

# Introduction to Active Research Data Management

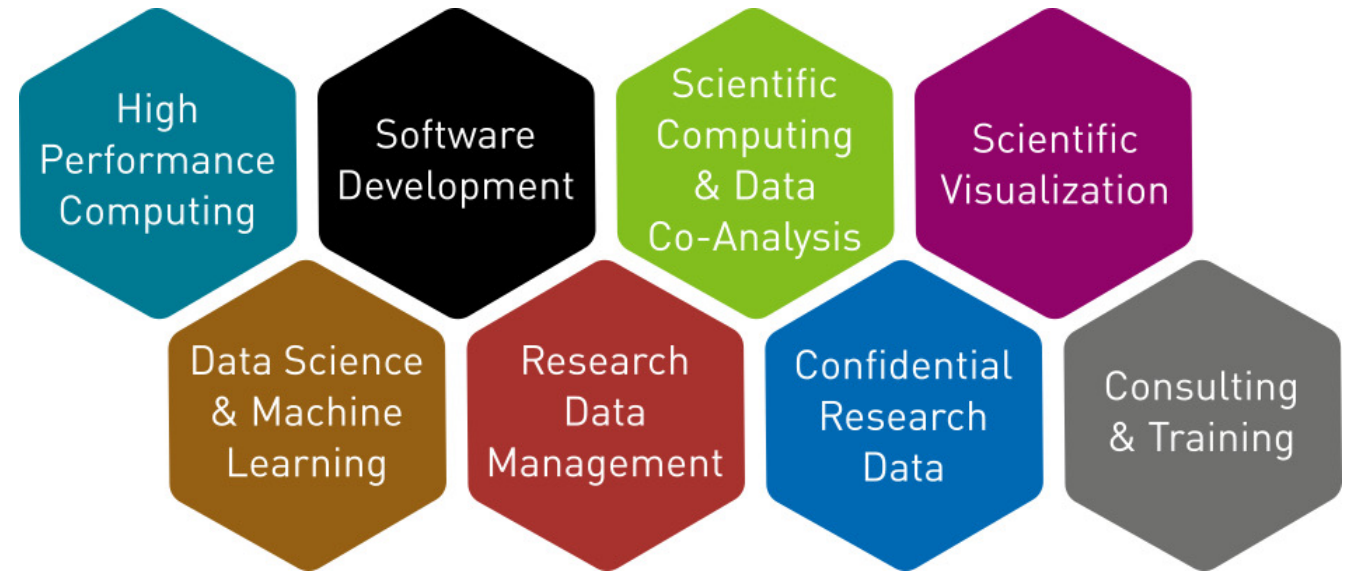
Caterina Barillari, Henry Lütcke  
*Scientific IT Services, ETH Zurich*

*October 26, 2022*





# Who is Scientific IT Services?



- ❑ A section of ETHZ IT Services
- ❑ Over 40 experts in various areas of scientific computing
- ❑ With a background in different areas of science

# Tell us a bit about yourselves

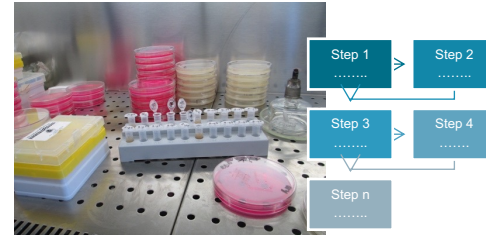
- ☐ Your affiliation
- ☐ Your research topic



# Overview of today's workshop



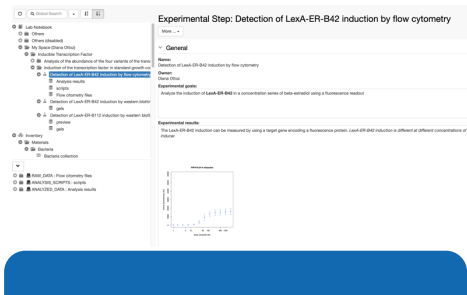
ARDM intro



Sample & protocol  
management



Management of  
Data & Metadata



ELNs



ETHZ SIS's RDM  
solution

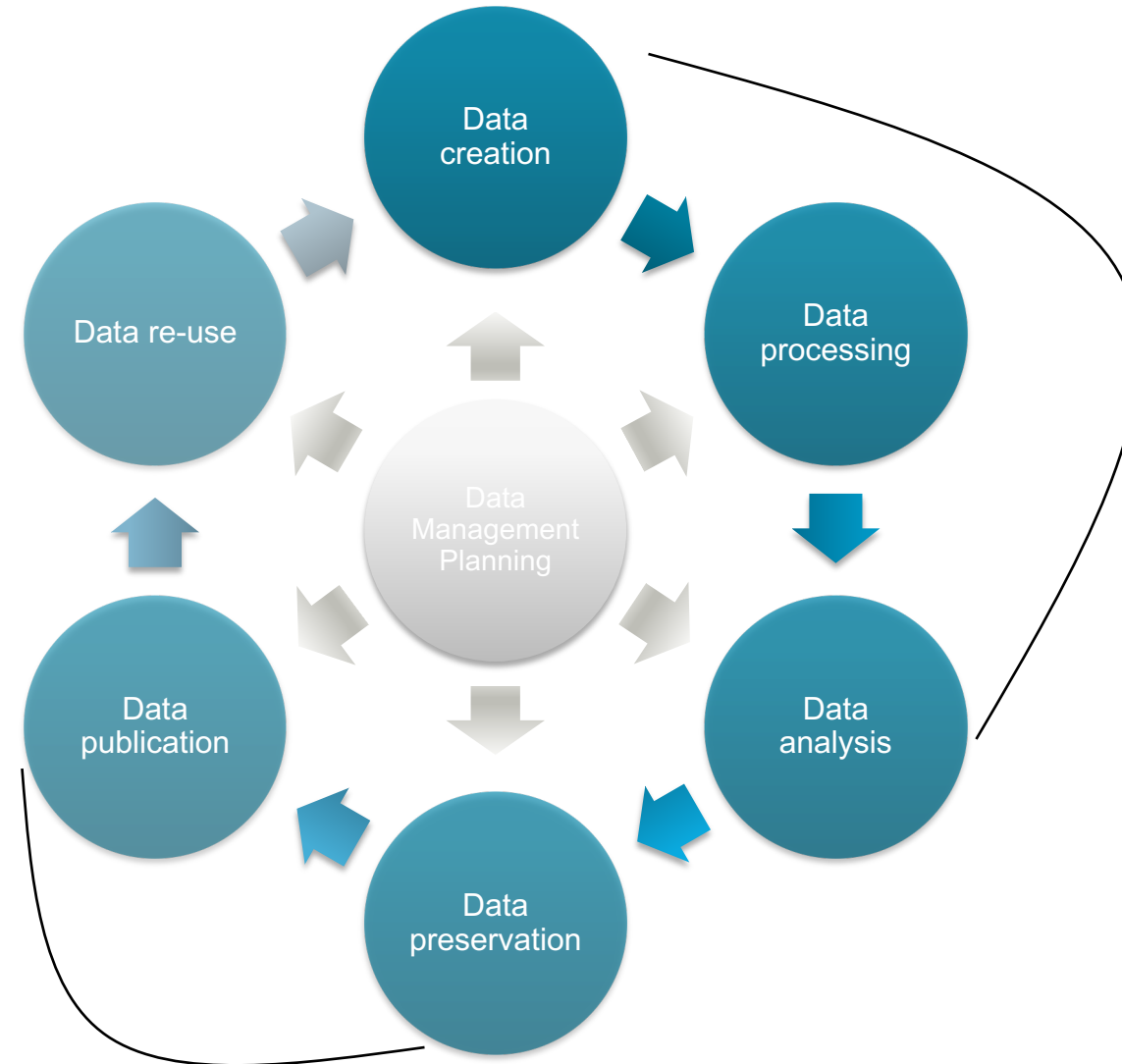




[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

# *Overview of active research data management*

# The data life cycle



**Active data management:**  
*Annotate, store, backup  
data while it is produced*

**Long term preservation:**  
*Annotate, store, backup data  
at the end of a project or after  
publication*



# Open Science and Open Research Data

Open Research Data



Requirement from funding agencies, journals, academic institutions

FAIR data

FAIR DATA PRINCIPLES

AH!



FINDABLE



Copyright © MMXV: Shep Hyken.

DO YOU  
A .X29 FILE?

ACCESSIBLE

INTEROPERABLE

REUSABLE

-HERE

Persistent Identifier;  
metadata

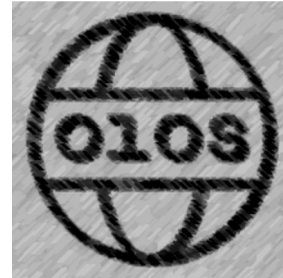
Open  
protocol

Data  
format

Metadata

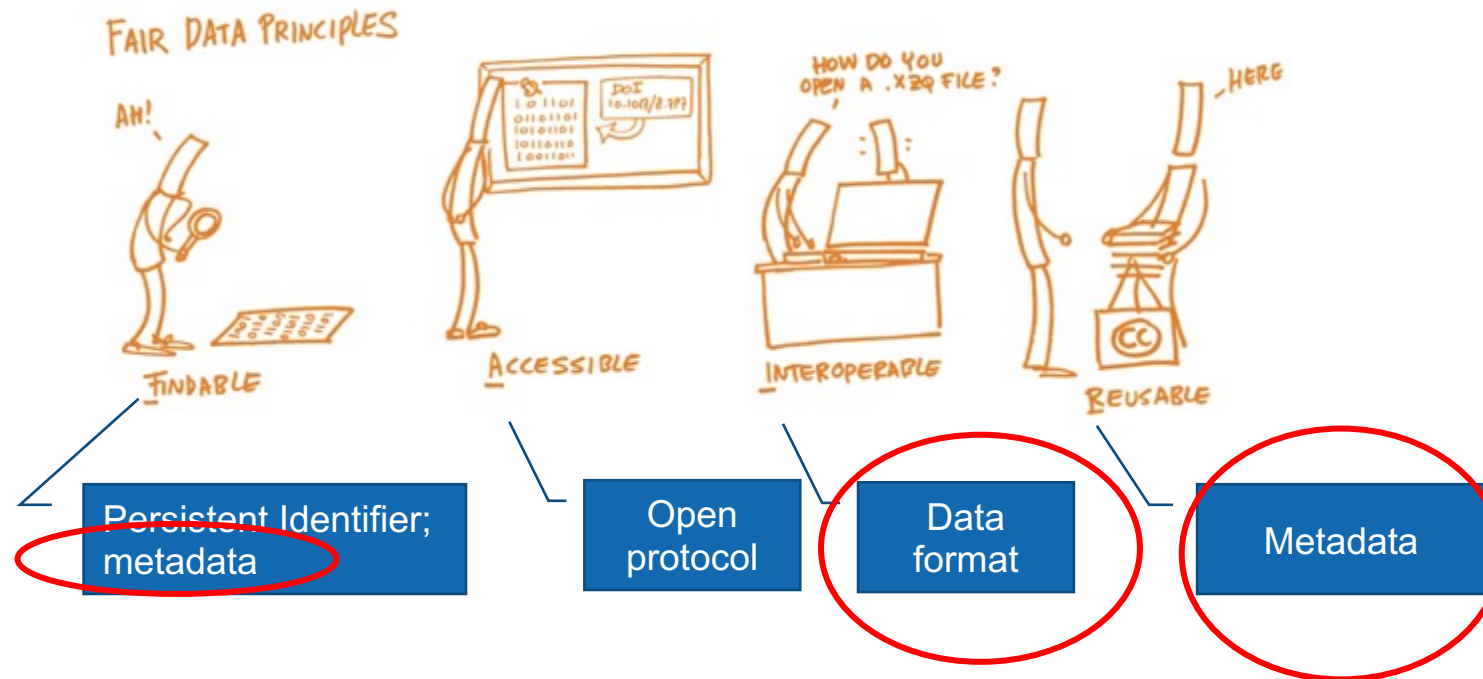
# How can we share data in a FAIR way?

A few generic [data repositories recommended by SNSF](#)

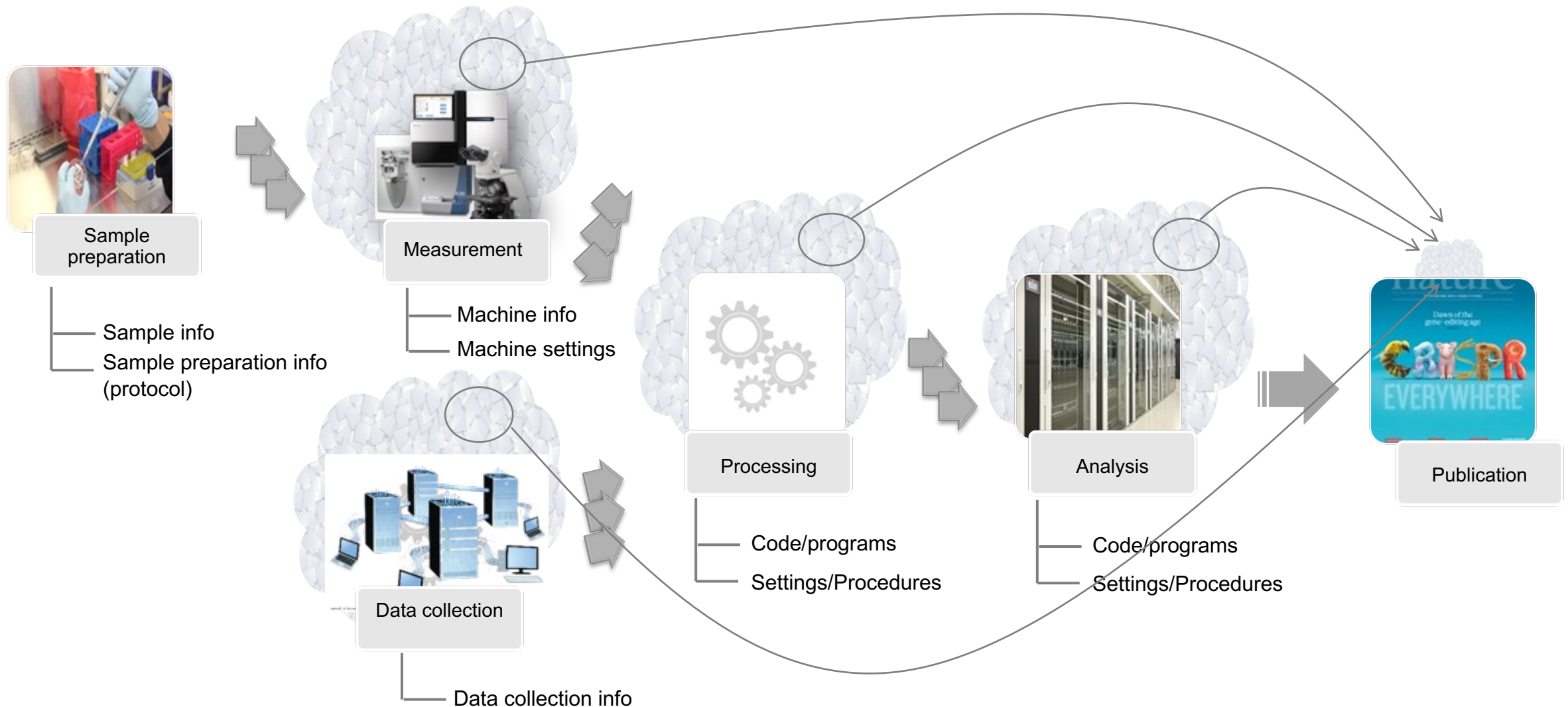




# Prepare to meet FAIR requirements when data are generated

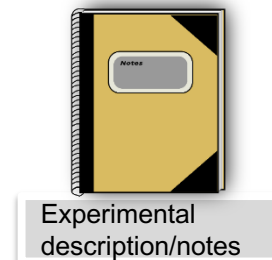
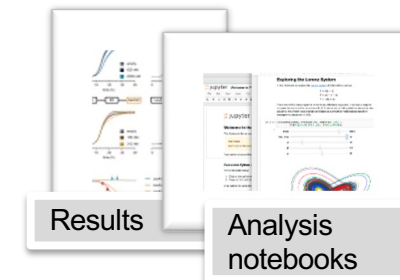
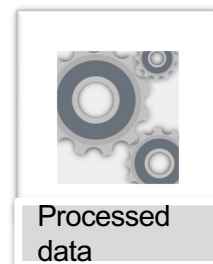
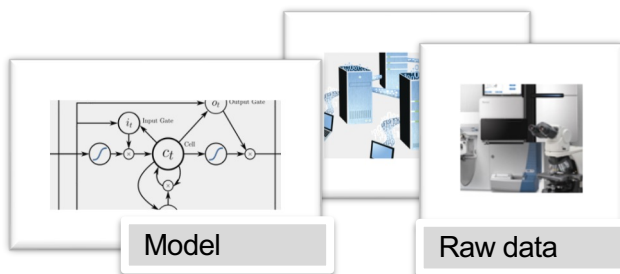
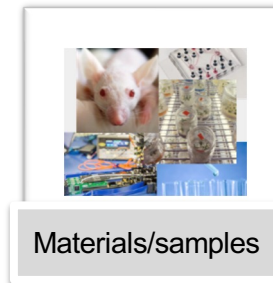
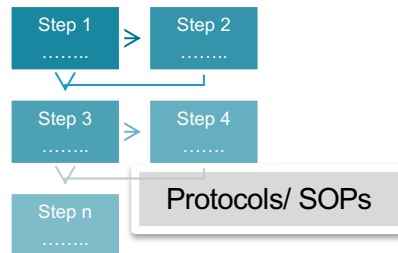


# Data and information generation during a research project

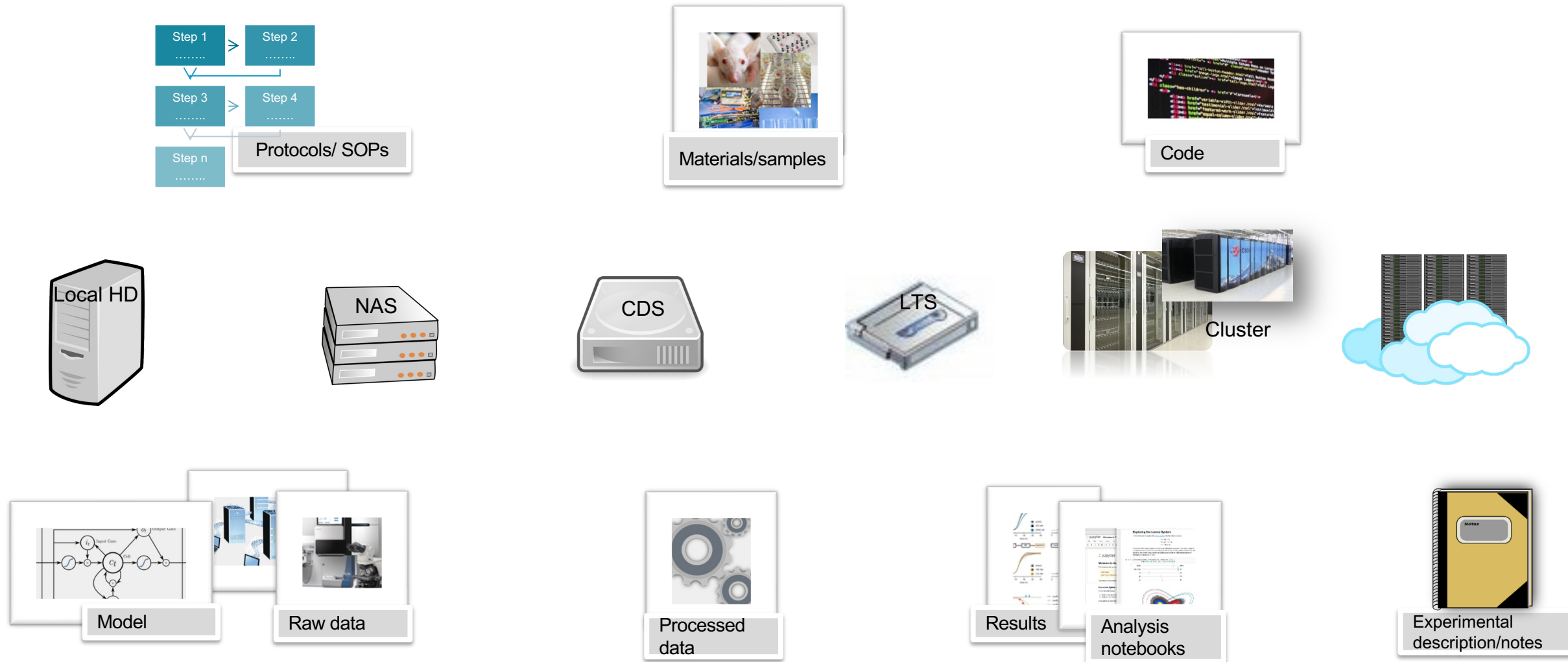




## The “*data spread*”: a common scenario in academic institutions

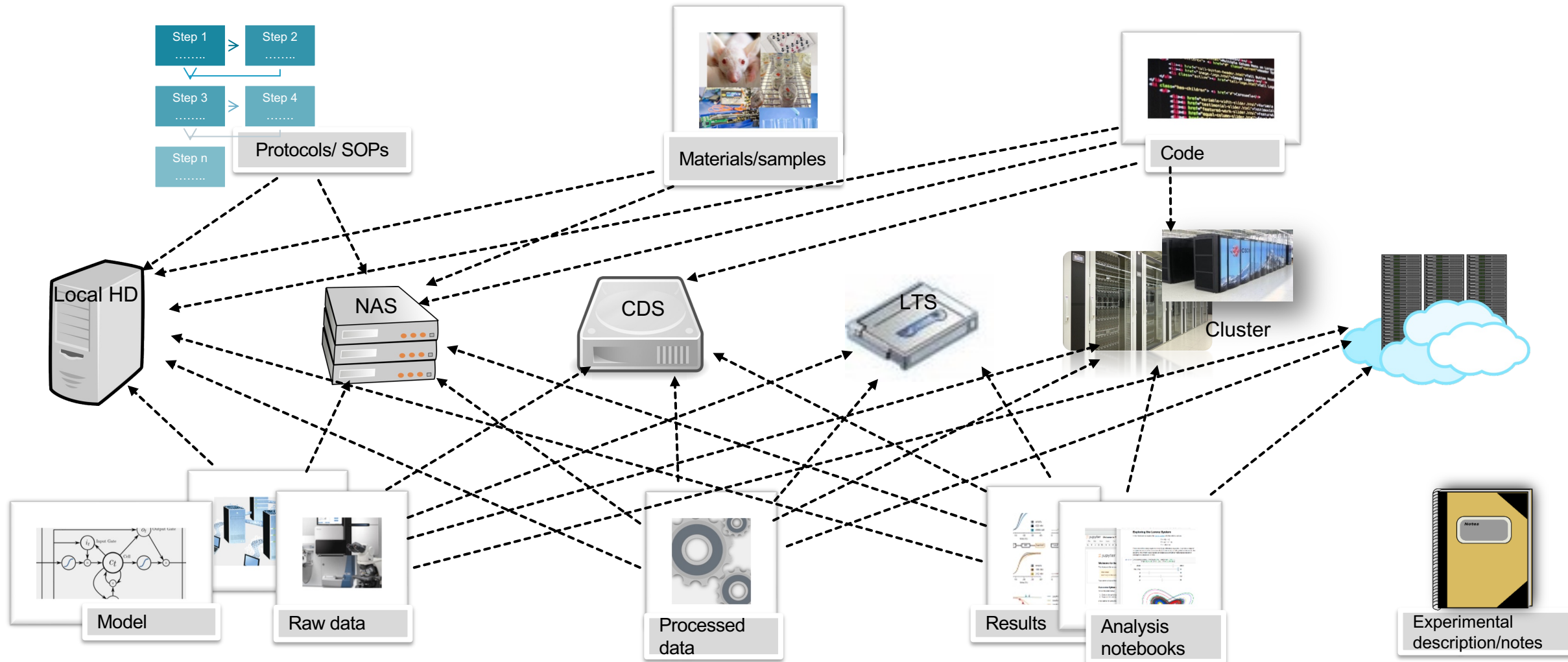


# The “*data spread*”: a common scenario in academic institutions



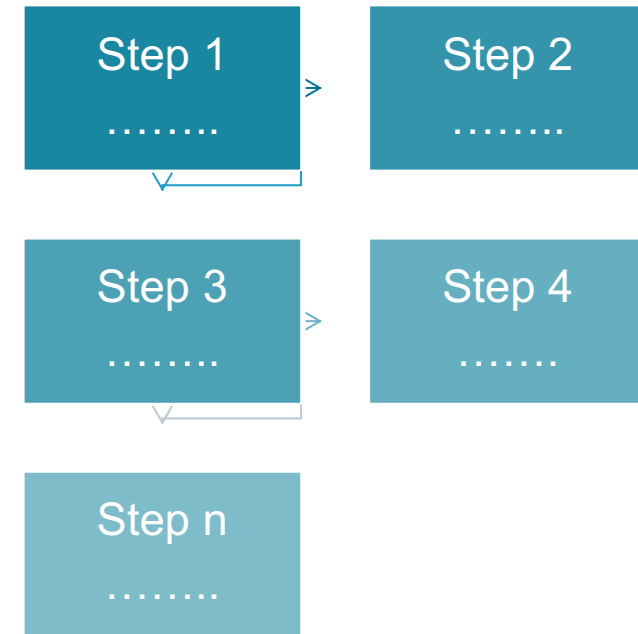
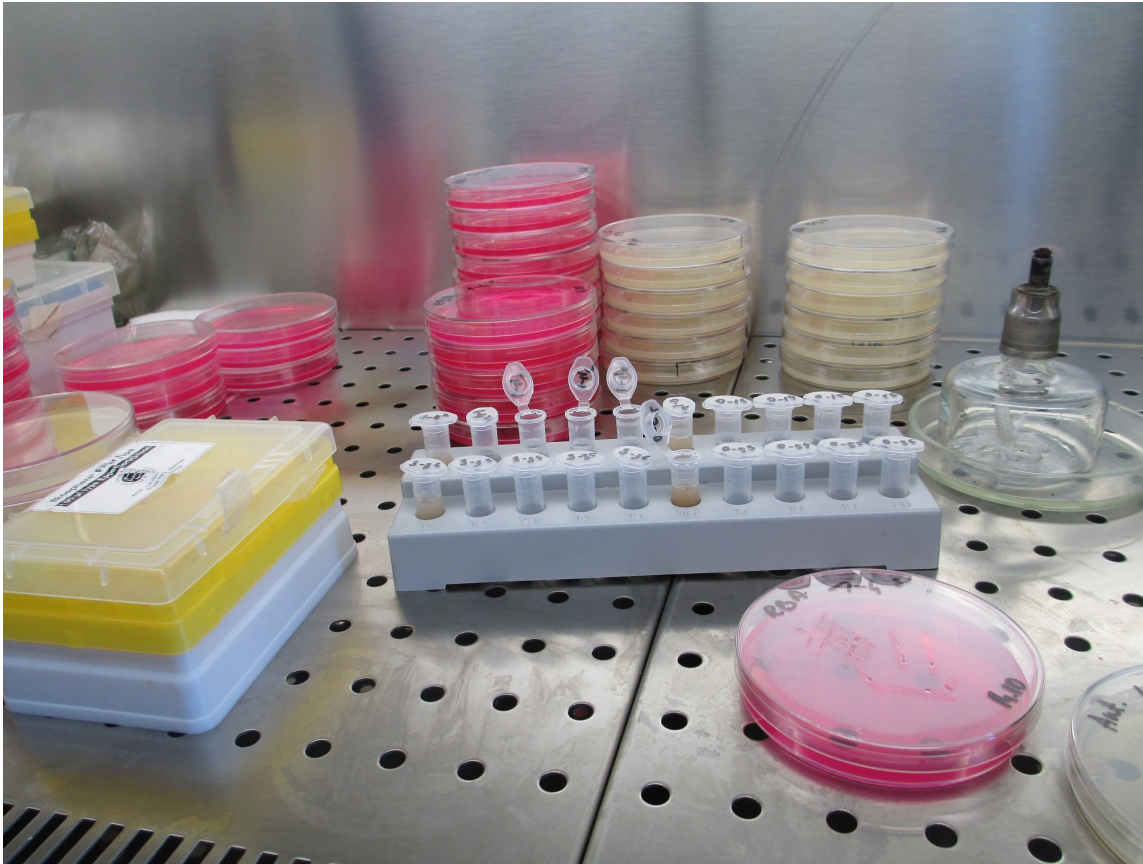


# The “*data spread*”: a common scenario in academic institutions



*How can we take care of the individual components and how can we bring things together?*





# *Management of samples and protocols*



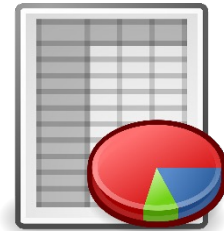
# Management of materials and samples



## What?

- ☐ Biological samples
- ☐ Chemical samples
- ☐ Materials
- ☐ Devices
- ☐ ....

## How?



### Spreadsheets

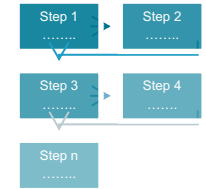
- *Not scalable*
- *No sharing*
- *No efficient search*
- *Easy to use*



### Database/ LIMS

- *Scalable*
- *Sharing*
- *Search functionality*
- *Require time for set up and maintenance*

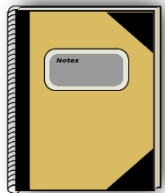
# Management of protocols



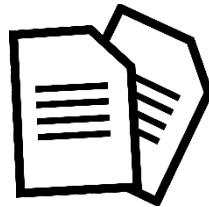
## What?

- ❑ Step by step description of procedure
- ❑ Experimental/computational parameters (e.g. temperature, time, etc.)
- ❑ Machine used (experimental)
- ❑ OS, program, version, etc. (calculation)

## How?



**Paper notebook**



**Text files**

- *Not scalable*
- *No sharing*
- *No efficient search*
- *Easy to use*

WIKI

- *Scalable*
- *Sharing*
- *Search functionality*
- *Versioning*



**Database/ LIMS**

- *Scalable*
- *Sharing*
- *Search functionality*
- *Require time for set up and maintenance*

# Laboratory Information Management System (LIMS)

- ❑ LIMS are software for managing laboratory operations:
  - **sample tracking** (*info about samples and about their storage*)
  - **sample data tracking** (*upload of data measured from samples, e.g. sequencing data, NMR, MS, etc.*)
  - **protocol management** (*info about sample preparation/handling*)
- ❑ LIMS were first used in companies for tracking the growing number of samples.
- ❑ LIMS were originally stand-alone solutions, which had to be integrated with other solutions such as ELNs. Nowadays **ELNs and LIMS are often combined** in one platform.



# Wikis at ETH

- ❑ ETH ITS provides wiki services:



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

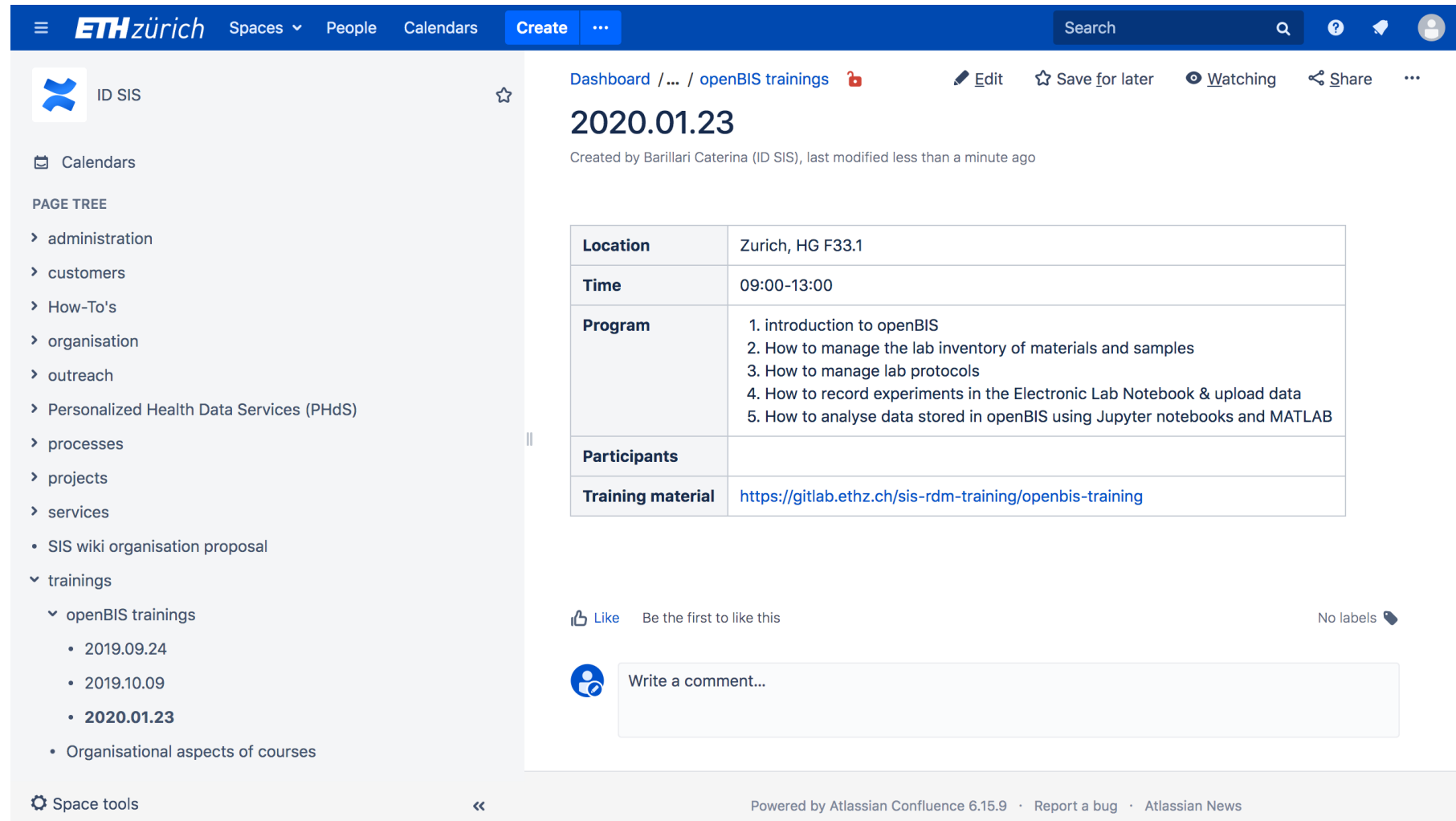


<https://gitlab.ethz.ch>

Size of Confluence Space/ git repository	Yearly price
<2 GB	free
<10 GB	300 CHF
<50 GB	1000 CHF

- ❑ Some departments and/or institutes host their own wiki. Always contact your ISG for info.

# Example of confluence wiki use in SIS



The screenshot shows a Confluence wiki page for the date '2020.01.23' within the 'ID SIS' space. The page is titled '2020.01.23' and was created by Barillari Caterina (ID SIS), last modified less than a minute ago. The page content is a table with the following details:

<b>Location</b>	Zurich, HG F33.1
<b>Time</b>	09:00-13:00
<b>Program</b>	<ol style="list-style-type: none"><li>1. introduction to openBIS</li><li>2. How to manage the lab inventory of materials and samples</li><li>3. How to manage lab protocols</li><li>4. How to record experiments in the Electronic Lab Notebook &amp; upload data</li><li>5. How to analyse data stored in openBIS using Jupyter notebooks and MATLAB</li></ol>
<b>Participants</b>	
<b>Training material</b>	<a href="https://gitlab.ethz.ch/sis-rdm-training/openbis-training">https://gitlab.ethz.ch/sis-rdm-training/openbis-training</a>


















The left sidebar shows the 'ID SIS' space with a 'trainings' section expanded, listing 'openBIS trainings' and several dates, including '2020.01.23'. The bottom of the page shows the footer: 'Powered by Atlassian Confluence 6.15.9 · Report a bug · Atlassian News'.

# Example of versioning in wiki

Pages / ... / 2020

Page History

Compare selected versions

	Version	Published	Changed By	Comment	Actions
<input type="checkbox"/>	<b>CURRENT (v. 17)</b>	<b>Jun 11, 2020 16:33</b>	 Luetcke Henry (ID SIS)		
<input type="checkbox"/>	v. 16	Jun 11, 2020 15:44	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 15	Jun 09, 2020 08:17	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 14	Jun 08, 2020 12:06	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 13	Jun 08, 2020 11:51	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 12	Jun 08, 2020 11:48	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 11	Jun 08, 2020 11:46	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 10	Jun 05, 2020 14:56	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 9	May 22, 2020 10:58	 Plamada Andrei Valentin (ID SIS)	email update	<a href="#">Restore</a>
<input type="checkbox"/>	v. 8	May 22, 2020 10:58	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 7	May 22, 2020 10:46	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 6	May 18, 2020 14:29	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 5	May 18, 2020 14:28	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 4	May 15, 2020 14:29	 Plamada Andrei Valentin (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 3	Feb 14, 2020 14:34	 Barillari Caterina (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 2	Feb 14, 2020 14:34	 Barillari Caterina (ID SIS)		<a href="#">Restore</a>
<input type="checkbox"/>	v. 1	Feb 14, 2020 14:32	 Barillari Caterina (ID SIS)		<a href="#">Restore</a>

[Return to Page Information](#)

# Questions on Management of Samples and Protocols?



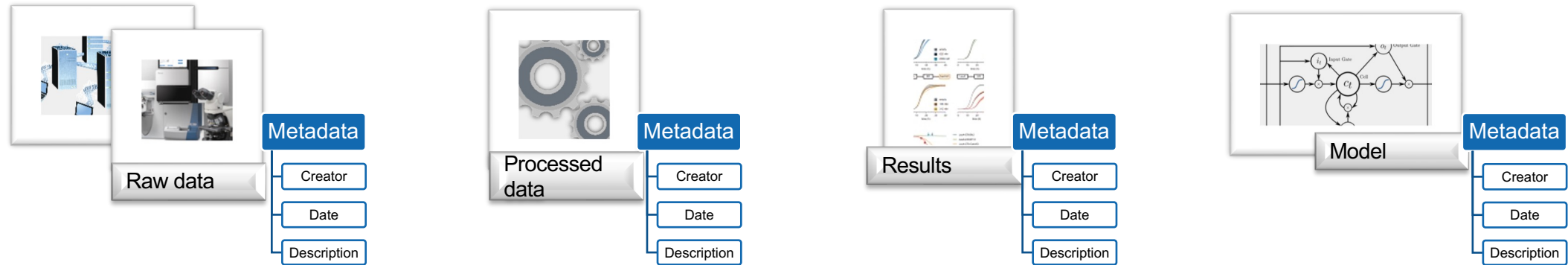




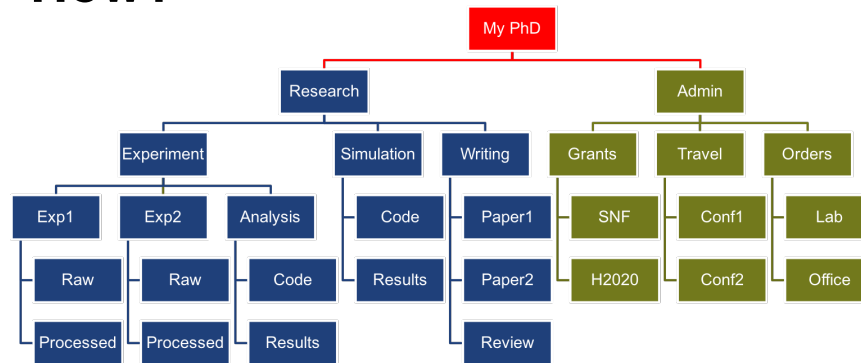
# *Management of Data & Metadata*

# Management of research data files

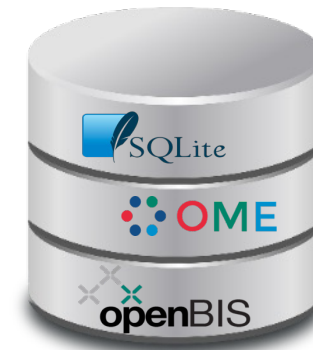
## What?



## How?



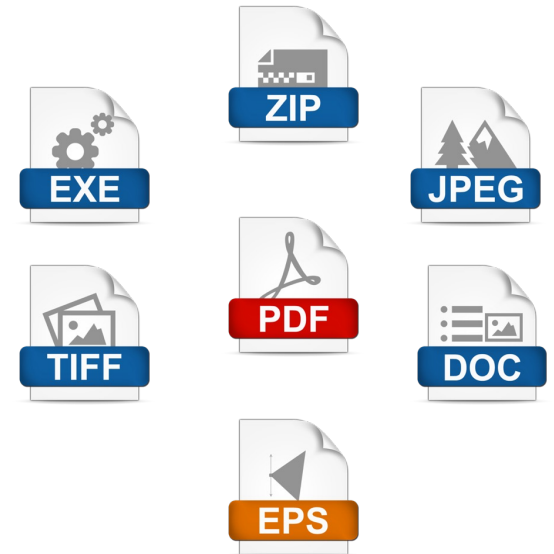
Files / folders hierarchy



Data management platform

# File types and formats

- ❑ File format: a **convention for encoding information** in a computer file
- ❑ Extensions typically *indicate* a specific file format
- ❑ Some file formats are preferable to others
- ❑ Stick to **non-proprietary** and **widely used** formats!
- ❑ Several general-purpose **scientific data formats** exist (e.g. HDF5, netCDF, FITS)



-	+
Binary	Text-based
Proprietary	Open
New kid on the block	Old as the hills
Compressed/encrypted	Uncompressed/unencrypted
Platform dependent	Interoperable
Complex	Simple

# File & folder organisation

## ❑ **Goals** of efficient file / folder organization:

- Easy to **find** something in the future (you, others)
- Easy to **file** something
- Save **disk space** (avoid multiple copies of files)
- **Reusable** components
- Avoid problems on different **operating systems**



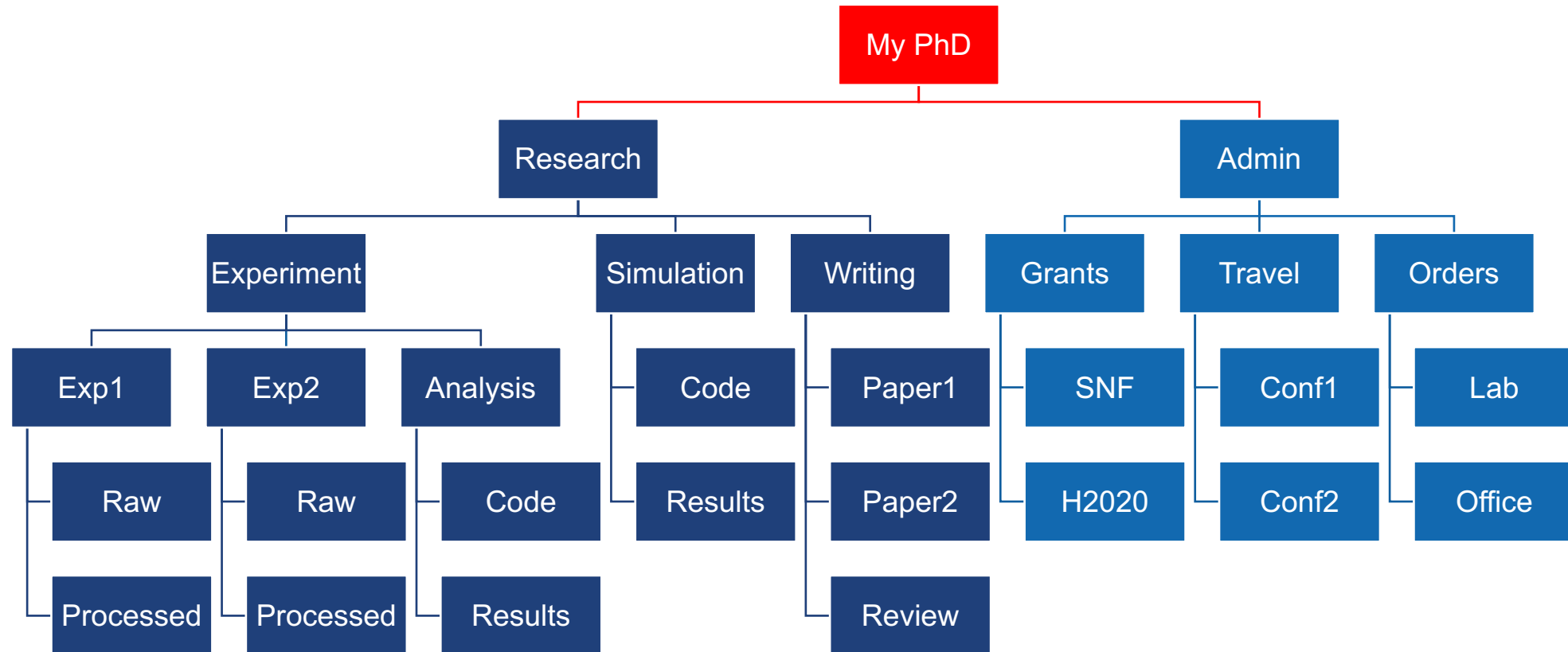
## ❑ Planning a good **folder hierarchy**

- **Who?** Individual, research group
- **Where?** Local disk, shared network drive
- Which operating system?
- What information are you going to **search** for?
- Avoid non-descriptive file and folder names (*figure\_02\_summary\_stats.png* and not *stats.png*)
- Add **descriptive text files** to folders (→ Metadata)
- **Document** your hierarchy and file naming convention



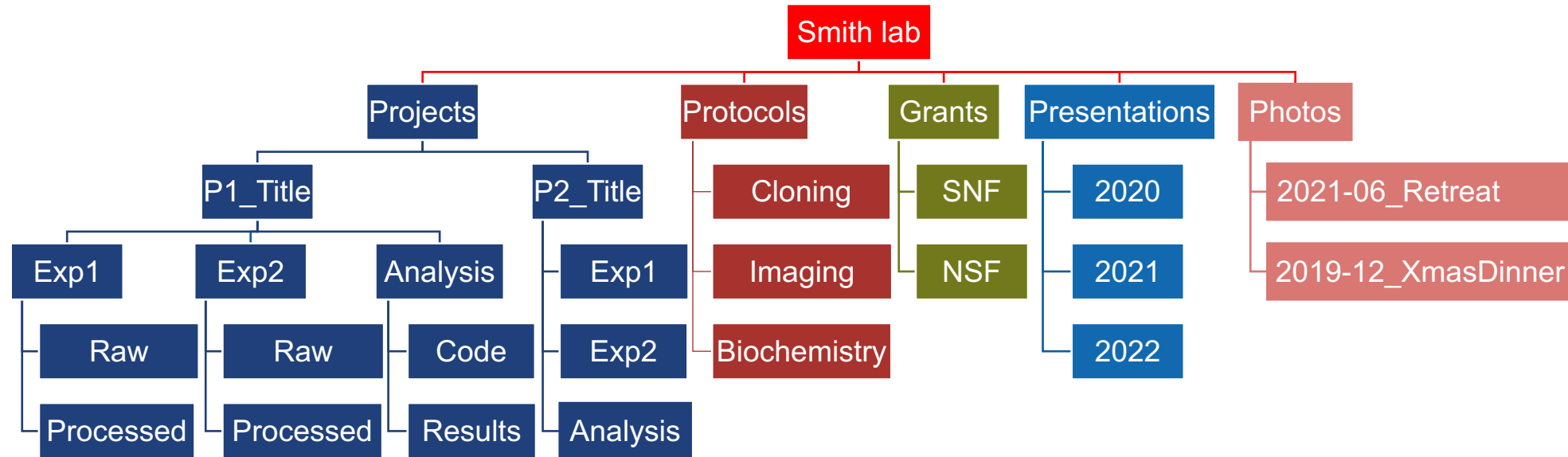
# File & folder organisation

## *Example hierarchy for a PhD project*



# File & folder organisation

## *Example hierarchy for a research group*



# File & folder organisation

## *The project directory (for a computational project)*

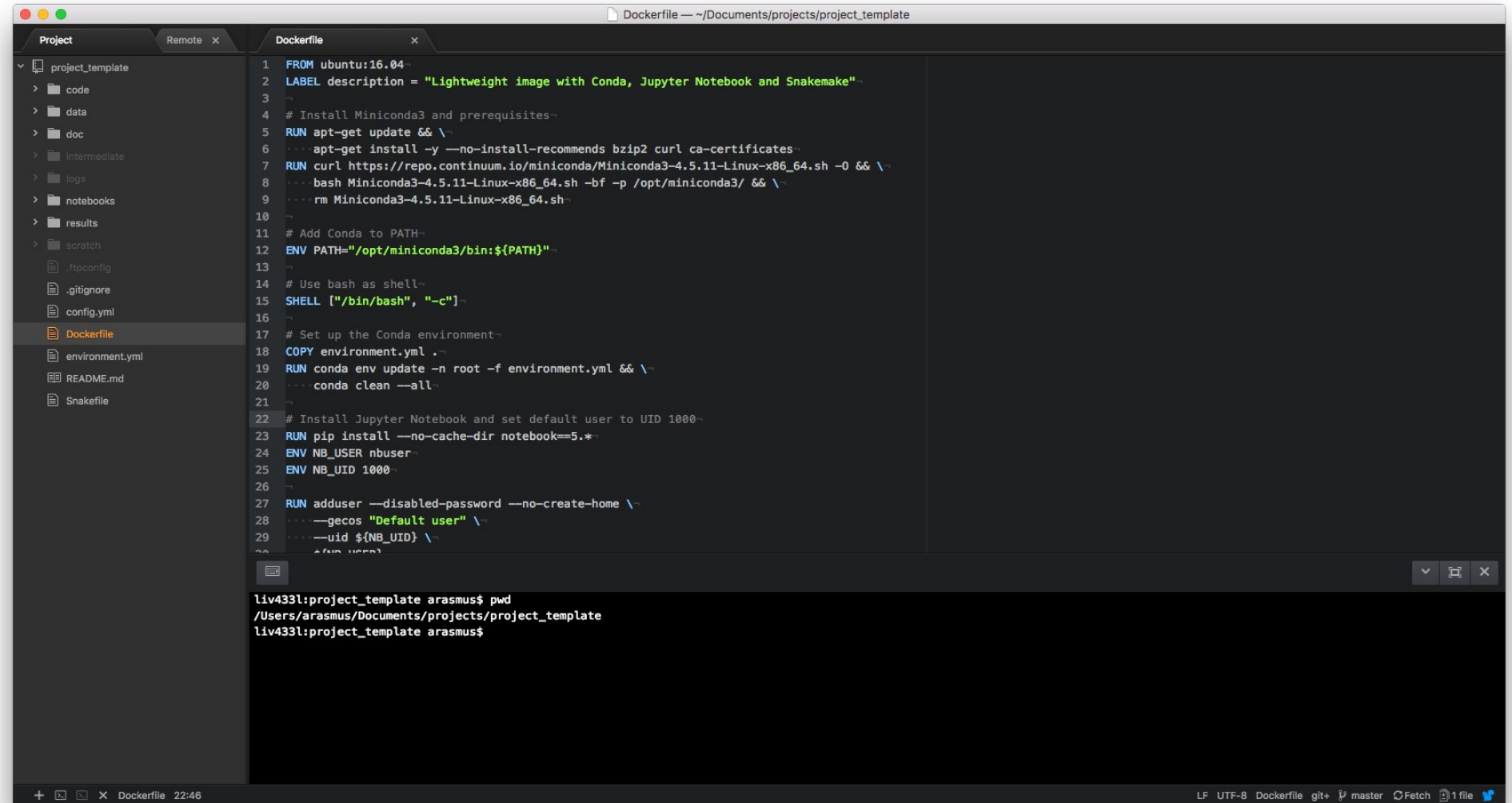
project	
- doc/	documentation for the study
- data/	raw and primary data, essentially all input files, never edit!
- raw_external/	
- raw_internal/	
- meta/	
- code/	all code needed to go from input files to final results
- notebooks/	notebooks that document your day-to-day work
- intermediate/	output files from different analysis steps, can be deleted
- scratch/	temporary files that can be safely deleted or lost
- logs/	logs from the different analysis steps
- results/	output from workflows and analyses
- figures/	
- tables/	
- reports/	
- Snakefile	project workflow, carries out analysis contained in code/
- config.yml	configuration of the project workflow
- environment.yml	software dependencies list, used to create a project environment
- Dockerfile	recipe to create a project container



[Noble WS \(2009\) A Quick Guide to Organizing Computational Biology Projects. PLoS Comput Biol 5\(7\): e1000424](#)

# File & folder organisation

## *A project in ATOM*



The screenshot shows the ATOM editor interface with a project named 'project\_template'. The left sidebar displays the project's file structure, including folders like 'code', 'data', 'doc', 'intermediate', 'logs', 'notebooks', 'results', and 'scratch', along with files like '.ftpconfig', '.gitignore', 'config.yml', 'Dockerfile', 'environment.yml', 'README.md', and 'Snakefile'. The 'Dockerfile' is selected and open in the main editor. The Dockerfile content is as follows:

```
1 FROM ubuntu:16.04~
2 LABEL description = "Lightweight image with Conda, Jupyter Notebook and Snakemake"~
3
4 # Install Miniconda3 and prerequisites~
5 RUN apt-get update && \~
6 ... apt-get install -y --no-install-recommends bzip2 curl ca-certificates~
7 RUN curl https://repo.continuum.io/miniconda/Miniconda3-4.5.11-Linux-x86_64.sh -O && \~
8 ... bash Miniconda3-4.5.11-Linux-x86_64.sh -bf -p /opt/miniconda3/ && \~
9 ... rm Miniconda3-4.5.11-Linux-x86_64.sh~
10
11 # Add Conda to PATH~
12 ENV PATH="/opt/miniconda3/bin:${PATH}"~
13
14 # Use bash as shell~
15 SHELL ["/bin/bash", "-c"]~
16
17 # Set up the Conda environment~
18 COPY environment.yml .~
19 RUN conda env update -n root -f environment.yml && \~
20 ... conda clean --all~
21
22 # Install Jupyter Notebook and set default user to UID 1000~
23 RUN pip install --no-cache-dir notebook==5.*~
24 ENV NB_USER nbuser~
25 ENV NB_UID 1000~
26
27 RUN adduser --disabled-password --no-create-home \~
28 ... --gecos "Default user" \~
29 ... --uid ${NB_UID} \~
30 ... & (for user)
```

Below the Dockerfile, a terminal window shows the execution of the Dockerfile commands:

```
liv4331:project_template arasmus$ pwd
/Users/arasmus/Documents/projects/project_template
liv4331:project_template arasmus$
```

The status bar at the bottom indicates the file is 'Dockerfile', the encoding is 'UTF-8', and the current branch is 'master'.

# File & folder organisation

❑ Keep path names short (< 256 characters)

❑ Recommendation for file names:

- Unique, reflect content (if possible)
- Use only ASCII characters
- Not include spaces
- Be aware of case sensitivity

❑ Bad examples:

- data%20management%20plan.docx
- sup figure 2.png
- lab meeting 19.10.2021.pptx

❑ Good examples:

- Data\_management\_plan\_SNF.docx
- sup\_figure\_02\_summary\_stats.png
- lab\_meeting\_2021-10-19.pptx

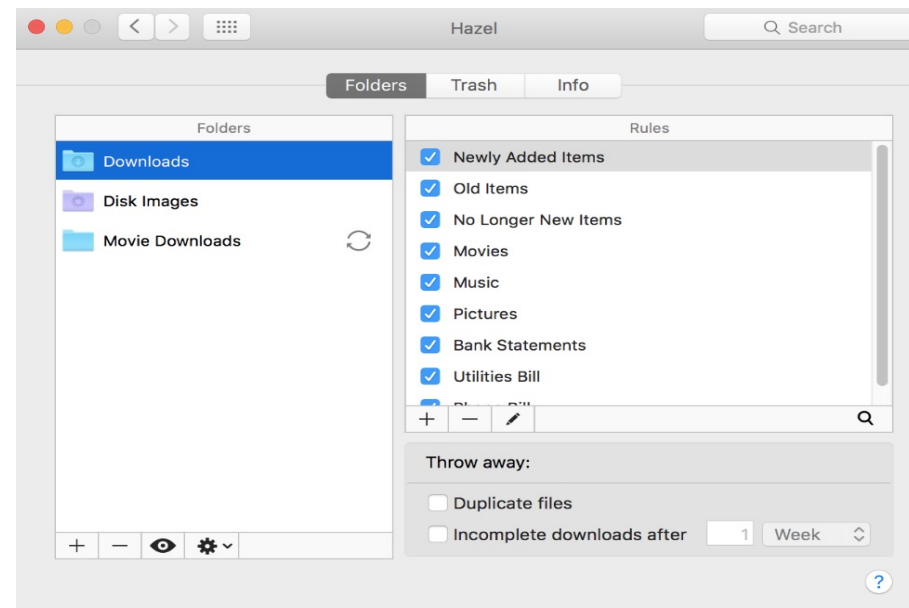
❑ Use links / shortcuts to avoid duplications

❑ Use tags for orthogonal classifications

❑ Create template folders

❑ Some tools for automated file organization:

- Mac: Hazel (<https://www.noodlesoft.com/>)
- PC: DropIt (<http://www.dropitproject.com/>)





# Batch renaming of files

## ❑ Windows

- Bulk Rename Utility ([www.bulkrenameutility.co.uk](http://www.bulkrenameutility.co.uk))
- Advanced Renamer ([www.advancedrenamer.com](http://www.advancedrenamer.com))
- Command prompt / PowerShell scripts

## ❑ macOS

- Finder rename functionality
- Automator
- Command line / scripts



## ❑ Linux

- Command line utility `rename`
- Métamorphose (<http://file-folder-ren.sourceforge.net/>)

Name	Date Modified	Size	Kind
file_00001.tif	4 May 2018, 14:16	532 KB	TIFF image
file_00002.tif	4 May 2018, 14:21	17.3 MB	TIFF image
file_00003.tif	4 May 2018, 14:24	17.3 MB	TIFF image
file_00004.tif	4 May 2018, 14:25	5.3 MB	TIFF image
file_00005.tif	4 May 2018, 14:27	3.4 MB	TIFF image
file_00006.tif	4 May 2018, 14:28	3.4 MB	TIFF image
file_00007.tif	4 May 2018, 14:29	3.4 MB	TIFF image
file_00008.tif	4 May 2018, 14:29	3.4 MB	TIFF image
file_00009.tif	4 May 2018, 14:29	3.4 MB	TIFF image
file_00010.tif	4 May 2018, 14:29	3.4 MB	TIFF image
file_00011.tif	4 May 2018, 14:29	3.4 MB	TIFF image
file_00012.tif	4 May 2018, 14:30	3.4 MB	TIFF image
file_00013.tif	4 May 2018, 14:30	3.4 MB	TIFF image
file_00014.tif	4 May 2018, 14:30	3.4 MB	TIFF image
file_00015.tif	4 May 2018, 14:30	3.4 MB	TIFF image
file_00016.tif	4 May 2018, 14:31	3.4 MB	TIFF image
file_00017.tif	4 May 2018, 14:31	3.4 MB	TIFF image
file_00018.tif	4 May 2018, 14:31	3.4 MB	TIFF image
file_00019.tif	4 May 2018, 14:31	3.4 MB	TIFF image
file_00020.tif	4 May 2018, 14:32	3.4 MB	TIFF image
file_00021.tif	4 May 2018, 14:32	3.4 MB	TIFF image
file_00022.tif	4 May 2018, 14:32	3.4 MB	TIFF image
file_00023.tif	4 May 2018, 14:32	3.4 MB	TIFF image
file_00024.tif	4 May 2018, 14:33	3.4 MB	TIFF image
file_00025.tif	4 May 2018, 14:33	3.4 MB	TIFF image
file_00026.tif	4 May 2018, 14:33	3.4 MB	TIFF image
file_00027.tif	4 May 2018, 14:34	3.4 MB	TIFF image
file_00028.tif	4 May 2018, 14:34	3.4 MB	TIFF image
file_00029.tif	4 May 2018, 14:34	3.4 MB	TIFF image
file_00030.tif	4 May 2018, 14:35	3.4 MB	TIFF image
file_00031.tif	4 May 2018, 14:35	3.4 MB	TIFF image
file_00032.tif	4 May 2018, 14:35	3.4 MB	TIFF image
file_00033.tif	4 May 2018, 14:35	3.4 MB	TIFF image
file_00034.tif	4 May 2018, 14:36	3.4 MB	TIFF image

# File validation and compression

## ❑ **Checksum** algorithms are useful to **verify data integrity**

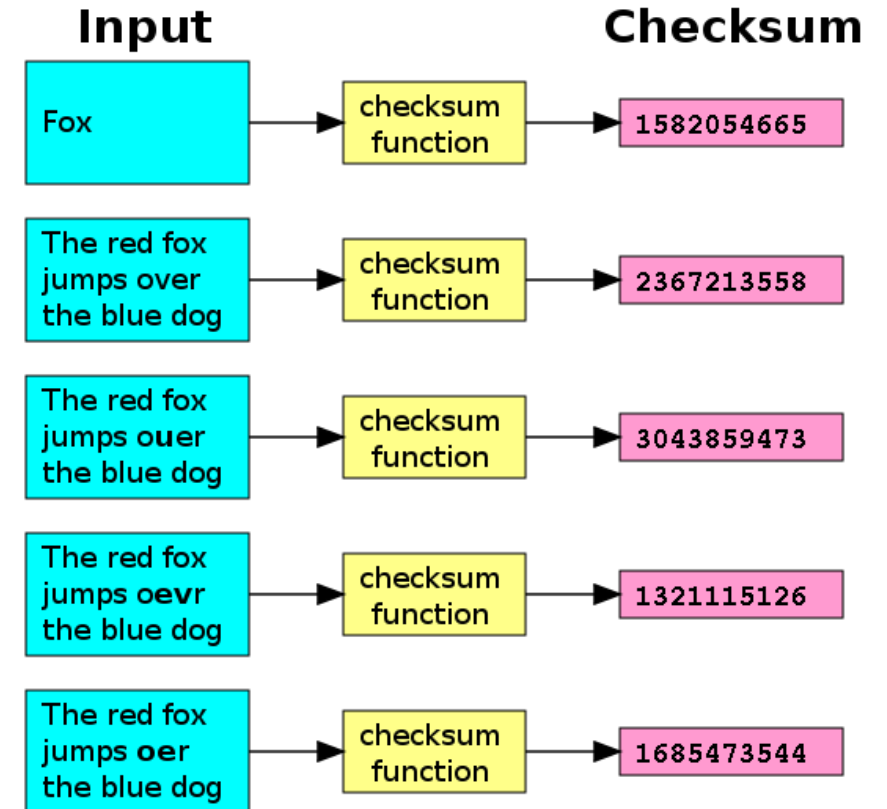
- For datasets stored over long time periods
- When transferring from one storage to another
- Examples: *cksum*, *md5sum*, *sha1sum*

## ❑ **Compression** algorithms encode information in fewer bits than the original (bit-rate reduction)

- Lossless vs. lossy compression
- Lossy compression may lead to distortion / loss in quality but compression rates are typically much higher
- Trade-off between processing time and disk space
- Examples: zip, gzip, specific formats (TIFF, HDF5)

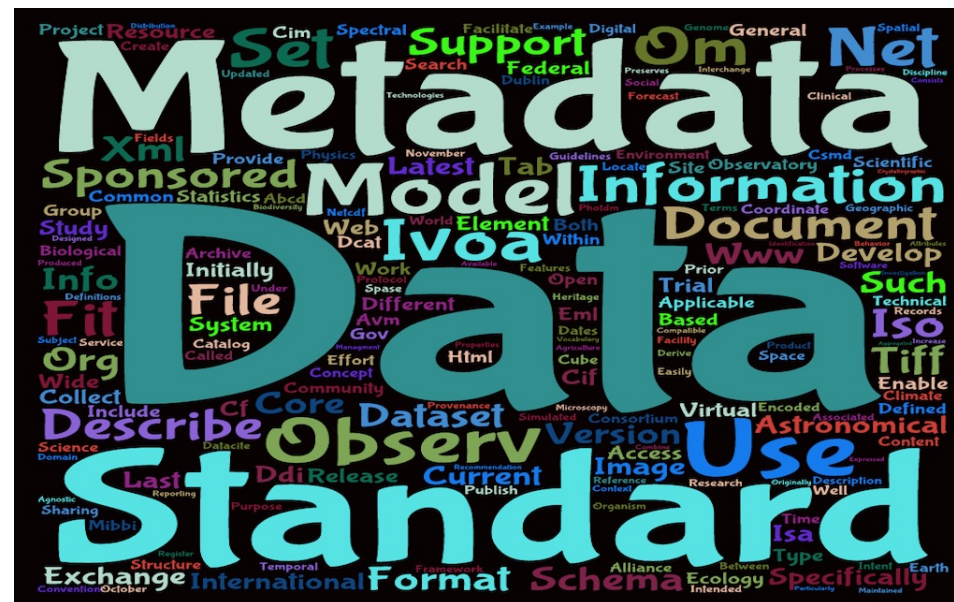
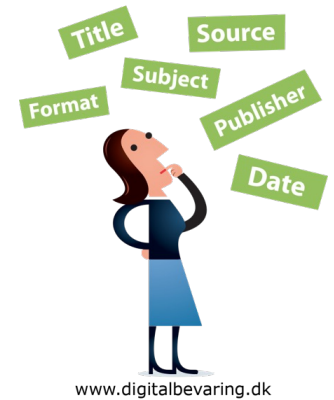
## ❑ **Archiving** combines many (small) files into a single (large) file

- Nowadays mainly used together with compression
- Required for long-term storage (LTS) and efficient data transfer
- Examples: Tape Archive format (tar)



# Metadata

- ❑ **Metadata** is the *data about your data* (a.k.a. **data model**)
- ❑ Use of structured metadata **facilitates data organization** and searches
- ❑ (Machine-readable) **Metadata** is a key element of the **FAIR data** principles
- ❑ Existing **metadata schemas** are preferred (can be extended, if necessary)



<https://rdamsc.bath.ac.uk/>

# Metadata

## Example for general research data: **DataCite Metadata schema**

Table 1: DataCite Mandatory Properties

<i>ID</i>	<i>Property</i>	<i>Obligation</i>
1	Identifier (with mandatory type sub-property)	M
2	Creator (with optional given name, family name, name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType (with mandatory general type description sub-property)	M

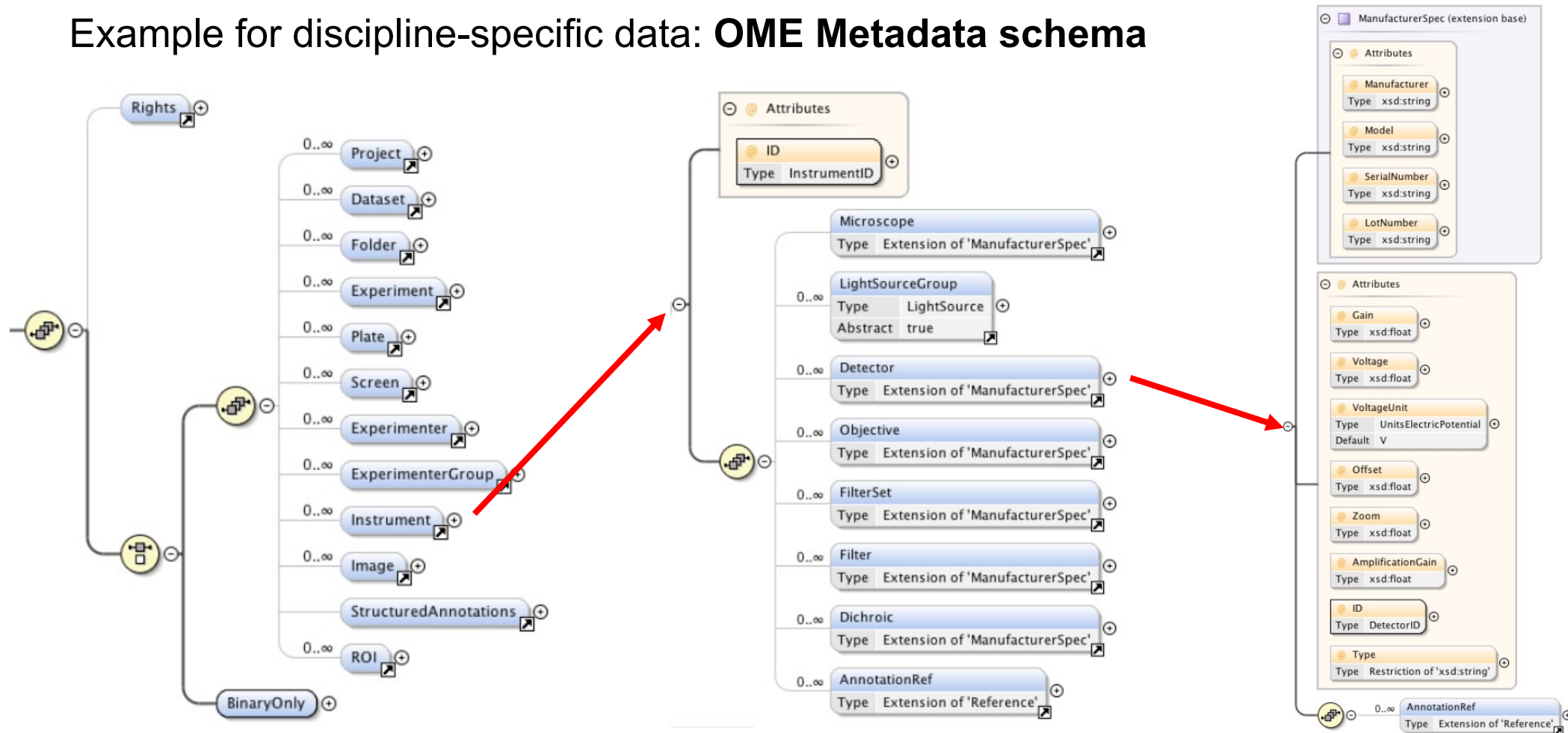
<https://schema.datacite.org/>

Table 2: DataCite Recommended and Optional Properties

<i>ID</i>	<i>Property</i>	<i>Obligation</i>
6	Subject (with scheme sub-property)	R
7	Contributor (with optional given name, family name, name identifier and affiliation sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point, box and polygon sub-properties)	R
19	FundingReference (with name, identifier, and award related sub-properties)	O

# Metadata

Example for discipline-specific data: **OME Metadata schema**



<https://www.openmicroscopy.org/Schemas/>



# Metadata

The screenshot displays the IDR Open Microscopy webclient interface. The main area features a grid of microscopy images, with a red rounded rectangle highlighting a portion of the grid. The sidebar on the right, outlined in blue, contains detailed metadata for the selected image. The metadata is organized into sections: General, Acquisition, and Preview. The General section includes fields for Image ID, Owner, Image Details, Cell Lines, Organism, and Others. The Acquisition section includes fields for Import Date, Dimensions (XY), Pixels Type, Pixels Size (XYZ), Z-sections/Timepoints, Channels, and ROI Count. The Preview section includes a list of Attachments, Comments, Tags, Ratings, and Others. The metadata is presented in a clean, modern layout with a light gray background and white text.

**Data**

**Metadata**



# Metadata Exercise – Discipline-specific Standards

The Research Data Alliance (RDA) has created a community-maintained catalog of metadata standards: <https://rdamsc.bath.ac.uk/>

The catalog contains metadata standards that have been established in different research fields as well as relevant tools and use cases. Consider the following short descriptions of hypothetical research projects. Based on the description, try to identify one or several suitable standards from the metadata catalog.

**Case 1: An online survey / questionnaire about attitudes towards climate change in different age groups. Documentation of the survey responses.**

# Metadata Exercise – Discipline-specific Standards

The Research Data Alliance (RDA) has created a community-maintained catalog of metadata standards: <https://rdamsc.bath.ac.uk/>

The catalog contains metadata standards that have been established in different research fields as well as relevant tools and use cases. Consider the following short descriptions of hypothetical research projects. Based on the description, try to identify one or several suitable standards from the metadata catalog.

**Case 1: An online survey / questionnaire about attitudes towards climate change in different age groups. Documentation of the survey responses.**

- ☐ [Data Documentation Initiative](#) (DDI), version 2 or 3
- ☐ An international standard for describing surveys, questionnaires, statistical data files, and social sciences study-level information

# Metadata Exercise – Discipline-specific Standards

**Case 2: A study on the diversity and geographic distribution of butterflies in the Swiss Alps. Documentation of the recorded occurrences.**

# Metadata Exercise – Discipline-specific Standards

**Case 2: A study on the diversity and geographic distribution of butterflies in the Swiss Alps. Documentation of the recorded occurrences.**

- ❑ [Darwin Core](#) (DwC)
- ❑ [Ecological Metadata Language](#) (EML)
- ❑ **Darwin Core** is a standard to facilitate the sharing of information about biological diversity. DwC includes a glossary of terms, and documentation providing reference definitions, examples, and commentary. It is an extension of Dublin Core.
- ❑ **EML** defines a vocabulary and a syntax for documenting research data. It is in widespread use in the earth and environmental sciences. EML includes modules for describing the spatial, temporal, taxonomic, and thematic extent of data, for describing research methods and protocols, for describing the structure and content of data, and for precisely annotating data with semantic vocabularies.

# Metadata Exercise – Discipline-specific Standards

**Case 3: The crystal structure of a novel viral protein determined by X-Ray crystallography. Documentation of the protein structure.**

# Metadata Exercise – Discipline-specific Standards

**Case 3: The crystal structure of a novel viral protein determined by X-Ray crystallography. Documentation of the protein structure.**

- ❑ Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework ([PDBx/mmCIF](#))



# Metadata Exercise – Discipline-specific Standards



**[FAIRsharing.org](https://fairsharing.org) provides an even more  
extensive resource of metadata standards,  
vocabularies, taxonomies etc.**

# Metadata

## ❑ Types of metadata

- Descriptive (Title, author, identifier)
- Administrative (License)
- Technical (File size, checksums)
- Structural (Relation to other data)

## ❑ Machine-readable metadata

- Annotation based on common standards
- Controlled vocabularies, taxonomies

The screenshot shows the 'Filters' section of the Europe PMC search interface. It includes several filter categories with checkboxes and dropdown menus:

- Sources:** Agricola (USDA/NAL), Chinese biological abstracts, CiteXplore records, Patents, Preprint records, PubMed/MEDLINE (NLM).
- Special Collections:** All BMJ, All manuscripts, EuroFIR, Europe PMC manuscripts.
- Full Text Availability:** In Europe PMC, Open Access.
- Publication Type:** A dropdown menu labeled 'Choose one Publication Type'.
- CC License:** A dropdown menu labeled 'Choose one License Type'.
- Article Sections:** A dropdown menu labeled 'Choose a section type'.
- Language:** A dropdown menu labeled 'Choose one Language' with a list of languages including Afrikaans, Albanian, Arabic, Armenian, Azeri, Bosnian, Bulgarian, Catalan, Chinese, Czech, Danish, Dutch, English (highlighted), Esperanto, Estonian, Finnish, French, and Georgian.
- Data Links and Data Citations:** A dropdown menu labeled 'Choose one Link/Citation type'.
- External Links:** A dropdown menu labeled 'Choose one External Links Provider'.

<https://europepmc.org/advancesearch>

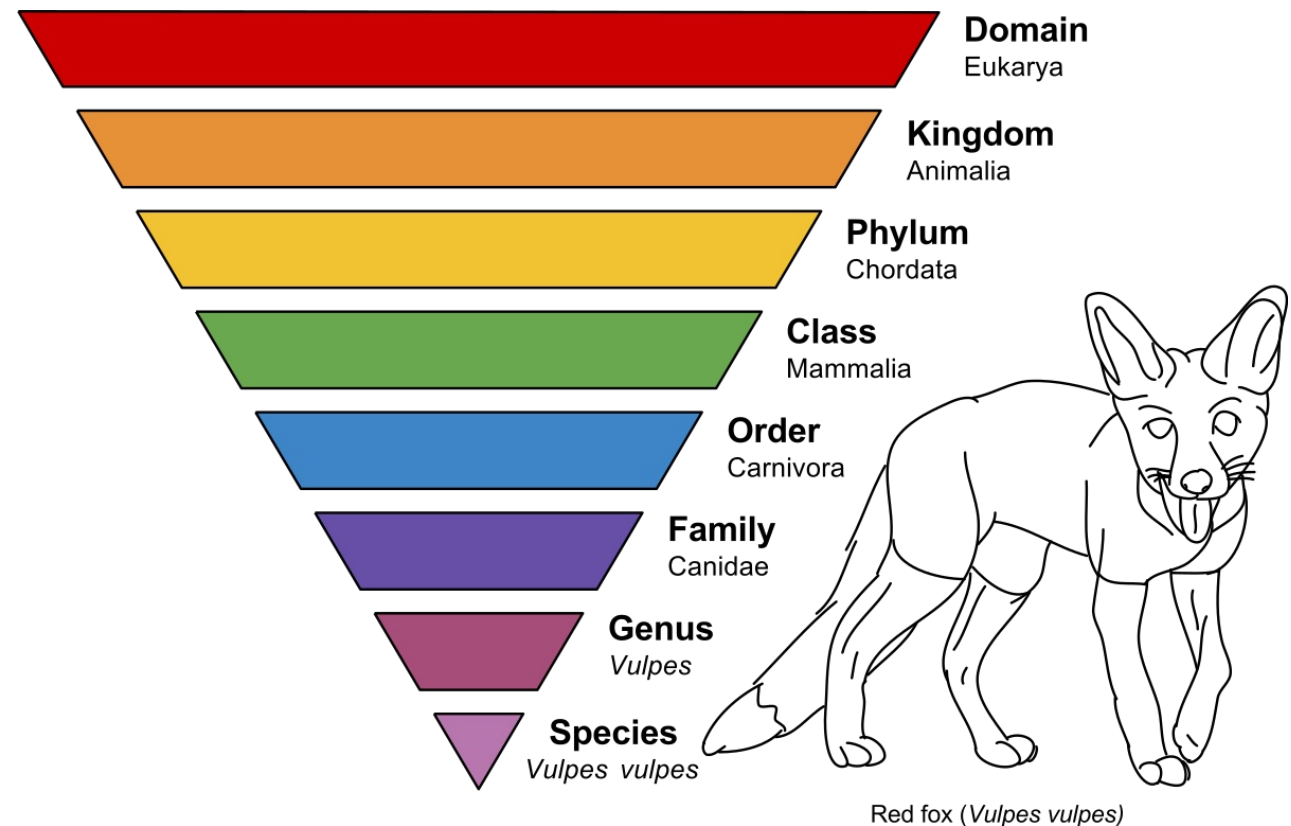
# Metadata

## ❑ Types of metadata

- Descriptive (Title, author, identifier)
- Administrative (License)
- Technical (File size, checksums)
- Structural (Relation to other data)

## ❑ Machine-readable metadata

- Annotation based on common standards
- Controlled vocabularies, taxonomies



[https://en.wikipedia.org/wiki/Domain\\_%28biology%29](https://en.wikipedia.org/wiki/Domain_%28biology%29)

# Metadata

## Types of metadata

- Descriptive (Title, author, identifier)
- Administrative (License)
- Technical (File size, checksums)
- Structural (Relation to other data)

## Machine-readable metadata

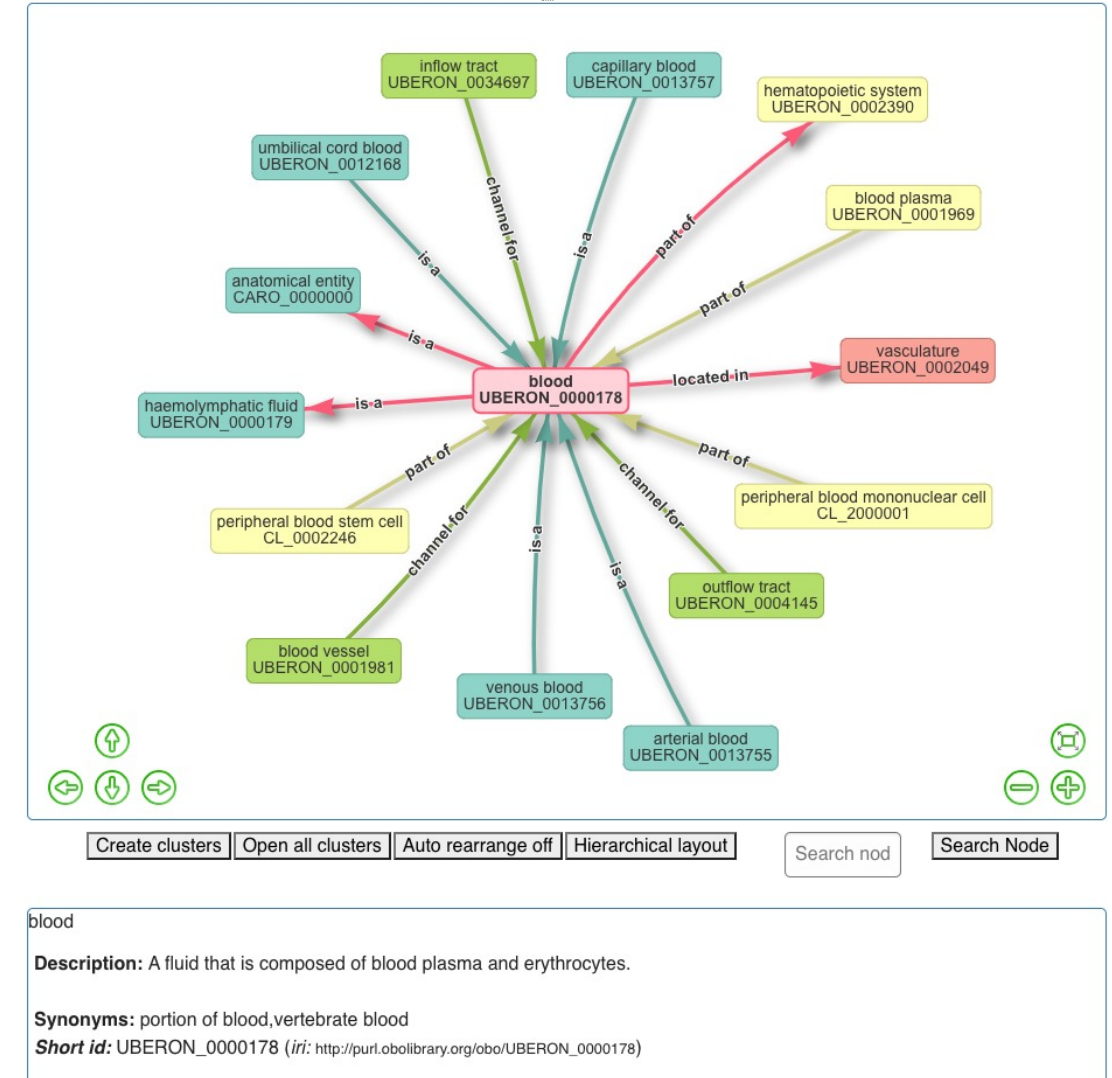
- Annotation based on common standards
- Controlled vocabularies, taxonomies
- Standardized metadata are the basis for ontologies (knowledge representations)

[Basic Register of Thesauri, Ontologies & Classifications](#)

[EMBL-EBI Ontology Lookup Service](#)

OLS > HCAO > UBERON\_0000178

Visualized term: blood ([http://purl.obolibrary.org/obo/UBERON\\_0000178](http://purl.obolibrary.org/obo/UBERON_0000178))



Knowledge graph for 'blood' from the [Human Cell Atlas Ontology](#)

# Metadata

## ❑ Types of metadata

- Descriptive (Title, author, identifier)
- Administrative (License)
- Technical (File size, checksums)
- Structural (Relation to other data)

## ❑ Machine-readable metadata

- Annotation based on common standards
- Controlled vocabularies, taxonomies
- Standardized metadata are the basis for ontologies (knowledge representations)

[Basic Register of Thesauri, Ontologies & Classifications](#)

[EMBL-EBI Ontology Lookup Service](#)

## ❑ How to apply metadata?

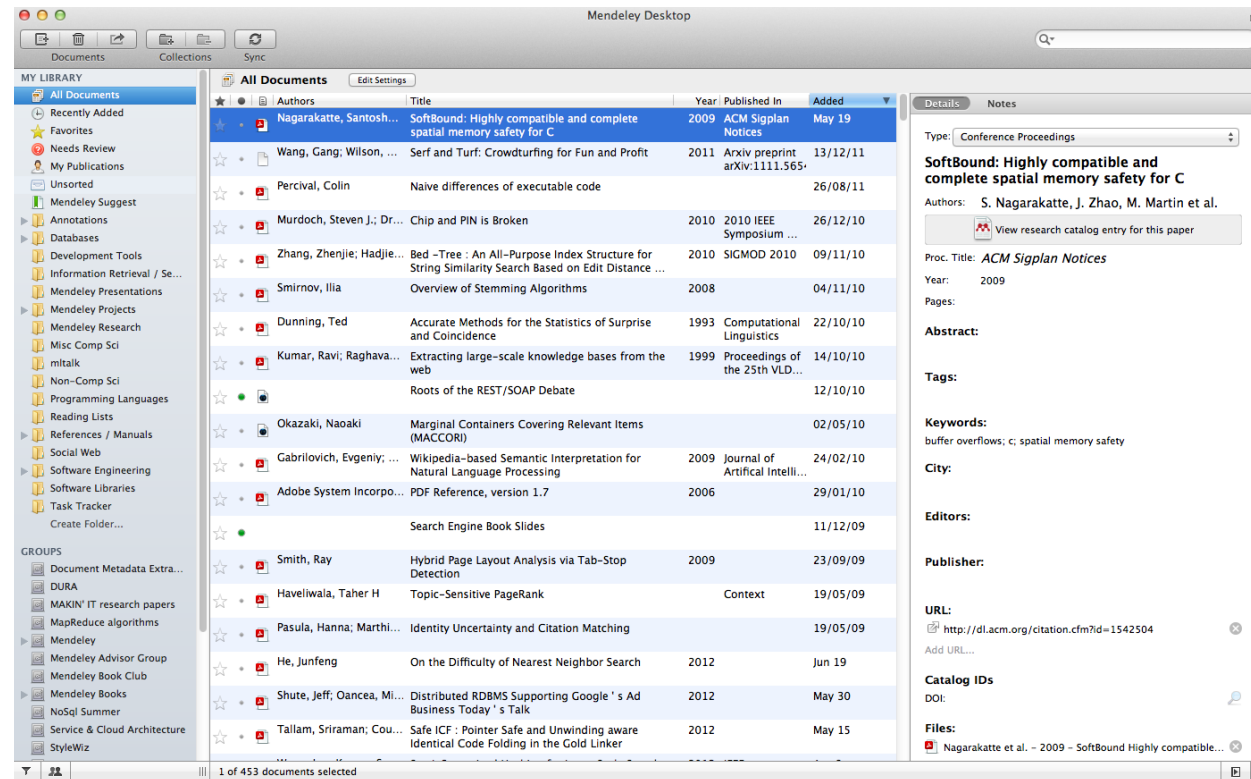
- File header (e.g. TIFF, netCDF)
- Sidecar files (e.g. XML, JSON)
- As columns in a database
- README text files

## ❑ **Apply metadata early** in your workflow (ideally during acquisition)

# Data management software



- ❑ System that allows **structured organization** of data
- ❑ Data is described by **metadata**
- ❑ Usually more FAIR-compliant than Files / Folders
- ❑ Searchable, scalable, flexible
- ❑ Allows user rights management
- ❑ Back up procedures are easily implemented





# Data management software

- ❑ System that allows **structured organization** of data
- ❑ Data is described by **metadata**
- ❑ Usually more FAIR-compliant than Files / Folders
- ❑ Searchable, scalable, flexible
- ❑ Allows user rights management
- ❑ Back up procedures are easily implemented
- ❑ Examples



## Generic

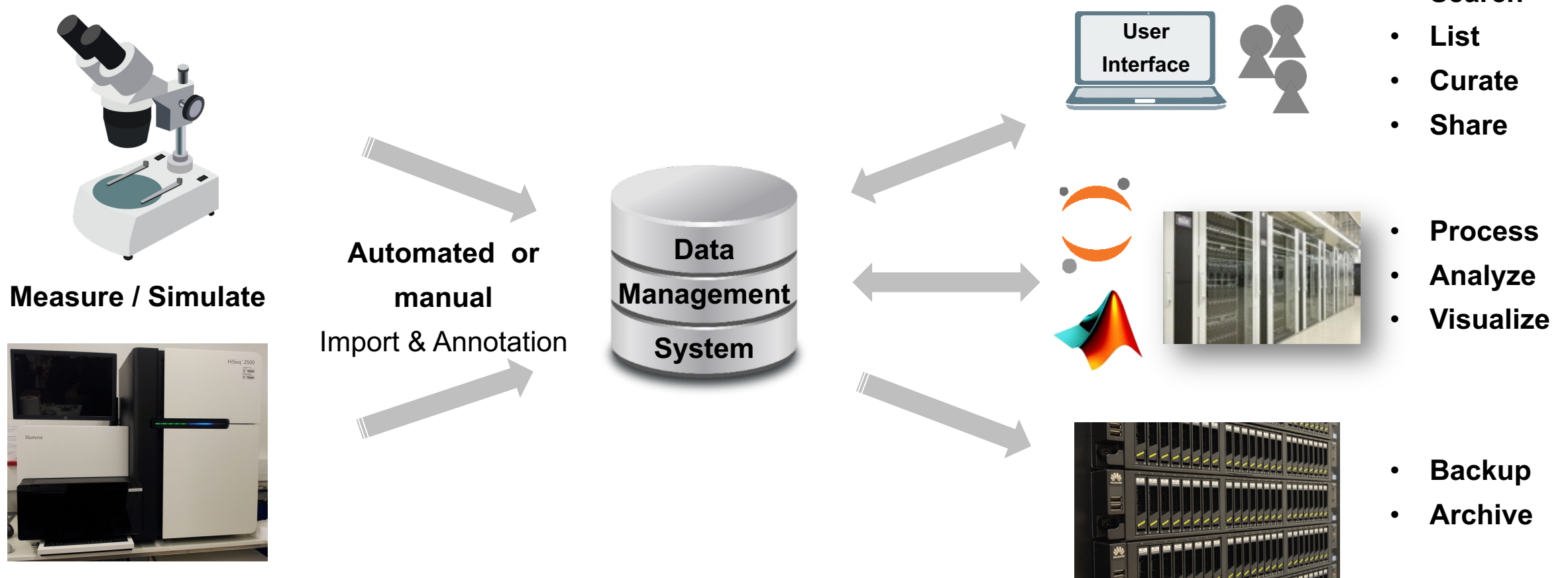


## Scientific



# Data management software

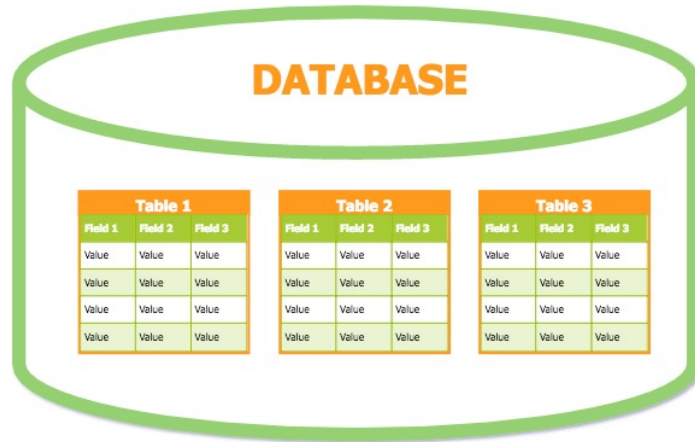
## Generic database workflow



# Data management software

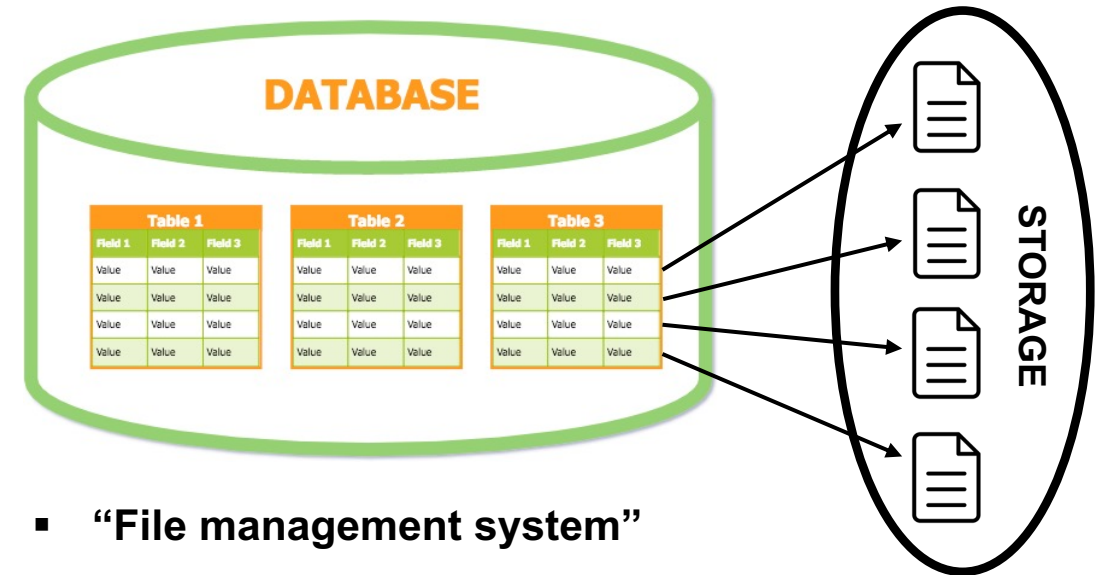
Where is my data???

## Metadata + Data



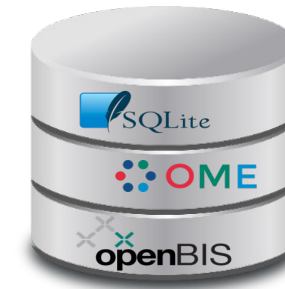
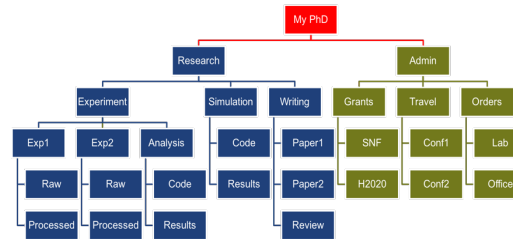
- Structured data required
- DB-specific data (array DBs)
- No conventional file access

## Metadata + Link to Data Files



- “File management system”
- Very large data volumes
- Conventional file access

# Comparison: folder hierarchy vs. databases



	File / folder Hierarchy	Database System
Easy to use	Yes	Initial learning curve
Flexibility	High	Limited
Scalability (data volume)	Limited	Yes
Scalability (users)	Limited	Yes
Versioning included	No	Yes
Backup included	No	Yes

# Data storage – Your options ...

- ❑ External hard disks are cheap but unreliable and don't scale!

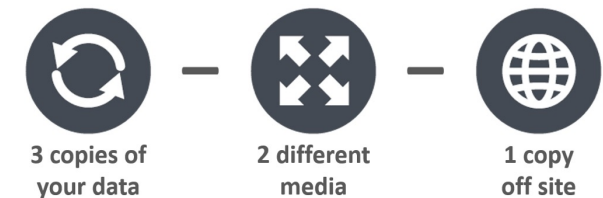
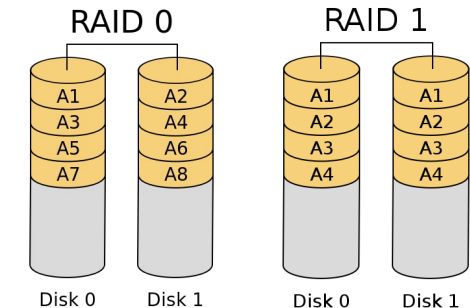


# Data storage – Your options ...

- ❑ External hard disks are cheap but unreliable and don't scale!
- ❑ Polybox / SwitchDrive
- ❑ Network Attached Storage (NAS)
  - NAS offering from IT Services – always check with your ISG first
- ❑ Cost Defined Storage (CDS)
  - For very large storage requirements (>100 TB)
- ❑ Long-term Storage (LTS)
  - Tape storage for data worthy of preservation (“Datenendlager”)
  - Data should be compressed and sized between 10 - 200 GB
- ❑ Also consider backups (ideally follow 3-2-1 rule)
  - RAID is not a backup!
- ❑ Details: <https://ethz.ch/services/en/it-services/catalogue/storage.html>
- ❑ ETH news on cloud storage regulations: <https://bit.ly/3BoQQD1>



**Redundant  
Array of  
Independent  
Disks**



# Questions on Management of Data Files?







Time for a break



# Experimental description / notes



## What to document?

- ☐ Goals
- ☐ Materials
- ☐ Methods
  - Experimental/computational procedure
  - Analysis procedures
- ☐ Results
- ☐ Links to data

## How?



**Paper laboratory notebook**



**Electronic laboratory notebook (ELN)**

# Definition of ELN & requirements

*An **electronic laboratory notebook** (also known as **electronic lab notebook** or **ELN**) is a software program or package designed to replace more traditional paper laboratory notebook. Laboratory notebooks in general are used by scientists and technicians to document, store, retrieve, and share fully electronic laboratory records in ways that meet all legal, regulatory, technical and scientific requirements.*

## Legal requirements

Electronic lab notebooks used for development or research in regulated industries, such as medical devices or pharmaceuticals, are expected to comply with the **21 CFR Part 11 FDA** regulations:

<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/cfrsearch.cfm?cfrpart=11>

- To our knowledge, in academia, there is no such requirement, but this can vary from one Institution to another.

Source: [https://www.limswiki.org/index.php/Electronic\\_laboratory\\_notebook](https://www.limswiki.org/index.php/Electronic_laboratory_notebook)

# ETHZ Regulations

❑ **ETHZ RDM guidelines** came into action on **01.07.2022**:  
<https://rechtssammlung.sp.ethz.ch/Dokumente/414.2en.pdf>

## Art. 5 Data collection and processing

<sup>1</sup> Digital data that are in the process of being collected or processed must be stored such that data can be restored in case of hardware problems or human error. Using storage systems centrally supplied by ETH fulfils this criterion. If such data are too big for being stored redundantly in the available infrastructure, researchers must be able to recreate them using existing code and documentation when required.

<sup>2</sup> Access rights to the storage system must be specified and regularly updated, e.g. upon the start of a new project. The scientific needs of all researchers working with the data must be considered by the access right system. Unauthorised access must be prevented by suitable technical and organisational measures, such as ETH Zurich's identity and access management system.

<sup>3</sup> In order to ensure long-term accessibility and reusability, the use of well-documented, non-proprietary file formats is recommended.

<sup>4</sup> The structure and the processing steps of all *Research Data* must be digitally documented in order to ensure adherence to the **FAIR principles**. Where documentation includes a lab journal, **Electronic Laboratory Notebooks (ELN)** are recommended.

# ELNs vs. paper notebook

## ❑ Advantages of ELNs over paper notebooks:

1. Sharing
2. Most ELNs have rights management
3. Most ELNs keep track of changes
4. Search functionality
5. Easier to link digital data
6. No issues with handwriting
7. Can be backed up

## ❑ Disadvantages of ELNs over paper notebooks:

1. Require change in working mode
2. Some ELNs have a learning curve

# Note-keeping applications



A few examples

cloud-based

self-hosted

- ☐ Most solutions are moving towards cloud-based services
- ☐ Straight replacement of paper notebooks with some added values (e.g. sharing, searching)
- ☐ Popular in academia due to ease of use
- ☐ Do not provide a solution for data management
- ☐ Do not comply with **21 CFR Part 11 FDA** regulations



# ELNs with database back end



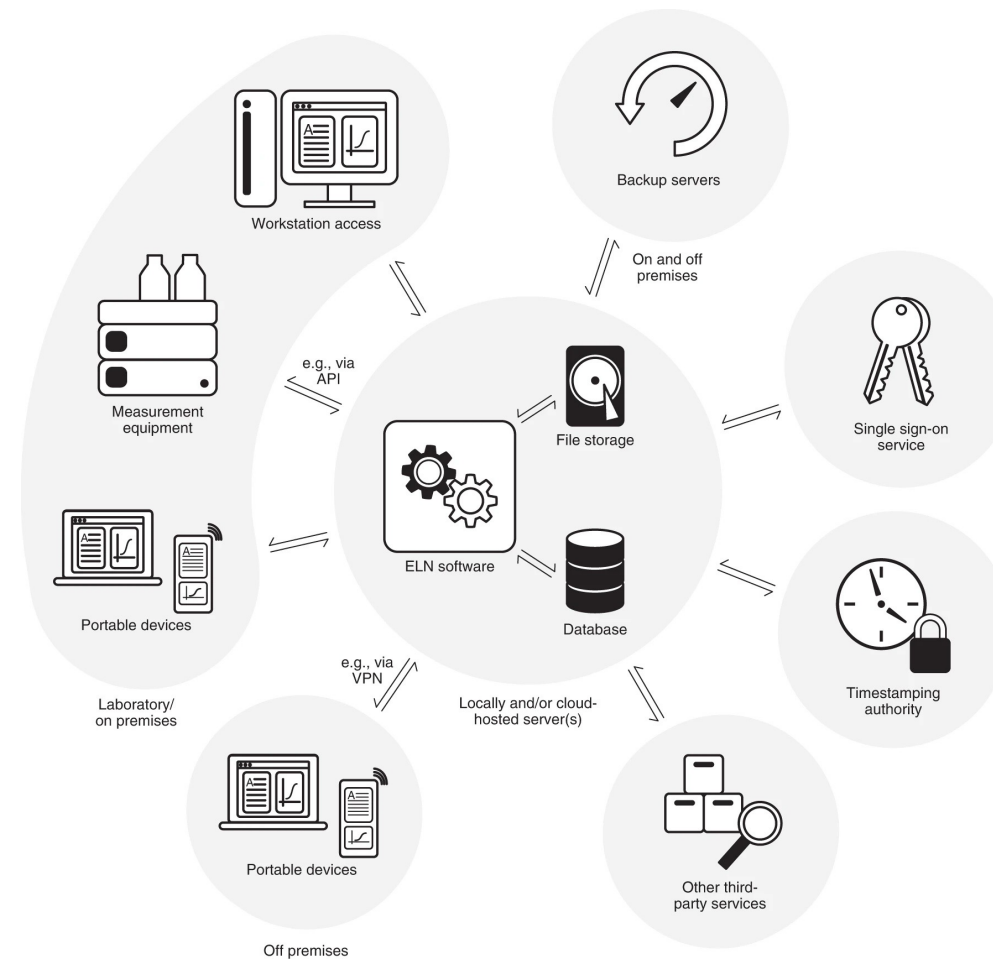
A few examples

cloud-based

self-hosted

- ☐ Structured ELNs compared to note-keeping applications
- ☐ Additional functionalities compared to note-keeping applications (e.g. workflow management, chemical structures drawing, etc)
- ☐ Can be discipline-specific or cross-disciplines
- ☐ Some systems offer an all-in-one solution for RDM
- ☐ Many systems comply with **21 CFR Part 11 FDA** regulations

# Simplified illustration of a commonly used ELN architecture



Considerations for implementing electronic laboratory notebooks in an academic research environment. *Nat Protoc* **17**, 179–189 (2022). <https://doi.org/10.1038/s41596-021-00645-8>

# Which ELN to choose?

❑ The ELNs offer on the market is very large. Answering some of these questions might help you restricting the choice:

1. Is it for personal use or group use?
  2. Can I/we use a cloud-based solution?
  3. Do I/we need specific features?
  4. What do I/we want to do with the ELN? (e.g. only write experimental descriptions, manage samples, manage data – how big?, etc.)
  5. Commercial v. open-source
  6. Budget?
  7. Can I export my data?
- *How to pick an Electronic Laboratory Notebook:* <https://www.nature.com/articles/d41586-018-05895-3>
  - *Harvard University Comparison Grid:* <https://datamanagement.hms.harvard.edu/electronic-lab-notebooks>

# ELN guidelines for academic research labs

- ❑ As part of the national DLCM project, ETH Zurich and EPFL prepared some ELN guidelines, available from: <https://www.dlcm.ch/services/dlcm-eln>

Stages of introduction of an ELN in an academic lab:

1. Initialization
2. Analysis of laboratory needs
3. Pilot phase
4. Production phase (*migration of existing databases or information; customization, integration with measuring instruments, integration with analysis pipelines and workflows*).

# What is needed for a successful introduction of an ELN in an academic lab?

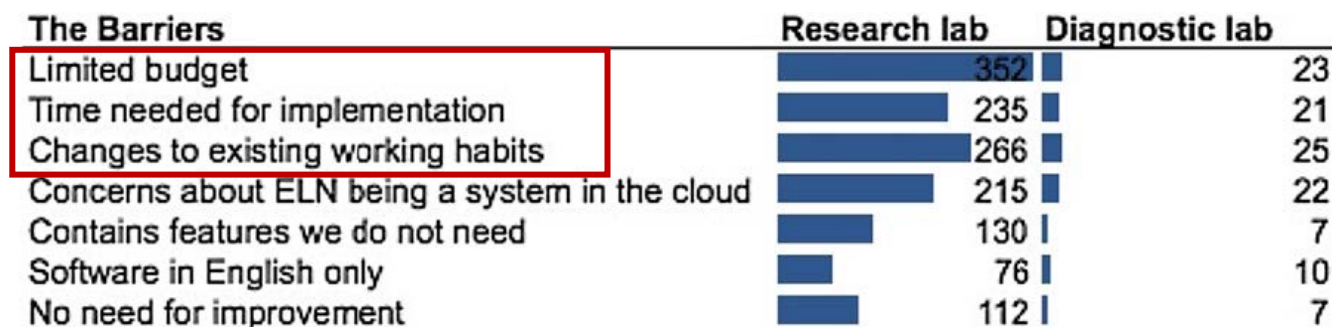
## ❑ Based on our experience at ETHZ:

1. Commitment from the PI
2. Lead person in the lab who drives the process
3. RDM expert support

## ❑ Examples of ELN implementation in academic institutions:

- Considerations for implementing electronic laboratory notebooks in an academic research environment. *Nat Protoc* **17**, 179–189 (2022). <https://doi.org/10.1038/s41596-021-00645-8>
- Institutional ELN/LIMS deployment. *EMBO Rep.* 2020 Mar 4; 21(3): e49862. doi: [10.15252/embr.201949862](https://doi.org/10.15252/embr.201949862)

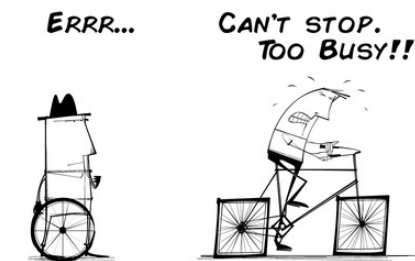
# Slow adoption of ELNs/LIMS in academia



Source: Potential uses of ELNs in Academia survey (University of Southampton). J. Cheminform (2017) 9:31.

In our experience at ETHZ, the main reasons are:

1. Change in working habits needed.
2. Time needed for introduction in a lab.
3. Concerns about data retrieval when leaving the lab.



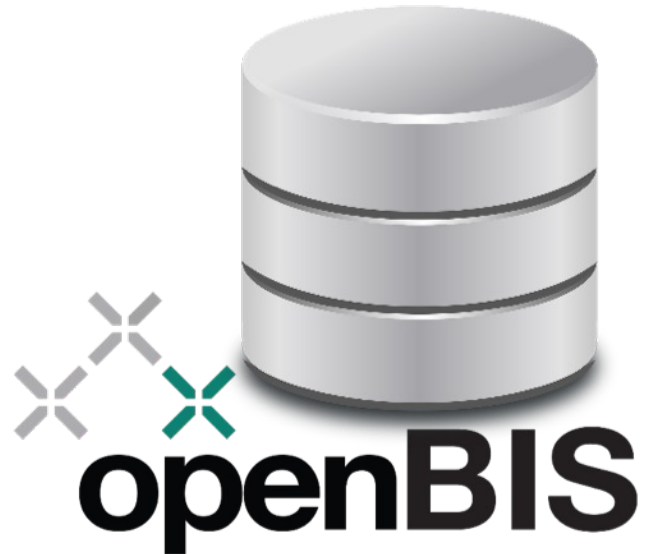
TOO BUSY TO IMPROVE?

WorkCompass

# Questions on Electronic Lab Notebooks?







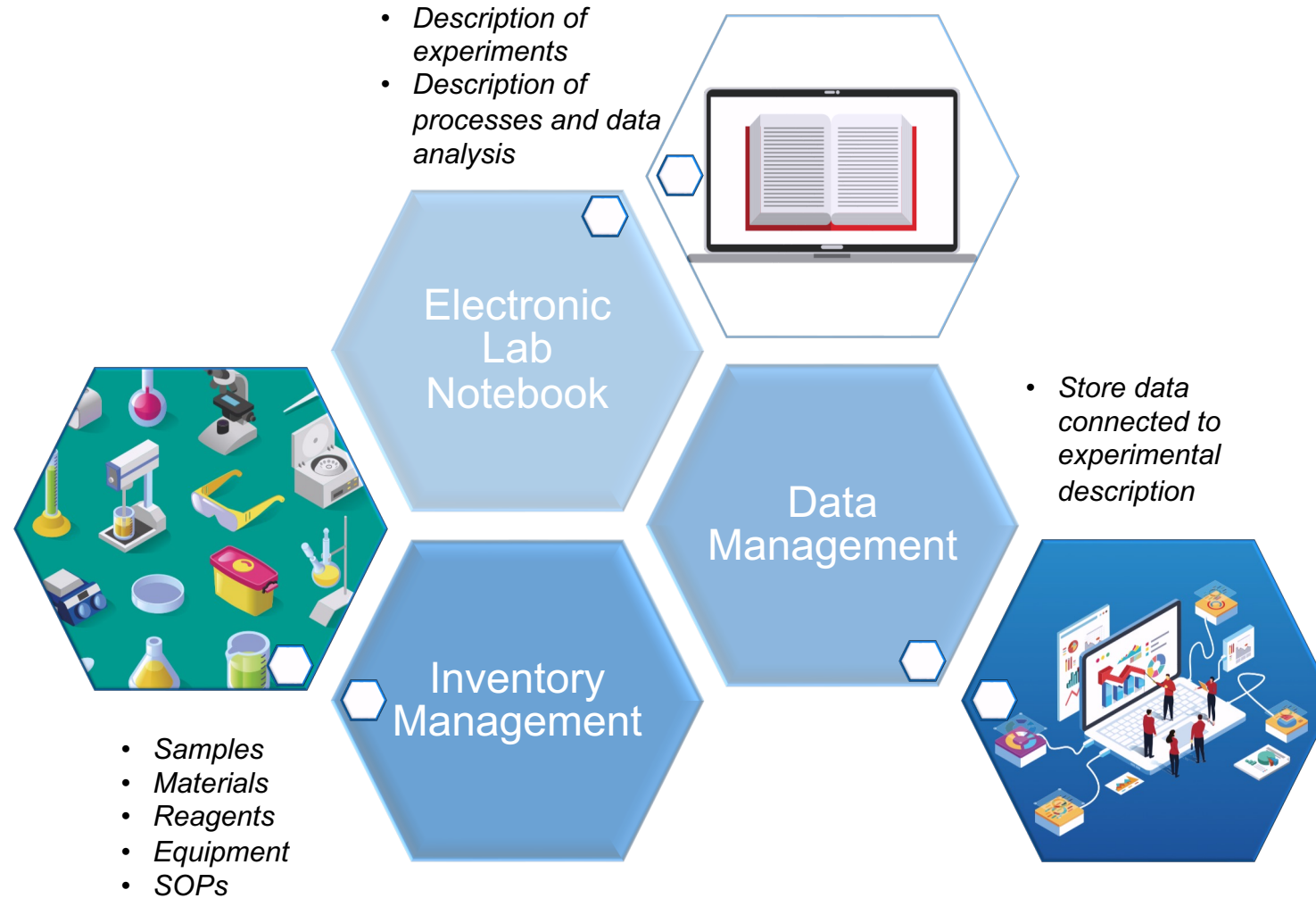
*The ETH Scientific IT Services data management solution for research groups*

# openBIS info

- ❑ Developed at **ETHZ** since 2007.
- ❑ Used in **different quantitative scientific disciplines** by users inside the **ETH domain** and **beyond**.



# openBIS: a complete solution towards FAIR data management



# Inventory management

## Lab equipment

## Lab samples & materials

Collection: Mixers

Rows per page: 10

Code	Name	Identifier	Mixing volume	Log Number	Year of registration	Notes	Type	Space	Parents
EQUIPMENT_MIXERS_6	Ensch 12.12 Plus	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_6	12 L	LOG 136-21.006	2018		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_7	Robert 10	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_7	9 L	LOG 136-22.001			Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_1	Twister evolution through mixer	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_1		LOG 136-28.008	2008		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_2	Twister evolution through mixer	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_2		LOG 308-29.007	2016		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_3	Ensch R 08 M	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_3	75 L	LOG 136-21.003	1991		Equipment Mixers	EQUIPMENT	EQUIPMENT-EQUIPMENT_MIXERS_3 (Change of 01 - example)
EQUIPMENT_MIXERS_4	Ensch RV 11	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_4	350 L	LOG 136-21.004	1998		Equipment Mixers	EQUIPMENT	
EQUIPMENT_MIXERS_5	Ensch R 08 M Space	EQUIPMENTPREPARATION-EQUIPMENT-EQUIPMENT_MIXERS_5	75 L	LOG 136-21.005	2012		Equipment Mixers	EQUIPMENT	

Collection: Chemical admixtures

Rows per page: 10

Code	Name	Identifier	Admixture type	Admixture type other	Date of reception	Manufacturer name	Production date	Mass volumetric density	Solids content	Notes	Type	Space
ADM6	Skagert 755 L	MATERIALS/RAIL-MATERIALS/ADM6		OTHER	2021-06-25 14:20:09 +0000	Sika		900.0			Admixture	WATERU
ADM7	S-0 STARIS S 2100 F SAP	MATERIALS/RAIL-MATERIALS/ADM7		OTHER	2020-04-20 21:00:00 +0000	SASF					Admixture	WATERU
ADM8	Cink 400	MATERIALS/RAIL-MATERIALS/ADM8		Retarder	2017-09-27 07:00:00 +0000						Admixture	WATERU
ADM9	Sika standard superplasticizer	MATERIALS/RAIL-MATERIALS/ADM9		Superplasticizer	2018-12-19 15:48:33 +0000	Sika	2020-01-02 15:48:33 +0000	1080.0			Admixture	WATERU
ADM10	Lithium Carbonate	MATERIALS/RAIL-MATERIALS/ADM10		Accelerator							Admixture	WATERU
ADM11	No W.A. 130	MATERIALS/RAIL-MATERIALS/ADM11		Superplasticizer							Admixture	WATERU

## Lab procedures

## Samples' storage manager

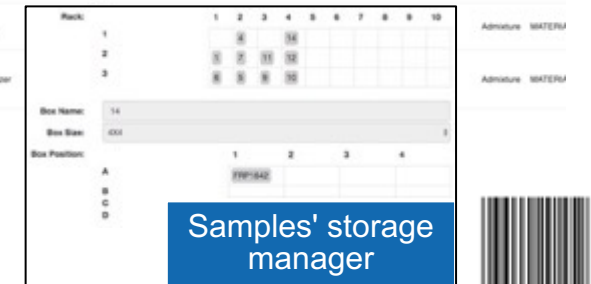
## Barcode reader



Collection: Shrinkage

Rows per page: 10

Code	Name	Identifier	Shrinkage dimensionality	Type of time scale (if measured values)	SCP ID	Notes	Type	Space	P
SHRINKAGE_PROTOCOL_1	SIA 2021, Appendix F	METHODS/MEASUREMENT-PROTOCOLS/SHRINKAGE_PROTOCOL_1	Linear shrinkage	Unstructured grid (varying time steps) linear scale	4003	Protocol for measuring linear shrinkage of concrete specimens according to the Swiss standard SIA 2621 (Appendix F)	Shrinkage Protocol	METHODS	



# Electronic Lab Notebook

Experimental Step: Detection of LexA-ER-B42 induction by flow cytometry

Name: Detection of LexA-ER-B42 induction by flow cytometry

Owner: Diana Ottoz

Experimental goals: Analyze the induction of LexA-ER-B42 in a concentration series of beta-estradiol using a fluorescence readout

Parents

Name	Identifier	Comments	Organism	Storage conditions	Stock concentration	Sterilization	Publication	Protocol type	Materials	Time requirement
1000X Cycloheximide	/MATERIALS/REAGENTS/FRSOB34	diluted 1/1000, treated for 1/2 hour		+4 degrees	1000X	none				
LexA-ER-B42 + target							Ottoz et al., Nucleic Acids Research, 2014			

Personal folder

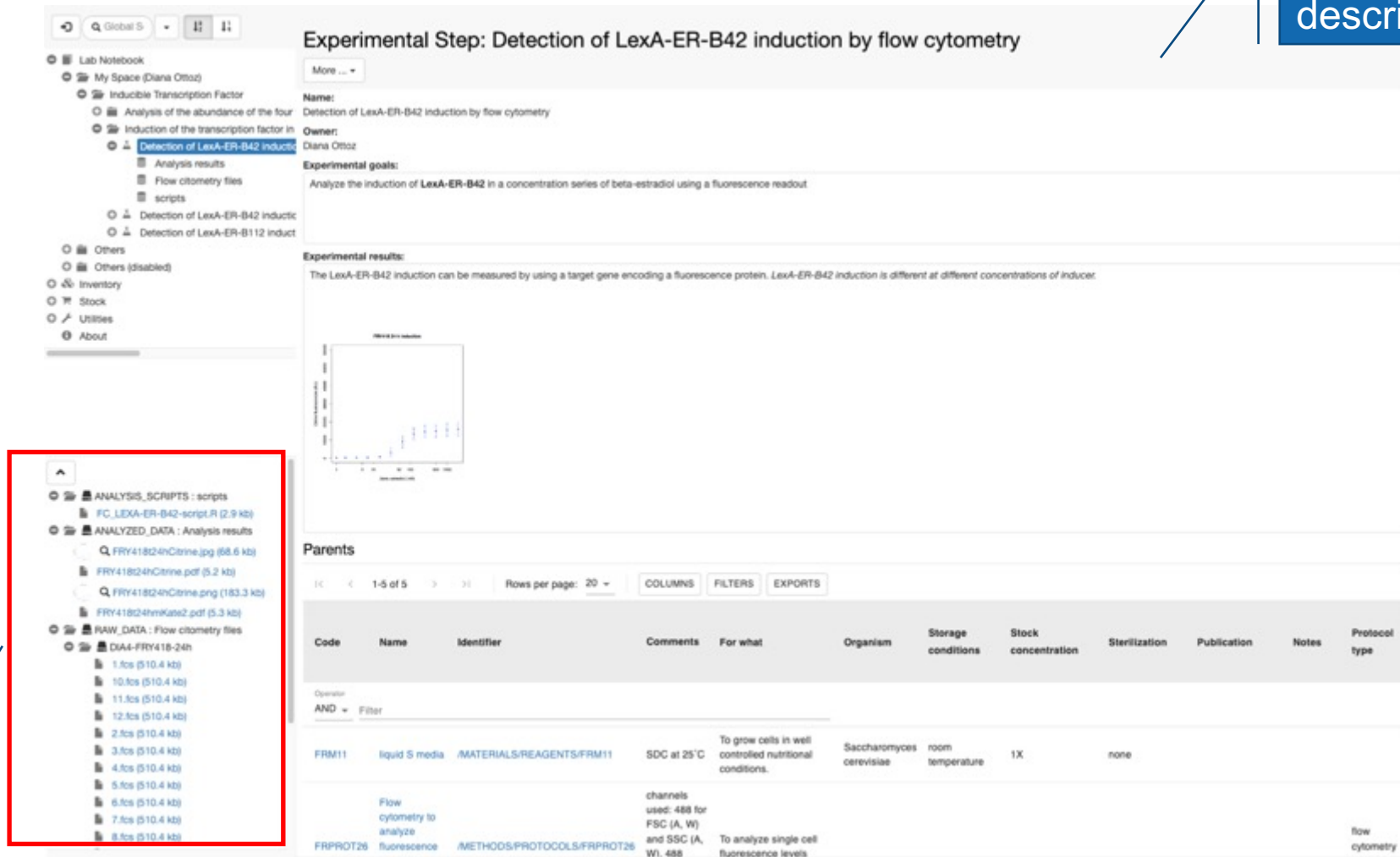
User rights management

Entities relations

# Data management

- ❑ Data are always connected to experimental descriptions

Experimental  
description



**Experimental Step: Detection of LexA-ER-B42 induction by flow cytometry**

**Name:** Detection of LexA-ER-B42 induction by flow cytometry  
**Owner:** Diana Ottol  
**Experimental goals:** Analyze the induction of LexA-ER-B42 in a concentration series of beta-estradiol using a fluorescence readout.

**Experimental results:** The LexA-ER-B42 induction can be measured by using a target gene encoding a fluorescence protein. LexA-ER-B42 induction is different at different concentrations of inducer.

**Parents**

Code	Name	Identifier	Comments	For what	Organism	Storage conditions	Stock concentration	Sterilization	Publication	Notes	Protocol type
FRM11	liquid S media	/MATERIALS/REAGENTS/FRM11	SDC at 25°C	To grow cells in well controlled nutritional conditions.	Saccharomyces cerevisiae	room temperature	1X	none			
FRPROT26	Flow cytometry to analyze fluorescence	/METHODS/PROTOCOLS/FRPROT26	channels used: 488 for FSC (A, W) and SSC (A, W), 488	To analyze single cell fluorescence levels							flow cytometry

**Data**

- ANALYSIS\_SCRIPTS : scripts
  - FC\_LEXA-ER-B42-script.R (2.9 kb)
- ANALYZED\_DATA : Analysis results
  - FRY41824hCitrine.jpg (88.6 kb)
  - FRY41824hCitrine.pdf (5.2 kb)
  - FRY41824hCitrine.png (183.3 kb)
  - FRY41824hCitrine2.pdf (5.3 kb)
- RAW\_DATA : Flow cytometry files
  - DIA4-FRY418-24h
    - 1.fcs (510.4 kb)
    - 10.fcs (510.4 kb)
    - 11.fcs (510.4 kb)
    - 12.fcs (510.4 kb)
    - 2.fcs (510.4 kb)
    - 3.fcs (510.4 kb)
    - 4.fcs (510.4 kb)
    - 5.fcs (510.4 kb)
    - 6.fcs (510.4 kb)
    - 7.fcs (510.4 kb)
    - 8.fcs (510.4 kb)

# Data archiving and unarchiving from LTS via openBIS (for ETHZ)

Archiving Helper

Request archiving

1. Search for the datasets you want to archive:

Search For:

Dataset

Using:

AND

Q

Field Type

Field Name

Comparator Operator

Field Value

+

All

-

2. Check all datasets you want to archive and click the 'Request archiving' button:

<< 1-3 of 3 >>

Rows per page: 10

COLUMNS

FILTERS

EXPORTS

<input type="checkbox"/>	Should be archived	Entity Kind	Name	Identifier	Entity Type	Code	Experiment/Collection	Object	Registrator	Registration Date	Modifier	Modification Date	Size
<input type="checkbox"/>	<input checked="" type="checkbox"/>	DataSet		20220223142603807-25	ANALYSIS_NOTEBOOK	20220223142603807-25	GENERAL_PROTOCOLS	GEN1	vkovtun	2022-02-23 14:26:04	vkovtun	2022-02-23 14:26:04	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	DataSet	demo data	20220303215515409-87	ATTACHMENT	20220303215515409-87	DEFAULT_EXPERIMENT		barillac	2022-03-03 21:55:15	barillac	2022-03-03 21:55:15	
<input type="checkbox"/>	<input type="checkbox"/>	DataSet	demo data 2	20220303215550279-88	RAW_DATA	20220303215550279-88	DEFAULT_EXPERIMENT	EXP3	barillac	2022-03-03 21:55:50	barillac	2022-03-03 21:55:50	

Operator

AND

Filter

Unarchiving Helper

Unarchive

ⓘ Data sets are usually archived together in bundles. Unarchiving one means that all data sets of the bundle are unarchived, too. Note, that all these data sets are unarchived on a scratch disk. They can be deleted without further notice. If they are needed again they have to be unarchived again.

1. Search for the datasets you want to unarchive:

Search For:

Dataset

Using:

AND

Q

Field Type

Field Name

Comparator Operator

Field Value

+

Property

Registrator [ATTR.REGISTRATOR]

thatContainsFirstName (First Name)

Elisa

-

2. Check all datasets you want to unarchive and click the 'Unarchive' button:

<< 1-10 of 80 >>

Rows per page: 10

COLUMNS

FILTERS

EXPORTS

<input type="checkbox"/>	Should be unarchived	Name	Entity Type	Code	Experiment/Collection	Object	Registrator	Registration Date	Modifier	Modification Date	Size	Datasets in Bundle	Bundle Size
<input type="checkbox"/>	<input type="checkbox"/>		RAW_DATA	20180705154320307-47214	ACETYLATION_IN_METABOLIC_STRESS_DEL_EXP2	EXP1224	edultz	2018-07-05 15:43:21	edultz	2018-07-05 15:43:21	2.8 Mb	20	12.6 Gb
<input type="checkbox"/>	<input type="checkbox"/>		RAW_DATA	20190401153645830-52592	MISCELLANEOUS_MEX67_CARINA	EXP1137	edultz	2019-04-01 15:36:46	edultz	2019-04-01 15:36:46	20.1 Mb	14	37.0 Gb
<input type="checkbox"/>	<input type="checkbox"/>	200505 Data sent to Lucien	RAW_DATA	20200509214136812-66780	SPECIALIZED_NPCS_EXP_2	ENTRY1	edultz	2020-05-09 21:41:37	edultz	2020-05-09 21:41:37	33.6 Mb	14	37.0 Gb
<input type="checkbox"/>	<input type="checkbox"/>		RAW_DATA	2018113214319106-50521	MISCELLANEOUS_MEX67_CARINA	EXP1215	edultz	2018-11-13 21:43:19	etiserver	2018-11-13 21:43:19	672.9 kb	3	13.0 Gb
<input type="checkbox"/>	<input type="checkbox"/>		RAW_DATA	20190618085503411-53794	ACETYLATION_IN_METABOLIC_STRESS_DEL_EXP2	EXP1912_3	edultz	2019-06-18 08:55:03	edultz	2019-06-18 08:55:03	5.8 Mb	13	48.7 Gb
<input type="checkbox"/>	<input type="checkbox"/>		RAW_DATA	20190618085523692-53795	ACETYLATION_IN_METABOLIC_STRESS_DEL_EXP2	EXP1912_2	edultz	2019-06-18 08:55:25	edultz	2019-06-18 08:55:25	13.2 Mb	13	48.7 Gb

Operator

AND

Filter



# Access to archived data (for ETHZ)

The screenshot shows a web interface with a sidebar on the left containing a navigation menu with items like Lab Notebook, Inventory, Stock, Utilities, Jupyter, New J, User F, Objec, Vocab, Advan, Archiv, Unarc, Exports, Storage Manager, User Manager, Trashcan, and Settings. The main area at the top is titled 'Experimental Step: 160129\_hexandiol\_microscopy' and includes buttons for '+ New', 'Edit', 'Upload', and 'More ...'. A large white error box in the center reads: 'Error Resource 'original/KWY5184\_Nucleolus/KWY5184\_Nop1\_0%\_10min.nd2' is currently unavailable. It might be archived.' The phrase 'It might be archived.' is enclosed in a red rectangular box. Below the error box, there are four small circular microscopy images. The bottom right section contains metadata for the experiment, including Name, Owner, Experimental goals, Experimental results, and Start date. On the left, a list of files is shown under the folder 'KWY5184\_Nucleolus', with the file 'KWY5184\_Nop1\_0%\_10min.nd2 (43.9 Mb)' highlighted. A blue arrow points from this file to the error message. A blue box on the far left contains the text 'Archived data still visible' with a line pointing to the file list.

Experimental Step: 160129\_hexandiol\_microscopy

**Error**

Resource 'original/KWY5184\_Nucleolus/KWY5184\_Nop1\_0%\_10min.nd2' is currently unavailable. It might be archived.

General

**Name:**  
160129\_hexandiol\_microscopy

**Owner:**  
Dultz, Elisa

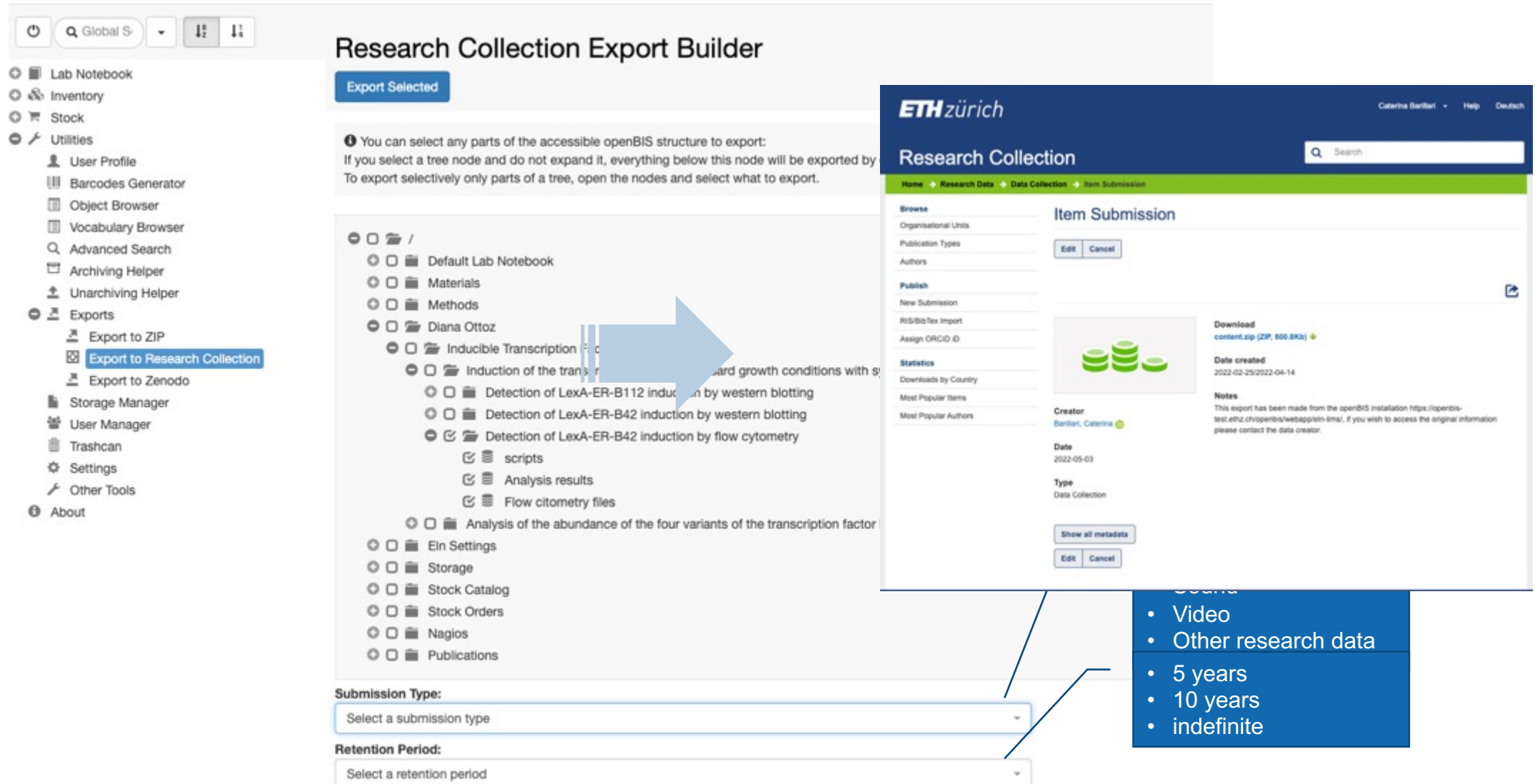
**Experimental goals:**  
test effect of hexanediol on position of gene locus (POA1 in Glucose or Galactose) and on nucleolus

**Experimental results:**  
Nucleolus: fragmentation of nucleolus in multiple small round spots in 5% hexanediol repeated from previous experiment; POA1 locus position: low count number especially in higher concentrations of hexanediol; this is most likely due to increased mishaping of the nucleus at higher concentrations; trend is towards more at the periphery in higher concentration as in previous experiment, however, this could also be an indirect result of the shape changes of the nucleus.

**Start date:**  
2016-01-26 13:02:00 +0100

Archived data still visible

# Data publication: export to ETH Research Collection (for ETHZ)



**Research Collection Export Builder**

Export Selected

You can select any parts of the accessible openBIS structure to export:  
If you select a tree node and do not expand it, everything below this node will be exported by default.  
To export selectively only parts of a tree, open the nodes and select what to export.

**Exports**

- Export to ZIP
- Export to Research Collection**
- Export to Zenodo

**Tree Structure:**

- Default Lab Notebook
- Materials
- Methods
- Diana Ottoz
  - Inducible Transcription Factor
    - Induction of the transcription factor under standard growth conditions with ...
      - Detection of LexA-ER-B112 induction by western blotting
      - Detection of LexA-ER-B42 induction by western blotting
      - Detection of LexA-ER-B42 induction by flow cytometry
        - scripts
        - Analysis results
        - Flow cytometry files
    - Analysis of the abundance of the four variants of the transcription factor

- Ein Settings
- Storage
- Stock Catalog
- Stock Orders
- Nagios
- Publications

**Submission Type:**  
Select a submission type

**Retention Period:**  
Select a retention period

**ETH zürich Research Collection**

Home → Research Data → Data Collection → Item Submission

**Item Submission**

Download  
content.zip (ZIP, 800.8Kb)

Date created  
2022-02-25/2022-04-14

Notes  
This export has been made from the openBIS installation https://openbis-test.ethz.ch/openbis/webapp/eth-ims/. If you wish to access the original information please contact the data creator.

**Submission Type:**  
Select a submission type

**Retention Period:**  
Select a retention period

- Video
- Other research data
- 5 years
- 10 years
- indefinite

# Data publication: export to Zenodo

**Global S**

- Lab Notebook
- Inventory
- Stock
- Utilities
  - User Profile
  - Barcodes Generator
  - Object Browser
  - Vocabulary Browser
  - Advanced Search
  - Archiving Helper
  - Unarchiving Helper
- Exports
  - Export to ZIP
  - Export to Research Collection
  - Export to Zenodo**
- Storage Manager
- User Manager
- Trashcan
- Settings
- Other Tools
- About

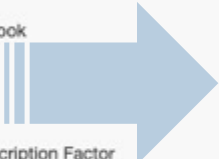
## Zenodo Export Builder

**Export Selected**

**Information:** You can select any parts of the accessible openBIS structure to export: If you select a tree node and do not expand it, everything below this node will be exported. To export selectively only parts of a tree, open the nodes and select what to export.

**Publication time constraint:** After the resource has been exported it should be published in Zenodo UI within 2 hours. Otherwise, the publication metadata will not be registered in openBIS.

- Default Lab Notebook
- Materials
- Methods
- Diana Ottoz
  - Inducible Transcription Factor
    - Induction of the transcription factor in standard growth conditions
      - Detection of LexA-ER-B112 induction by western blotting
      - Detection of LexA-ER-B42 induction by western blotting
      - ☒ Detection of LexA-ER-B42 induction by flow cytometry
        - scripts
        - Analysis results
        - Flow cytometry files
    - Analysis of the abundance of the four variants of the transcription factor



**Submission Title (\*):**

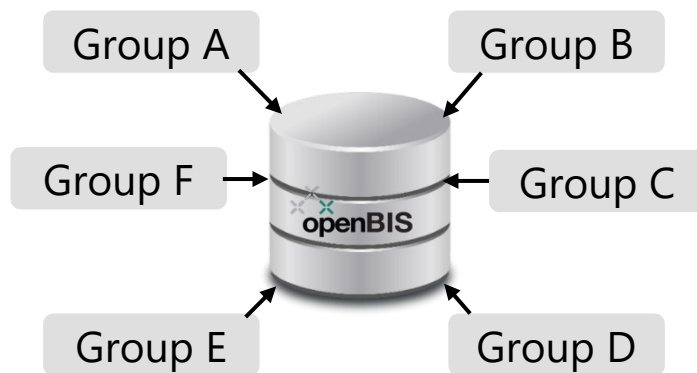
• Title



# RDM services @ ETHZ

- ❑ ETH-wide services provided since 2018

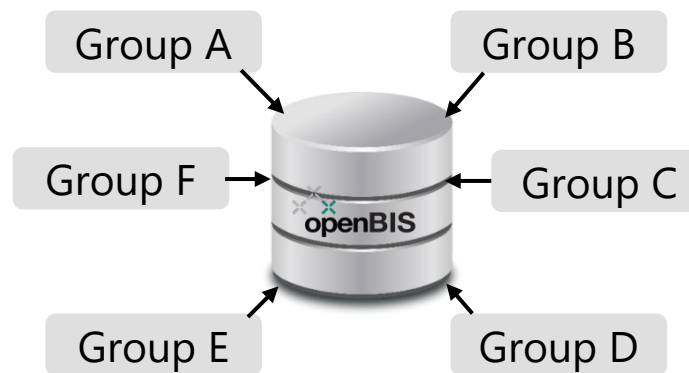
## ETH Research Data Hub (RDH)



For all ETH groups

- Centrally managed
- Shared resource
- Limited Customization
- No sensitive data

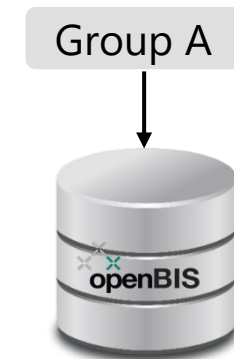
## ETH Departmental Data Hub (DDH)



For groups of a certain department/institute

- Centrally managed
- Shared resource
- Department customization

## ETH Research Data Node (RDN)



For single ETH research groups

- Individually managed
- Dedicated resource
- Individual customization

# A national RDM service for the academic community

- ❑ Service establishment funded by a swissuniversities P5 project
- ❑ Project: 2018-2020
- ❑ Service start: 2021



## Cloud-hosted openBIS

- Virtual servers per research group, institute or institution
- Optionally with JupyterHub server for analytics



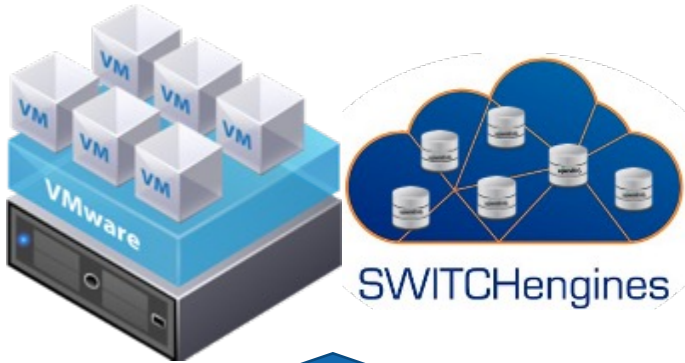
## Self-hosted openBIS

Support for set up on local IT infrastructure



**Current customers**

# What do the SIS's RDM services provide?



Installation on ETH ID infrastructure or SWITCHEngines



Maintenance + upgrades



Consulting



Tailored data modelling



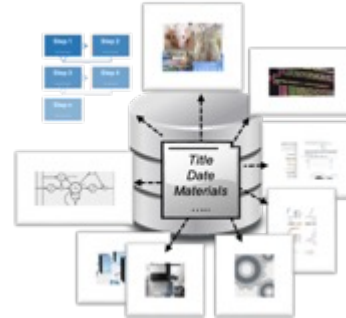
- Trainings
- Users' support



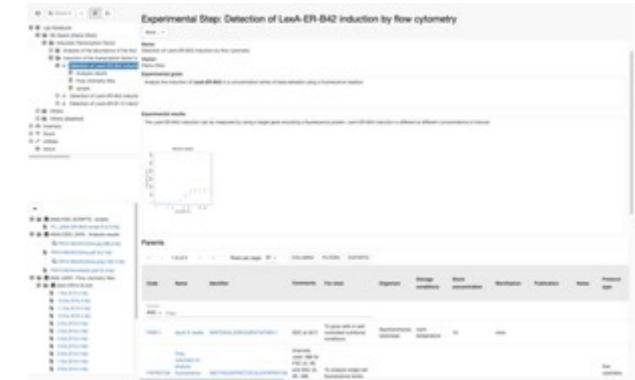
# How do the SIS's RDM services help researchers?



## Support for all phases of the data life cycle



## Single platform for data & information management



## Structured & standardized data organization



## Easy access to LTS (ETHZ)



Research Collection



## Simplified data publication & sharing



## Support for DMP writing

<https://sis.id.ethz.ch/services/rdm/SNSF-DMP-openBIS-template.pdf>



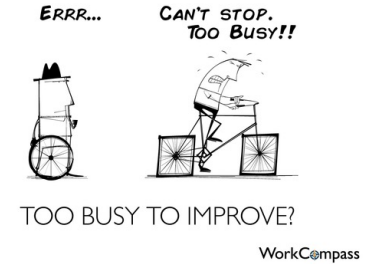
# Questions on openBIS and RDM services provided by SIS?



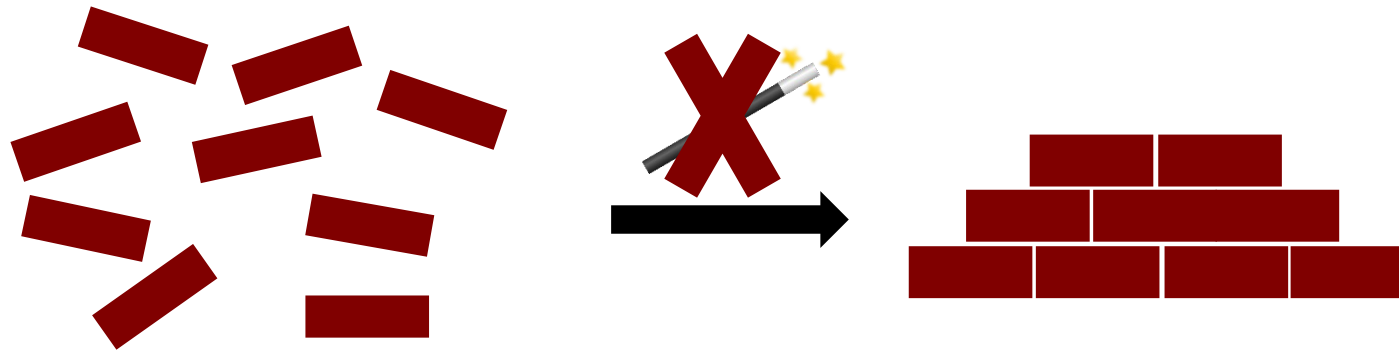
# Code management

- ❑ Code management is an integral part of active data management
- ❑ We did not forget about it, it is a topic for next week's workshop on ***Reproducible scientific computing and data analysis***

# Take home messages



- ❑ Data management should be an integral part of your work
- ❑ Try to establish standard procedures in the group, if not yet in place
- ❑ RDM requires **WORK & TIME**, but the time spent on this is an **investment** for the future!



**Contact us for consultations / trainings on:** data management, version control, reproducible computational workflows or data science support

[sis.helpdesk@ethz.ch](mailto:sis.helpdesk@ethz.ch)



# Why is this all important for me?

Ok, time to start writing up my **PhD thesis**. Wow, 3 years of work. Let's begin..

I have a request from a reader to check the analysis I presented in this paper. Now, where are they?

These results will make an excellent **publication**! Let me find all the data I need



My ex-PhD supervisor wants to know where my data are, so their new student can carry on my work

❑ If you have well annotated, organized and backed up your data over time, all this will be easy!

# Contacts

**Caterina Barillari**

caterina.barillari@id.ethz.ch

**Henry Lütcke**

henry.lutcke@id.ethz.ch

**sis.helpdesk@ethz.ch**

<https://sis.id.ethz.ch/>



[https://twitter.com/ETH\\_SIS?lang=en](https://twitter.com/ETH_SIS?lang=en)

Any final questions on what we have seen today?



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