- e.g. ETH Data Archive

Open access publication and publication of research data in a FAIR repository

- Research Collection
- Zenodo
- etc.
- Use CC-Licenses



Writing a DMP

- Data collection and documentation
- Ethics, legal and security issues
- Data storage and preservation
- Data sharing and reuse

Active (daily) data management

- ELN-LIMS (e.g. openBIS, ETH RDH)
- Versioning (e.g. git)
- File formats (already consider LTP?)
- Consistent file/folder naming schemes
- Annotation and Metadata
- Backup
- Keep in mind: what would others need to resuse your data?

Active (daily) data management

- **Git-lfs**
  - Extension for storing larger files in an efficient way in a repository
  - Versioning of code and smaller to medium size amounts of data (max. 10s – 100s of GB)
  - gitlab.ethz.ch supports git-lfs
  - Management of code and data + versioning + backup
  - e.g. SDSC uses git-lfs as data store in Renku

- **External Data Storage**
  - we recommend the ETH NAS offers
  - https://www.ethz.ch/services/de/it-services/katalog.html, «Storage»)
  - The use of external harddrives is not recommended

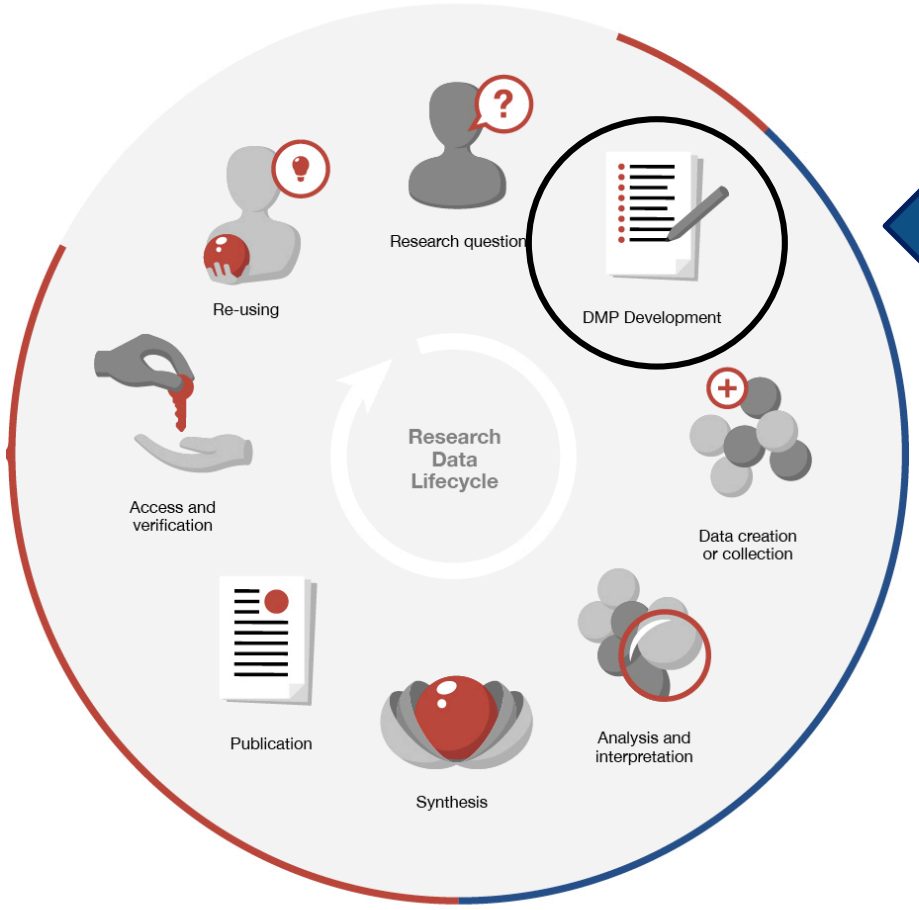
- **Cloud:**
  - Products for private end users (Dropbox, Google One etc.) → see cloud computing regulations at ETH
  - Professional (business-to-business) offers (e.g. AWS, Azure) → still under debate

The data life cycle

Publication and preservation: annotate, share, publish, preserve data at the end of the project/publication

Research Support Services

- Publication services
- Preservation services
- Training and consulting



Active data management: annotate, store, backup data while it is produced

Scientific IT Services

- Active research data management services
- Soft- and hardware
- Training and consulting

Why spend time and effort on this?

Benefits

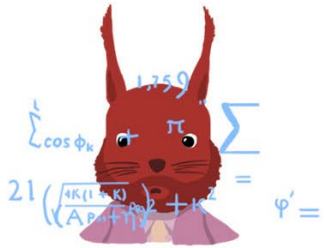
Preserve non replicable data

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

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

Questions?

RDM and Digital Curation @ ETH Library

www.library.ethz.ch/RDM
data-management@library.ethz.ch

Dr. Malin Ziehmer
044 633 86 32
malin.ziehmer@library.ethz.ch

RDM @ ETH Zurich

www.ethz.ch/researchdata
researchdata@ethz.ch

Jointly curated by the
ETH Library and
ID Scientific IT Services