


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](#)



Research Data Management Workshop at RCS18



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**



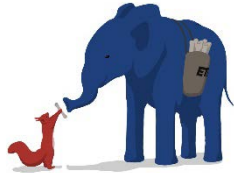
The screenshot shows the ETH Library website with a red header. The main content area features a video player titled "Digital Curation at ETH Zurich" with a thumbnail showing "ETH Library Digital Curation Services for Scientists". To the right is a navigation menu with links like "Why digital curation?", "Research data", "Digital collections", "Our services", "Contact persons", "News", "Instructions and Downloads", and "Links". At the bottom of the page, there is a red banner with the URL <http://www.library.ethz.ch/Digital-Curation>.



Let's get to know you a bit better...

«Cross the line»

What we are going to do...



Long-term preservation



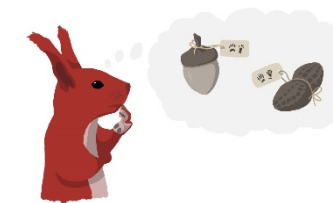
What is data management and why should it concern you?



Data sharing



Regulations, intellectual property, privacy and access rights



Data Management Planning



What is data management and why should it concern you?

An introduction



What is data?

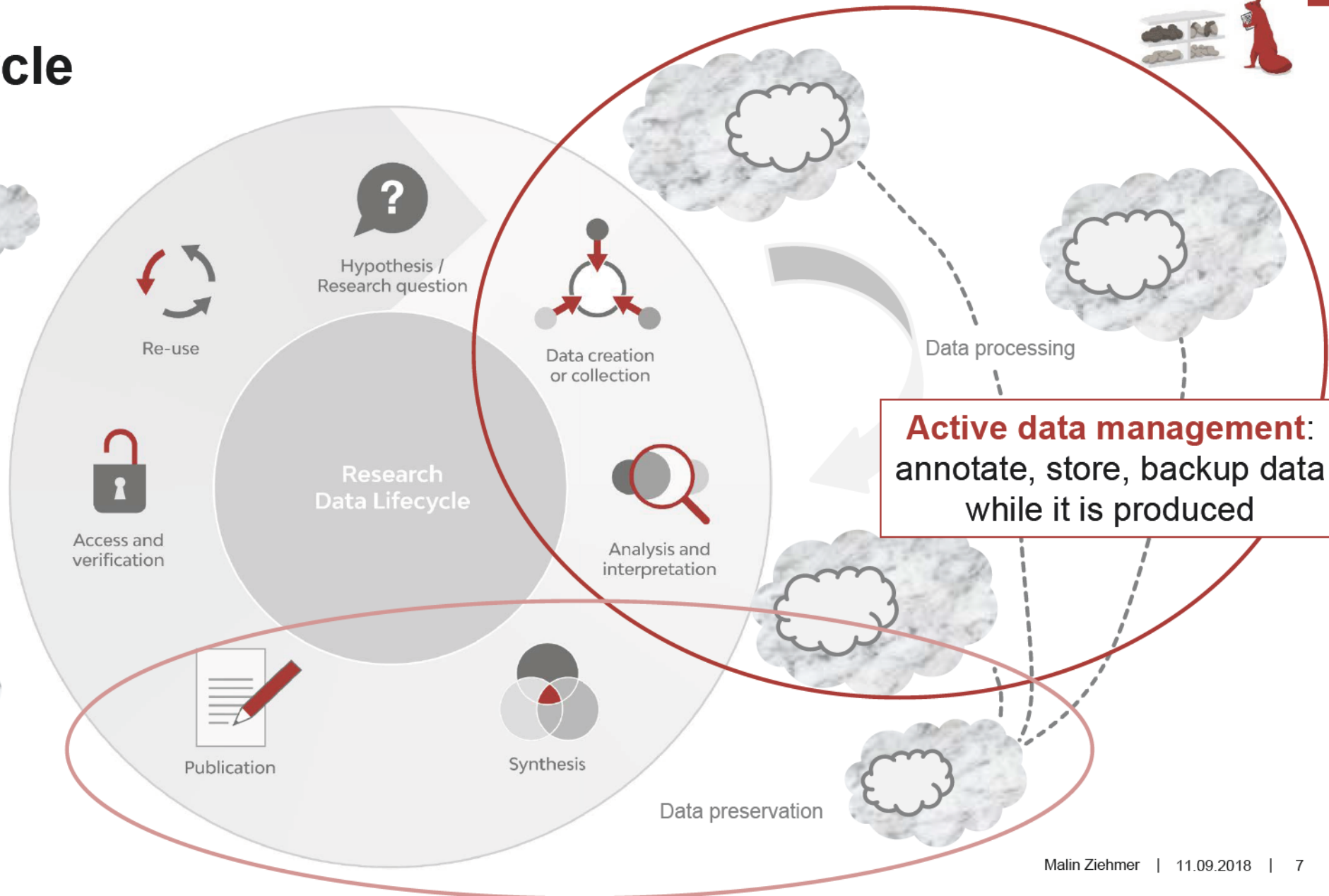


“A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.”

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...

The collage features three documents:

- Left Document (Blue background):**
 - Header: ETH zürich
 - Title: **Compliance Guide**
 - Text: English version 2015
 - Quote: GOOD PRACTICE BETTER PRACTICE BEST PRACTICE
 - Illustration: A stick figure pointing upwards.
- Middle Document (Red background):**
 - Text: Edition May 2018
 - Title: **COMPLIANCE GUIDE**
 - Illustration: A cartoon character with wings flying over a sign that says 'EPFL', with sheep in the background.
 - Logo: EPFL ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
- Right Document (White background):**
 - Header: ETH ZÜRICH
 - Title: **GUIDELINES FOR RESEARCH INTEGRITY AND GOOD SCIENTIFIC PRACTICE AT THE EPFL**
 - Section: **FOREWORD**
 - Text:

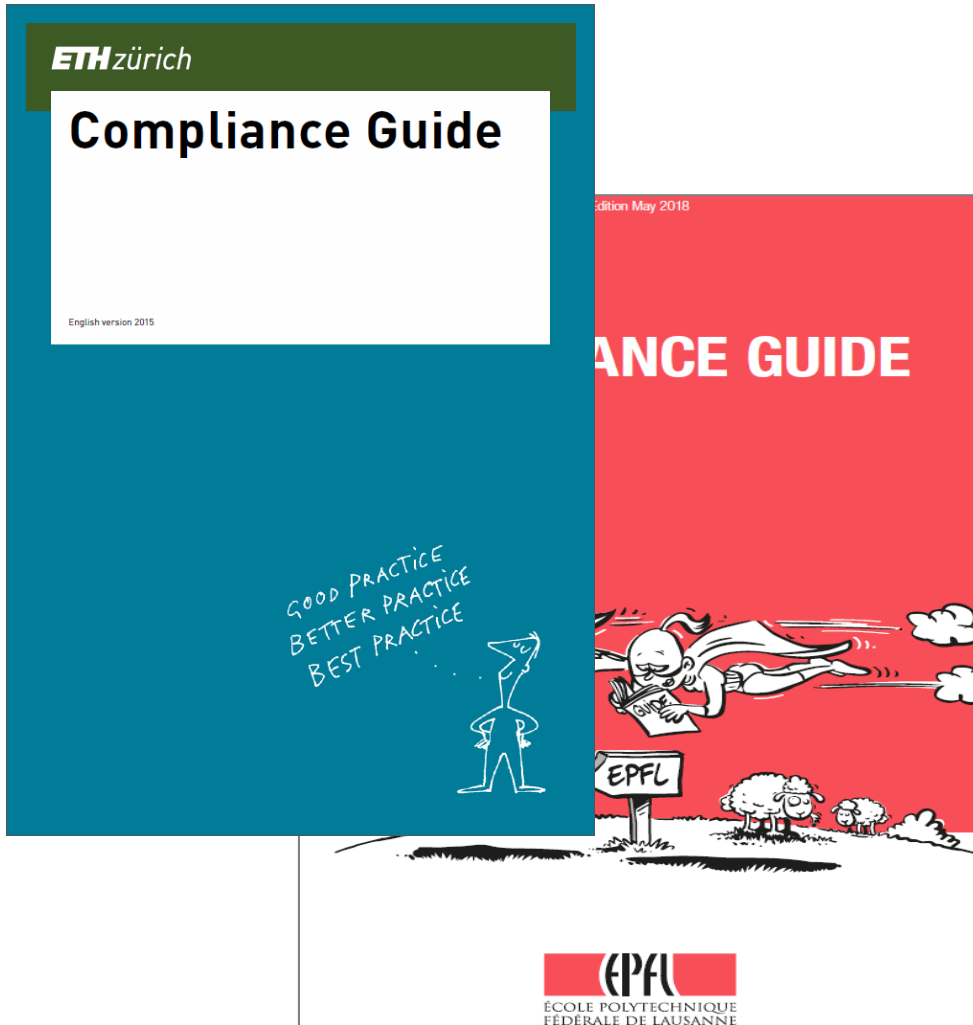
Integrity is a significant feature of all societies and represents major commitment, in the public or private sector. Results and findings form the basis for policy decisions at the levels of government and the private sector. Therefore, it is of paramount importance that research itself is conducted with integrity, in a responsible manner and in accordance with high ethical standards.

Integrity defines the commitment of researchers to adhere to the fundamental rules of good scientific practice. Truth and transparency, self discipline, self criticism and fairness are essential elements of the behaviour of integrity. They represent the basis of all scientific activity and are the basis for the credibility and acceptance of science. It is the responsibility of all researchers at the EPFL to uphold the good reputation of the school, to observe its rules, policies and to undertake activities in accordance with them. All members of the EPFL are expected to conduct research with integrity. The experience shows that, in day-to-day research, researchers may have some questions about the best way of properly conducting research.

The 'Guidelines for Research Integrity and Good Scientific Practice' define the principles that apply to all researchers at the EPFL in all disciplines. These guidelines are intended to serve as guiding principles for the planning, execution, presentation and reporting of research work at the EPFL. In this way, they help ensuring that intellectually honest research will continue to be conducted at our institution.
 - Text: Prof. Bernard Rossier, Ombudperson
 - Page: Page 1



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

ETH ZÜRICH

RESEARCH
AND SCIENTIFIC
ETHICS

Richtlinien für
Integrität in der Forschung

Guidelines for
Research Integrity

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Monday, June 15, 2009 Page 1

- 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.**



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



There is NO CLOUD, just other people's computers

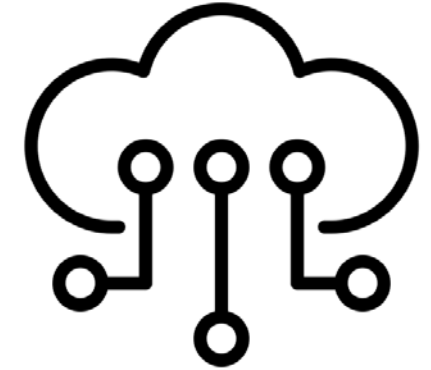
"There is no cloud just other people's computers" (4.9.2018)
by Markus Meier
CC BY-SA 4.0

Cloud computing @ ETH Zurich

Rules and Regulations



- 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**.



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

- 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)

(<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) (<https://www.admin.ch/opc/en/classified-compilation/19920153/index.html>)

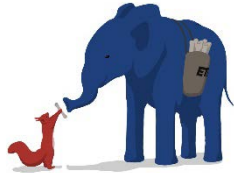
[Swiss Criminal Code](https://www.admin.ch/opc/en/classified-compilation/19370083/index.html) (<https://www.admin.ch/opc/en/classified-compilation/19370083/index.html>)

- 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?



Data sharing



Regulations, intellectual property, privacy and access rights



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/>

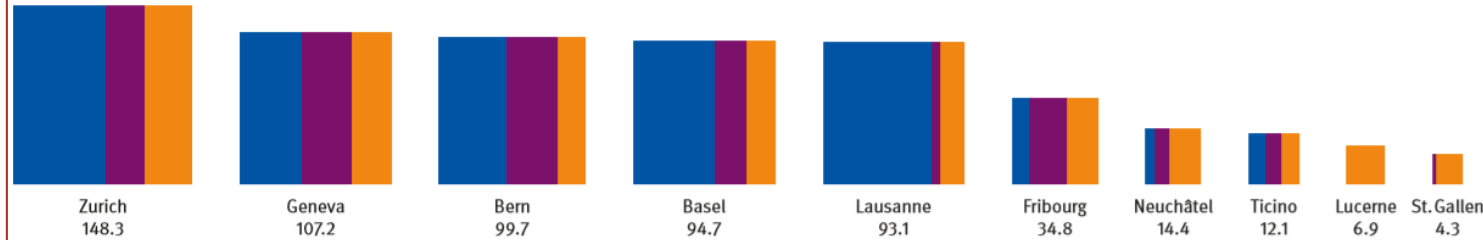


SNSF Statistics 2017

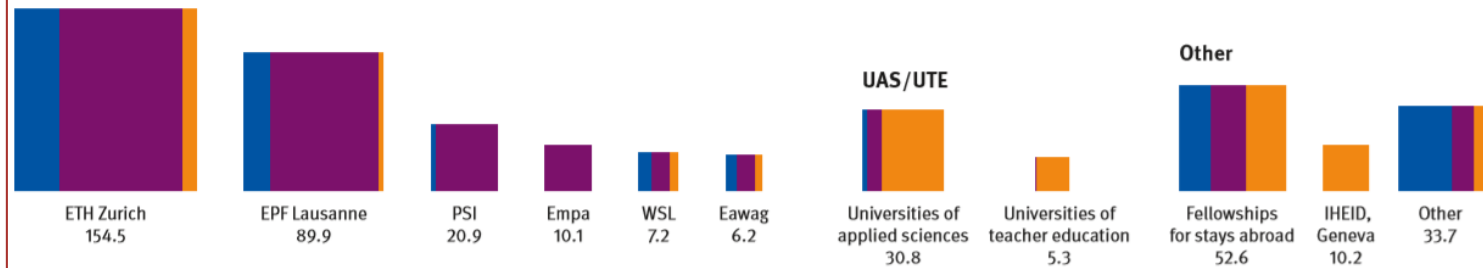
By institution and discipline

in CHF million

Universities (incl. university hospitals)



ETH Domain



39%

Biology and medicine

38%

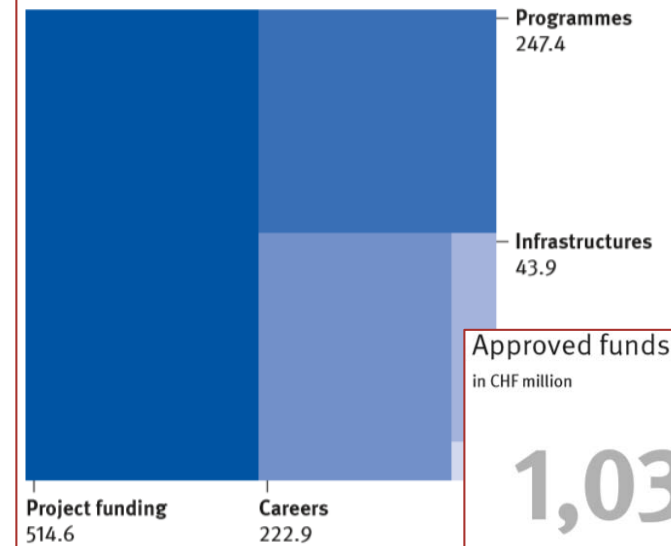
Mathematics, natural and engineering sciences

23%

Humanities and social sciences

By funding category

in CHF million



Approved funds

in CHF million

1,037

Approved applications

2,971

Submitted applications: 6,041



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**

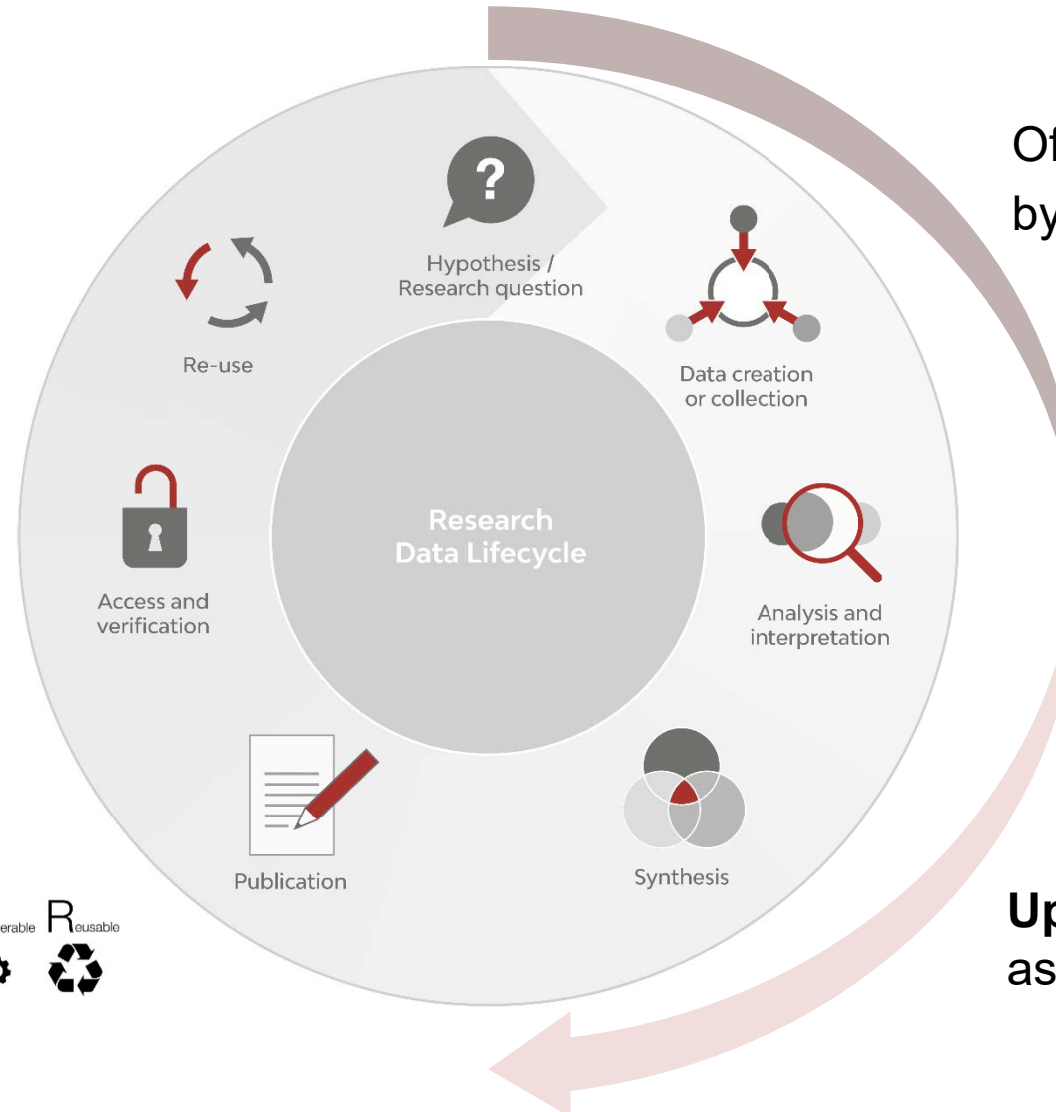
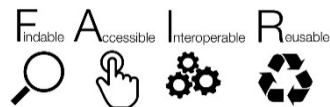
Please also be aware of SNSF's updated Open Access Policy for Publications and changes to the General implementation regulations for the Funding Regulations!

<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



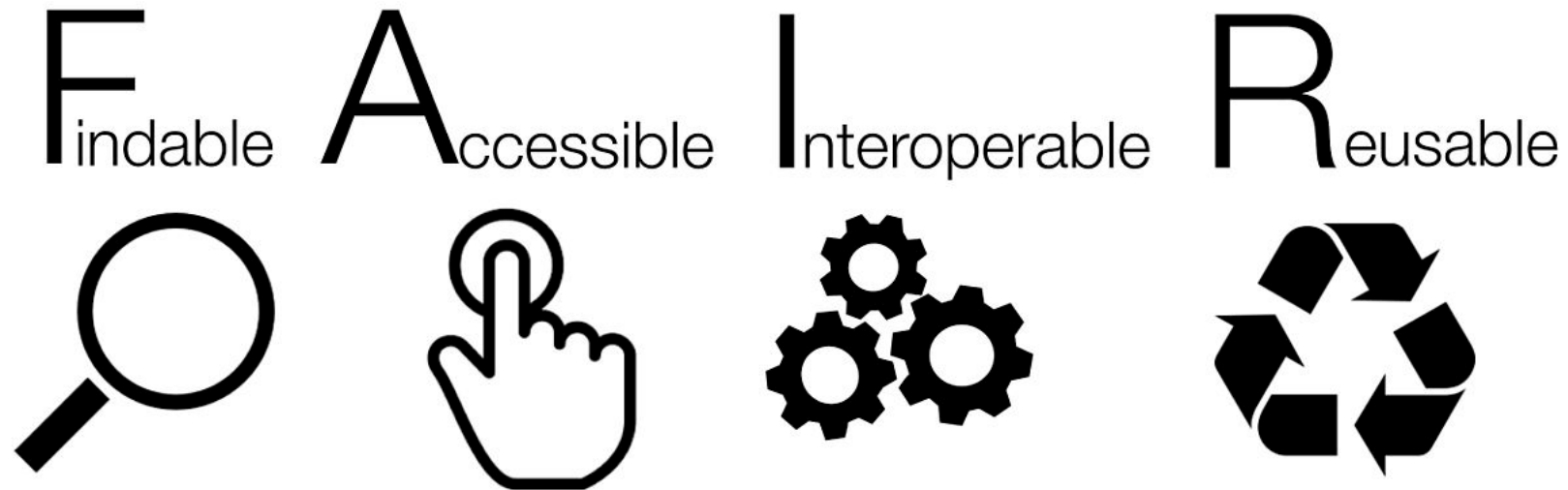
Offering a **long-term perspective** by outlining how the data will be:

- 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](https://doi.org/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>
- **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>

What we offer:

- (individual) **training** on DMP
- **consultation** on data management planning
- **review** of DMPs





Research Group Policy



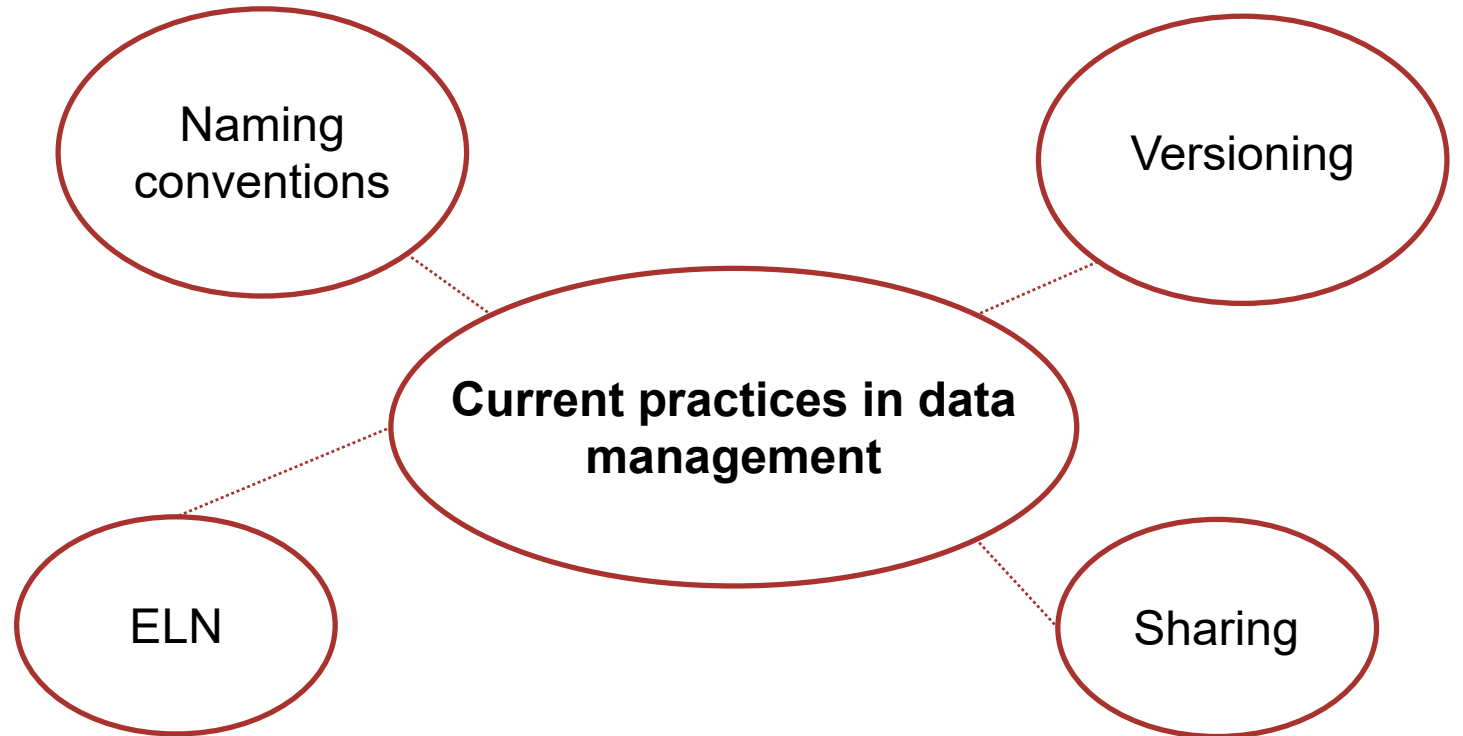
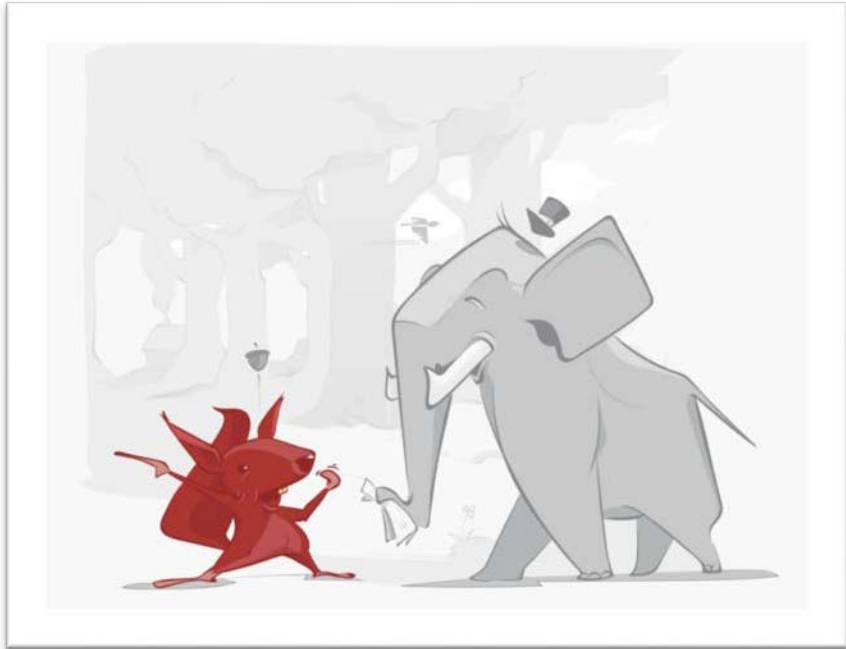
- **Self-critical questions:**

- 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



EPFL & ETHZ Summer School on Reproducibility in Computational Sciences 2018 (Magliano, Torino (CH))
 Data Management in Research (Workshop, Tuesday, 11.09.2018, 15:30 - 17:30), Malin Ziehmer (ETH Zürich, ETH Library)
 Exercises on current practices in active research data management

Naming Conventions

Filename: date generated (e.g. timestamp_directory) General information about data (HACAM) file for how to use folder for new data Folder for processed data (including documentation on processing)	Usually, between words and symbols for e.g. digits with minor and minor L/C2... Also dates are pyyyyymm	Interspersing	Conventions used in the field - ie. protein names	No file naming convention One folder names
name	date_yyyyymmdd_hhmmss For the experiments Name should give you what's about responsibility acronym	Major	When there is data in a folder, components in the folder name appear twice and a date or description	On OpenMML, ID number + human-understandable labels + version number for each data set R packages HPC workflow folders Year Host site
I prefer to have single letters & a specific description, but not...	We have for test data, sometimes: timestamp, source, measured	Minor, pyyyyy-mm-dd, descriptive Group naming	Water chemistry experiments have Both identifiers, but sometimes a name of site for	
Start measurement names with date E.g. 20180811... Add project for your acronym in last E.g. ICD4 (group)	Date, email characters, name of experiment (important part)	On demand	Not existing	
Name recognition Name pyyyyy-mm-dd Name1, Name2...	Write down the name of independent variables and dependent variables	Yes, per project, not general Try to follow established conventions	Prevent hash based workflow instability EIS not sure how it's called, my program just needs to find it	
All small case letters Date or acronym as a part of name of the folder or that it's folder structure E.g. 2018-01-01_experiment1-run-01	When changing a certain parameter: usually save it like: date_parameter_value for that parameter	Project folder with lowercase and date (usually yyyy-mm-dd)	Not yet	

Versioning

OpenMML	version_..._final_..._v1 -> not working !! Code: git -> automatically (+), not everything uses it (-)	Git Google about typescripts	SVN Others...
Git for version control research Custom system for data (includes old data)	Git ->3	Git: BUT: everyone is afraid to merge -> too many branches right now, not nice assigned as team responsible	Research Git: F.S. Overkill for collaboration with non-chemists
Git (code)	Folder naming	GitLab On manual versioning using annotations and small HACAM in each	GitLab for code GitLab for manuscript Others: to give each user a GitLab who access from home
v1, v2, v3 Clearly not optimal if you have v1, v2, v3, v4... Impossible to know which version point to which database	Different tables on PostgreSQL, one for each version of the data All for code	Write a HACAM file and version folder name with descriptions	Folder name: number the versions with numbers at the end of the name using leading zeros e.g. 01.01.01.01.01.01.01.01
Local git repository	Git - however, not enforced	git	I make back data from a simulation (PRO, ACT) and how they simulated a new version with different the types and column names. Can't use that or all should end up months for use the new version.
For test data, there is a convention, which however is not enforced to change over time Simulation data up to the researcher	GitLab (code only)	GitLab (code only)	Using SVN instead of git Cloning the gitlab.

ELNs - Electronic Laboratory Notebooks

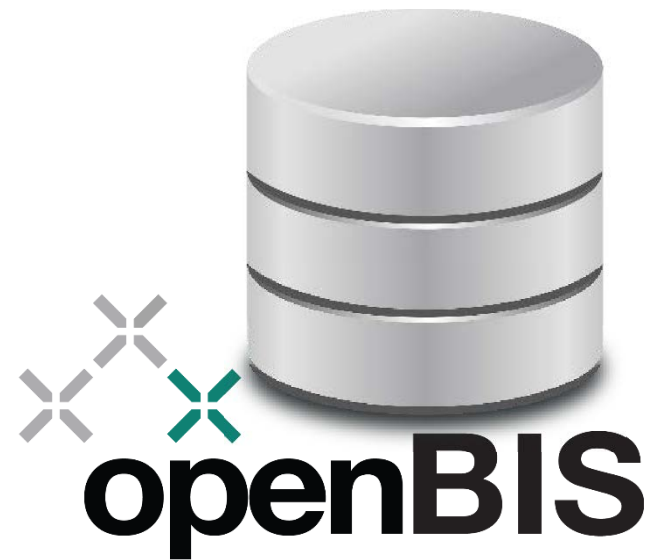
How often to handle clean room notebooks?	I usually use Google Docs and save them with the date of the document and adding what I did that day	There are and they are used daily	No Start Lab Book BUT: who of experiments are stored in metadata of dataset
Applyer	OpenMML/ELN Applyer	No	Applyer No, back-up my notes on OneDrive with general thing system and systems of notebooks
No	repro	Evernote	ELN - no, this is here to me I only use gDoc notebooks
Flat with GoodNotes (private backed only)	No idea	OneDrive	Evernote Type notes are the closest
			Lab Book on SharePoint Using a mobile and learning it (not enforced)

Sharing

SSH cVIM? Scal RAS	Cloud Desktop Institutional exchange server (lightly used) Scal RAS	ICED repository Internal network (not used) SVN	Overleaf Sharelatex Chorus 2018 Acorns tools for simulation data sharing
GitLab Shared internet drive Monthly Data	Dropbox (internal) Google drive (external)	Mainly OneDrive than for my share	Mail Git Chorus storage
Shared server with 1 PlexServer database	Local RAS + 2018 VTN with internal access rights based on projects	Dropbox Google drive	OneDrive Google drive OneDrive I only have 100 GB data through direct download for main gitlab. I have 1 OneDrive.
Git OpenMML S.F. page + CIMAN	OpenMML Collab for train datasets	Shared folders via PlexServer	Nextcloud Pangeo (ETH) LARCH server (for internal) Shardflow (for internal) Building up an solution to share data
We stored data on cluster and share it from there	Playbook Group Notebook One	OneDrive Google drive Shared server	GitLab Dropbox OneDrive OneDrive OneDrive Dropbox OneDrive OneDrive OneDrive OneDrive OneDrive

Current practices in active research data management

11.09.2018

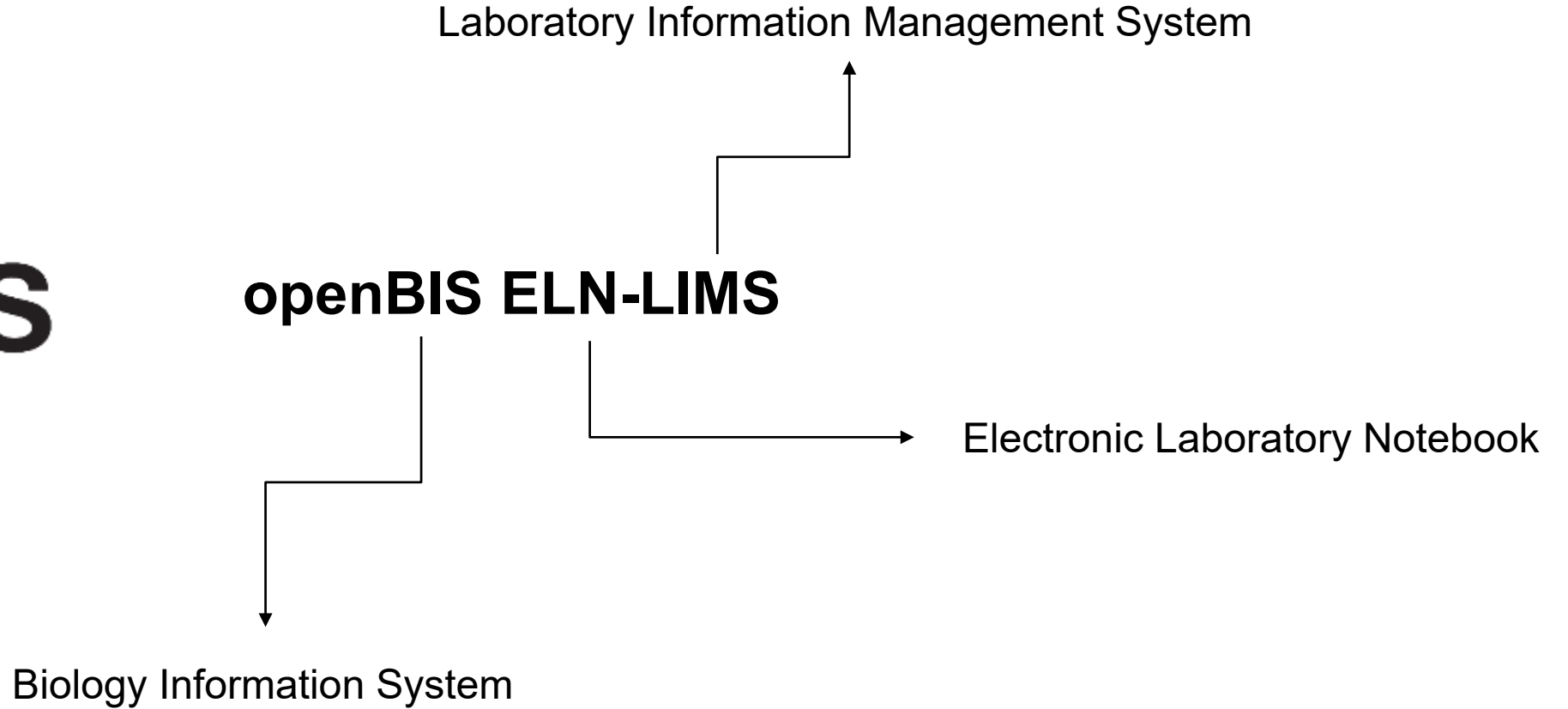


The ETH Scientific IT Services data management solution

for active research data management

An example for ARDM

What is openBIS?

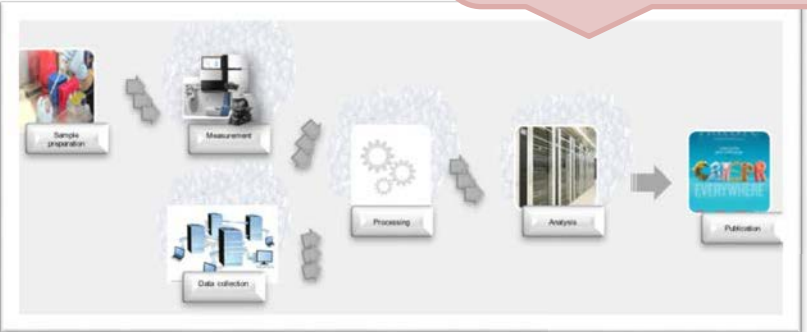
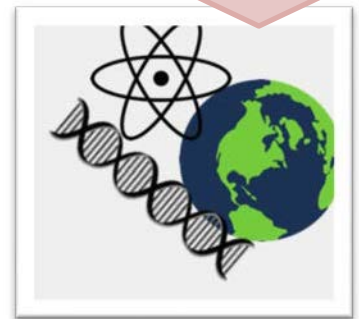


openBIS facts

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

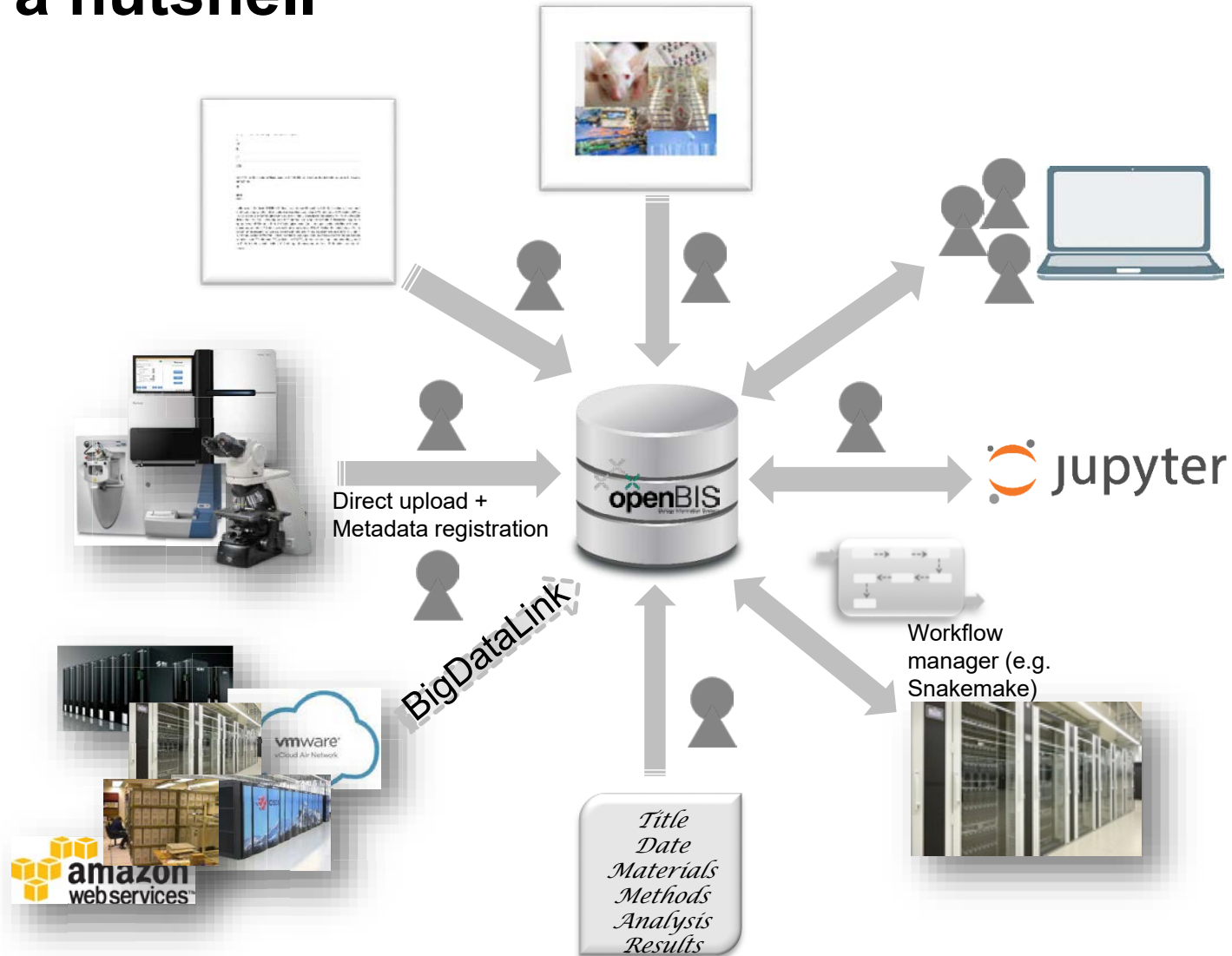
Can be used in most quantitative science fields (e.g. life sciences, physics, env. sciences, etc)

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

openBIS in a nutshell





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

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)

Excursion: File sharing tools



polybox.ethz.ch



cifex.ethz.ch/



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

recommended

- Data stored in Switzerland
- Security regulations fulfilled



www.dropbox.com



www.wetransfer.com

only conditionally
recommended

- 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

ETH zürich
polybox.ethz.ch

polybox.ethz.ch

Sign in to CIFEX

User:

Password:

Login

<https://cifex.ethz.ch/>

SWITCH

www.switch.ch/driv

SWITCHfilesender SWITCH

Menu Max. upload 50 GB / File

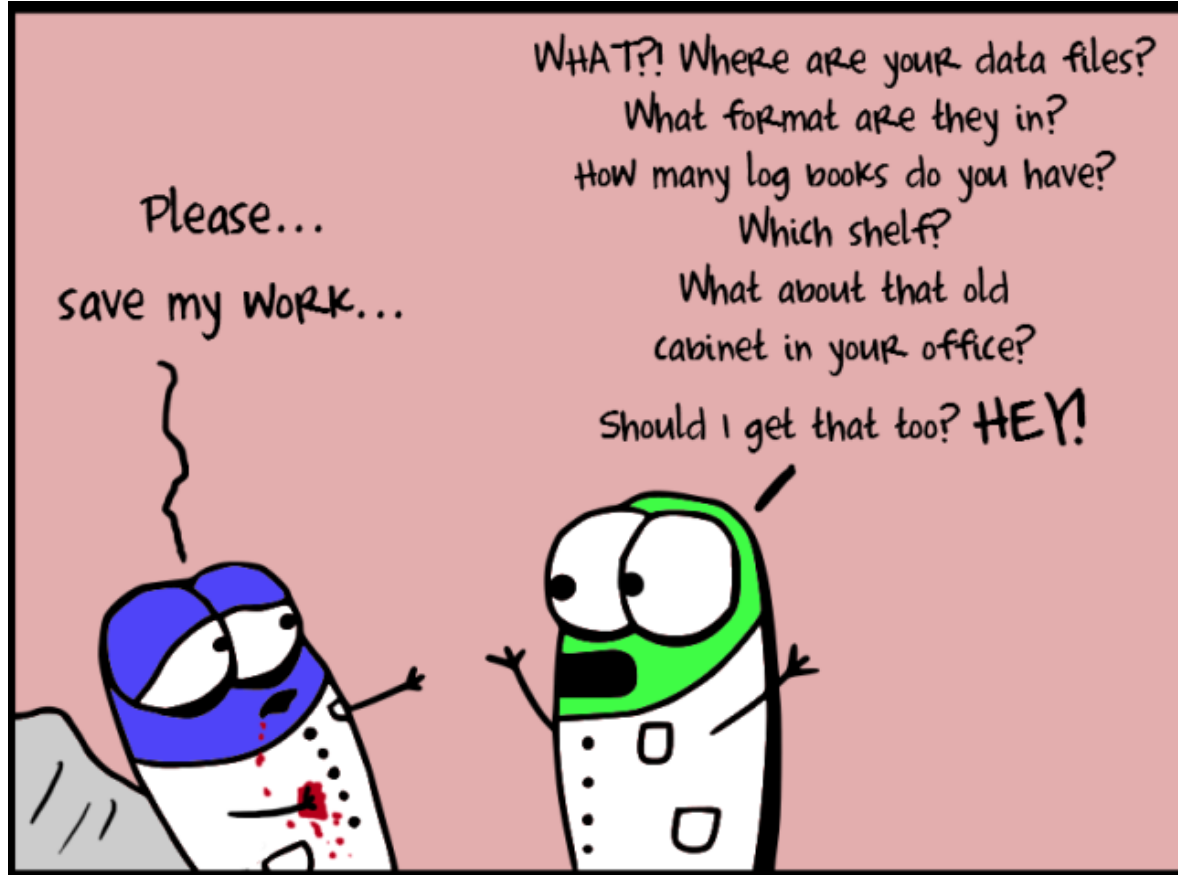
Willkommen bei SWITCHfilesender

SWITCHfilesender ist eine sichere Methode, grosse Dateien zu verschicken. Loggen Sie sich ein, um Dokumente weiterzugeben oder um jemanden dazu einzuladen.

Einloggen

To share files permanently, use SWITCHdrive

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



Step 2

Add branches to your map



Step 4

Color your branches

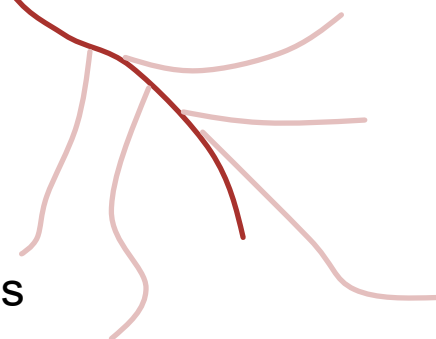


Step 3

Add key words

Key words

Key theme



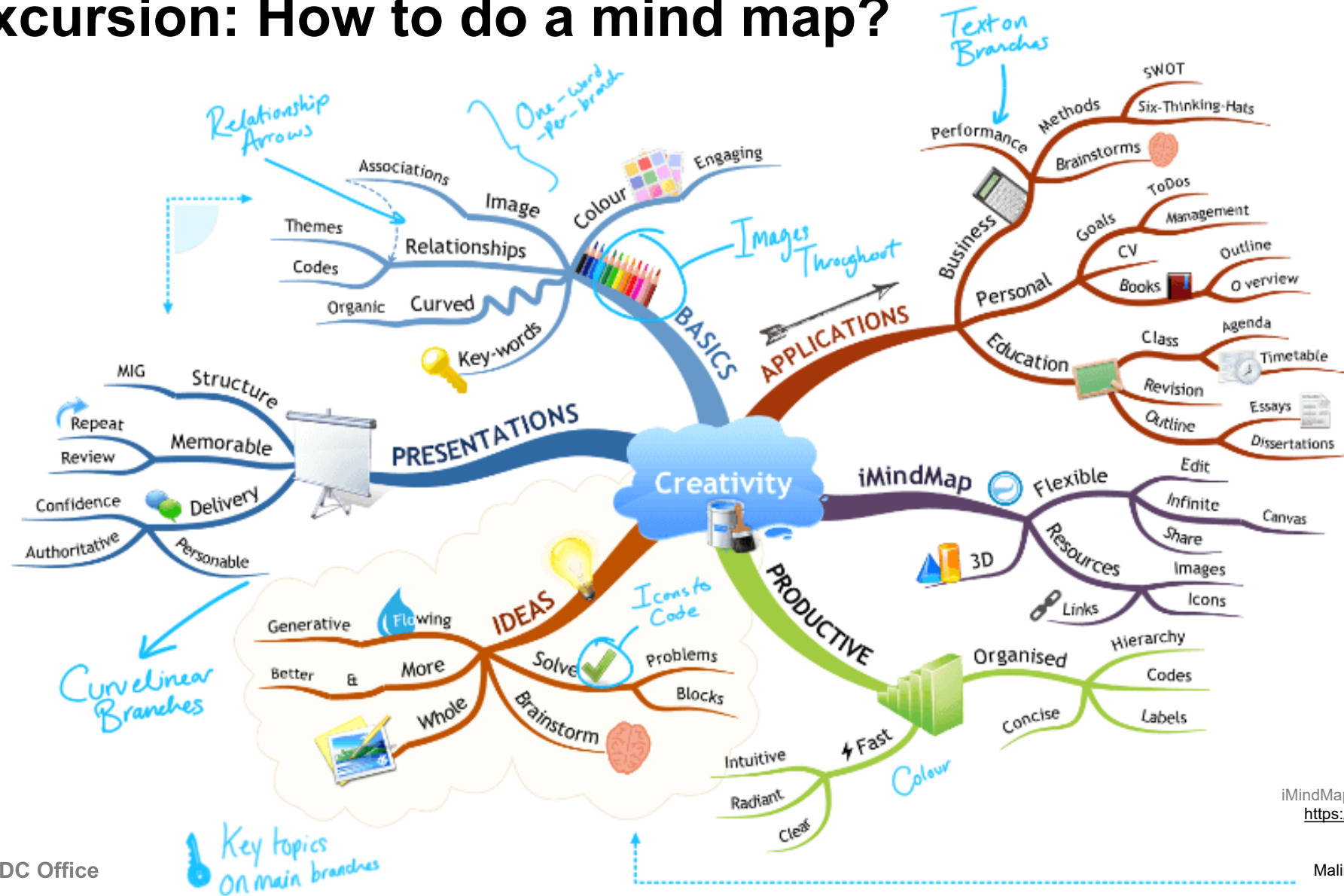
Step 5

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?



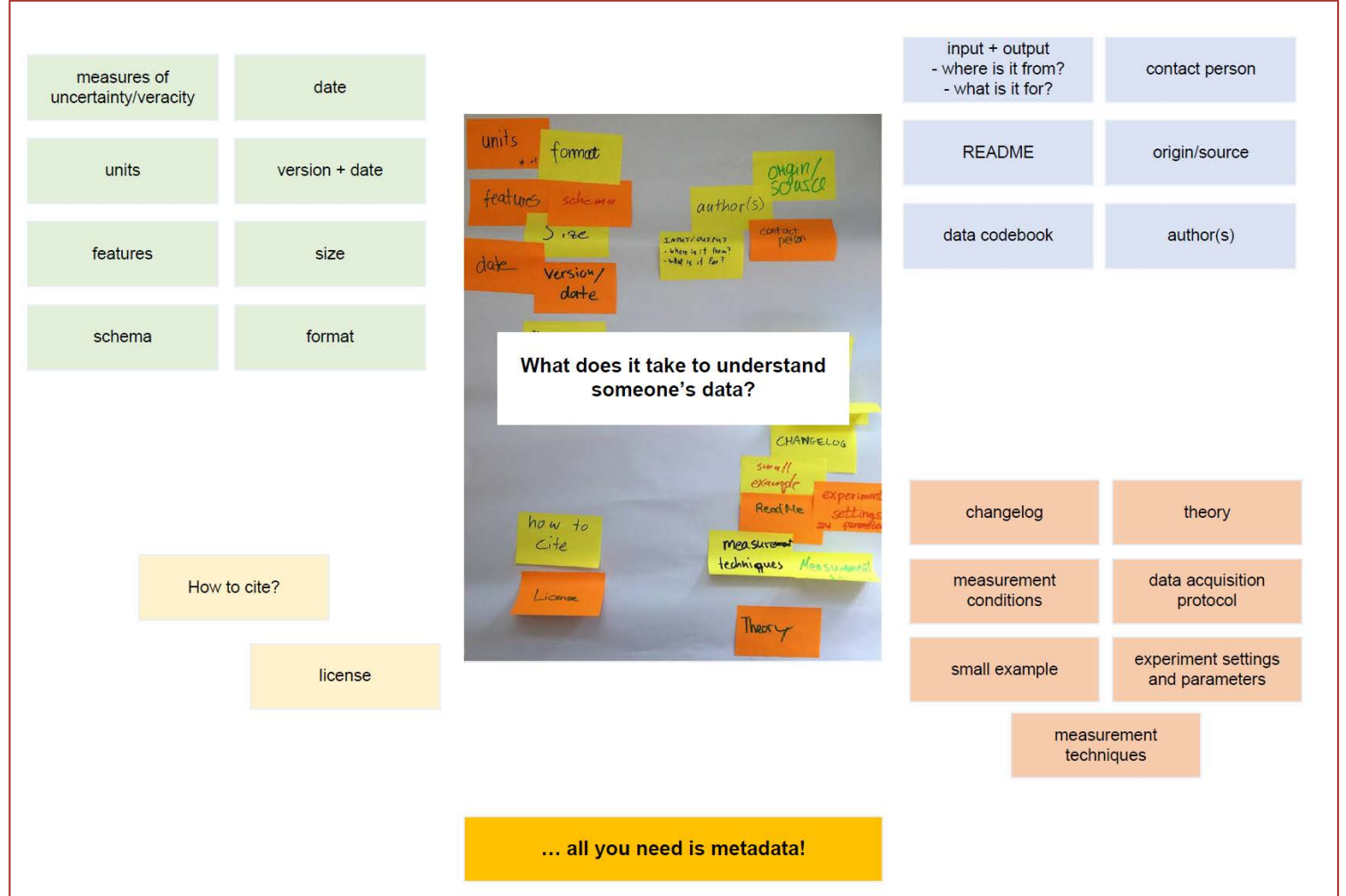
iMindMap: "What is a Mind Map" (4.9.2018)
<https://imindmap.com/how-to-mind-map/>

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



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

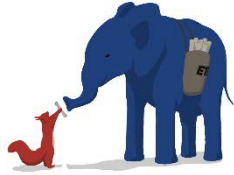
<http://www.dcc.ac.uk/resources/metadata-standards>

Practical Data Management

- right after this session!
- led by Anna Krystalli



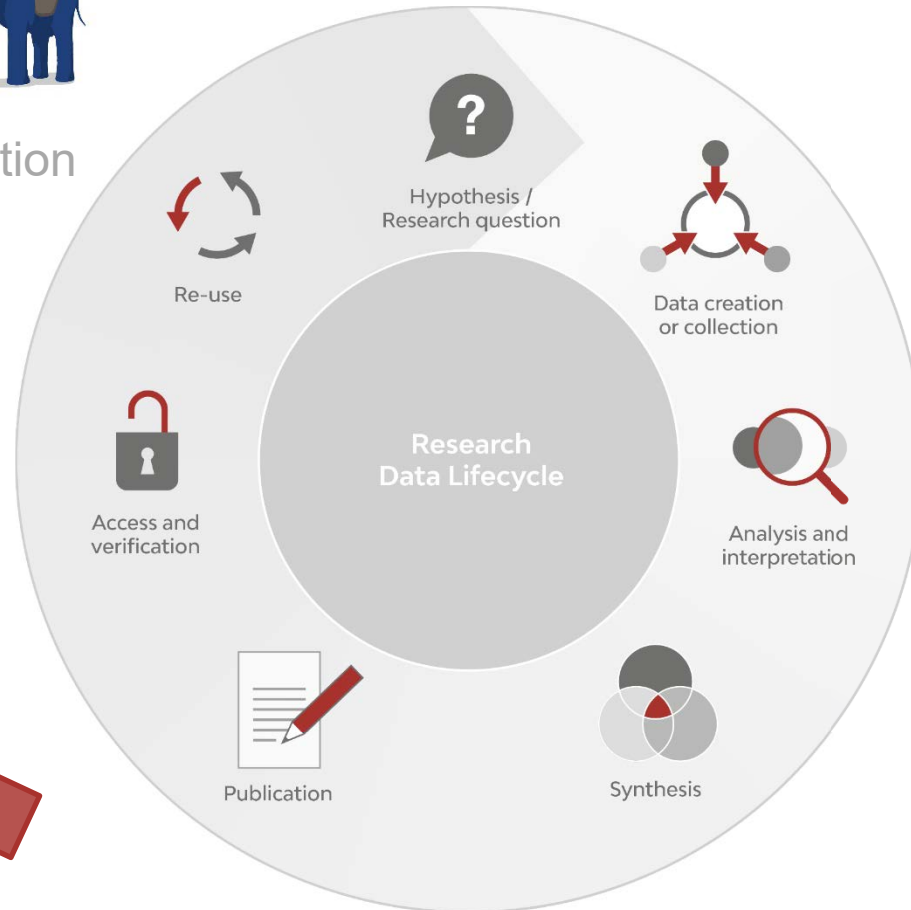
What's next...



Long-term preservation



What is data management and why should it concern you?



Data sharing



Regulations, intellectual property, privacy and access rights

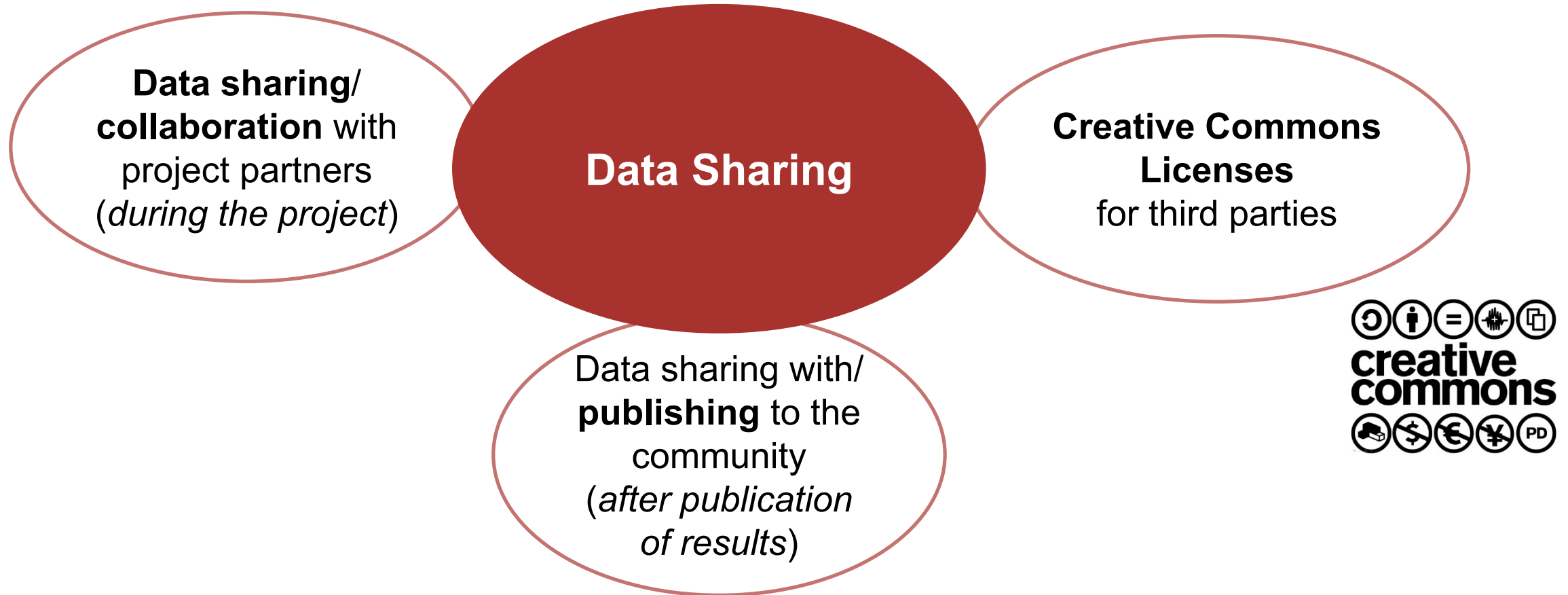


Data Management Planning

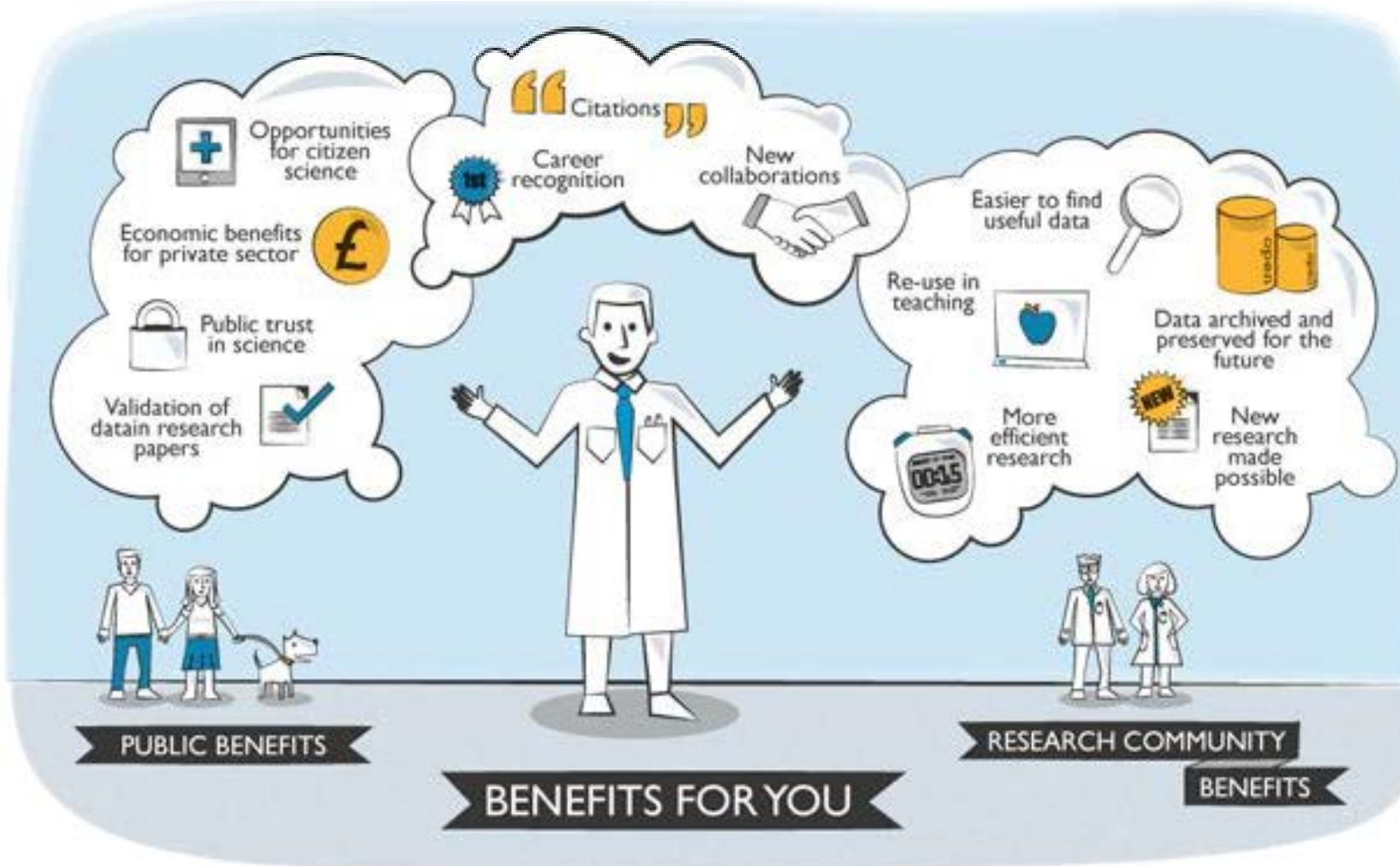




Data sharing



Benefits of data sharing



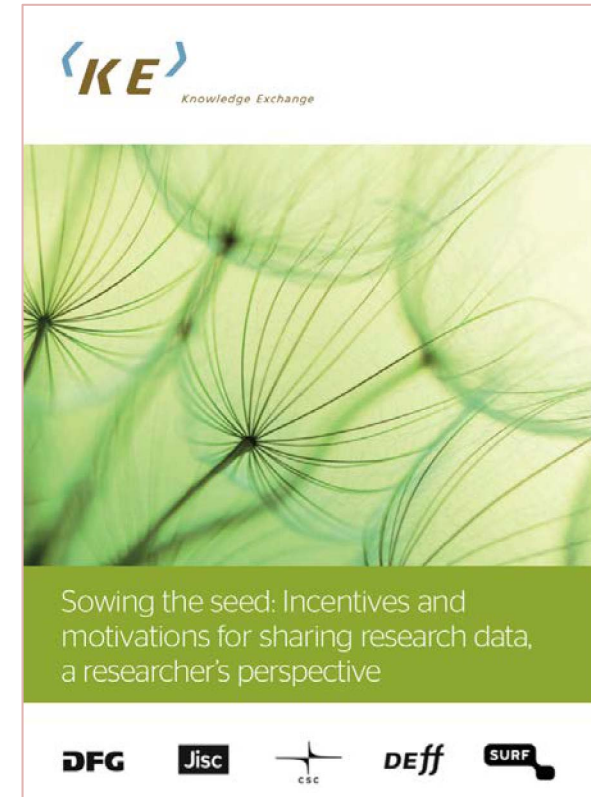
© Neil Chue Hong <http://dx.doi.org/10.6084/m9.figshare.942289> (4.9.2018)



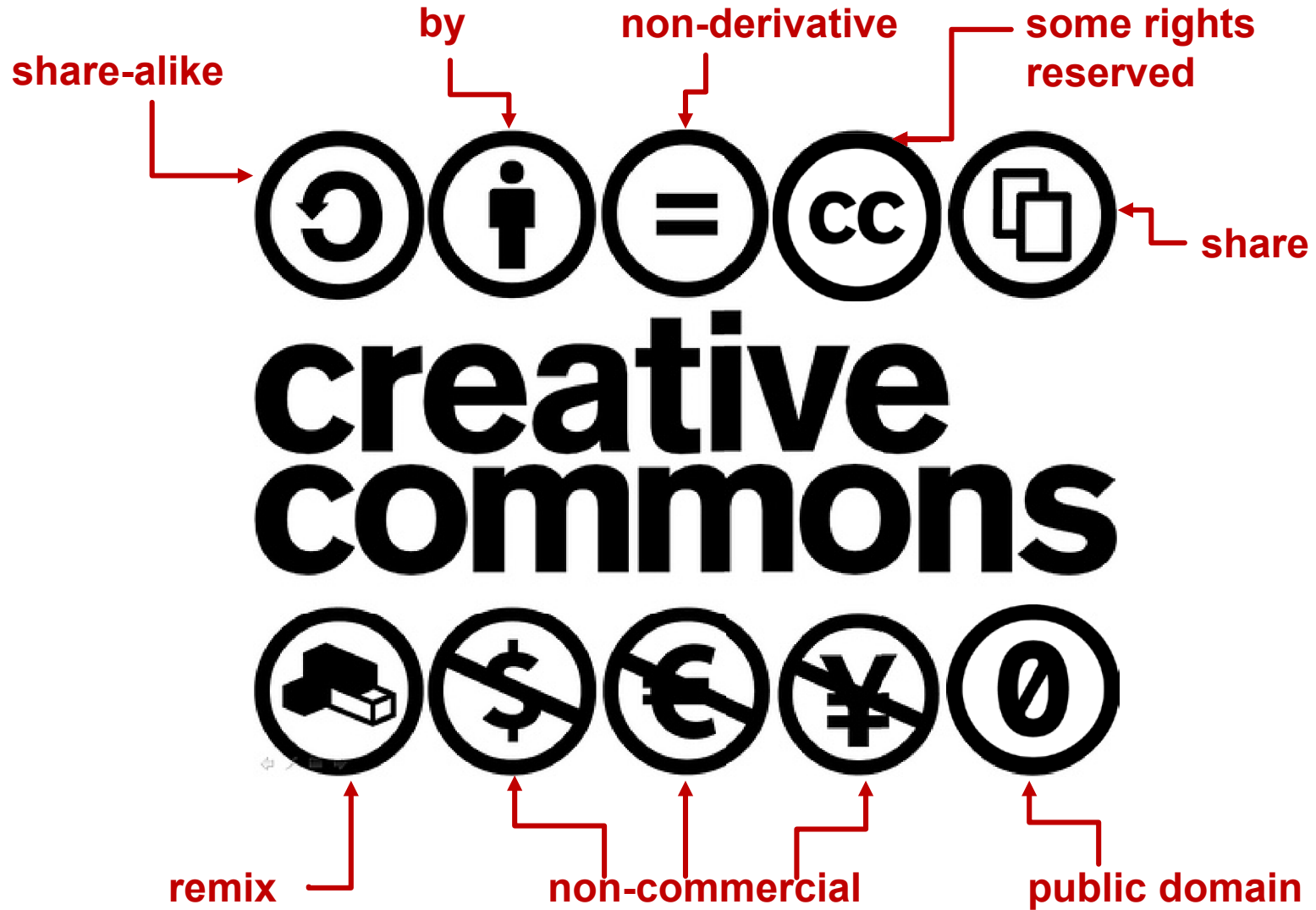
Benefits of Open Data: Impact and longevity

“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



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

CREATIVE COMMONS LIMITATIONS



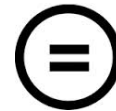
NC Non-Commercial

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



www.openaire.eu/search/data-providers



zenodo.org



datadryad.org



figshare.com*



**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



The image displays two screenshots of the re3data.org website. The left screenshot shows the homepage with a search bar and news items. The right screenshot shows the 'Browse by subject' page with a circular subject hierarchy diagram.

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Browse by subject

Graphical Text

click to zoom into subjects or to select a bottommost subject in the hierarchy as filter for the re3data search page
ctrl + click on a top subject to select it as filter

2,000 Data Repositories and Science Europe's Framework for Discipline-specific Research Data Management

Three new DOI Fabrica features to simplify account management

One step closer to instant DOI search results



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)

ETH zürich Research Collection

Suche

Ergebnisse

- Soliton mododlocking with normal dispersion via adiabatic excitation of $\chi(2)$ solitons in apodized fanout PPLN devices
Phillips, Chris R., Mayer, Alvin S., Klenner, Alexander, et al. (2015)
Proceedings of the 2015 Conference on Lasers and Electro-Optics (CLEO)
Conference Paper

ETH zürich Departement Bau, Umwelt und Geomatik | Institut für Bau- und Infrastrukturmanagement

Infrastrukturmanagement

Publikationen

Publikationen 1-10 von 137

- GPU-Accelerated Rendering Methods to Visually Analyze
Magnus Heitzler, Juan Carlos Lam, Jürgen Hackl, Bryan T. Adey and Lorenz Hurni
- A Process to Enable the Automation of Road Asset Management
Bryan T. Adey
- Multi-Layer Spatially Embedded Random Network Model for Complex Transportation Networks
Jürgen Hackl and Bryan T. Adey
- A Methodology to Determine Optimal Intervention Programs for Multiple Urban Infrastructure Networks with a Consistent Representation of Service Levels
Clemens Kellhauser and Bryan T. Adey

Annual Academic Achievements (AAA)

Wichtige Publikationen wurden im Kalenderjahr 2015 von Mitgliedern der Gruppe veröffentlicht

Publ.	Autoren	Titel	Erschienen in	Publikationsart
pub-15001	Döhrenwehde, Alexander, Walter	Some Statistics on Semimetric Components	Journal of the American Chemical Society, Volume 137, 2015-2025	Zeitschriftenbeitrag
pub-15002	Emeth, S., McCue, L.L., Beecher, C.J., Xu, D., Chen, C.Y., Zeng, S.L.	522-01: A Benzocyclohexane with Unusually Flexible 24 Ring System	Dalton Transactions, Volume 44, 4288-4295	Zeitschriftenbeitrag
pub-15003	Pinar, A.B., McCullen, L.B., Beecher, C.J., Schmidt, J., Huang, S.-J., Davis, R.E., Zeng, S.L.	Location of Ge and other framework species in the zeolite ITQ-24	Applied Physics Letters, Volume 106, 133605	Zeitschriftenbeitrag
pub-15004	Wolke, Janusz, Kaczka, Paweł, Stępnik, Krzysztof	Periodically distributed objects with paraxial optical diffraction pattern	Solar Energy, Materials and Solar Cells, Volume 138, 129-138	Zeitschriftenbeitrag
pub-15005	Chen, Shao, Anagnostou, Stavros, Vassilios, Vassilios, Lisciani, Michael, Christakos, Loukas, Burg	High-efficiency quantum sensing for solar photovoltaics	Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography, Volume 71, 133-140	Zeitschriftenbeitrag
pub-15006	Shenker, Walter, Döhrenwehde, Alexander	High-resolution X-ray diffraction of a composite crystal in a five-dimensional lattice	Surface & Coatings Technology, Volume 283, 24-26	Zeitschriftenbeitrag
pub-15007	Witz, G., Schiller, F., Steiner, W., Bachmann, S., Essmann, R.P.	High-temperature interaction of alpha-irradiated porous coatings with Ca-Mg-O-2015/02 (2015) Journal	Physical Review B, Volume 91, 194101	Zeitschriftenbeitrag
pub-15008	Garza, Lorenz, Mariani, César, Fabian, Philipp, Haas, Michael, Bury, Shalini Parth, Parth, Shree Prasad, Binu, Binu, Binu, Binu	Long-range modulation of a composite crystal in a five-dimensional lattice	Zeitschrift für Kristallographie, Volume 230, 300-309	Zeitschriftenbeitrag
pub-15009	Ge, Peng, Wang, Wei, McCullen, Lorenz, Beecher, Christian, Zeng, Jianping	On the relationship between walls and channel systems in high silica zeolites with the butterfly projection	Journal of Synchrotron Radiation, Volume 22, 70-70	Zeitschriftenbeitrag
pub-15010	Trank, P., Degen, C., Hahn, M., Patsche, F., Pralle, S.J., Radonić, V., Seifert, S.A., Völkl, D.A., Bruneau, C.	Beam-line specific site correlations for Pt(L2,3) detection	Journal of Applied Crystallography, Volume 48, 104-104	Zeitschriftenbeitrag
pub-15011	Dreier, Catherine, Smith, S.M., Beecher, Christian, Tamara, Robinson, Patricia, Philip, Akiba, Rabele, McCullen, Lorenz	Serial synchrotron crystallography for materials science with Synchrotron	Acta Crystallographica A, Volume 71, 133-140	Zeitschriftenbeitrag
pub-15012	Castro, Pascal, Wolke, Janusz	Generalized Fermion ring as a quantifier for designed paracrystal structure analysis	Zeitschrift für Kristallographie, Volume 230, 18-43	Zeitschriftenbeitrag
pub-15013	Schäfer, Alexander, Walter	Single-phase high-entropy alloy - an overview	Zeitschrift für Kristallographie, Volume 230, 100-109	Zeitschriftenbeitrag
pub-15014	Witz, G., Steiner, W., Bachmann, S., Essmann, R.P., Parth, Shree Prasad, Binu, Binu, Binu, Binu	On the relationship between walls and channel systems in high silica zeolites with the butterfly projection	Journal of Applied Crystallography	Tagungsbeitrag
pub-15015	Emeth, Janusz, Kaczka, Paweł, Stępnik, Krzysztof, Olszewski, Marcin	Fast structure of CaB2O7 influenced on diffuse scattering data with the SPCoat XRD method	Journal of Applied Crystallography	Tagungsbeitrag
pub-15016	Aljanabi, Rana, Aze, B., Helgeson, Kenneth, Ferguson, Stephen J.	The effect of water on the mechanical properties of soluble and insoluble ceramic ceramics	Journal of the Mechanical Behavior of Biomedical Materials, Volume 11, 39-48	Zeitschriftenbeitrag
pub-15017	Su, Lei, Dong, Jianping, McCullen, Lorenz B.	Structural Synthesis and Structure of a New Layered Strontium Phosphate	Organic Chemistry, Volume 54, 7953-7958	Zeitschriftenbeitrag

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

OXFORD ACADEMIC University of Zurich Sign In Register

European Review of Agricultural Economics EAAEP Foundation

Issues JEL More Content Submit Purchase About

Potential effects of the income stabilisation tool (IST) in Swiss agriculture

Nadja El Benni, Robert Finger, Miranda P.M. Meuwissen

European Review of Agricultural Economics, Volume 43, Issue 3, 1 May 2016, 475–502, <https://doi.org/10.1093/erae/jbv023>

Published: 30 May 2016 Article history

Views PDF Cite Permissions Share

Abstract

We extend the existing literature on the income stabilisation tool (IST) by investigating the influence of farm and farmers' characteristics on potential indemnification applying double-hurdle models on a rich panel data set on Swiss farms. We find more likely and higher indemnifications for part-time and low-income farmers. Thus, the IST might become a new transfer instrument hampering the structural change. Even though the estimated costs of the IST are low compared with the current direct payment level in Switzerland, both policy measures are partly substitutes with respect to income risk reductions implying more frequent and higher indemnifications through the IST if direct payments are reduced.

Keywords: income stabilisation tool, double-hurdle model, Switzerland

JEL classification: D8, G22, Q12, Q18

1. Introduction

Open Access version

ETH zürich Anmelden Hilfe English

Research Collection Suche

Start Journal Contributions Journal Article Trefferanzeige

Browsen Organisationseinheiten Publikationstypen Autoren Publizieren Meine Publikationen Statistik Downloads nach Land Beliebteste Publikationen Beliebteste Autoren

Potential effects of the Income Stabilisation tool (IST) in Swiss agriculture

Open access

Autor(in) El Benni, Nadja Finger, Robert Meuwissen, Miranda

Datum 2016-05-30

Typ Journal Article

Altmetrics

Im Viewer öffnen Full text (accepted version) Download Full text (accepted version) Rechte / Lizenz In Copyright - Non-Commercial

Persistenter Link <https://doi.org/10.3929/ethz-b-000000000>

Publikationsstatus published

Externe Links <https://doi.org/10.1093/erae/jbv023> Voltext via SFX

Zeitschrift / Serie European Review of Agricultural Economics

Potential effects of the Income Stabilization Tool (IST) in Swiss agriculture

Abstract

We extend the existing literature on the Income Stabilization Tool (IST) by investigating the influence of farm and farmers' characteristics on potential indemnifications applying double-hurdle models on a rich panel data set on Swiss farms. We find more likely and higher indemnifications for part-time and low-income farmers. Thus, the IST might become a new transfer instrument hampering the structural change. Even though the estimated costs of the IST are low compared to the current direct payment level in Switzerland, both policy measures are partly substitutes with respect to income risk reductions implying more frequent and higher indemnifications through the IST if direct payments are reduced.

Keywords: Income Stabilization Tool, double-hurdle model, Switzerland

1 Introduction


Income risks in agriculture are of increasing importance and have received particular attention by policy makers and other stakeholders in the last years (e.g. Agreste et al., 2012; Binda, Diaz-Curiel et al., 2008; Calvo et al., 2007; EC, 2001; Haase et al., 2004; Meuwissen et al., 2008; Vercauteren, 2007). To a large extent, this increasing interest has been caused by characteristics of a higher exposure of farmers to production and market risks (Crisan, 2011; Fox et al., 2013). For instance, world market price volatility has changed over recent decades (FAO, 2008) and some experts believe that the volatility is expected to rise and could prove since late 2006 will remain (Gomez and O'Connor, 2009; Janssens et al., 2010; Gilbert and Kilgus, 2010; Sauer, 2009). This increasing price volatility is also of increasing importance for European farmers (EC, 2011; van Wassen et al., 2011; Driesselt et al., 2002). This is due to the fact that most European countries have switched from market-based support to decoupled direct payments which not only decreased prices but also expose producers to (higher) world market price volatility. Besides changes in (global) price volatility, also input prices are likely to be affected (Chang and Gohmert, 2006; Meuwissen et al., 2003; FOAG, 2009; Yang et al., 2004; Berg, 2002; Stahl, 2003; Schickel and Sauer, 2005). Perceptions of increasing agricultural production risks are closely connected to changes in climate conditions and are expected to be of even higher relevance in the future (see e.g. Olesen et al., 2011, for an overview). Moreover, also agricultural policy has influenced developments in production risks. For instance, cross-compliance regulations and agro-environmental schemes have resulted in a one-straw-cow-estate production system (see e.g. El Benni and Lohmann, 2016; Finger and El Benni, 2013). Even though low-intensification systems may lead to more shock-resilient and thus less volatile systems (e.g. Schuster et al., 2002), it has




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

LevelDynamics_Scilab/Xcos_v541_sources.7z [↗](#)



Open access



Download
[LevelDynamics_ScilabXcos_v541_sources.7z \(Unknown, 25.32Mb\)](#) ↓

Rechte / Lizenz
[Creative Commons Attribution-ShareAlike 4.0 International](#) ↗

Abstract
 Source files of models used in ' Free Surface Water Flow - Modeling, Simulation and Control ' developed in Scilab/Xcos v. 5.4.1

Persistenter Link
<https://doi.org/10.3929/ethz-b-000175543>

Beteiligte
 Kontaktperson: Glattfelder, Adolf Hermann

Verlag
 ETH Zurich

Software
 Scilab/Xcos

Organisationseinheit
 02650 - Institut für Automatik (IfA) / Automatic Control Laboratory (IfA)

Zugehörige Publikationen und Daten
 Is supplement to: <https://doi.org/10.3929/ethz-b-000175527> ↗

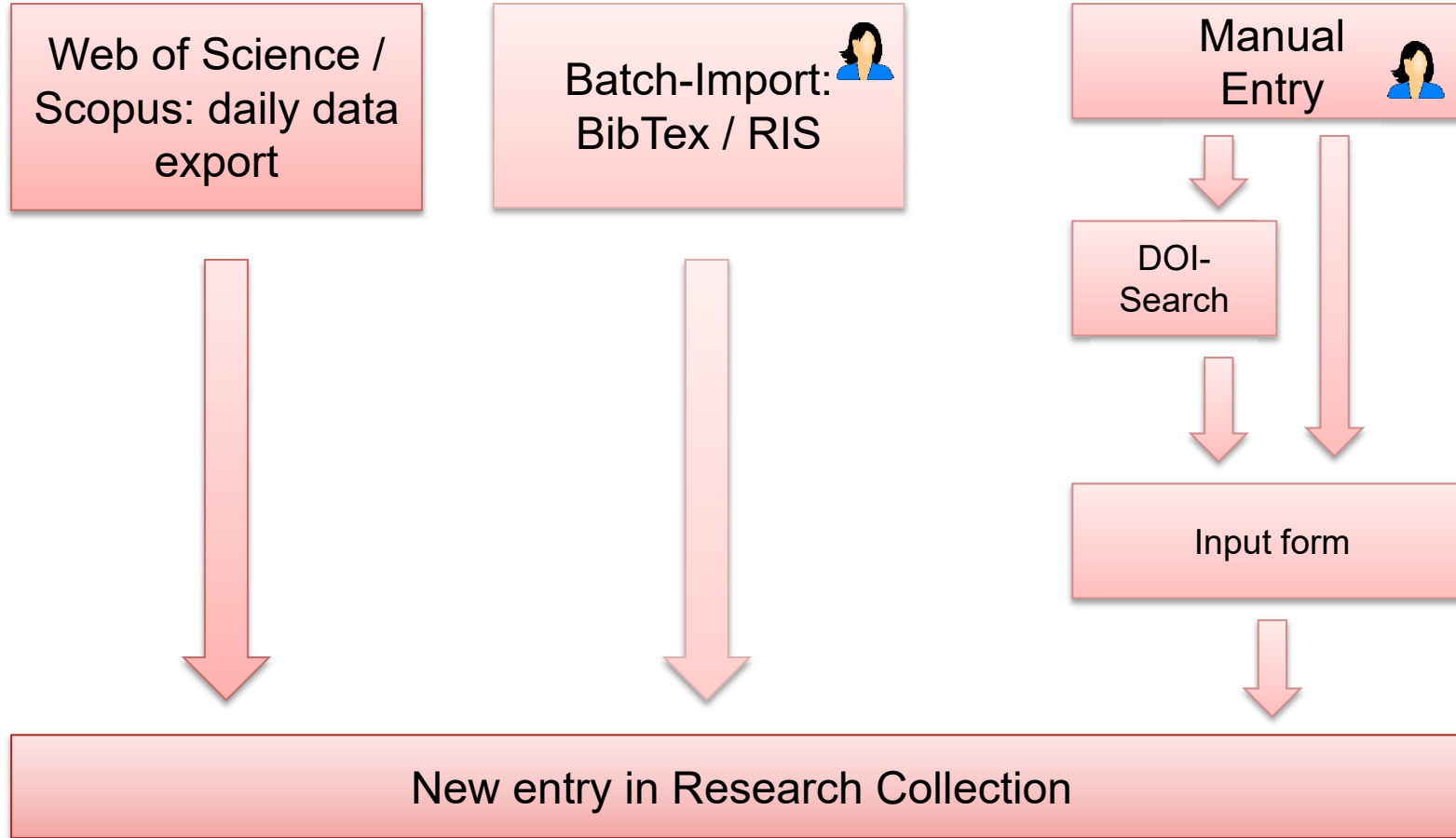
Anmerkungen
 7-zip 16.04 (64 bit) container with zero compression

Produzent(in)
 Glattfelder, Adolf Hermann

Datum
 2017-08-09

Typ
 Model

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: Advice and support by ETH Library

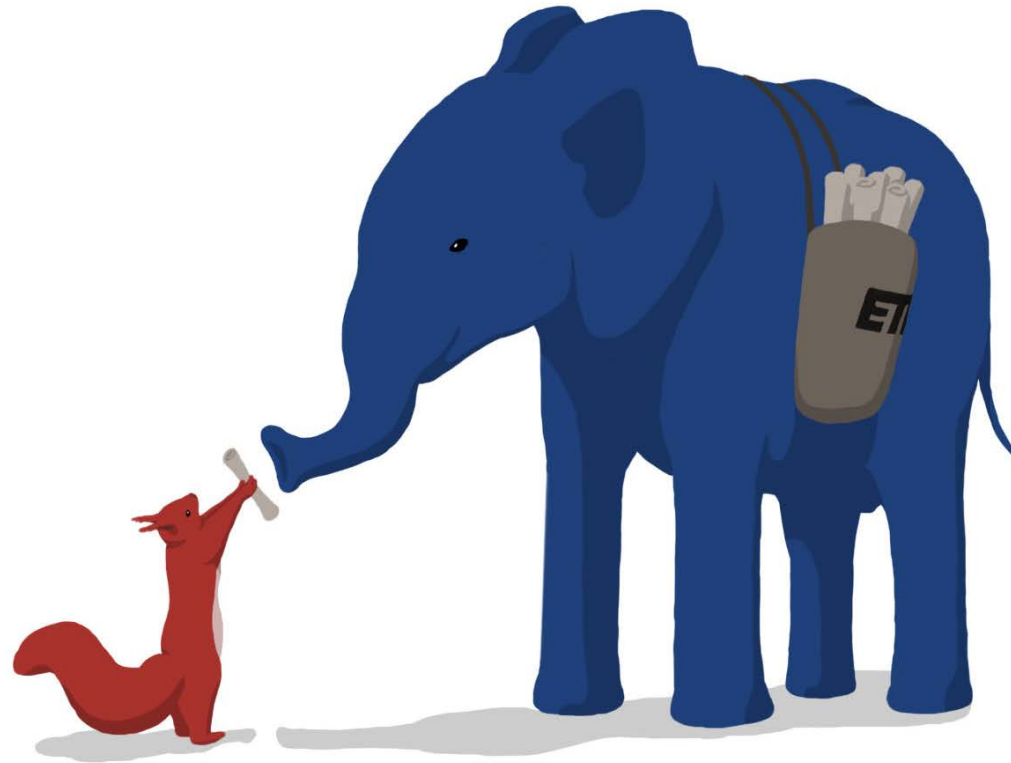
- Legal issues in Open Access publishing
- Open Access- and guidelines of research funders (SNSF, EU)
- Data management and digital curation
- ORCID support

www.research-collection.ethz.ch

Mail: research-collection@library.ethz.ch

Tel. 27 222

The screenshot shows the ETH Research Collection website. The header includes the ETH zürich logo and navigation links (Anmelden, Hilfe, English). Below the header is a search bar labeled 'Suche'. The main content area is titled 'Repositorium für Publikationen und Forschungsdaten' and includes a sidebar with navigation options like 'Browsen', 'Publizieren', and 'Statistik'. A large red arrow points from the 'Aktuell' section towards the 'Publizieren' and 'Handbuch' buttons on the right side of the page.



Long-term preservation of data

What does long-term mean?

Different time horizons and purposes



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”)

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

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

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

- **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**



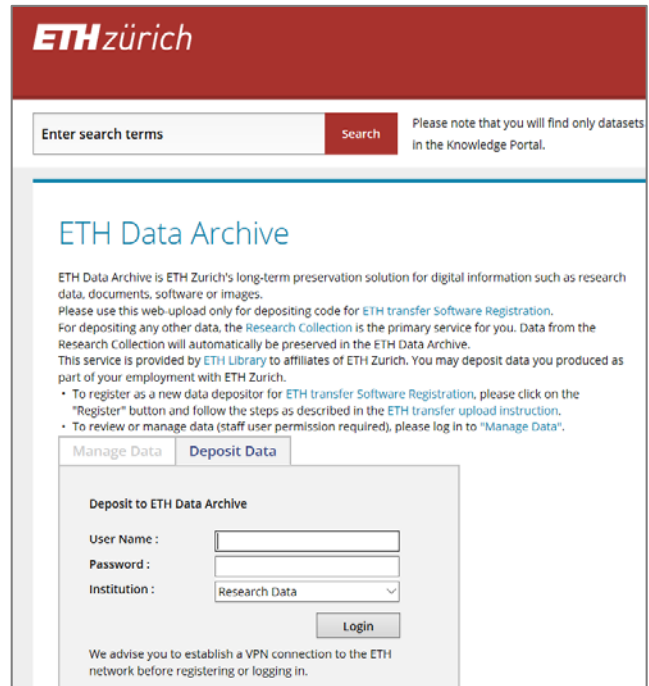
ETH Data Archive

Digital preservation solution for ETH Zurich, operated by ETH Library



Research Collection

→ automatically archiving



Heritage content from ETH University Archives and ETH Library

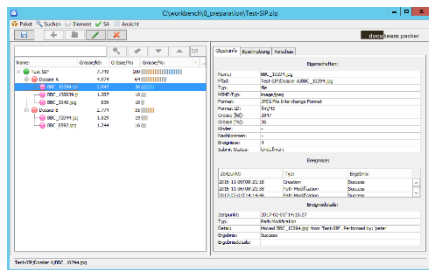
← automatically archiving

«Software Disclosure» workflow for ETH transfer

↔ software disclosure workflow

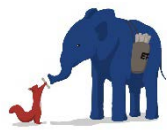


Data



Docuteam packer

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

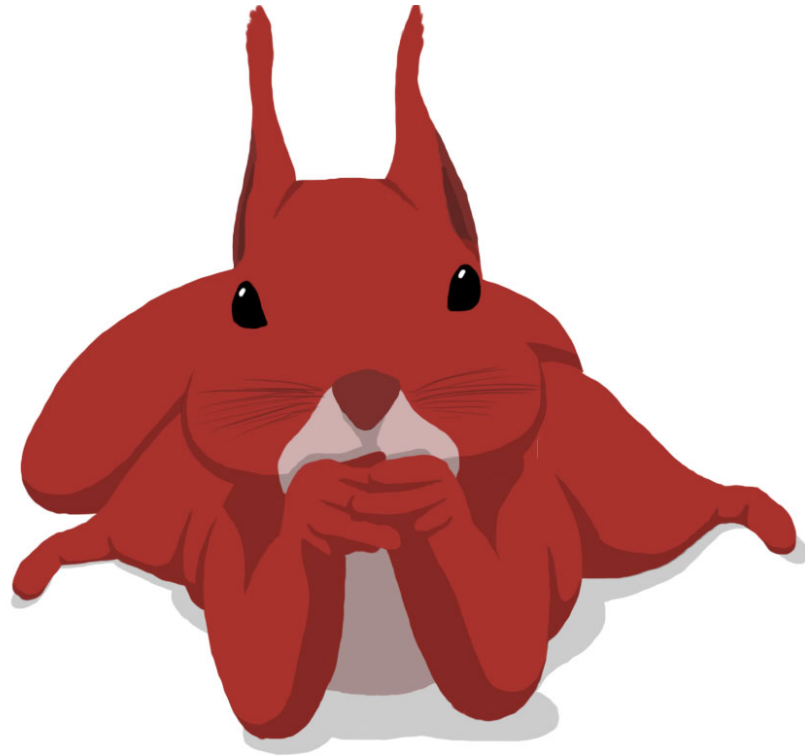
Dr. Ana Sesartic Petrus
RDM/DMP Consulting and Training
ETH Library
Rämistrasse 101
8092 Zurich
044 632 73 76
ana.petrus@library.ethz.ch



Research Data

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

We need your feedback!



Please fill out the course evaluation form – Thank you!

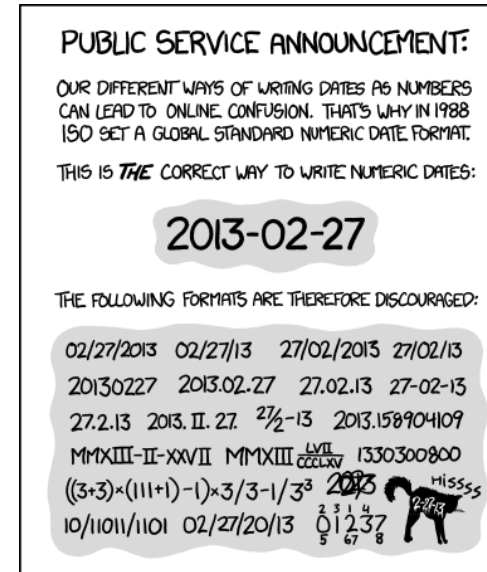
<https://www.umfrageonline.ch/s/a13b937>

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**

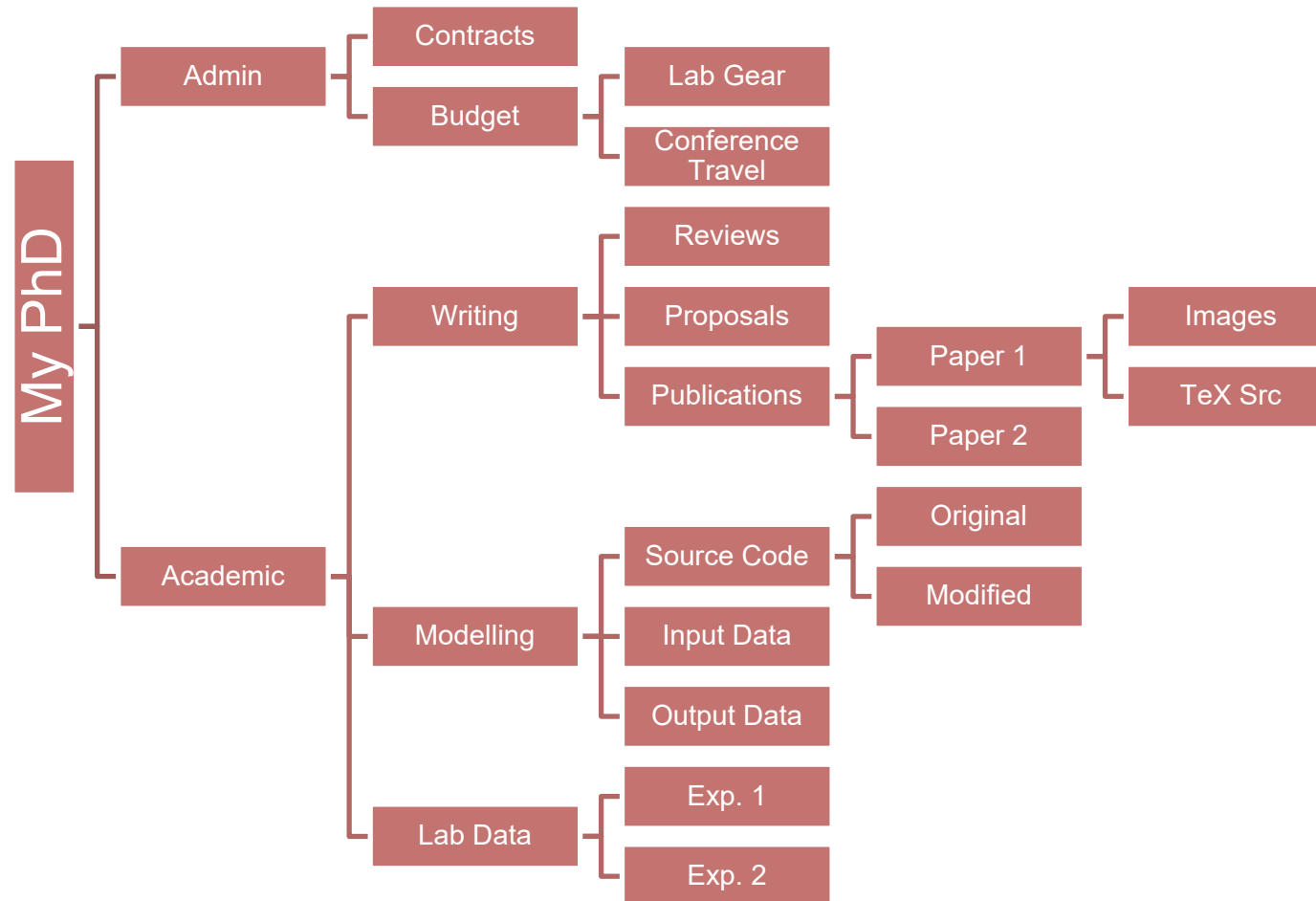


© XKCD
<https://xkcd.com/1179/>

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

**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.

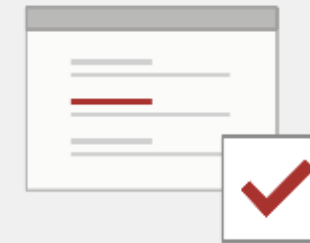
**4. Development of innovative applications and services**

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

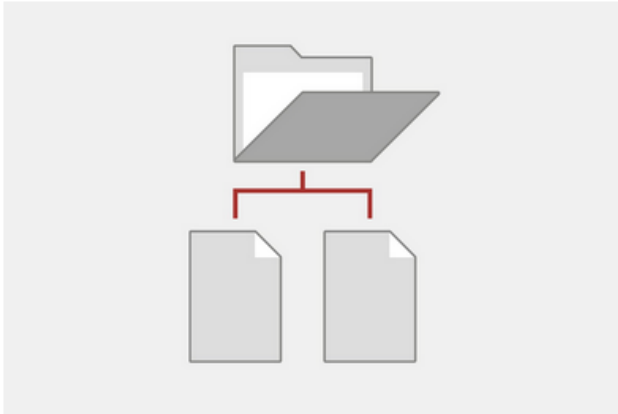
**5. Creation of new business models**

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

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

**6. Traceability in search engines**

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

**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.

**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.

**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

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...

