ETH zürich

ETH Library

Research Data Management Workshop at RCS18

Educational Material

Author(s):
Ziehmer, Malin Michelle

Publication date:
2018-09-09

Permanent link:
https://doi.org/10.3929/ethz-b-000296565

Rights / license:
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

ETH zürich

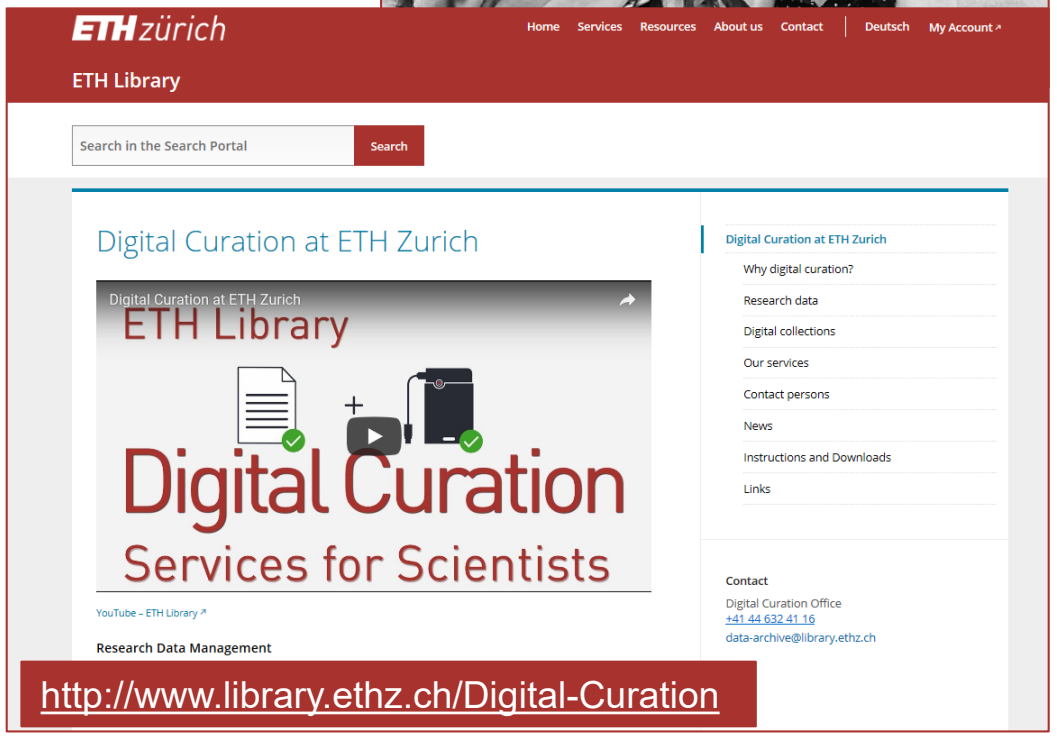
Research Data Management Workshop at RCS18

CC BY NC SA

Malin Ziehmer

Nice to meet you, we're..

- From the
  - **Research Data Management and Digital Curation Office** 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 re-use**

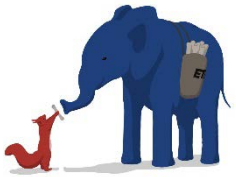


Let’s get to know you a bit better…

«Cross the line»

What we are going to do…

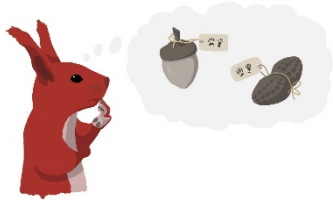
What is data management and why should it concern you?



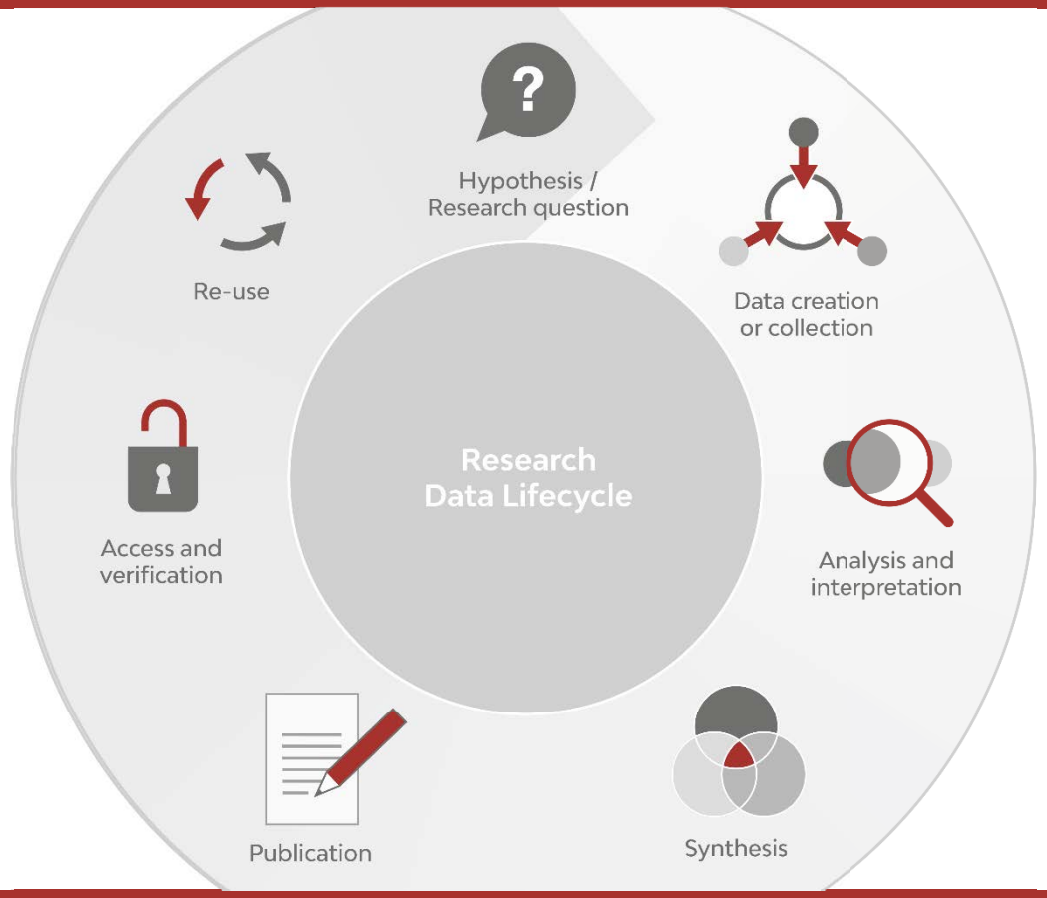
Long-term preservation

Regulations, intellectual property, privacy and access rights

Data sharing



Data Management Planning



What is data management and why should it concern you?

An introduction

What is data?

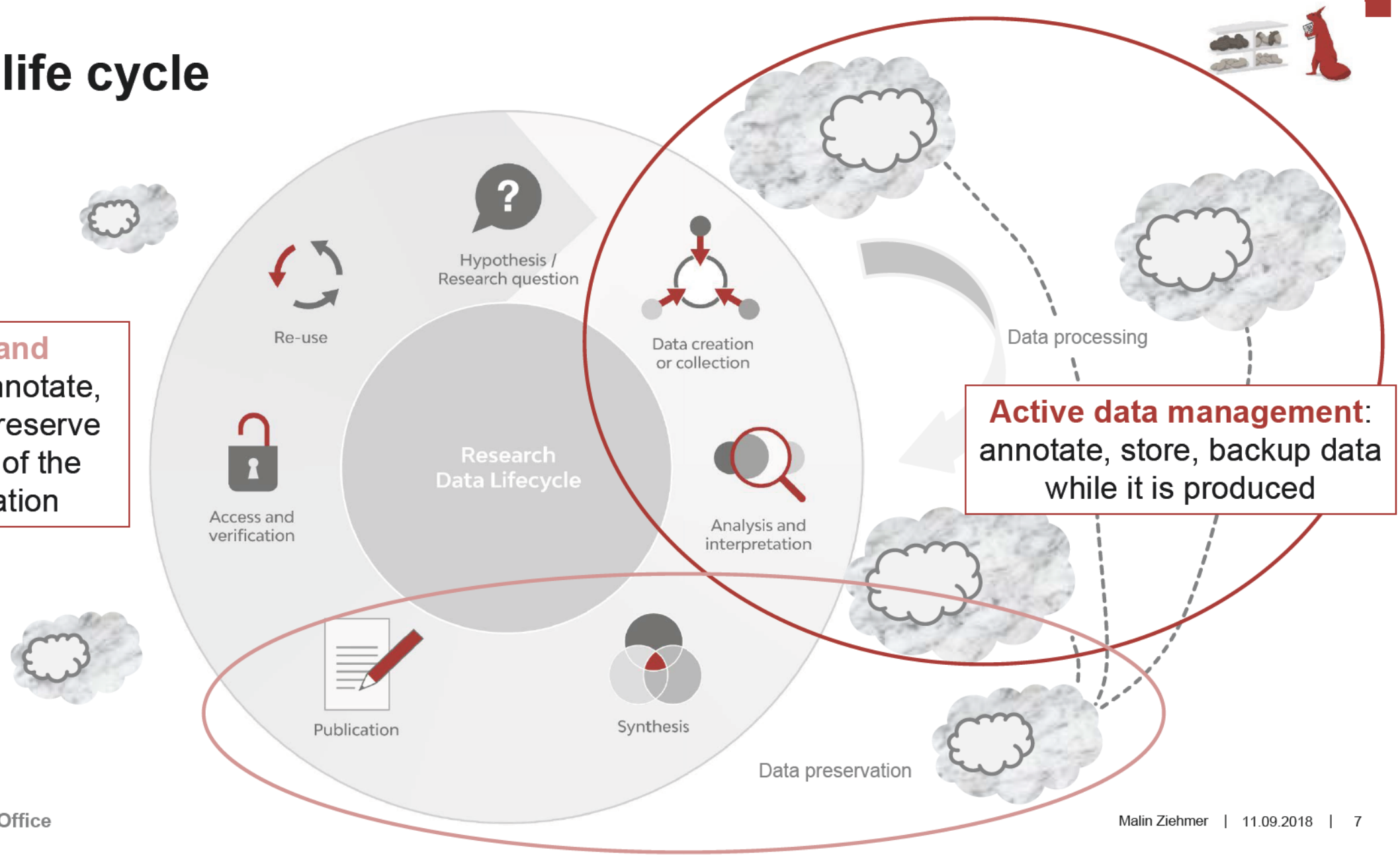
"A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing."

Digital Curation Centre

Digital Curation Centre
Slide adapted from the PrePARe Project – CC BY-SA

The data life cycle

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



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

Why spend time and effort on this?

Your benefit

- **Preserve data that cannot be replicated** (e.g. observational data)
- **Avoid redundant data creation/collection**
- **Highlight patterns or connections** that might otherwise be missed
- Enable data **re-use and sharing** – even for yourself
- **Facilitate collaboration**
- **Raise your impact**: your data can be cited

Your duty

- **Meet** funders' and institutional **requirements**
  - SNSF asks for data management plans as of October 2017
  - EU Horizon 2020 asking for data management plans
- Keep work in accordance to **good scientific practice**, transparency and validity
- You may be able to **influence the discussion** in your community, in your institution and **with funders**

Regulations, intellectual property, privacy and access rights

An overview

What you (should have) received at the beginning of your PhD…

Compliance Guide

- […] 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**.
- From:
  https://rechtssammlung.sp.ethz.ch/Dokumente/133_en.pdf

  https://direction.epfl.ch/files/content/sites/polylex/files/recueil_pdf/ENG/ComplianceGuideEPFL_EN.pdf

Guidelines for Research Integrity

- 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.

- https://doi.org/10.3929/ethz-b-000179298
- https://research-office.epfl.ch/research-ethics-integrity/research-integrity

Roles and responsibilities

- 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.

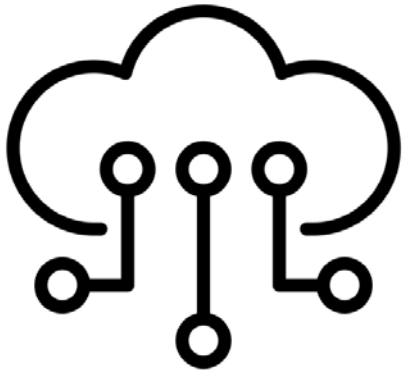
- Project **Members**:
  - **adhere to the principles of good scientific practice** and the guidelines for Research Integrity at ETH.
  - All steps of treatment of primary data must be **documented in a form appropriate to the discipline** and results must be **reproducible**.

See Guidelines for Research Integrity: https://doi.org/10.3929/ethz-b-000179298

Do you know where your data is and who has access to it?

"There is no cloud just other people's computers" (4.9.2018)
by Markus Meier
CC BY-SA 4.0

Cloud computing @ ETH Zurich
Rules and Regulations



«Cloud Computing» (4.9.2018)
by Symbolon / CC BY

- 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

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 (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 (www.transfer.ethz.ch)

Privacy

- 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)

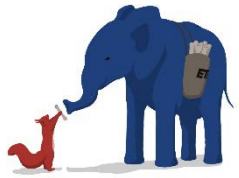
  Federal Act on Data Protection (https://www.admin.ch/opc/en/classified-compilation/19920153/index.html)

  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-gruppenkommissionen/ethikkommission.html

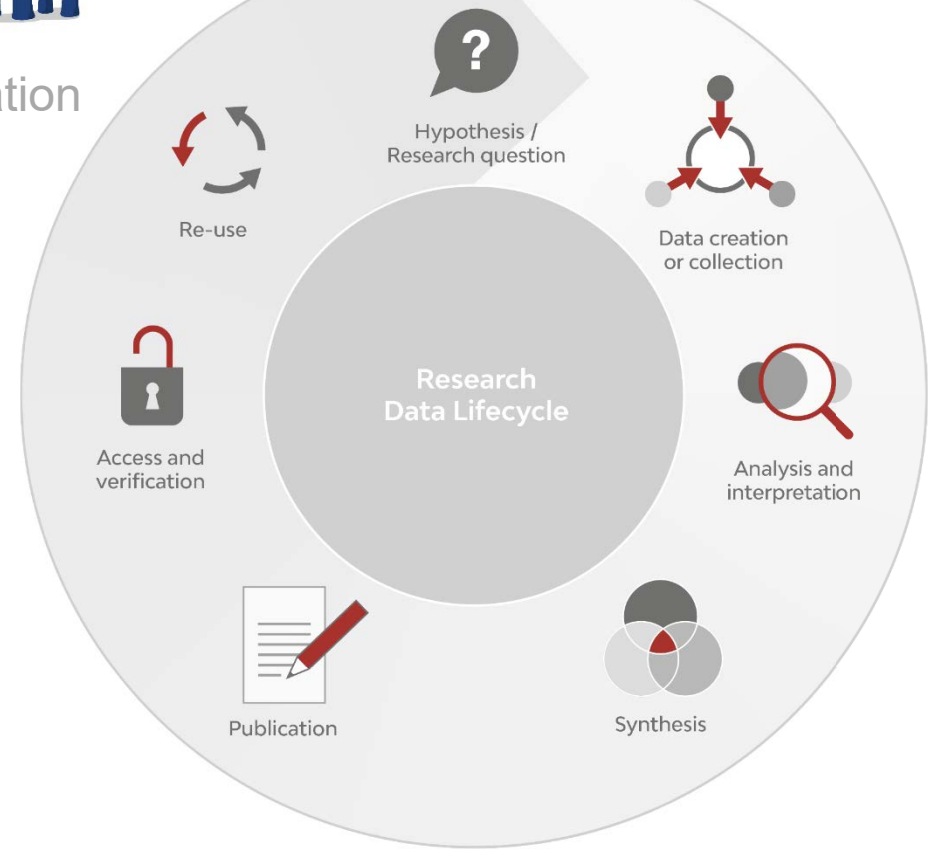
«Incognito» (4.9.2018) by Hea Poh Lin / CC BY

What's next?



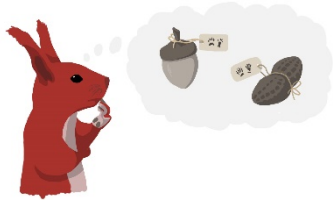
Long-term preservation

What is data management and why should it concern you?



Regulations, intellectual property, privacy and access rights

Data sharing



Data Management Planning

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?

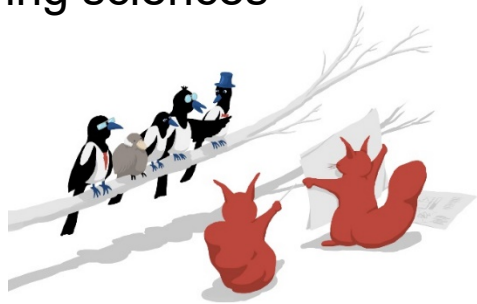
DMPs are e.g. demanded by:

SNSF from October 2017 on
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

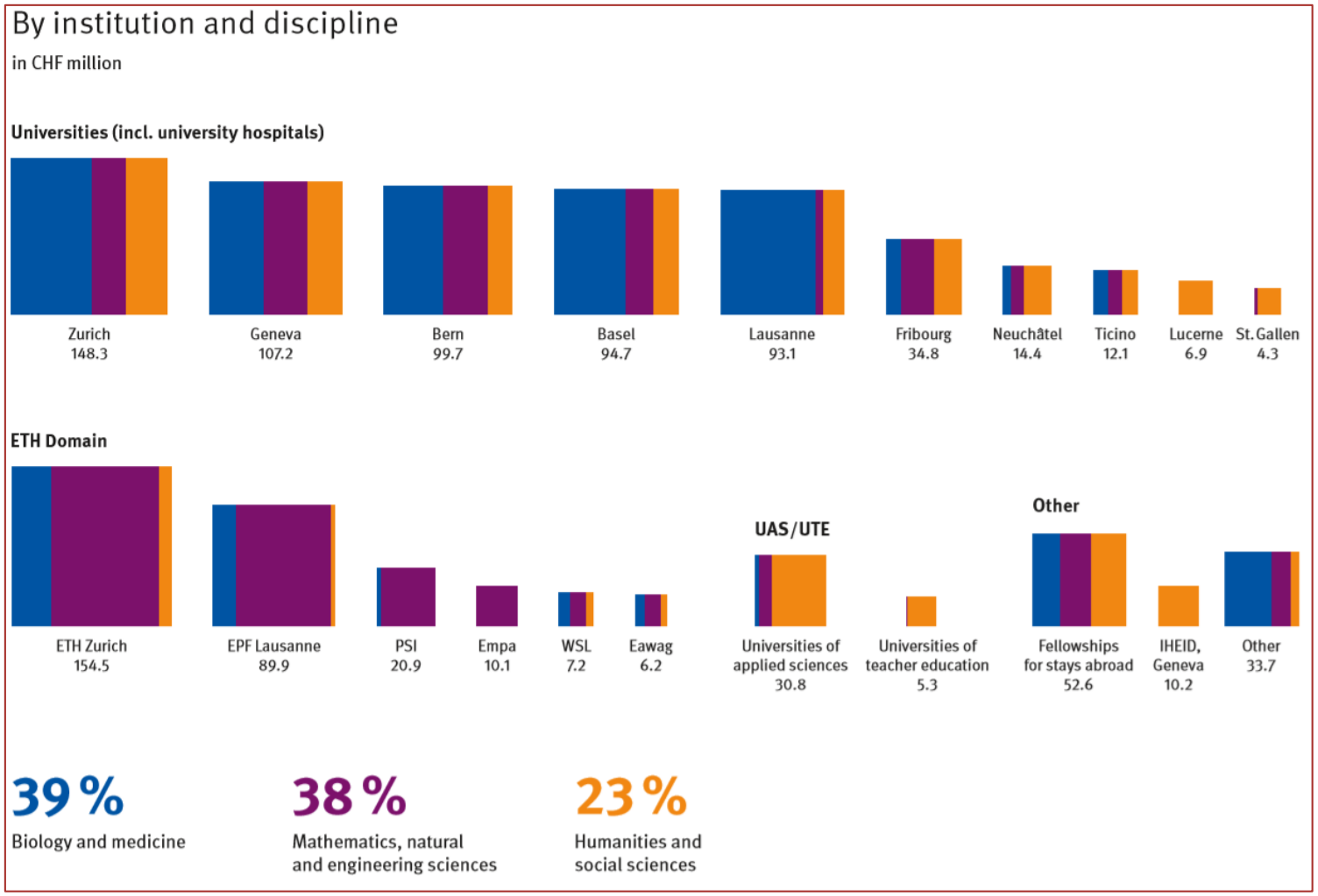
Who is the SNSF?

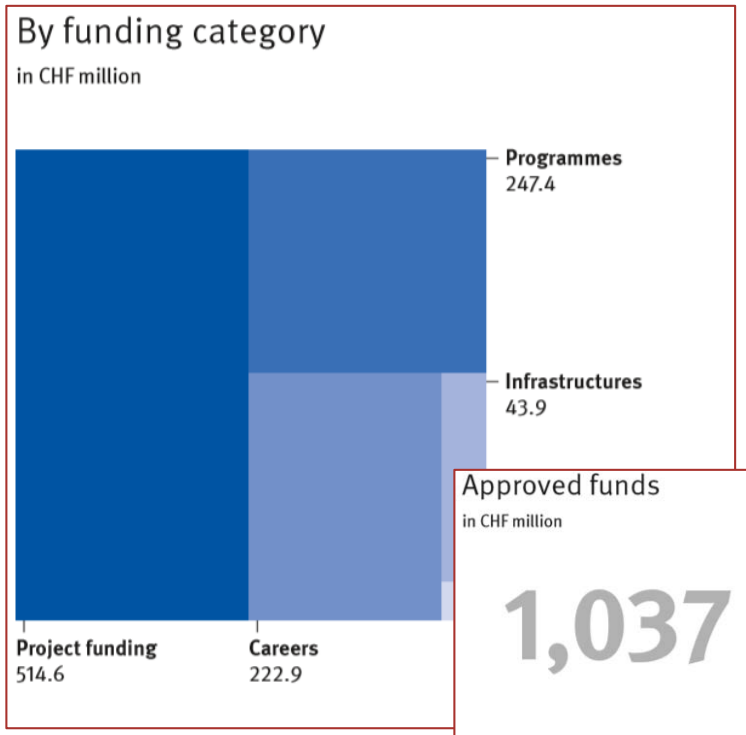
- Based on a government mandate, the **Swiss National Science Foundation (SNSF)** supports scientific research
- The SNSF **supports all academic disciplines** from history to medicine and the engineering sciences
- At the end of 2017, the SNSF was funding **5800 projects involving 16,000 researchers**
- To ensure independence, the SNSF was established as a **private institution in 1952**
- **Core task: evaluation of research proposals**
- In 2017, the SNSF awarded **CH 1037 million** to the most promising project proposals
- By allocating public research money based on **the principle of competition**, the SNSF contributes to the high quality of research in Switzerland
- **Particular attention to the promotion of young researchers**

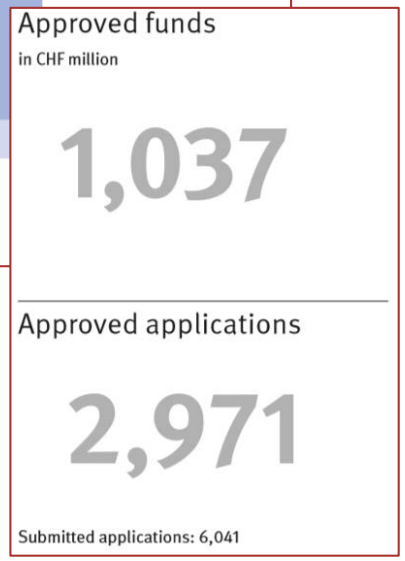


http://www.snf.ch/en/

SNSF Statistics 2017







© SNSF: Statistics 2017 http://www.snf.ch/en/theSNSF/profile/facts_figures/statistics (5.10.2018)

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.

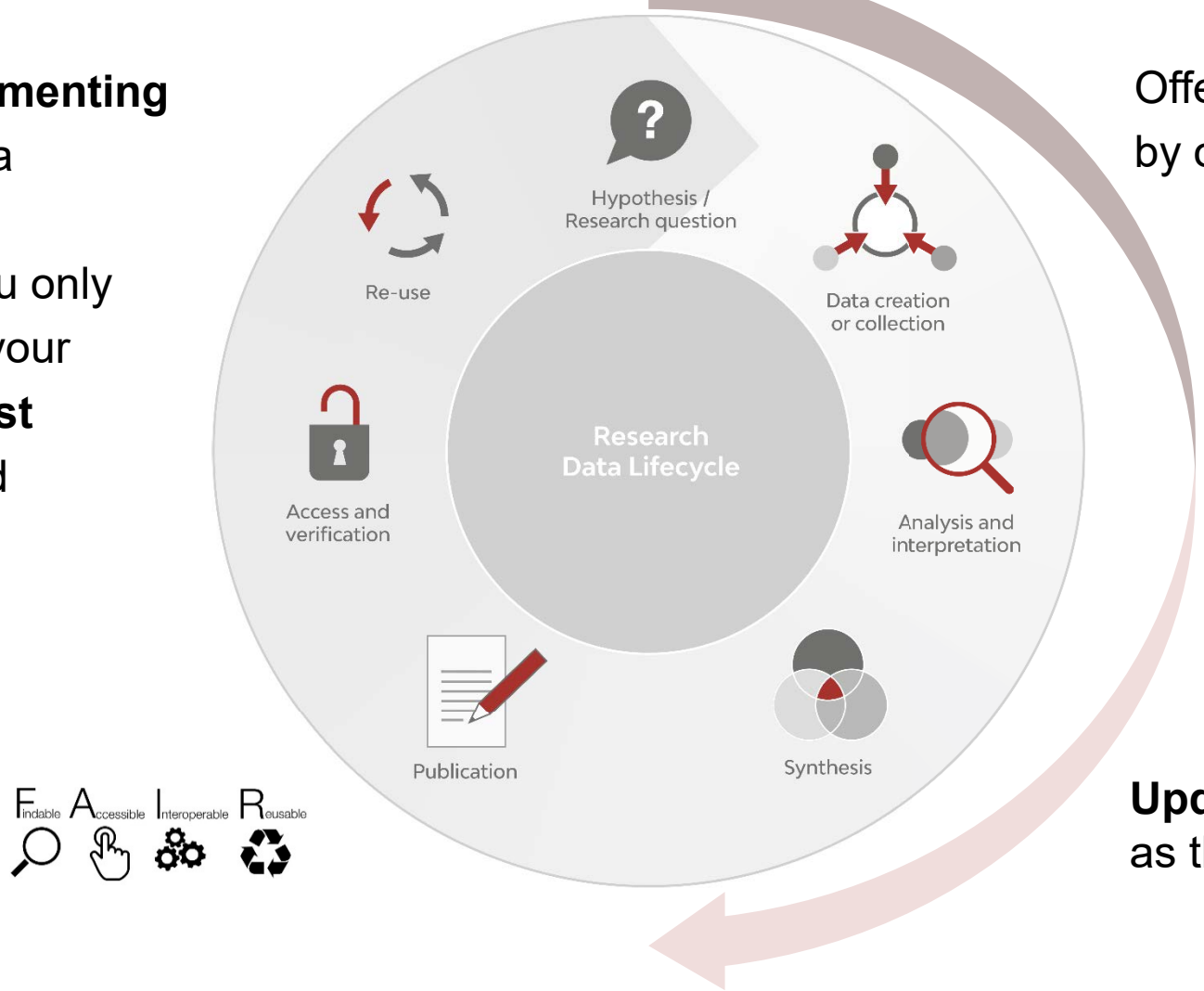
Article 47 of the Funding Regulations
(1 Jan 2016, http://www.snf.ch/SiteCollectionDocuments/allg_reglement_16_e.pdf):
"[…] the data collected with the aid of an SNSF grant must also be made available to other researchers for further research and integrated into recognised scientific data pools […]"

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

Please also be aware of SNSF's updated Open Access Policy for Publications and changes to the General implementation regulations for the Funding Regulations!
http://www.snf.ch/en/theSNSF/research-policies/open-access/

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

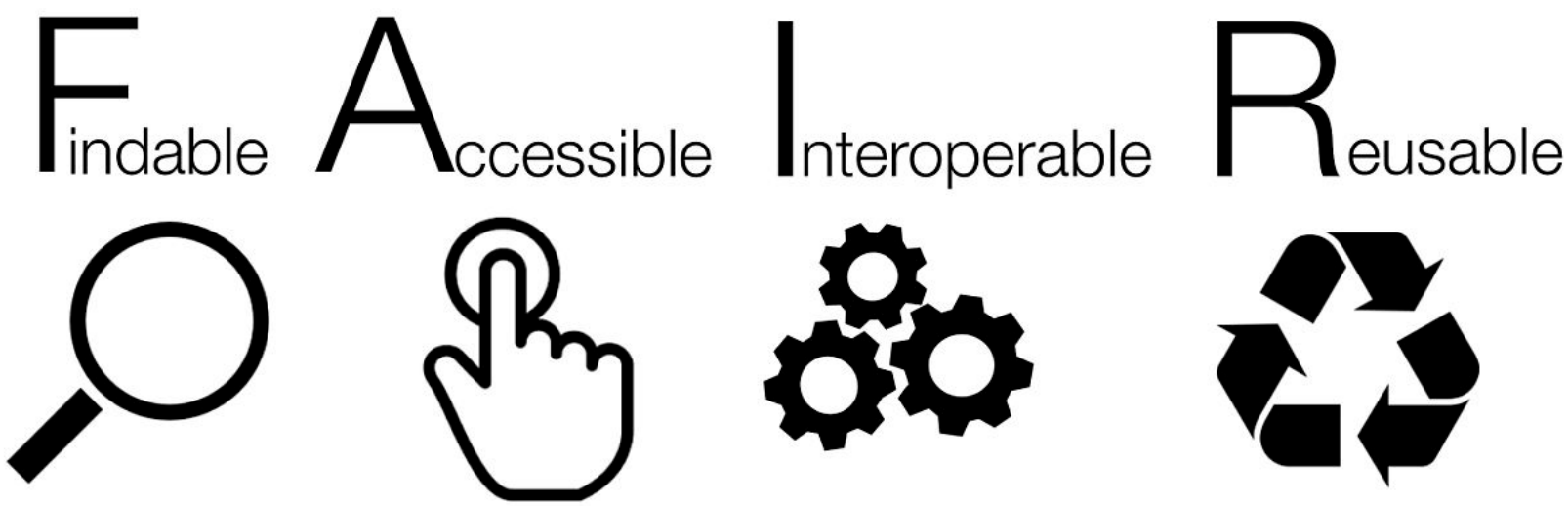


Offering a **long-term perspective** by outlining how the data will be:

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

Updating the plan
as the project progresses

Making research data FAIR



FAIR image (4.9.2018) by Sangya Pundir / CC BY-SA 4.0

The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data, Issue 3, 2016. 10.1038/sdata.2016.18.*

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

What to do for other funders?

- **Data Management Checklist by ETH and EPFL**
  - Supports you in the creation of a DMP or in discussing data management in general, even if you don't need to do it to comply with funders
  - https://documentation.library.ethz.ch/display/DD/Data+Management+Checklist

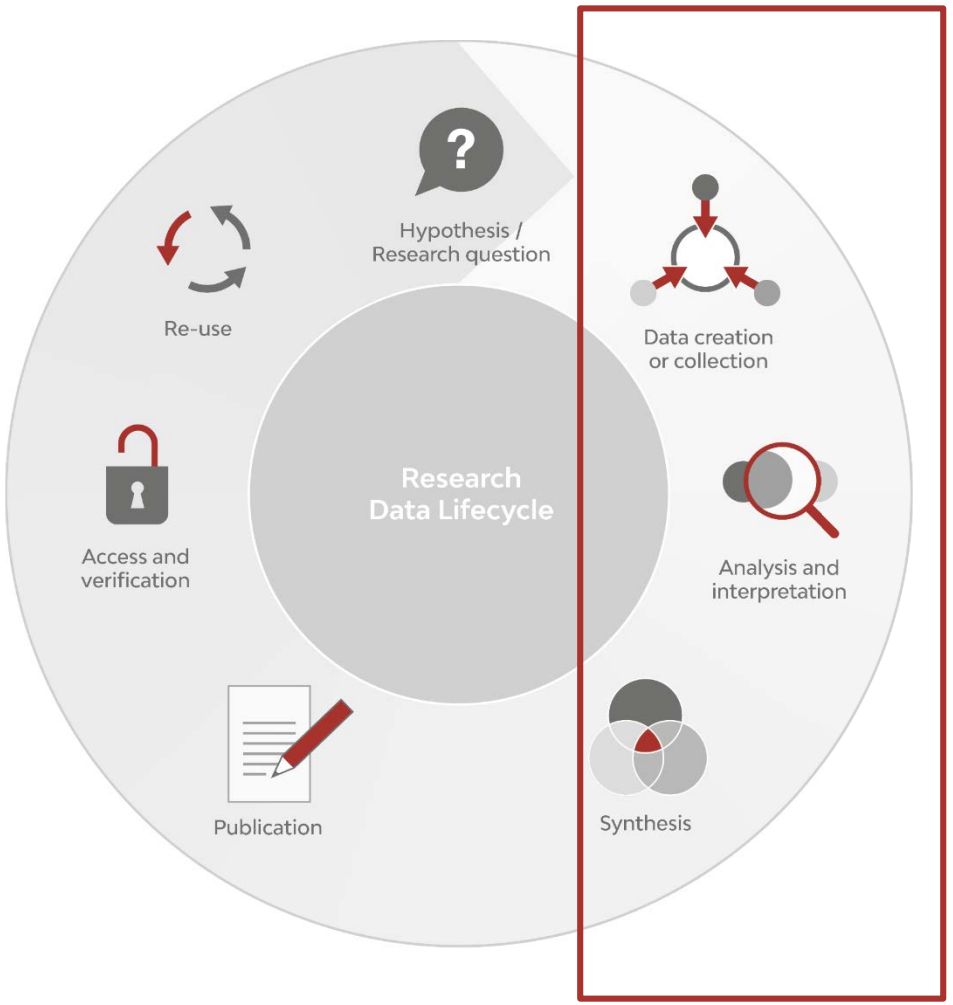
- **Collection of DMP examples**
  - http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples

What we offer:
- (individual) **training** on DMP
- **consultation** on data management planning
- **review** of DMPs

- **H2020 Information by EU GrantsAccess**
  - http://grantsaccess.ethz.ch/en/servicesupport/uzh-eth-zurich-support/open-access-publications-data/

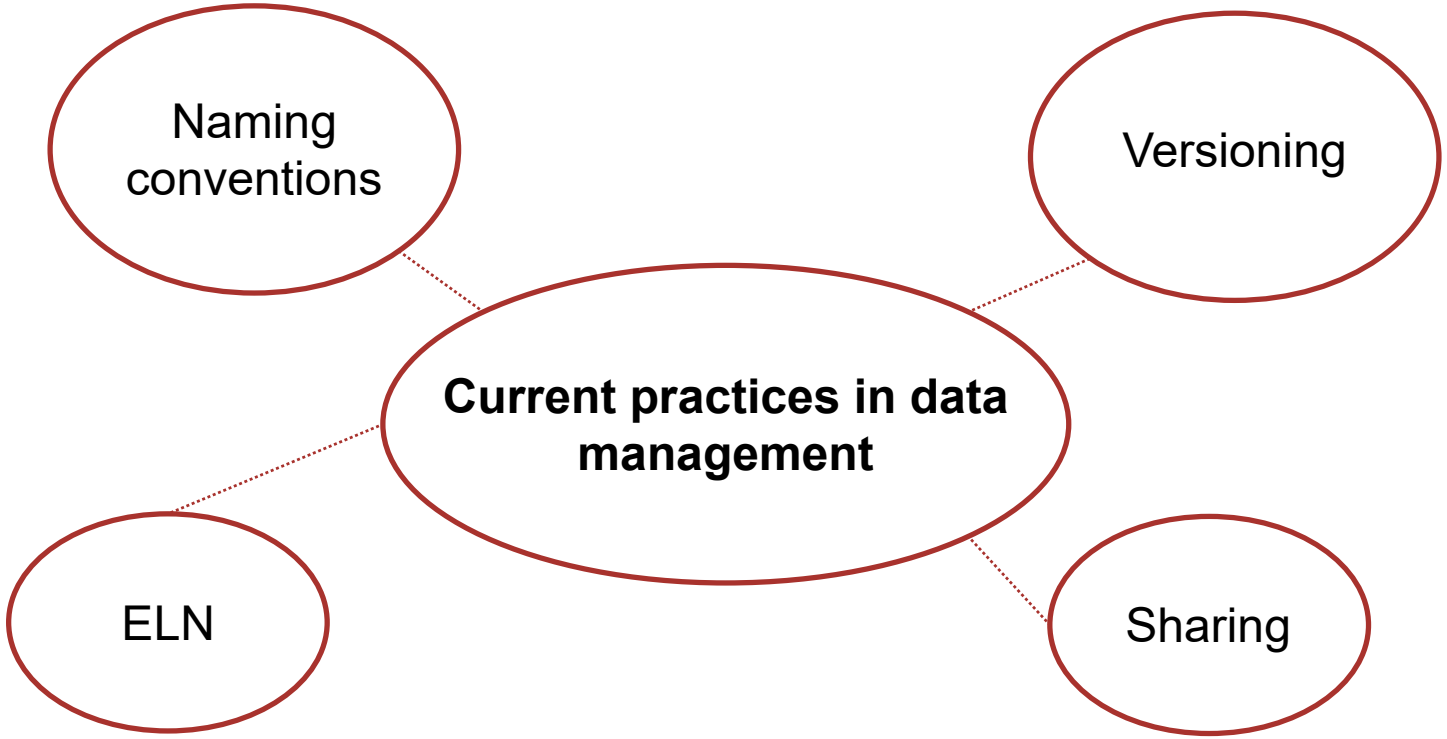
- **DMPOnline**
  - A tool by the UK Digital Curation Centre that helps you create Horizon 2020 compliant data management plans, by answering a questionnaire
  - https://dmponline.dcc.ac.uk

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:**
  - Agree on **binding rules**
  - Define **data management responsible (DMR)** within the group
  - **Discuss** and **document rules** (in writing) with DMR



Current practices in data management – Active Research Data Management (ARDM)

What are your best practices? **Post it!**

Current practices in data management

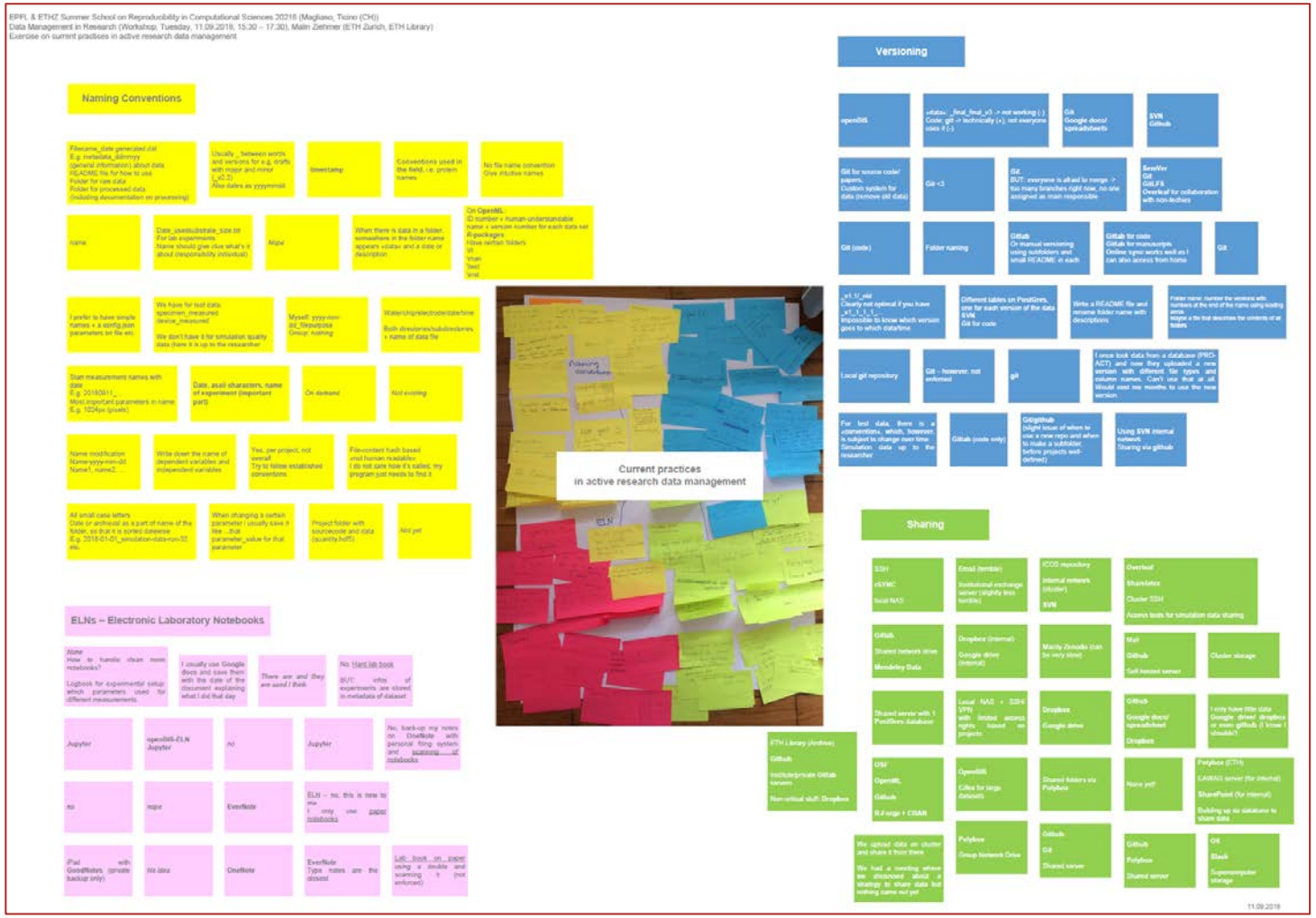
- **Naming conventions**:
  Do you have any? Which rules apply?

- **Versioning**:
  How do you currently handle it? What works well? What went wrong?

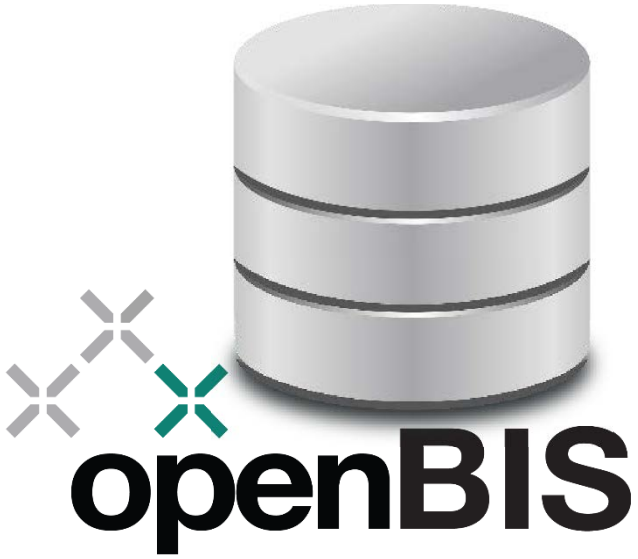
- **Electronic Laboratory Notebooks (ELN)**:
  Do you have experience with any?

- **Sharing**:
  Which tools or services do you use? What are your experiences?

Current practices in data management



see attached PDF for details



The ETH Scientific IT Services data management solution

for active research data management

An example for ARDM

What is openBIS?

Laboratory Information Management System

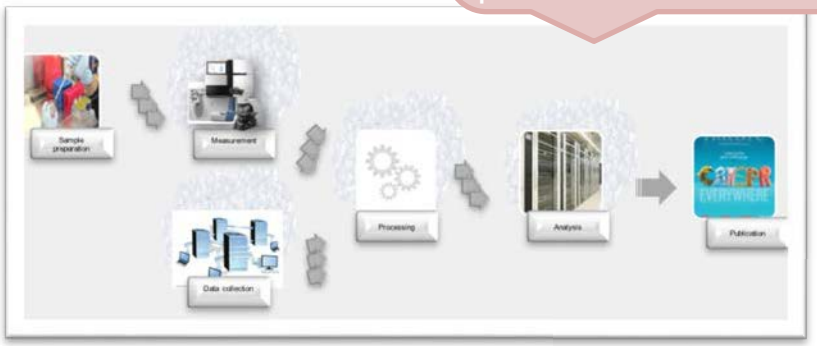
openBIS ELN-LIMS

Electronic Laboratory Notebook

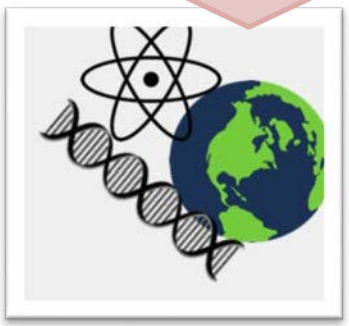
Biology Information System

openBIS facts

Platform for managing scientific information and supporting research data workflows from “bench” to publication



Can be used in most quantitative science fields (e.g. life sciences, physics, env. sciences, etc)



Used by research groups and facilities @ ETH, Swiss & European Universities, a few companies

Summer 2007
- openBIS development start (SystemsX)

April 2008
- first openBIS release (v08.04)

Summer 2009
- SystemsX projects start using openBIS

Summer 2013
- openBIS ELN-LIMS UI start

Spring 2014
- first ELN-LIMS beta version

May 2015
- first downloadable ELN-LIMS plugin

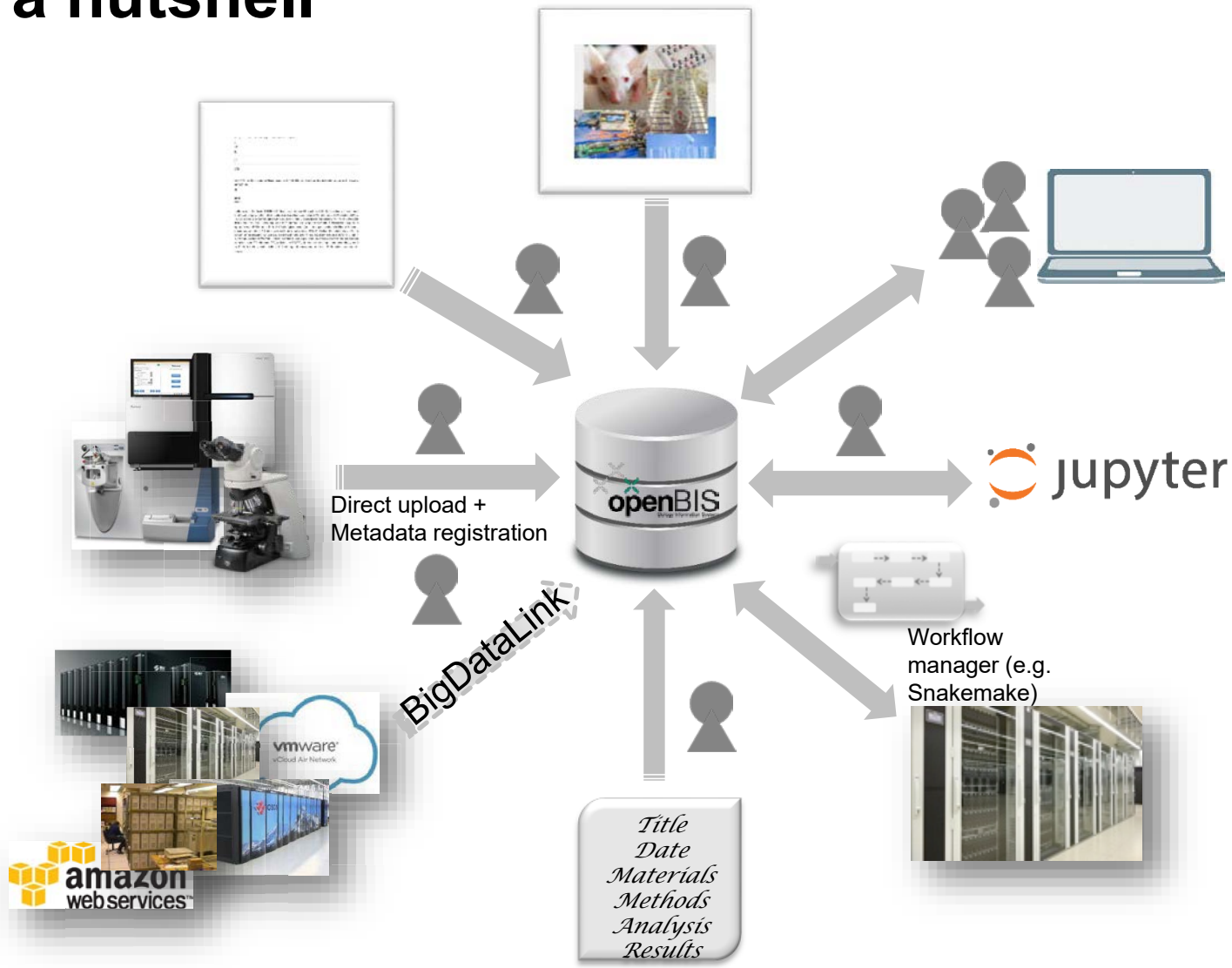
May 2016
- first ELN-LIMS official release

May 2017
- BigDataLink v.1

December 2017
- JupyterHub integration

Slides provided by Caterina Barillari (ETH ID-SIS)

openBIS in a nutshell



Slides provided by Caterina Barillari (ETH ID-SIS)

openBIS as a service from ID-SIS at ETH

- From **2018**, SIS has the **mandate to provide active data management services to all ETH as well as to all scientists within the Swiss research community (openRDM.swiss)**

- **Basic service for research groups**:
  - Provide openBIS to research groups (central instance, private instances)
  - Initial training
  - Continuous support
  - Prefilled DMP template for openBIS users

- **Additional services (on demand)**
  - Database customization
  - Migration of existing databases
  - Instrument integration for direct data upload
  - Upload of existing historic raw data

Scientific IT Services

https://sis.id.ethz.ch/
sis.helpdesk@ethz.ch

Slides provided by Caterina Barillari (ETH ID-SIS)

NEWS: ETH Research Data Hub (ETH-RDH)

- Available from **September 2018**
- versatile **data management solution** specially designed for ETH research groups working in **quantitative scientific disciplines**
- **based on** the powerful and well-established **openBIS platform**, developed by ETH Scientific IT Services (https://sis.id.ethz.ch)

- ETH-RDH allows scientists to **document** and **annotate their research data** from initial acquisition onwards
- **create inventories** of materials and protocols
- **facilitating knowledge sharing** and **transfer** within the research group
- ETH-RDH also enables **sharing with collaborators**
- The first **100GB of storage** in ETH-RDH **is free** of charge for each ETH research group (additional storage beyond that will be priced competitively)

https://blogs.ethz.ch/its/

Excursion: File sharing tools

recommended

polybox.ethz.ch

cifex.ethz.ch/

www.switch.ch/drive/
www.switch.ch/filesender

→ **Data stored in Switzerland**

→ **Security regulations fulfilled**

only conditionally recommended

www.dropbox.com

www.wetransfer.com

→ **Data stored in EU/USA**

→ **Security regulations only partially fulfilled**

→ **Never store sensitive / private data there!**

Excursion: File sharing tools

A closer look at ETH and SWITCH Tools

ETH zürich

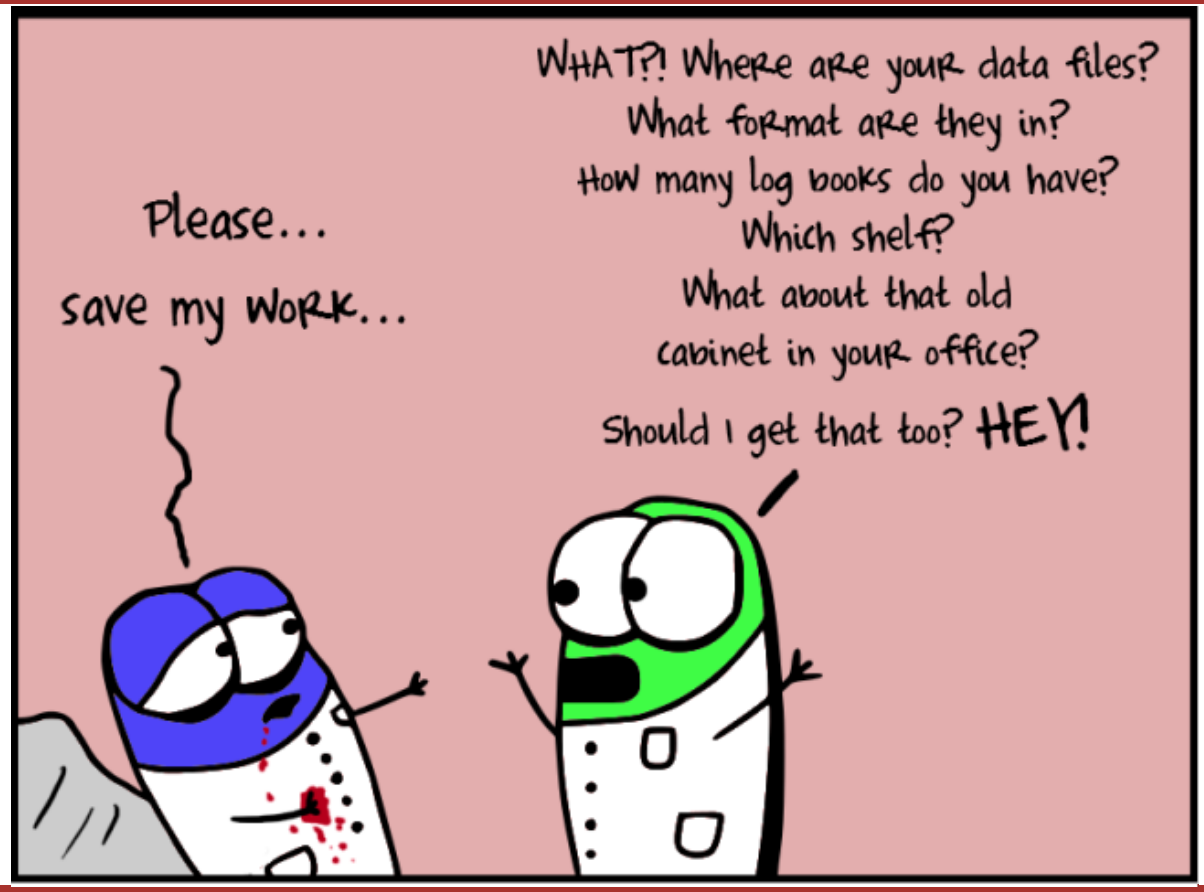
polybox.ethz.ch

https://cifex.ethz.ch/

SWITCH

www.switch.ch/driv

www.switch.ch/filesender



"Real vs movie scientist 3" (detail, 4.9.2018)
by Nik Papageorgiou
CC BY-NC-ND

What it takes to understand someone's data

Mindmapping

Short excursion: How to do a mind map?

Step 1

Create a central idea

Step 2
Add branches to your map

Key themes

Key theme

Step 3
Add key words

Key words

Step 4
Color your branches

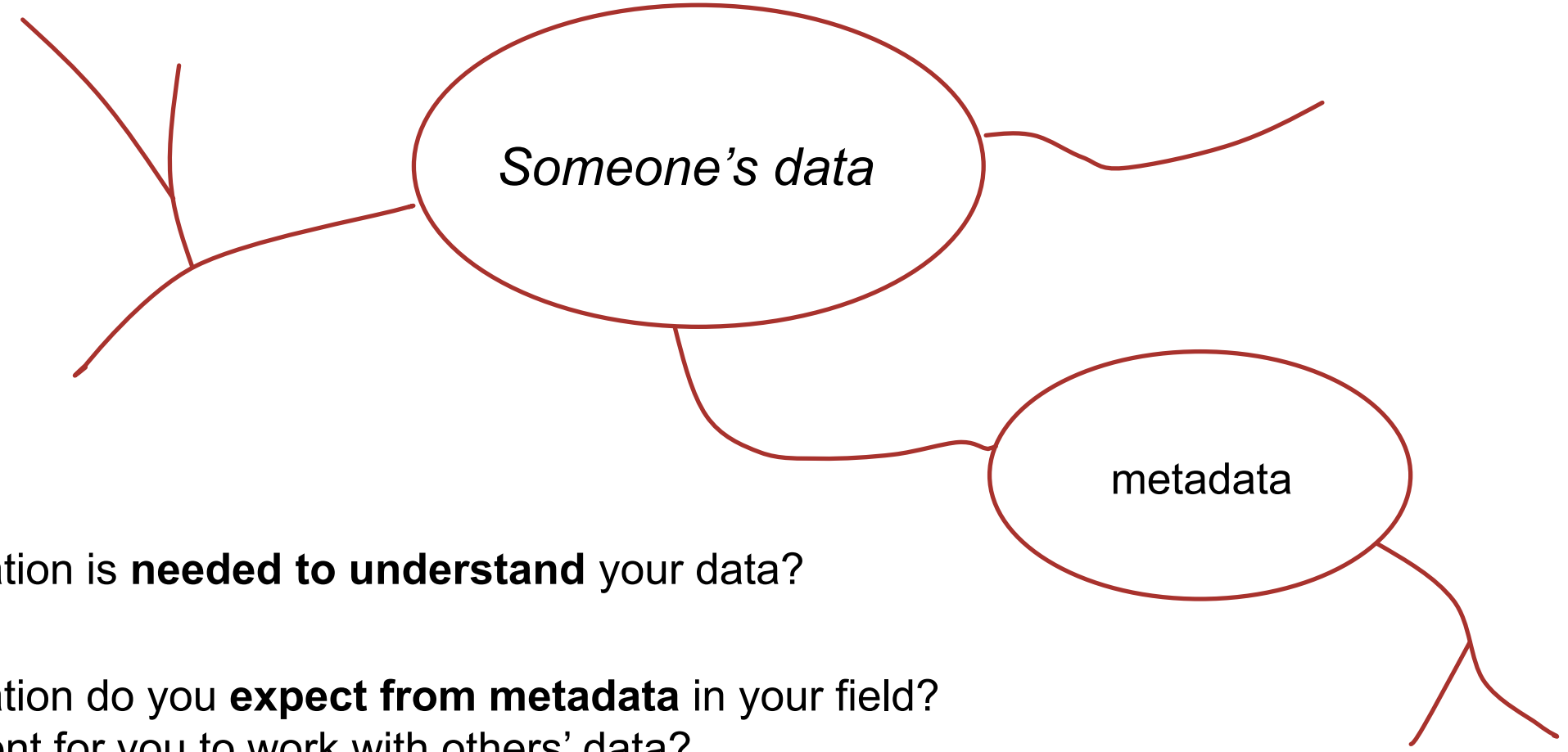
Step 5
Include images

"Dialog-information on" (4.9.2018) from the Tango! project
CC BY-SA 2.5

Short excursion: How to do a mind map?

iMindMap: "What is a Mind Map" (4.9.2018)
https://imindmap.com/how-to-mind-map/

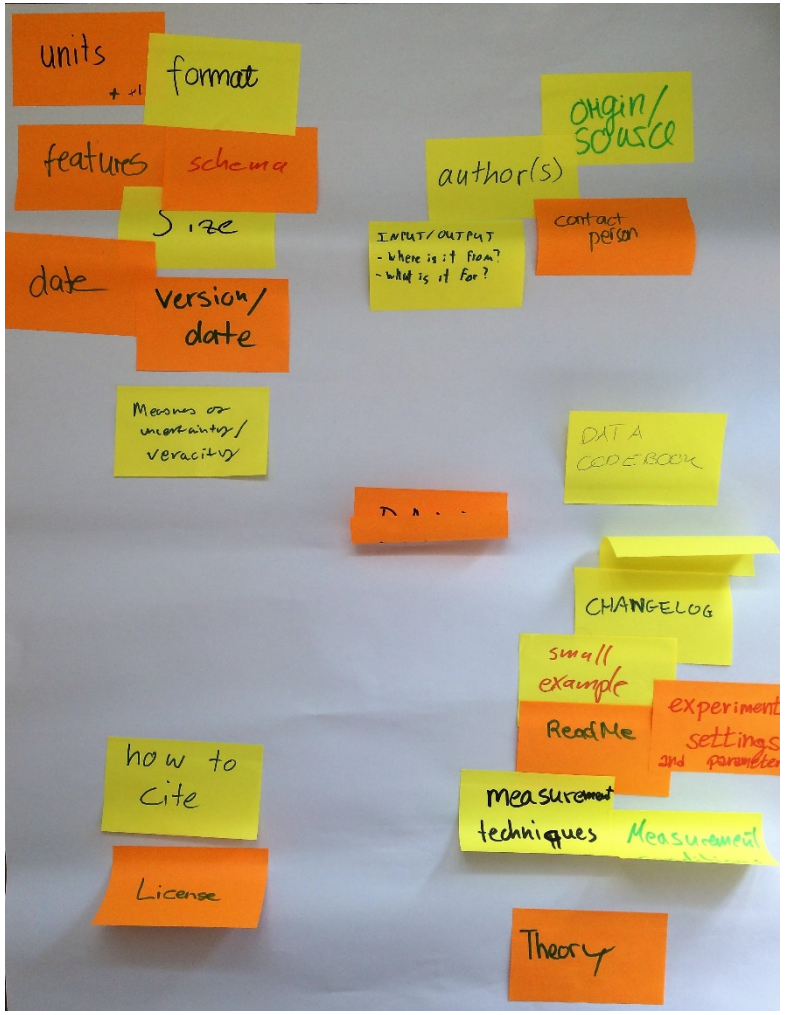
What it takes to understand someone's data – Mind map



1. What information is **needed to understand** your data?

2. What information do you **expect from metadata** in your field?
   Is this sufficient for you to work with others' data?

What it takes to understand someone's data – Mind map



measures of uncertainty/veracity | date

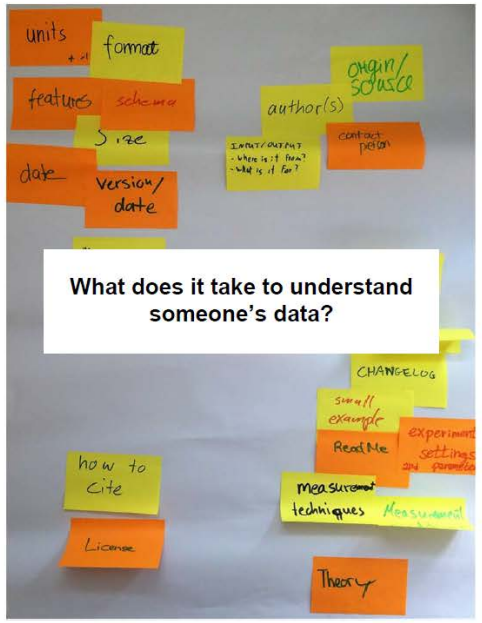
units | version + date

features | size

schema | format

How to cite?

license



input + output
- where is it from?
- what is it for?

contact person

README

origin/source

data codebook

author(s)

changelog

theory

measurement conditions

data acquisition protocol

small example

experiment settings and parameters

measurement techniques

… all you need is metadata!

see attached PDF for details

Critically re-thinking the (re-)use of data

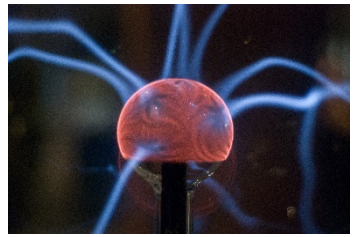
- **Data**, **metadata** and **context** are needed to **properly understand** a data set
- Data management does not start with your own data, but also includes a **critical view** on other people's data you use:
  - Do you understand **how** they were **produced**?
  - Do you have **enough information** on evaluating their **reliability**?
  - Are you **comfortable** with using data **without talking** to its producers?
  - Will you know in a few months time **which data** you **re-used** from other researchers?
  - Do you know **how to cite** the data you use?

Metadata options

- Develop your **own metadata scheme** or **use existing standards** (preferred)
- **Apply** metadata as **early as possible** in data life cycle (i.e. during acquisition)

Biology

Earth Science



Physical Science

Social Science & Humanities

General Research Data

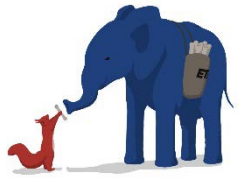
http://www.dcc.ac.uk/resources/metadata-standards

Practical Data Management

→ right after this session!

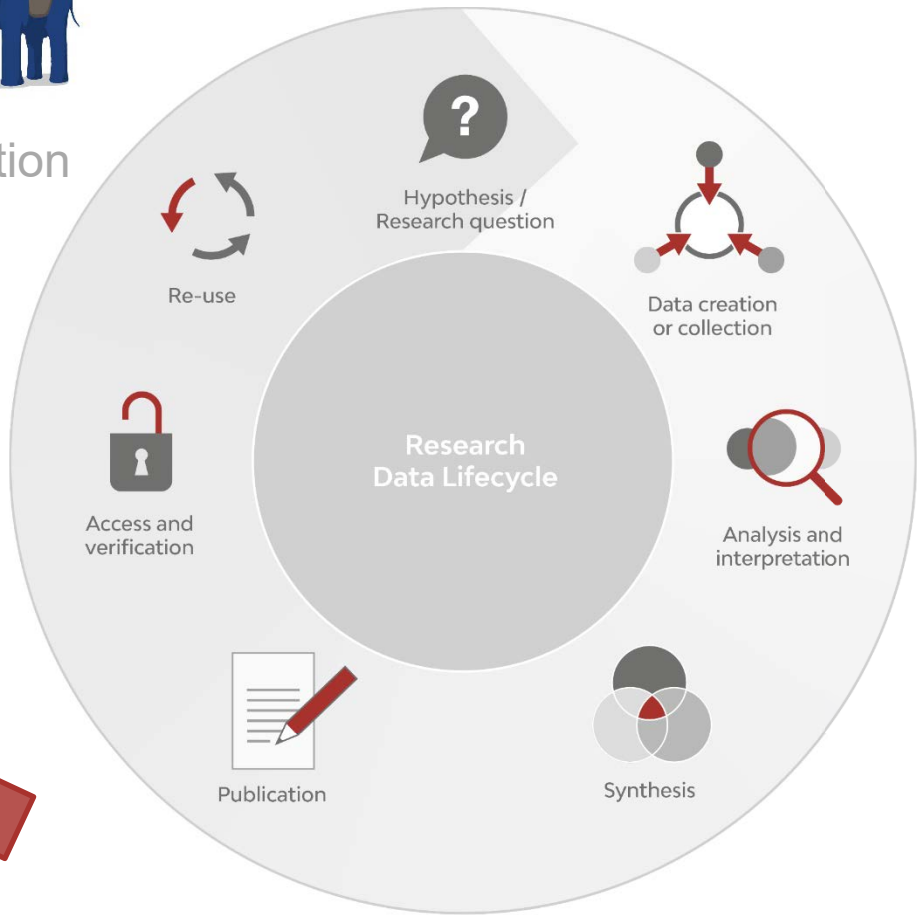
→ led by Anna Krystalli

What's next…



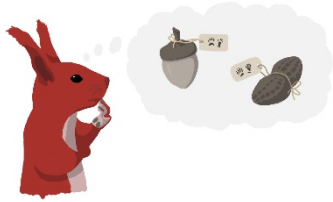
Long-term preservation

What is data management and why should it concern you?



Regulations, intellectual property, privacy and access rights

Data sharing



Data Management Planning

Data sharing

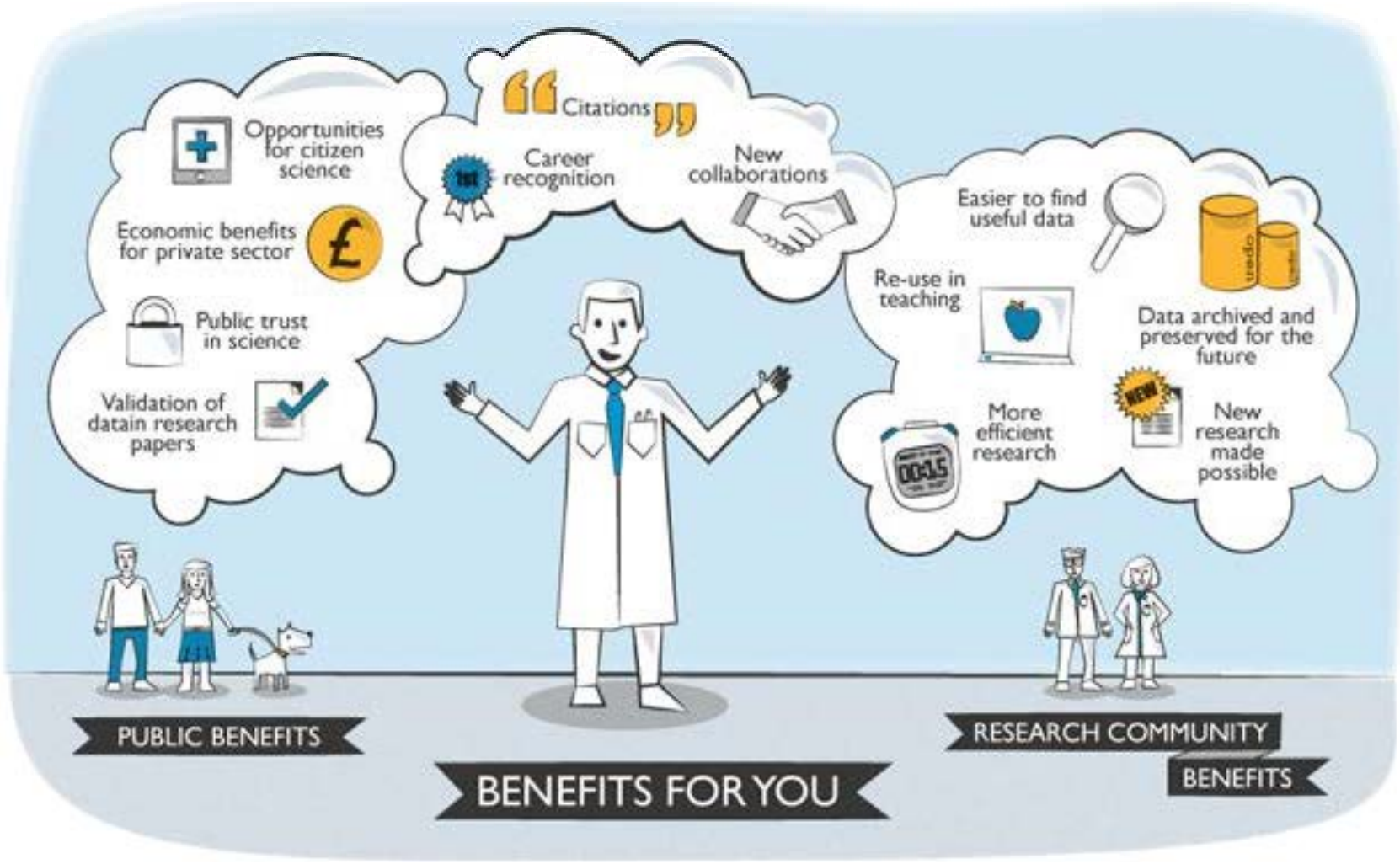
Data sharing/ collaboration with project partners (*during the project*)

Data Sharing

Creative Commons Licenses for third parties

Data sharing with/ **publishing** to the community (*after publication of results*)

Benefits of data sharing

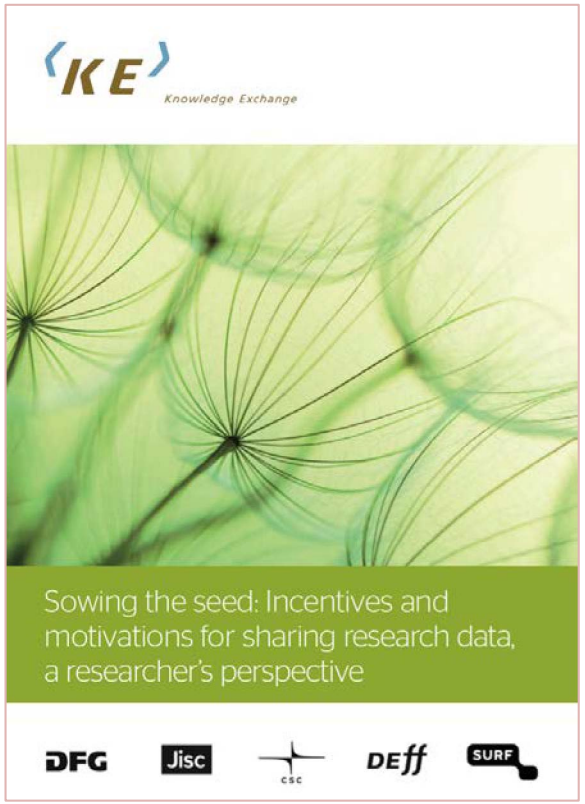


© Neil Chue Hong http://dx.doi.org/10.6084/m9.figshare.942289 (4.9.2018)

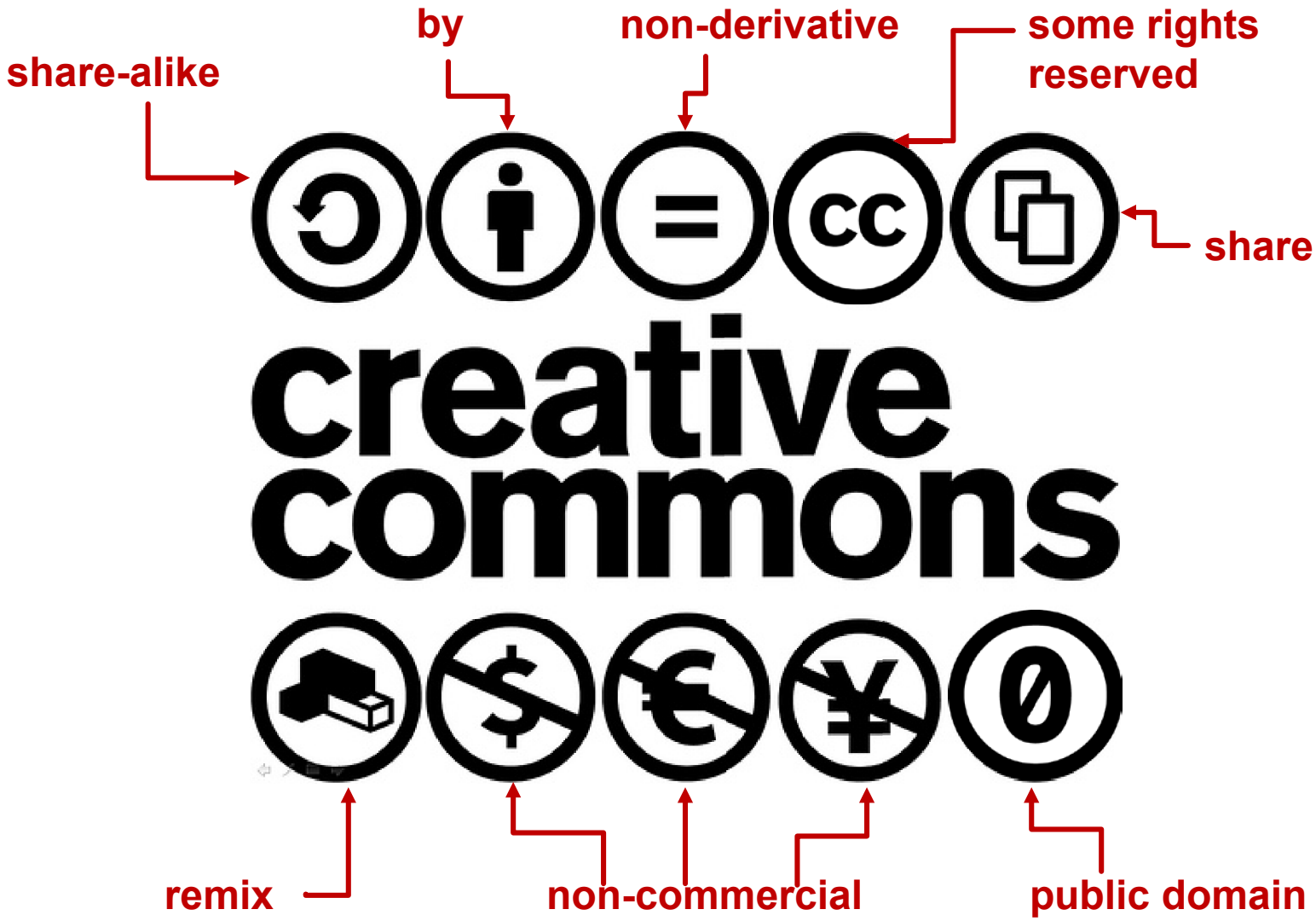
Benefits of Open Data: Impact and longevity

"In genomics research, a large-scale analysis of data sharing shows that studies that made data available in repositories received **9% more citations**, when controlling for other variables; and that whilst self-reuse citation declines steeply after two years, **reuse by third parties increases even after six years**."

(Piwowar and Vision, 2013)

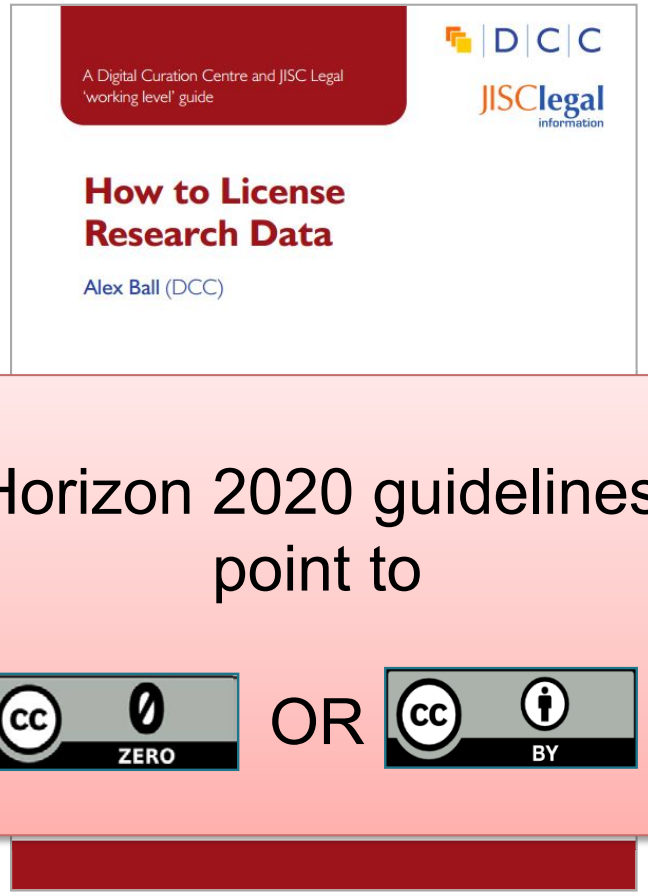


Van den Eynden, V. and Bishop, L. (2014). Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective. A Knowledge Exchange Report, http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf



"Creative Commons" (4.9.2018)
by Michael Porter
CC BY-NC-ND 2.0

Licensing research data



Horizon 2020 guidelines point to

CC ZERO OR CC BY

Outlines pros and cons of each approach and gives practical advice on how to implement your licence

CREATIVE COMMONS LIMITATIONS

NC Non-Commercial
What counts as commercial?

SA Share Alike
Reduces interoperability

ND No Derivatives
Severely restricts use

www.dcc.ac.uk/resources/how-guides/license-research-data

Deposit in a repository – but in which one?

Repositories and registries

www.re3data.org

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

datadryad.org

www.openaire.eu/search/data-providers

figshare.com*

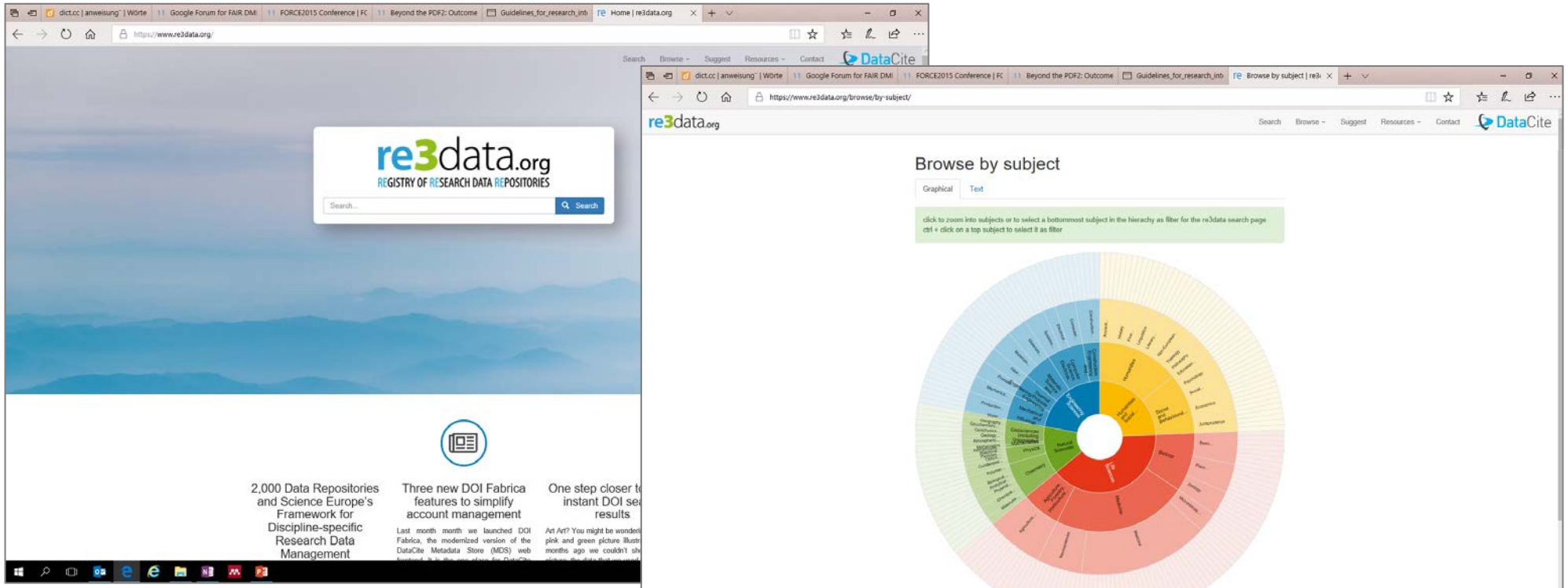
zenodo.org

zenodo

*Only partially recommendable as according to their Terms of Use, figshare is allowed to delete data anytime and without notice

Deposit in a repository – but in which one?

www.re3data.org

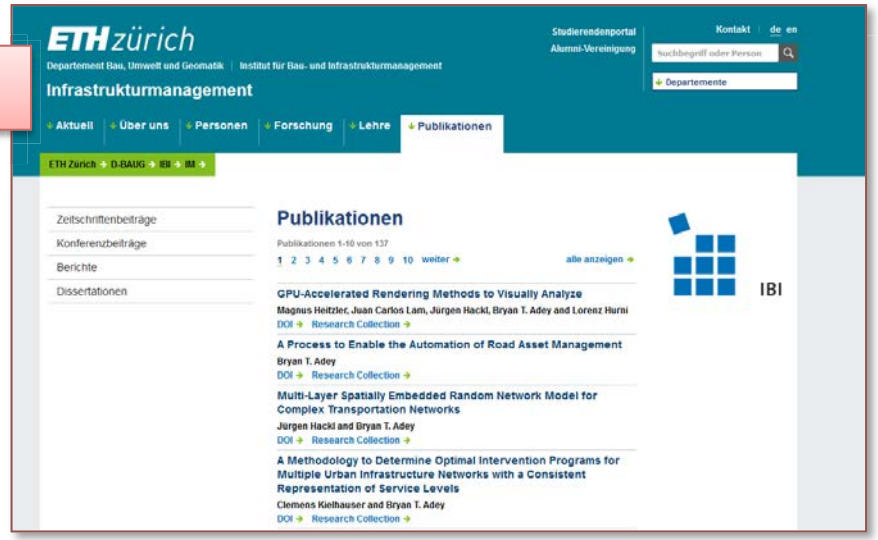


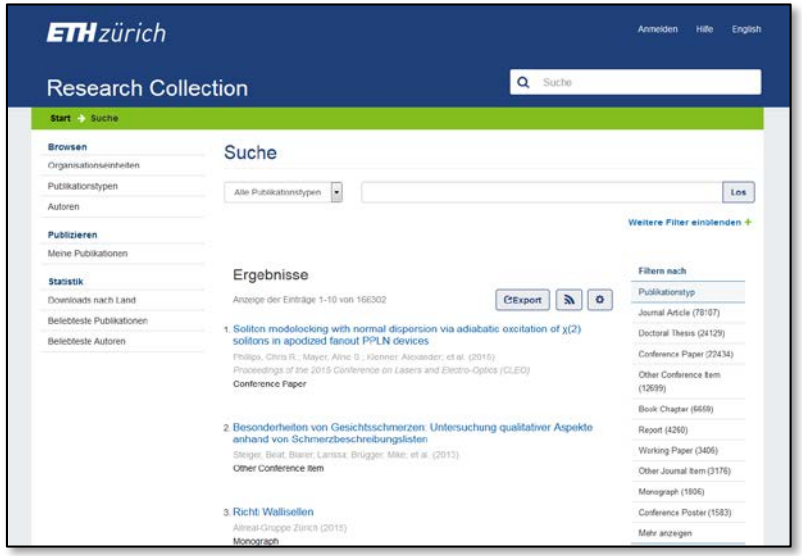
ETH Research Collection

- **New one-stop-shop for depositing research output**
  ETH Research Collection (https://www.research-collection.ethz.ch)
  - Publications, **Research Data**
  - Web upload, **DOI-reservation** and registration, ORCID, export to OpenAire…
  - Long-term preservation in **ETH Data Archive** (http://www.library.ethz.ch/Digital-Curation)
- **Metadata is always public, access to content may be delayed or restricted**
- **Aligned with FAIR principles** (Findable – Accessible – Interoperable – Re-usable) according to SNSF guidelines

Registry of publications / University bibliography

Web pages (AEM)





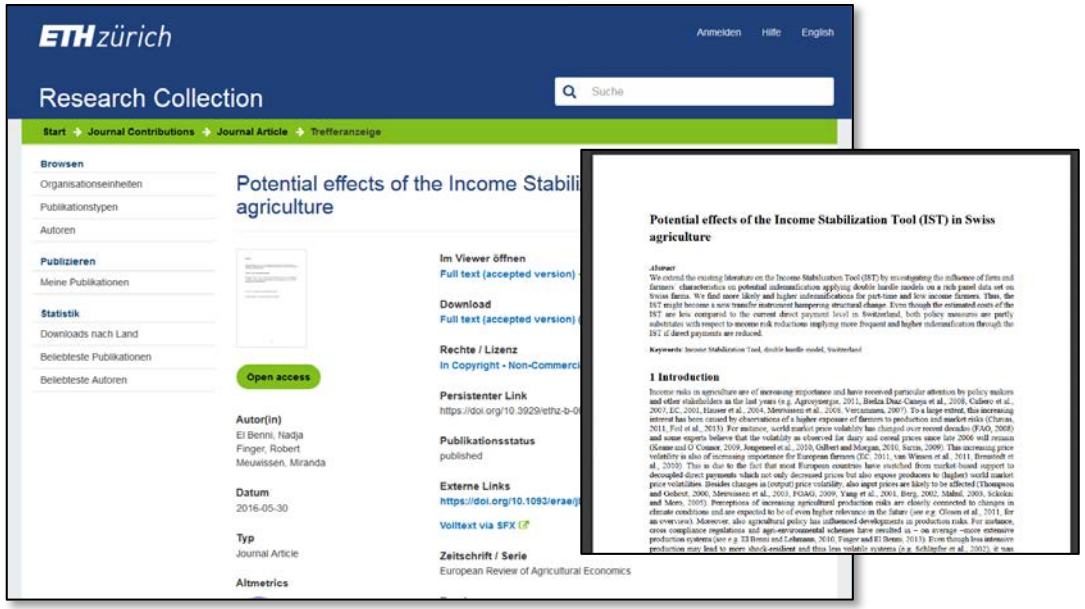
Annual Academic Achievements

Open Access repository

- Primary publication of reports, presentations, dissertations etc.
- Secondary publication of scientific papers (Green Road to Open Access)

Publisher's version

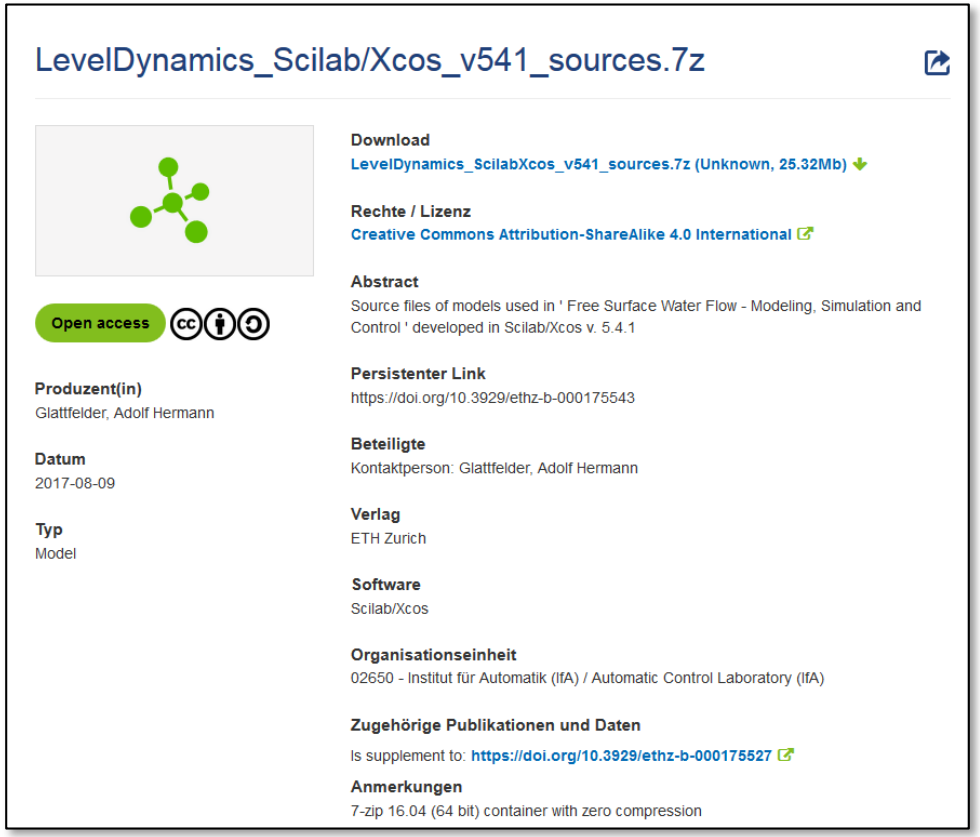
Open Access version



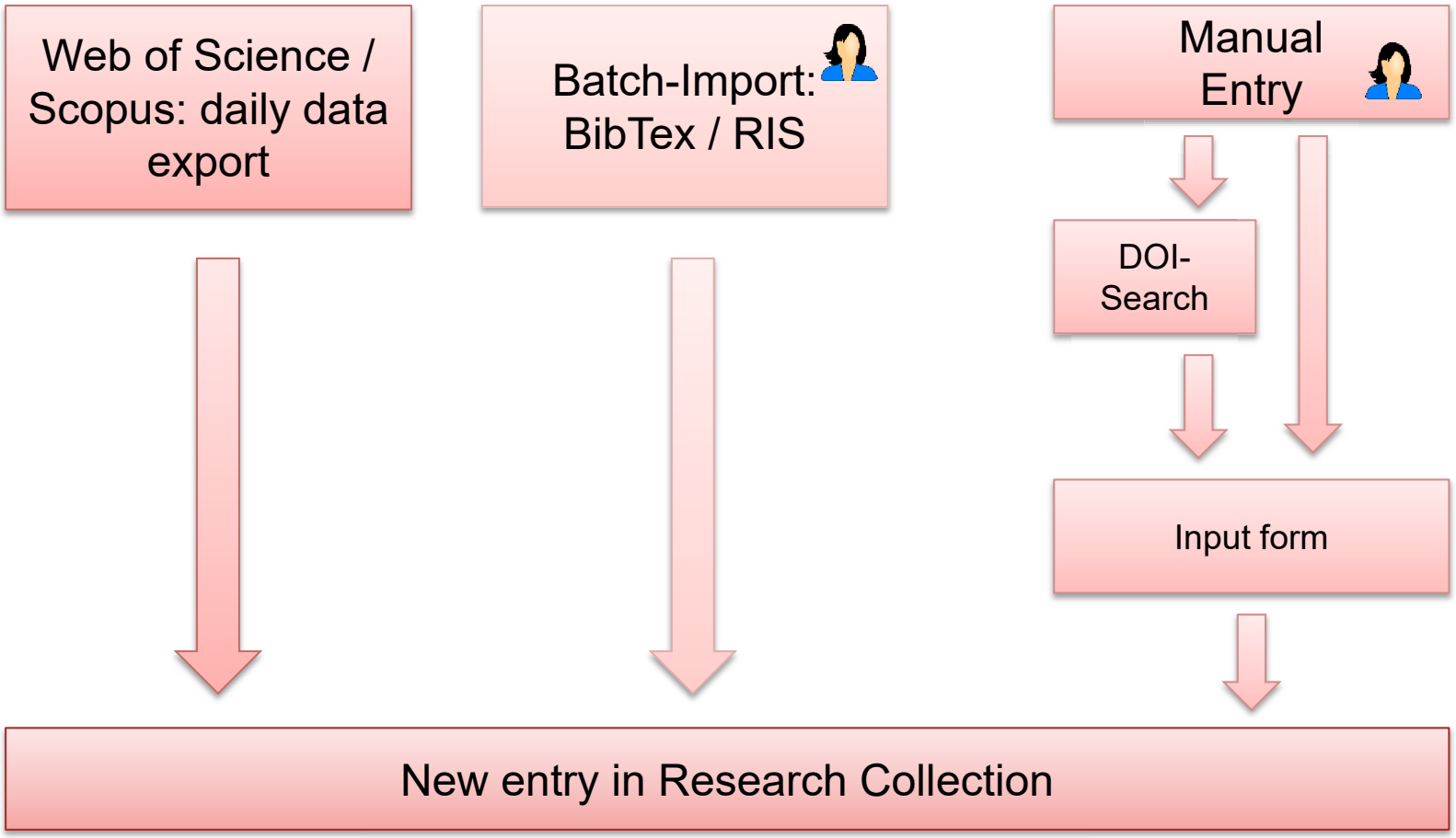
Research data repository

- Publication of research data as supplementary material or stand alone
- Access limited to selected users
- Deposit for preservation only

- All file formats permitted
- Retention periods:
  10 years / 15 years / unlimited



3 Ways for importing data



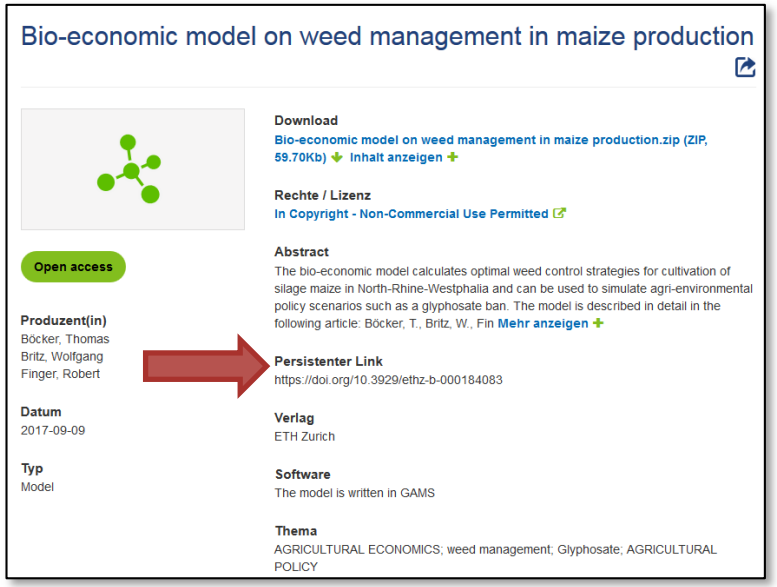
ETH Research Collection

Selection of access rights for full texts / data

	Open Access	Embargoed	ETHZ users	Selected users	Closed access
Publications	✓	✓			
Research data	✓	✓	✓	✓	✓

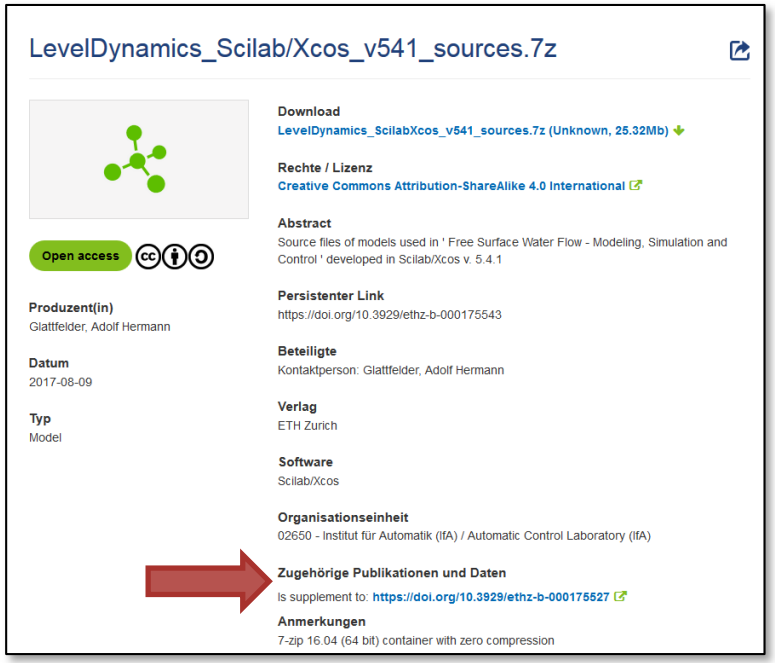
ETH Research Collection: Benefits of the Research Collection

Citable DOIs & possibility to reserve a DOI



Citation numbers / altmetrics / download statistics

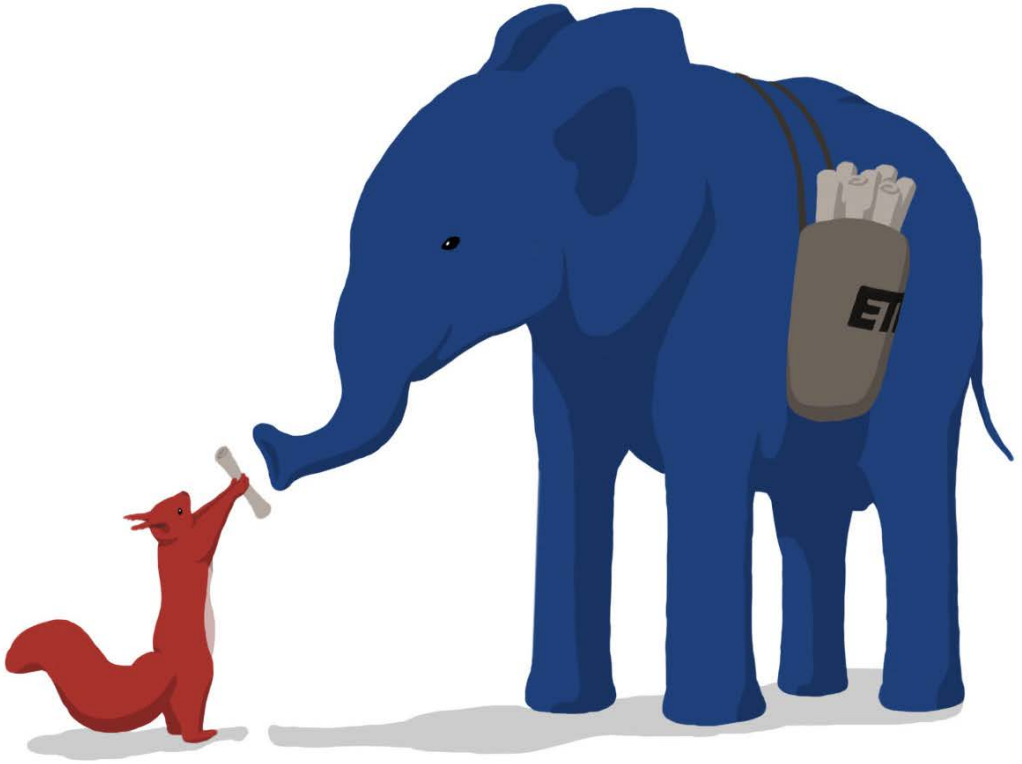
Linking between data set and publication



ETH Research Collection: Advice and support by ETH Library

- Legal issues in Open Acess publishing
- Open Access- and guidelines of research funders (SNSF, EU)
- Data management and digital curation
- ORCID support

www.research-collection.ethz.ch
Mail: research-collection@library.ethz.ch
Tel. 27 222



Long-term preservation of data

What does long-term mean?

Different time horizons and purposes

- Potentially **unlimited retention of data with permanent value** (e.g. long running series of observational data)

short term | up to 10 years | 10 years to permanent

Keeping data for at least **ten years** to ensure **accountability** if results are challenged (as defined in the ETH "Guidelines for Research Integrity")

- **Permanent retention of published data which is considered as part of the scientific record** and is expected to remain available just like articles and journals are
- In general **"long-term" signifies any time period which spans technological changes** in the way data is being used

How does this relate to data management?

Proper data management or its absence determine if presevation of data will be possible

If data is to be kept and used **for longer periods:**

short term | up to 10 years | 10 years to permanent

For a period of ten years, data management alone *might* suffice, but thinking further ahead is useful

- **Data should be as self-contained as possible**, including documentation of any tools used or better: the tools themselves;
  remember e.g. including **reference outputs** for model algorithms
- **More care is required in the choice and use of file formats**

Preferences for file formats

- **Open standards** (non-proprietary)
  - If proprietary, convert or if not possible include data viewer
- **Well documented**
- **Widely used and supported** by many tools
- **Uncompressed** (or at least losslessly compressed)
- **Unencrypted**
- When in doubt, **keep original and create a copy** in an open or exchange format
- Don’t rely on **file extensions**
- Consider that **data might be used in different operating systems**

Examples

Data	File format
Images	Uncompressed TIFF, JPEG2000
Text	ASCII, including XML etc.
Text (page-based)	PDF/A1-b, (PDF)
Data from spreadsheets	CSV
Spreadsheets	(CSV), (ODF, OOXML)

Add **encoding** information and **dependencies** such as stylesheets or TeX-libraries!

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

Note

- **This does not mean you «must not» keep data in other formats**
- Just be aware that **proprietary or undocumented formats** (even your own!) **might cause trouble in the future**
- Think about adding an **alternative format** (yes, redundantly) **for a proprietary one…**
- **… and add any context information** you yourself would like to have on your own formats in a few years time in a **ReadMe-file, an accompanying document** or as **metadata**

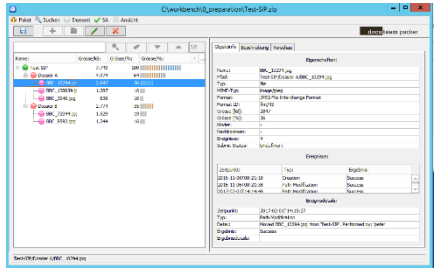
ETH Data Archive

Digital preservation solution for ETH Zurich, operated by ETH Library

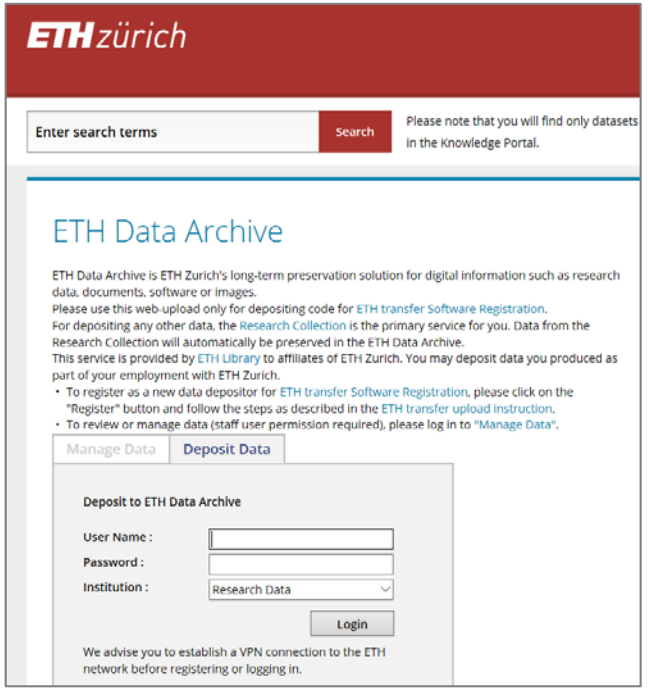
Research Collection

automatically archiving

Data



Docuteam packer



Heritage content from ETH University Archives and ETH Library

automatically archiving

«Software Disclosure» workflow for ETH transfer

software disclosure workflow

Yael Fitzpatrick: Science Cover Vol 331 (4.9.2018)
http://science.sciencemag.org/content/331/6018/728

ETH Data Archive

- **Digital preservation** solution for ETH Zurich, operated by ETH Library
- Automatically archives **content from Research Collection** and also **heritage content** from ETH University Archives and ETH Library
- Handles «**Software Disclosure**» workflow for ETH transfer
- For certain **automated use cases**, Research Data can also be submitted directly to ETH Data Archive via dedicated interfaces
- **Data** previously **organised in docuteam packer** will also be submitted to ETH Data Archive
- More information: https://www.library.ethz.ch/Digital-Curation

What messages are you taking home with you?

Take home message

- **Think about what you do!**
- **Start early**
- Agree on clean concepts and simple tools
- You do not need the latest sophisticated apps – but there are useful tools
- **Talk to colleagues**
- Check what your local service providers can offer
- «Keep it as simple as possible – but distrust it!»

Thank you!

Questions?

Research Data Management and Digital Curation

http://www.library.ethz.ch/RDM
data-archive@library.ethz.ch

Dr. Malin Ziehmer
RDM/DMP Consulting and Training
ETH Library
Rämistrasse 101
8092 Zurich
044 632 60 32
malin.ziehmer@library.ethz.ch

Dr. Ana Sesartic Petrus
RDM/DMP Consulting and Training
ETH Library
Rämistrasse 101
8092 Zurich
044 632 73 76
ana.petrus@library.ethz.ch

Research Data

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

We need your feedback!

Please fill out the course evaluation form – Thank you!
https://www.umfrageonline.ch/s/a13b937

Additional information

File organisation tips

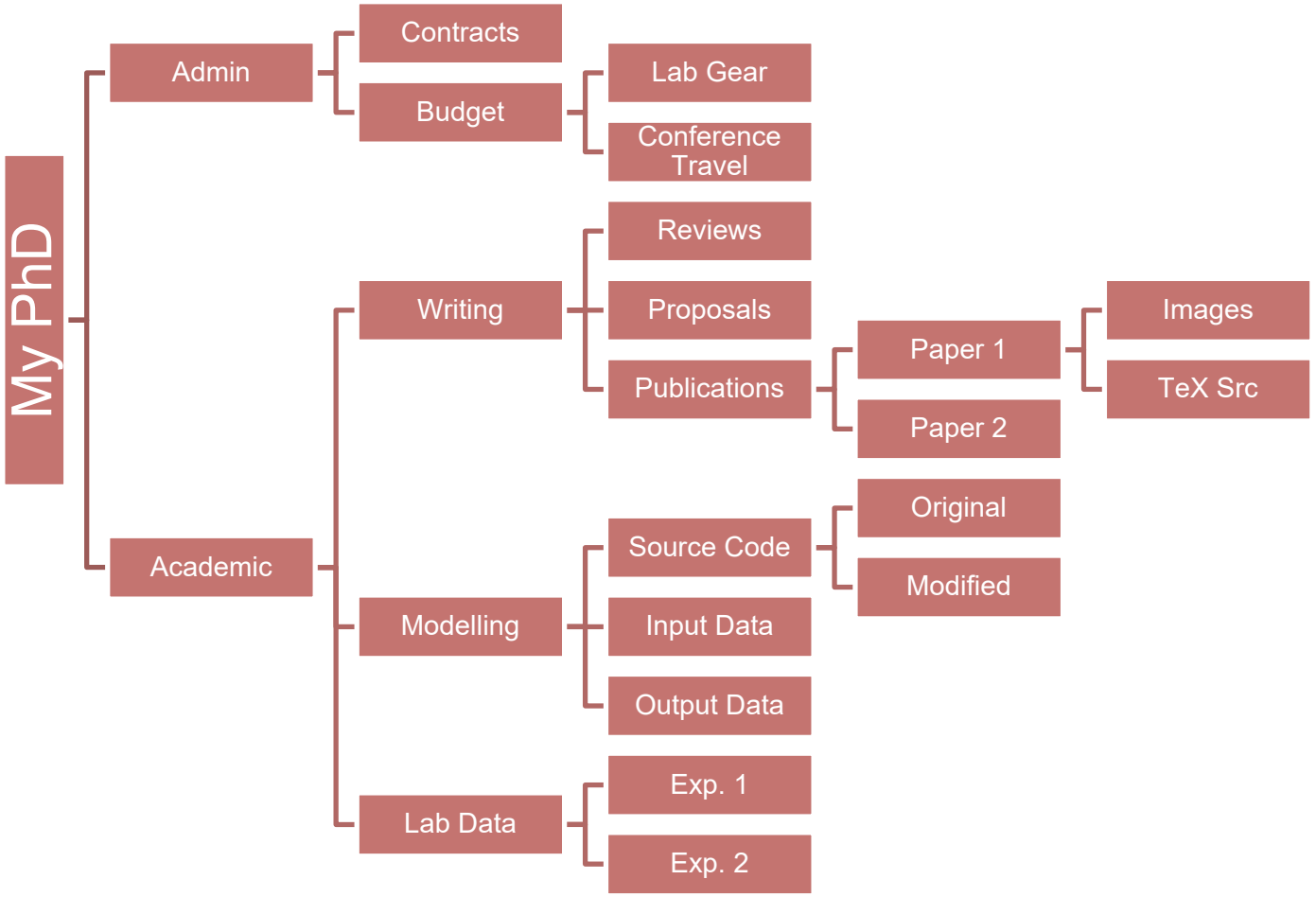
- **Keep stuff together that belongs together**
- Keep **path names** short
  - < 255 characters
- **File names** should
  - Reflect content and be unique
  - Use only ASCII characters (no diacritic characters)
  - No spaces
  - Lowercase or camel case (LikeThis)
- Careful! Not all systems are **case sensitive**!
  - UNIX: case sensitive
  - Win/Mac: mostly case insensitive
  - Assume that this, THIS and tHiS are the same.
- Document your structure and file naming conventions in a README text file
- Write **dates** like this: **YYYY-MM-DD**

© XKCD
https://xkcd.com/1179/

For **further file and folder organisation tips**, see:

- http://www.data.cam.ac.uk/data-management-guide/organising-your-data
- http://www.wur.nl/en/Expertise-Services/DataManagement-Support-Hub/Browse-by-Subject/Organising-files-and-folders.htm
- http://datalib.edina.ac.uk/mantra/organisingdata/

A possible folder structure during your Phd could look like…



Metadata & Standards

- **Metadata** is the *data about your data*
- Use of structured metadata facilitates data organization and searches
- Examples of metadata*:*
  - *Investigator*
  - *Date*
  - *Title*
  - *Description*
- Several metadata schemas are available.
  - For info, check the DCC website

"Metadata" by Jørgen Stamp / CC BY 2.5

- **Standards** (taxonomies, synonyms, ontologies) are important to guarantee consistency
- **General standards**:
  - ISO 8601 for dates (YYYY-MM-DD or YYYYMMDD)
  - ISO 6709 for latitude/longitude
  - standards for SI base units (meters, kilograms, etc.)
- **Scientific standards** examples:
  - Biology -> Gene ontology, NCBI taxonomy, etc.
  - Physical sciences -> IUPAC, InChI
  - Earth science and ecology -> USGS Thesaurus, GIS dictionary, etc.
  - Math & computer science -> Mathematics Subject Classification, ACM Computing Classification System

Visible and promoting innovation – advantages of open data

1. Transparency

Users can swiftly and easily get an idea of the kind, scope and content of open data.

2. Accessibility

Open data is swift and straightforward to access.

3. Contribution toward the global information infrastructure

The efficiency is increased as data only needs to be collected once, thereby halving the work.

4. Development of innovative applications and services

Open data simplifies and accelerates the development of new services as there is no need to clarify legal issues.

5. Creation of new business models

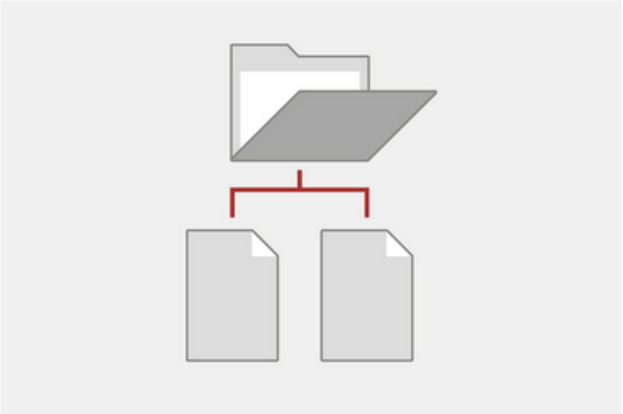
Thanks to the concept of open data, new business models are born.

6. Traceability in search engines

Open data is indexed and thus displayed in the list of hits during web searches.

Source: https://doi.org/10.22010/ethz-exp-0004-en

Six easy tips to keep your data safe



1. Organise and standardise

Establish a file and folder structure that works for you and use it consistently.

2. Identify

Determine which files need to be preserved.

3. Automate backups

Create automated backups and keep them both locally and off-site.

4. Know the lifespan

Know the lifespan of your data carriers and re-copy your data to new ones in time.

5. Use simple tools

When collaborating, agree on simple workflows and backup tools. Don't forget to document the context of your data.

6. Use open file formats

Use open file formats and don't compress data to ensure its compatibility with different operating systems.

Source: https://doi.org/10.22010/ethz-exp-0002-en

ETH Library's services all around data

1. Support and training

ETH Library offers training courses on various topics.

2. Consulting

Individual support on all things related to research data management.

3. Open data

ETH Library provides various kinds of data openly accessible.

4. Preservation

Expertise in data preservation is important to ETH Library.

5. Publication

Publish your data in ETH Zurich's Research Collection.

6. Informational material

Find checklists and reference material on ETH Library's website.

ETH Library's services regarding open access

1. Open access repository

Members of ETH Zurich can self-archive their papers in the Research Collection.

2. APC funding

ETH Zurich covers the publication costs for members of ETH Zurich at certain publishing houses.

3. Open access policy

ETH Library is the contact point for questions concerning ETH Zurich's open access policy.

4. Copyright

Advice on legal questions related to open access publications

5. Research funders

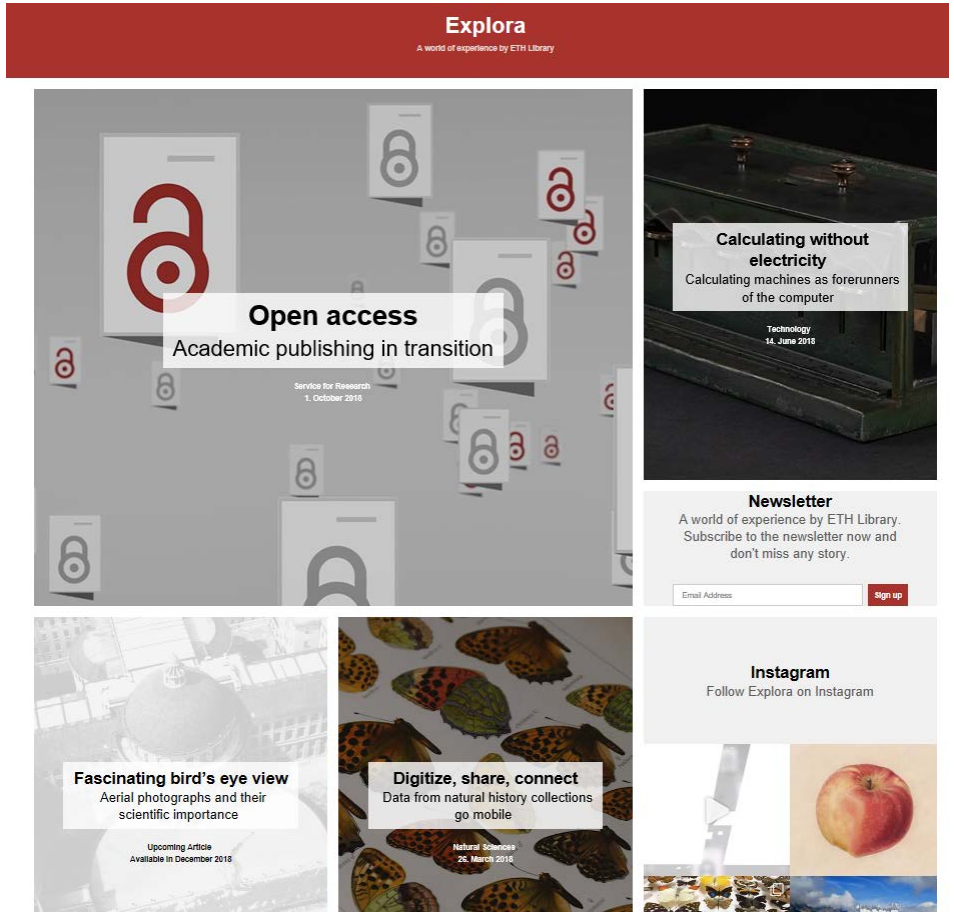
Advice on the research sponsors' open access rules

6. Courses / Book a Librarian

Workshops, webinars and individual consultations

Source: https://doi.org/10.22010/ethz-exp-0009-en

Visit us on Explora – A world of experience by ETH Library

- https://www.explora.ethz.ch/en/

- **serialized stories** on
  - **Open access** – Academic publishing in transition
  - We love data – Why **data management** matters
  - **Open Data** – Transparency for everyone
  - Surfing the sea of data – Handling **data storage** and preventing degradation
  - and many more…