04.04.2023

Research Data Management – The Basics

Cantini, Federico
Felder, Fabian
Minotti, Carlo

These are your trainers today!

Federico Cantini

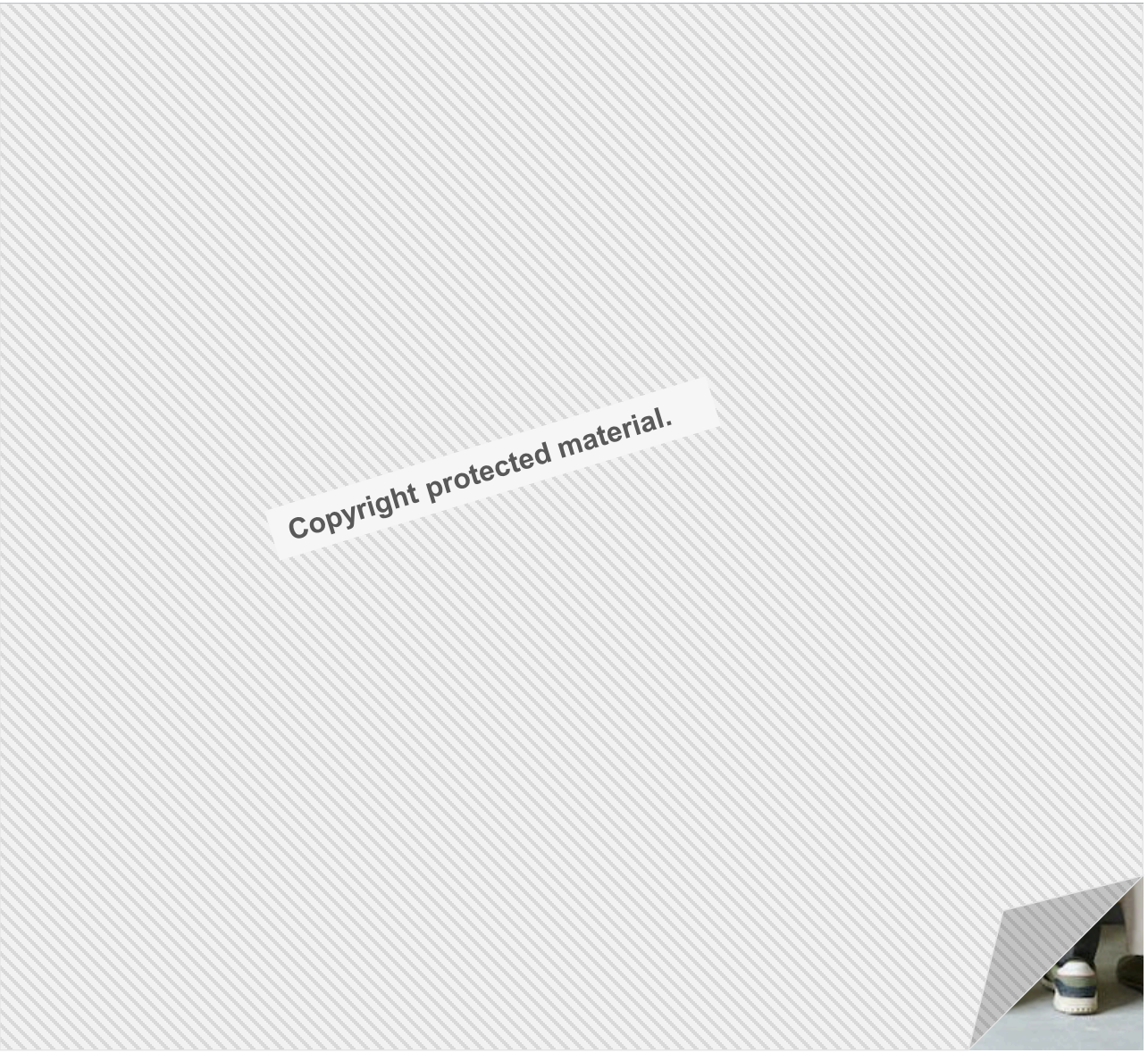
- Software Developer
- Technical Lead at Lib4RI

Fabian Felder

- Open Science specialist
- Group Leader IT services and E-resources at Lib4RI

Carlo Minotti

- Software Engineer
- PSI Data Management Group



Who are you and why are you here?

https://www.pexels.com/photo/group-of-people-standing-indoors-3184396/

Learning Aims

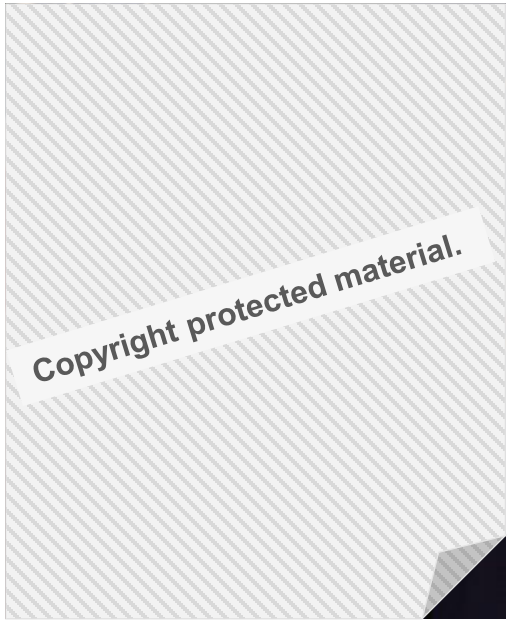
- Life cycle of research data
- Adequate metadata documentation for your code and data
- Storing and publishing data
- Using OpenBIS (ELN) and writing Data Management Plans (DMP)

Program

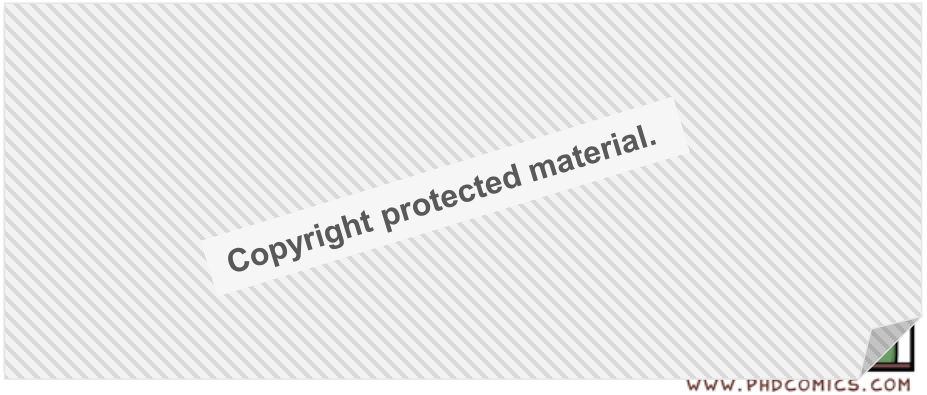
Topic	Speaker	Time
Introduction	Fabian Felder	9.00 - 9.15
Policies, Incentives & the Research Data Life Cycle	Fabian Felder	9.15 - 9.45
Collect & Store	Federico Cantini	9.45 - 10.15
Break		10.15 - 10.30
Evaluate & Archive	Fabian Felder	10.30 - 10.40
RDM Services & Support at PSI	Carlo Minotti	10.40 - 11.00
Plan & Design	Everyone	11.00 - 11.45

Why is data and associated metadata important?

Why is data and associated metadata important?



«We kill people based on metadata» (2014), Michael V. Haden, director of CIA 2006-2009



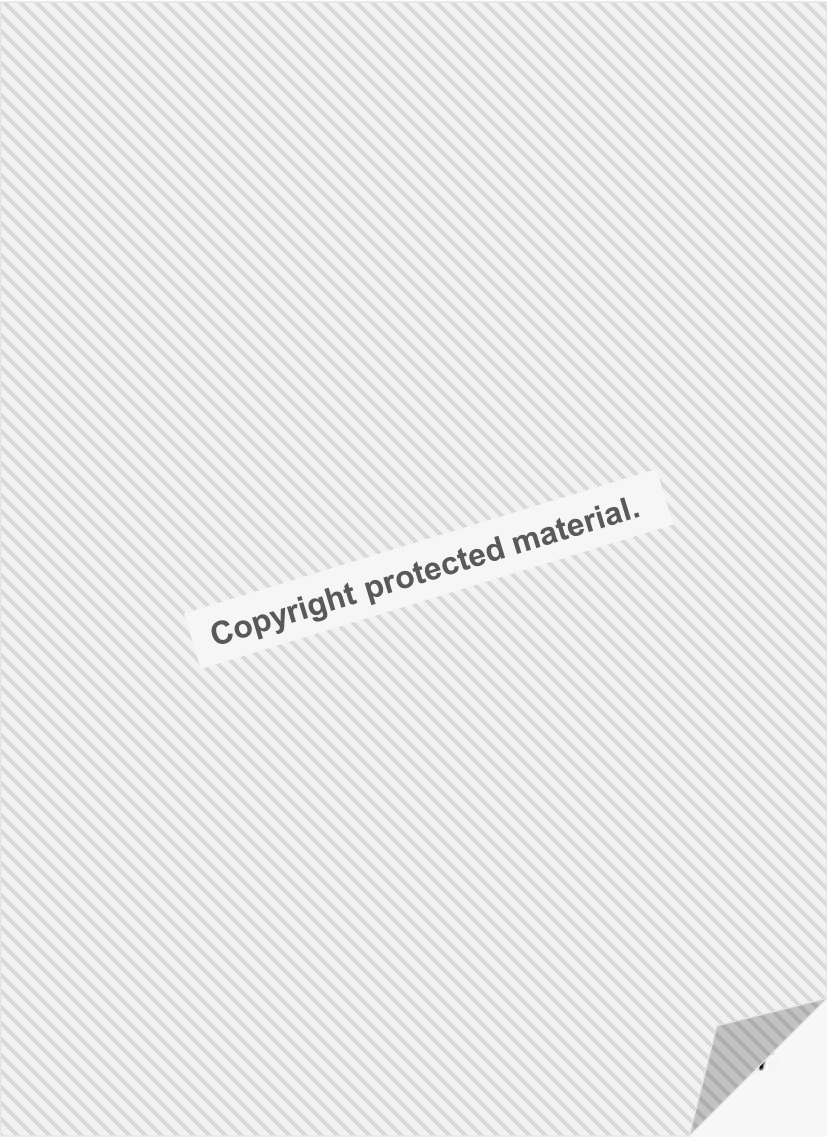
Cham, J. G., «Scratch: A context-changing framework for contextualizing nano informatic structures» (2014), International Journal of Temporal Deflective Behaviour, 4 (1689), p. 432.

Why is data and associated metadata important?

No clean metadata

=

Limited access to Data



Source: www.fosteropenscience.eu/project

Why is data and associated metadata important?

Proper metadata tagging

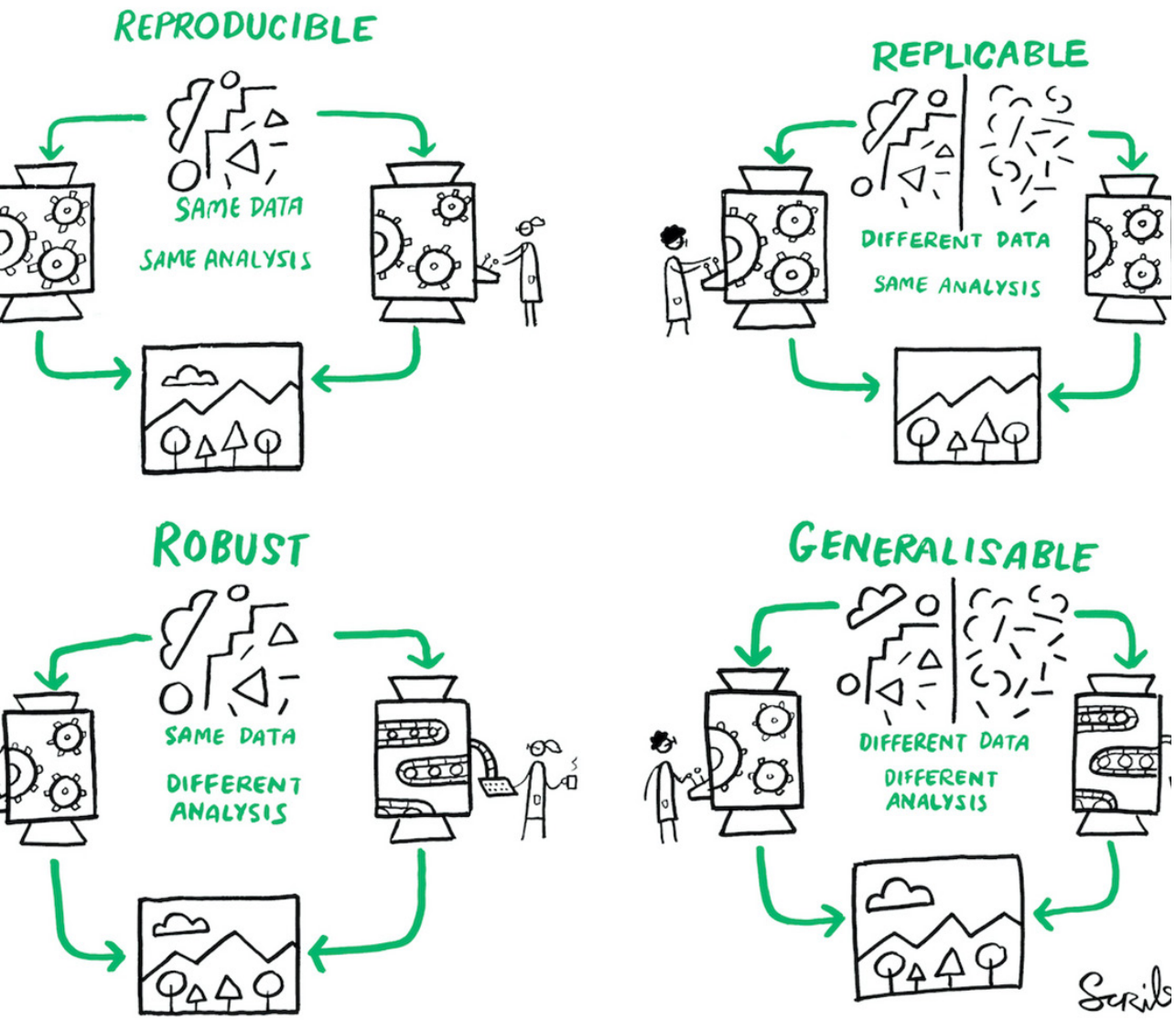
and

research description

is

time consuming

Reproducibility



Reporducibility

Scriberia, "Reproducible Research", *The Turing Way*, CC-BY, DOI: 10.5281/zenodo.3332807

A Handbook for Reproducible Data Science, https://the-turing-way.netlify.app/welcome.html

LEGO® Metadata for Reproducibility

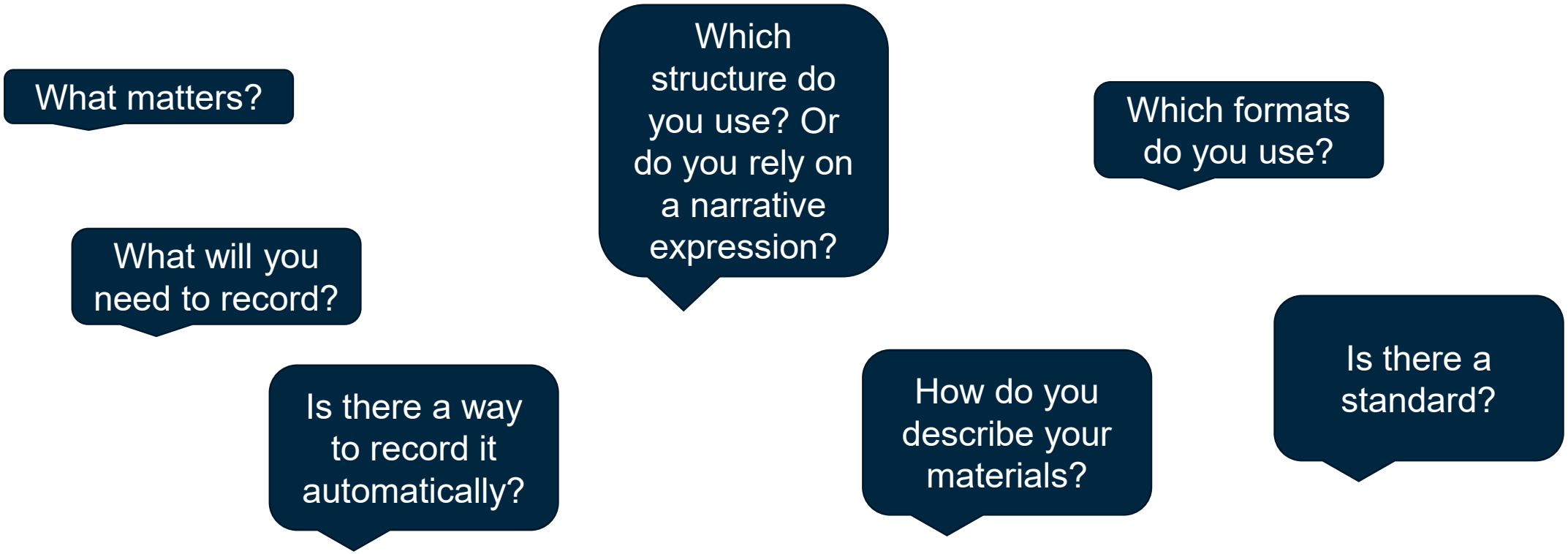
Group A builds Car

Group A documents build

Group B rebuilds the car

Donaldson, Mary and Matt Mahon, «Lego: Metadata for reproducibility», 10.5281/zenodo.3685685.

LEGO® Metadata for Reproducibility



Donaldson, Mary and Matt Mahon, «Lego: Metadata for reproducibility», 10.5281/zenodo.3685685.

FAIR principles – A lot of Metadata

Findable

F1 (Meta)data are assigned a globally unique and persistent identifier
F2 Data are described with rich metadata
F3 Metadata clearly and explicitly includes the identifier
F4 (Meta)data are registered or indexed in a searchable resource

Accessible

A1 (Meta)data are retrievable by their identifier using a standardised communications protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorisation procedure, when necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperale

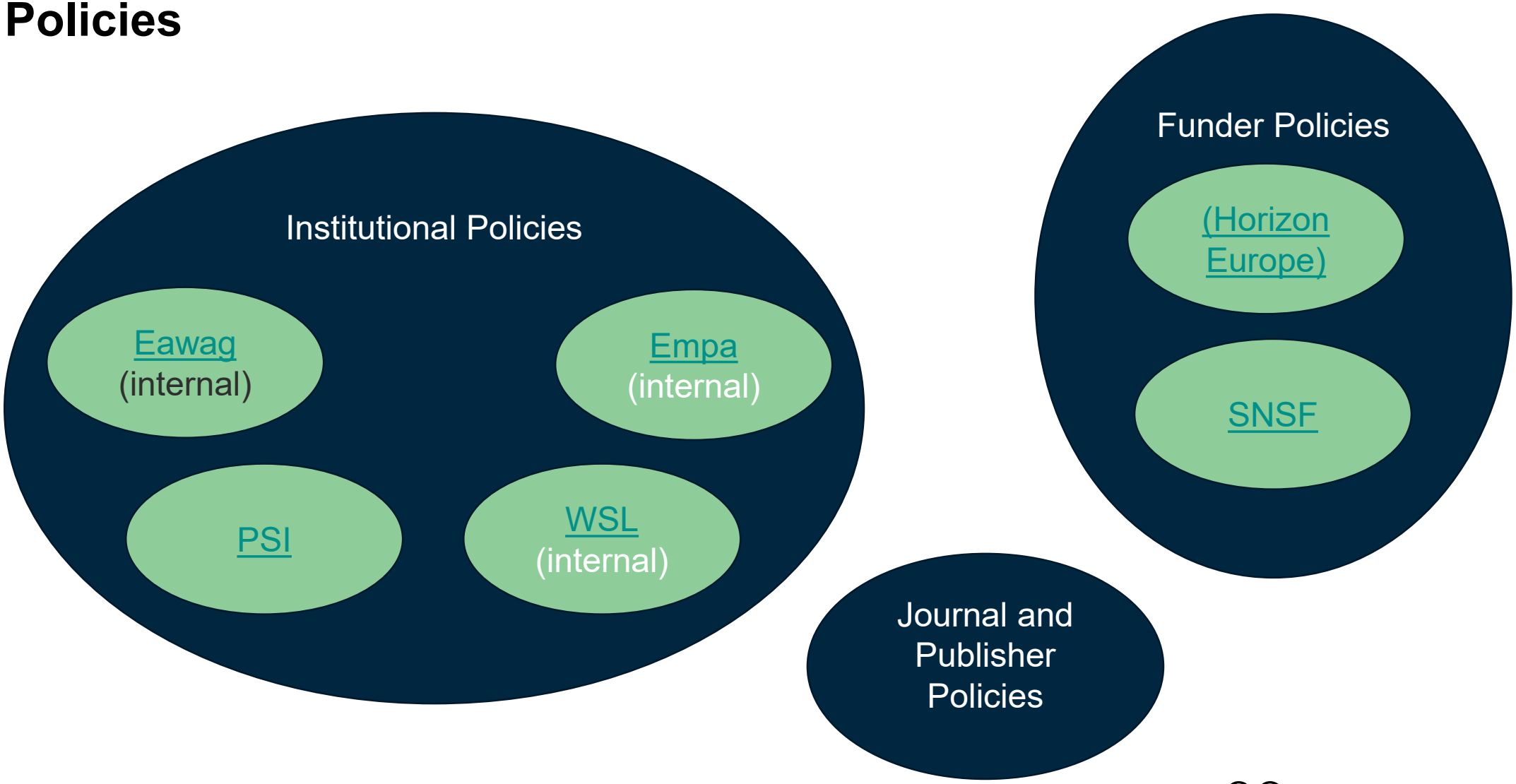
I1 (Meta)data use formal, accessible, shared, and broadly applicable language for knowledge representation
I2 (Meta)data use vocabularies that follow FAIR principles
I3 (Meta)data indlude qualified references to other (meta)data

Reusable

R1 (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1 (Meta)data are released with a clear and accessible data usage license
- R1.2 (Meta)data are associated with a detailed provenance
- R1.3 (Meta)data meet domain-relevant community standards

Policies

Policies


Institutional Policies
Eawag (internal)
Empa (internal)
PSI
WSL (internal)
Funder Policies
(Horizon Europe)
SNSF
Journal and Publisher Policies

Policies

Compliance

Project Manager/
Group Leader

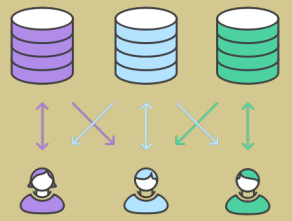
DMP

Required for
Funders

As open as
possible,

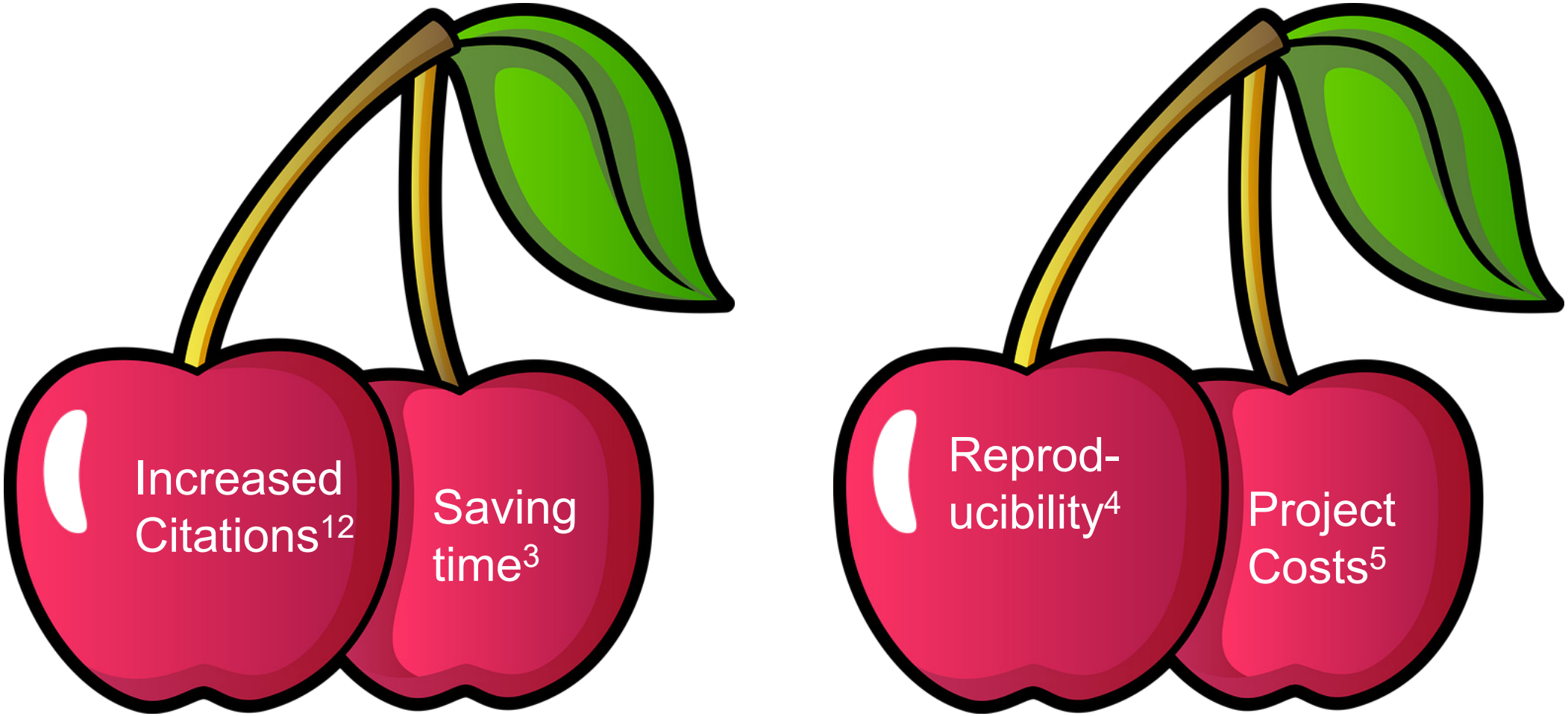
as closed as
necessary.

Repositories



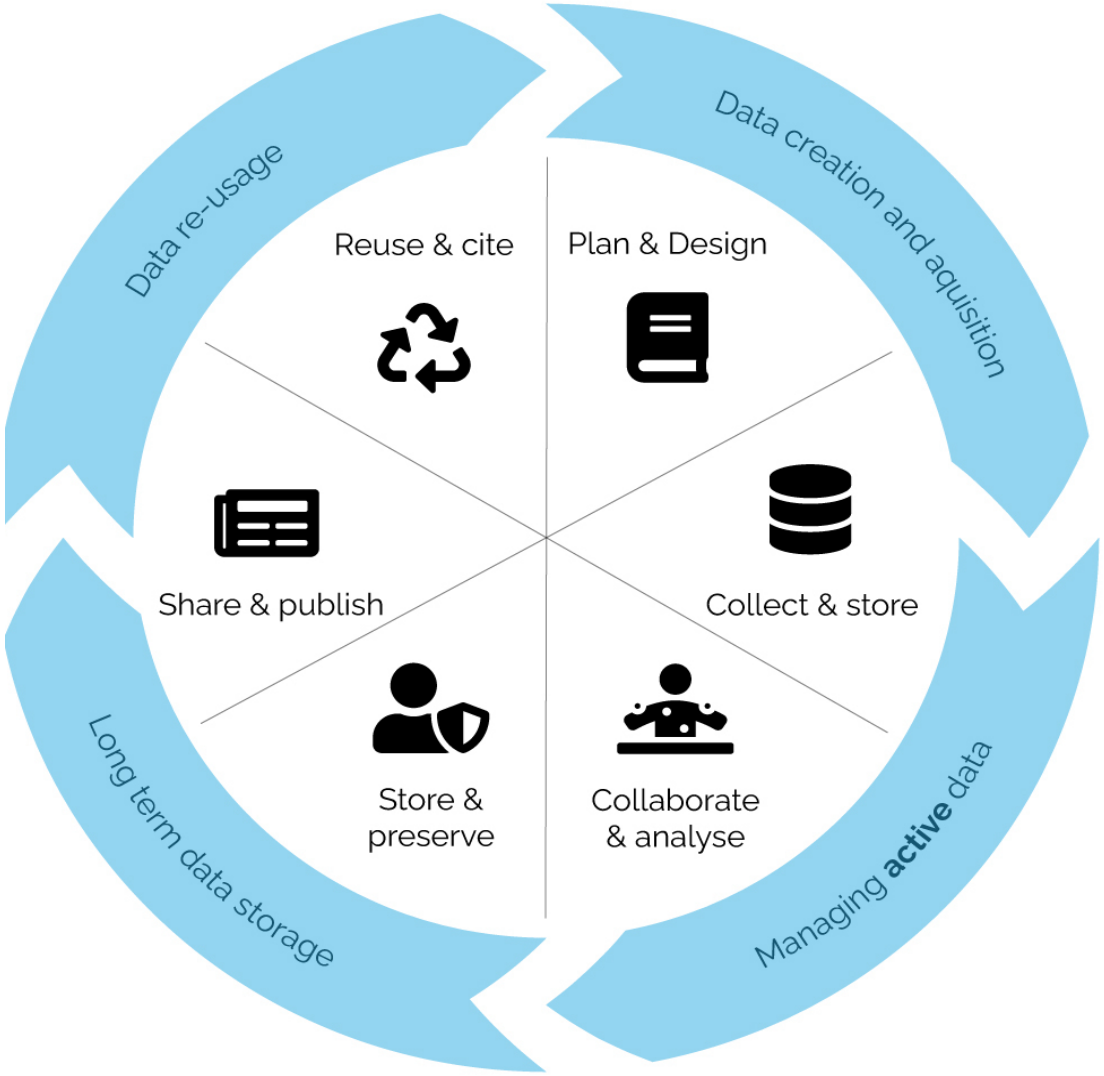
FAIR and non-
commercial

Policies


Increased Citations[12]
Saving time[3]
Reprod-ucibility[4]
Project Costs[5]

Research Data Life Cycle

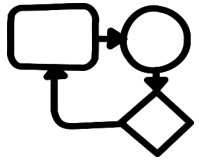
Research Data Life Cycle



Collect & Store

Collect & Store: Research Data Types

- Data: observational, experimental, simulation, canonical…



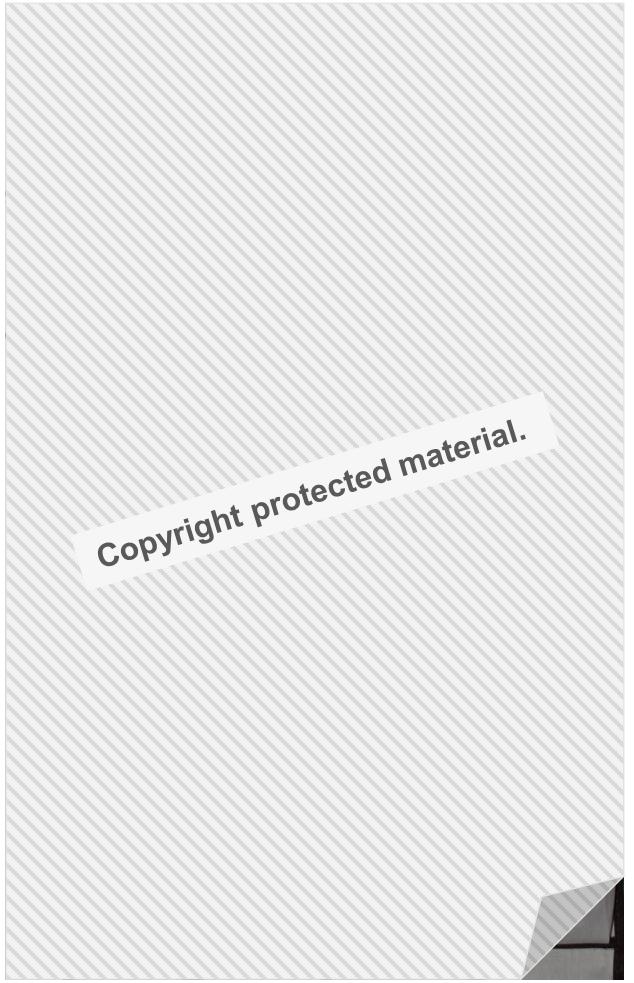
- Software: applications, software and analysis scripts…

- Metadata: Structured information associated with data (Read-me files, Publication keywords, File and folder names)

Collect and Store

Organize data so that:

- **You** can find it
- **Your coworkers** can find it
- You can easily **share** it
- It's **ready** for **archiving/publishing**



Collect and Store: File Formats (recommendation)

Data type	Recommended file formats
Text	• PDF/A • Plain Text coded as ACII. UTF-8 or UTF-16 • XML
Spreadsheet	• CSV
Images	• TIFF (uncompressed or lossless compressed) • PNG
Code	• Languages with free environments (e.g. Py or R UTF-8 format of ASCII text)
Audio	• FLAC • Wav

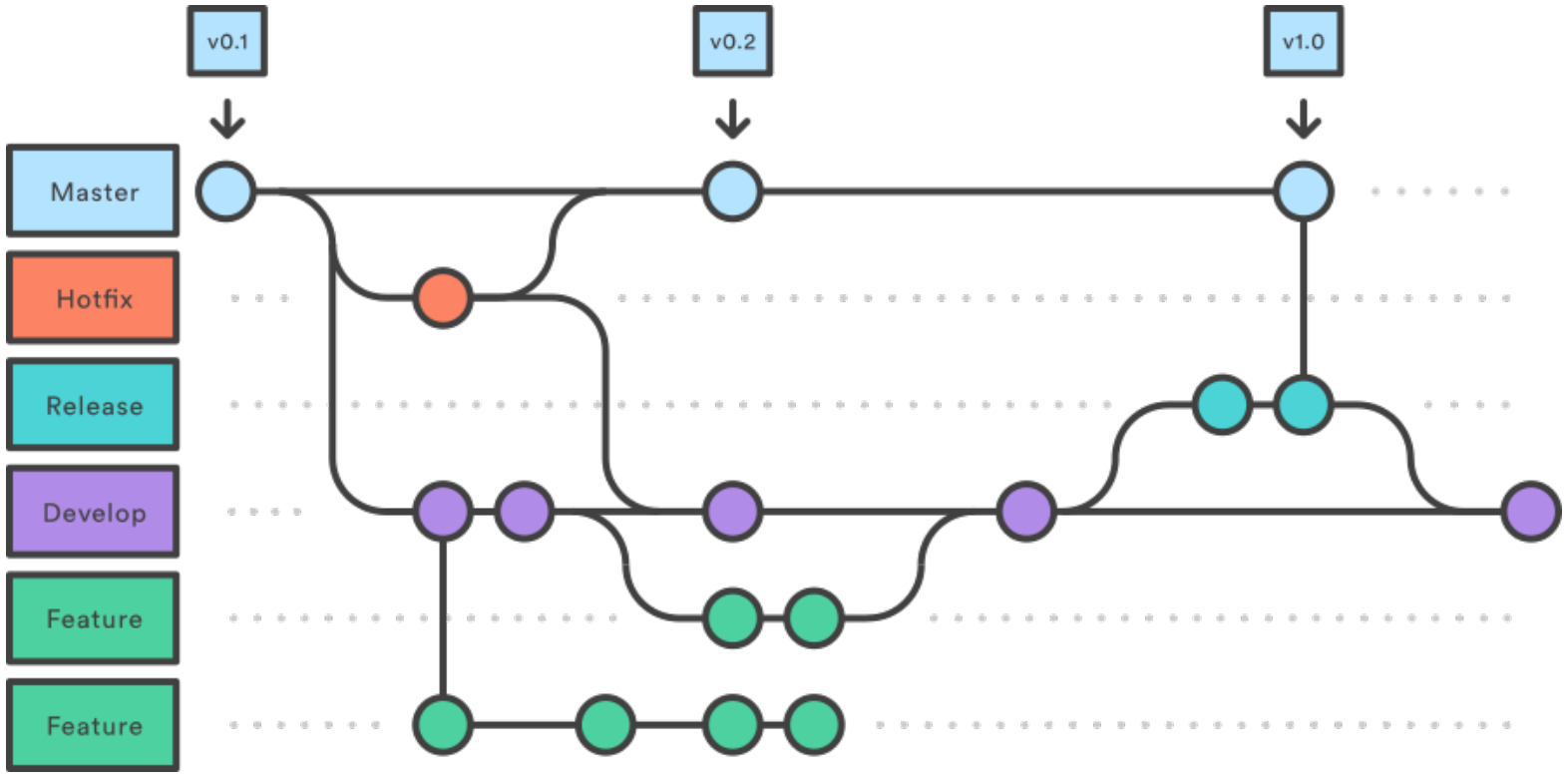
Open and lossless formats
If you are using a proprietary format, think about adding an additional format

Collect and Store: File Naming

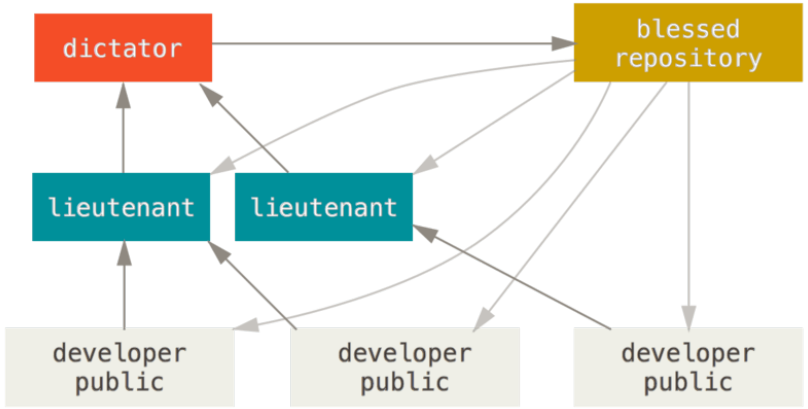
- Use unique names referencing content
- Limit to 42 characters (preferably less)
- Use ASCII characters, no spaces, points or special characters, e.g. ~!@#$%^&*()[]{}<>’;,’»/
- Include dates and label versions
- Use names to order files:
  - Either, use Dates YYYY-MM-DD or YYYYMMDD (according to ISO 8601) at the beginning to enable chronological order
  - Or, use Versioning with leading zeroes to enable numerical order (enables versions to go beyond 9 without disrupting order)

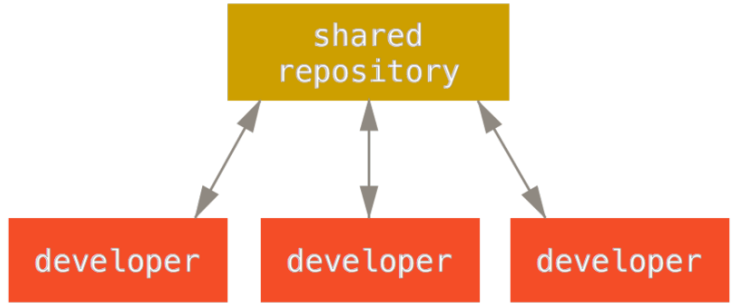
- If you have started with your project use *Bulk Rename Utility* (Windows) or *Renamer 6* (Mac), *Rename/Thunar Bulk Rename* (GNU/Linux)

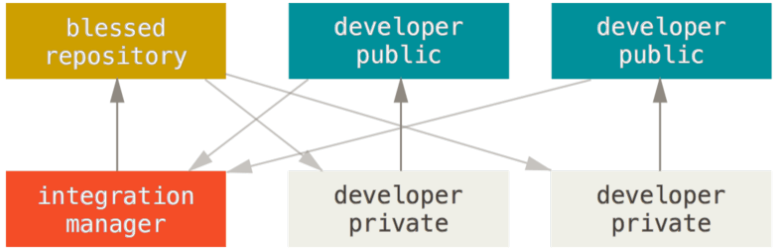
Collect and Store: Software version control



Collect and Store: Software version control







Collect and Store: Software version control

Your own server

Workstation

https://git-scm.com/

Bitbucket

Codeberg

Collect & Store: Metadata Standards

- Definition: Structured data that contains information about other data, but is not the content of the data.

- Metadata is very subject specific. The following directories are helpful:
  - Digital Curation Centre (https://www.dcc.ac.uk/guidance/standards)
  - RDA Metadata Standards (https://rdamsc.bath.ac.uk/)
  - Fairsharing (https://fairsharing.org/)

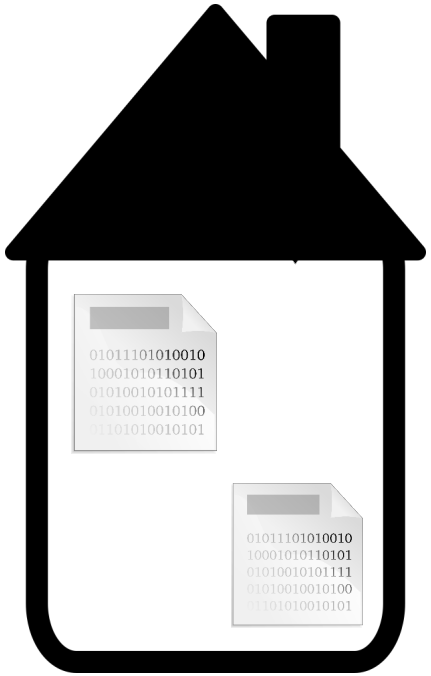
- Recommendation: Stick to a list of defined terms (controlled vocabulary) and don’t use synonyms to describe the same object (e.g. picture or image)

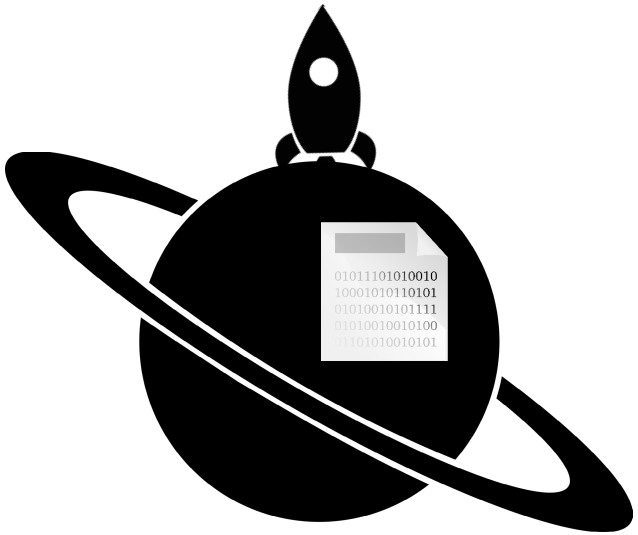
Collect & Store: README File

General information	• Title of the dataset • Contact information principal investigator • Date of data collection • Geographic location
Data and file overview	• Short discription for each file name • Date
Sharing and access informations	• Licenses or restrictions
Methodological information	• Description of methods for data collection or generation • Description of methods used for data processing
Data specific information (repeat for each dataset)	• Variable list, including names and definitions • Units of measuments • Definition for codes or symbols to record missing data

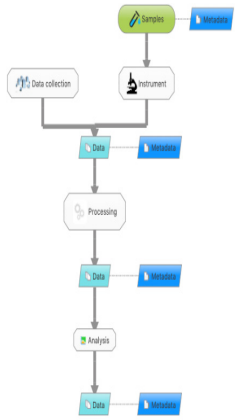
Cornell University: Minimal viable content. For recommended visit: https://data.research.cornell.edu/content/readme

Collect and Store: 3 – 2 – 1 backup



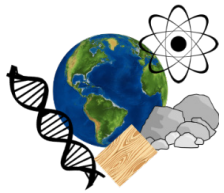


Collect & Store: Electronic Lab Notebook OpenBIS



Development started for biology – Now it can be used in most quantitative science fields

(*e.g. life sciences, biomedical sciences, physics, env. sciences, material sciences, etc*)



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

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

Collect & Store: Electronic Lab Notebook OpenBIS

openBIS

Electronic Lab Notebook

Description of
- *Experiments*
- *Measurements*
- *Processes*
- *Data analysis*

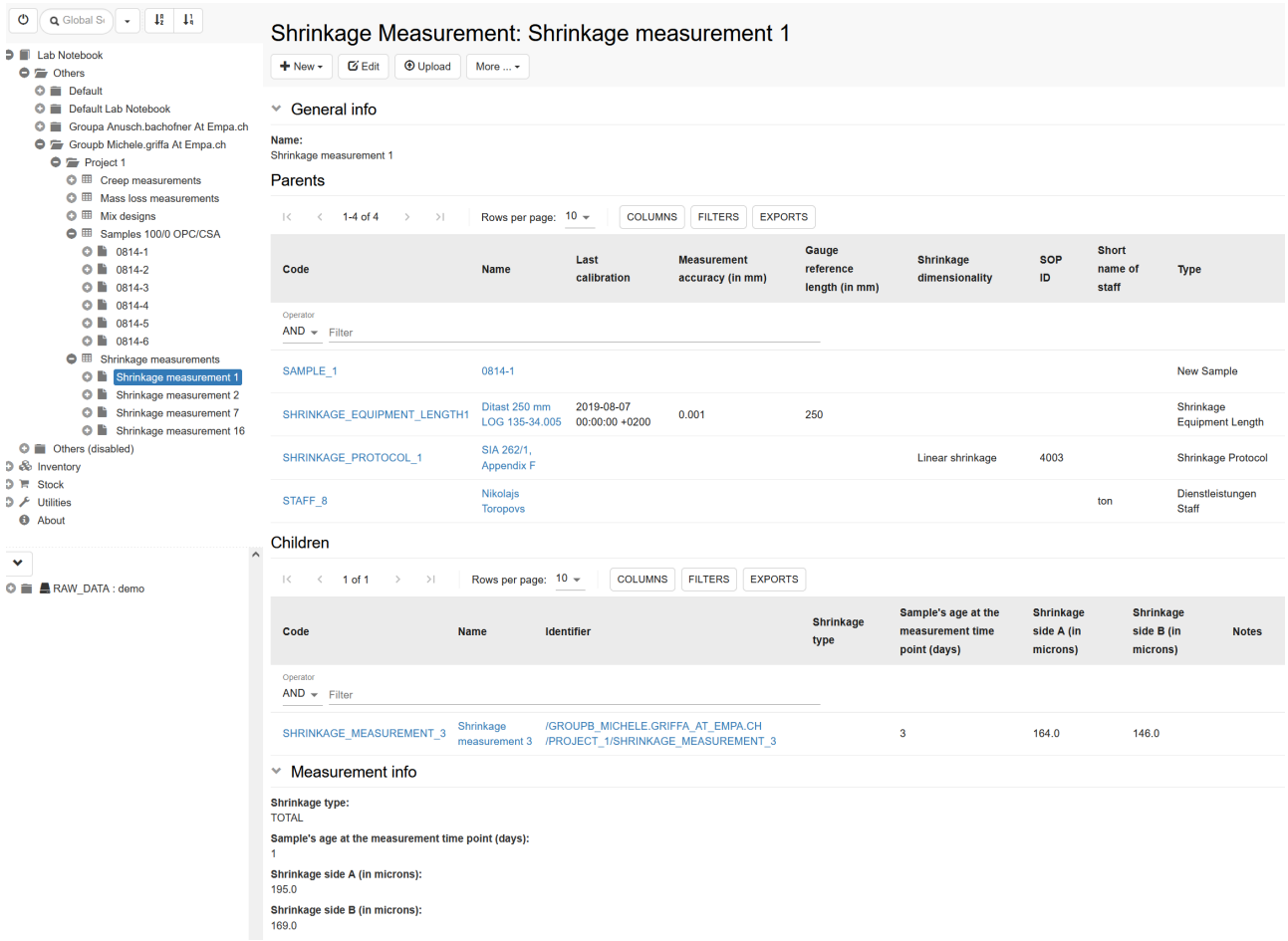
Inventory of
- *Samples*
- *Materials*
- *SOPsReagents*
- *Equipment*

Inventory Management

Data Management

Storage of research data connected to the experiments described in the electronic lab notebook

Collect & Store: Electronic Lab Notebook OpenBIS



Evaluate & Archive

Evaluate & Archive: Data Protection

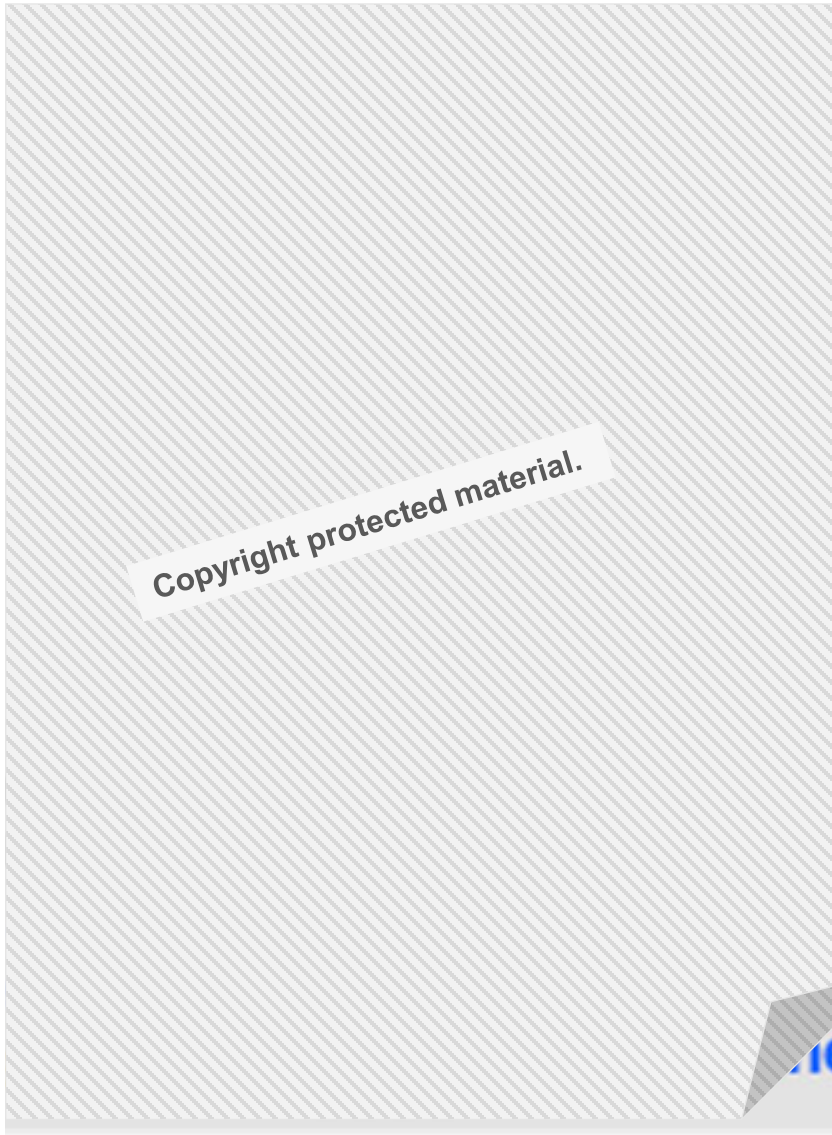
- **Relates to identified or identifiable person**
- **Solutions (https://dmlawtool.ccdigitallaw.ch/) :**
  - Identity irrelevant -> anonymisation
  - Identity relevant -> Ask for consent

    -> Pseudoanomization

    -> Manage access rights

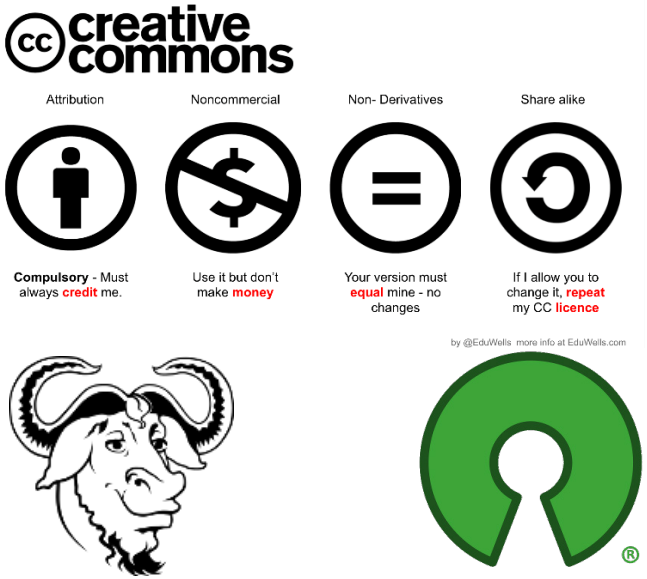
    -> Ability to address

    subject's rights
- **Always contact Data Protection Officers at your Research Institute if your research involves personal data**



Evaluate & Archive: Data Protection

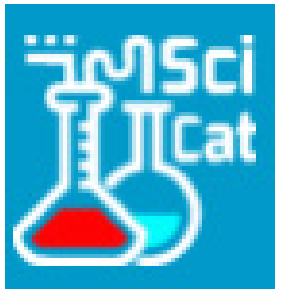
- Processed Data has copyright according to Swiss law
- Use CC licences when publishing factual data on data repositories (ideally CC 0)
- For software use licences specifically designed for software:
- Free Software (Open Source) licences like GPL, Apache, BSD and MIT.
- Exceptions! If you collaborated with external partners in your research project, you need to clarify together with them how and if data can be published.
- Contact the legal teams at your research institute if you feel lost.

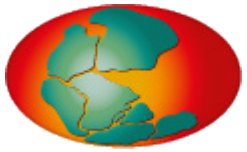


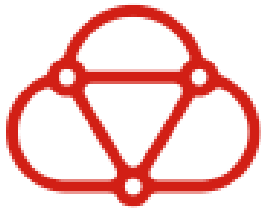
Share & Disseminate

Share & Disseminate: The Choice of Data Repository

ERIC open



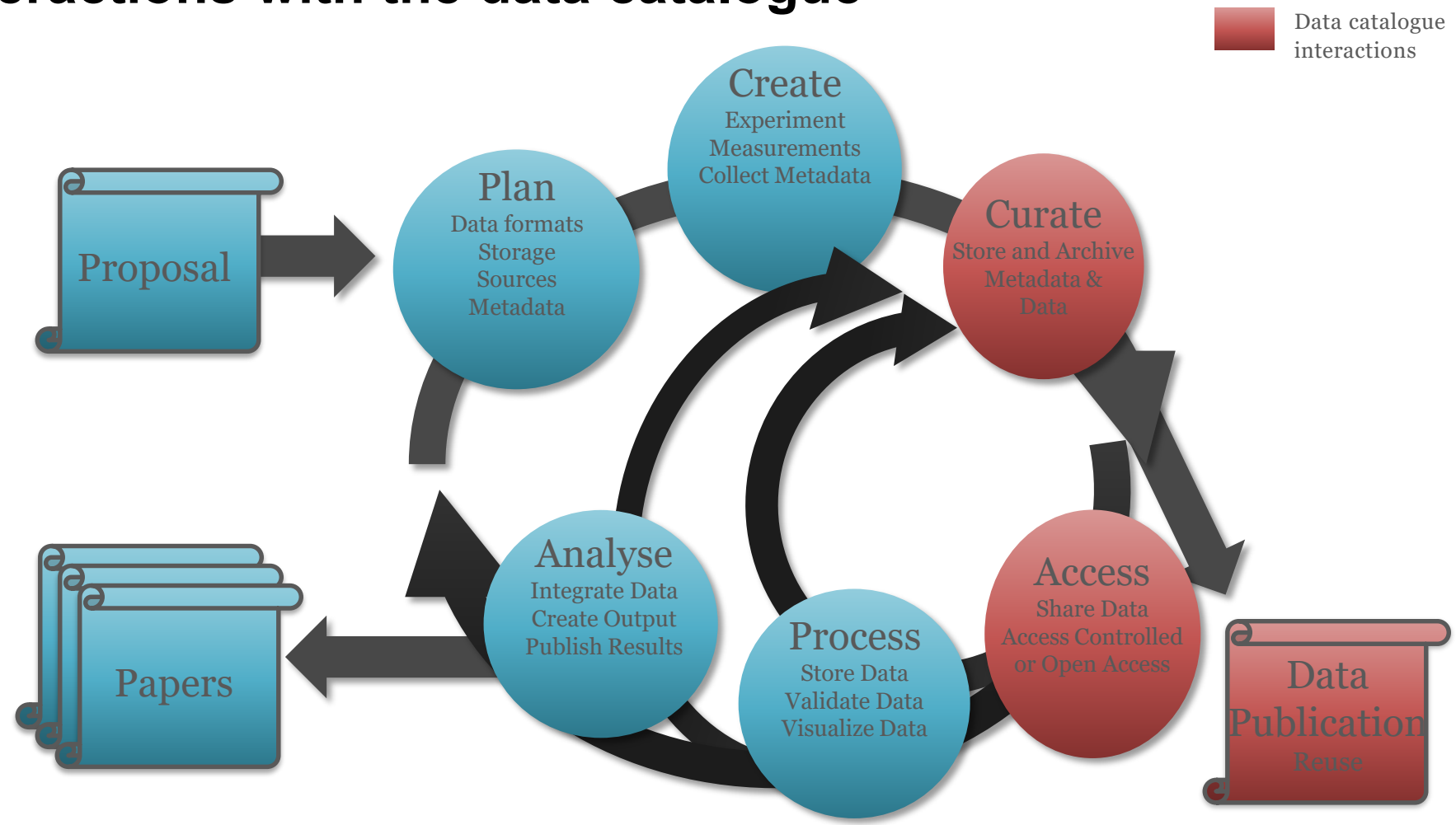




For alternatives: https://www.re3data.org/

RDM Services and Support at PSI

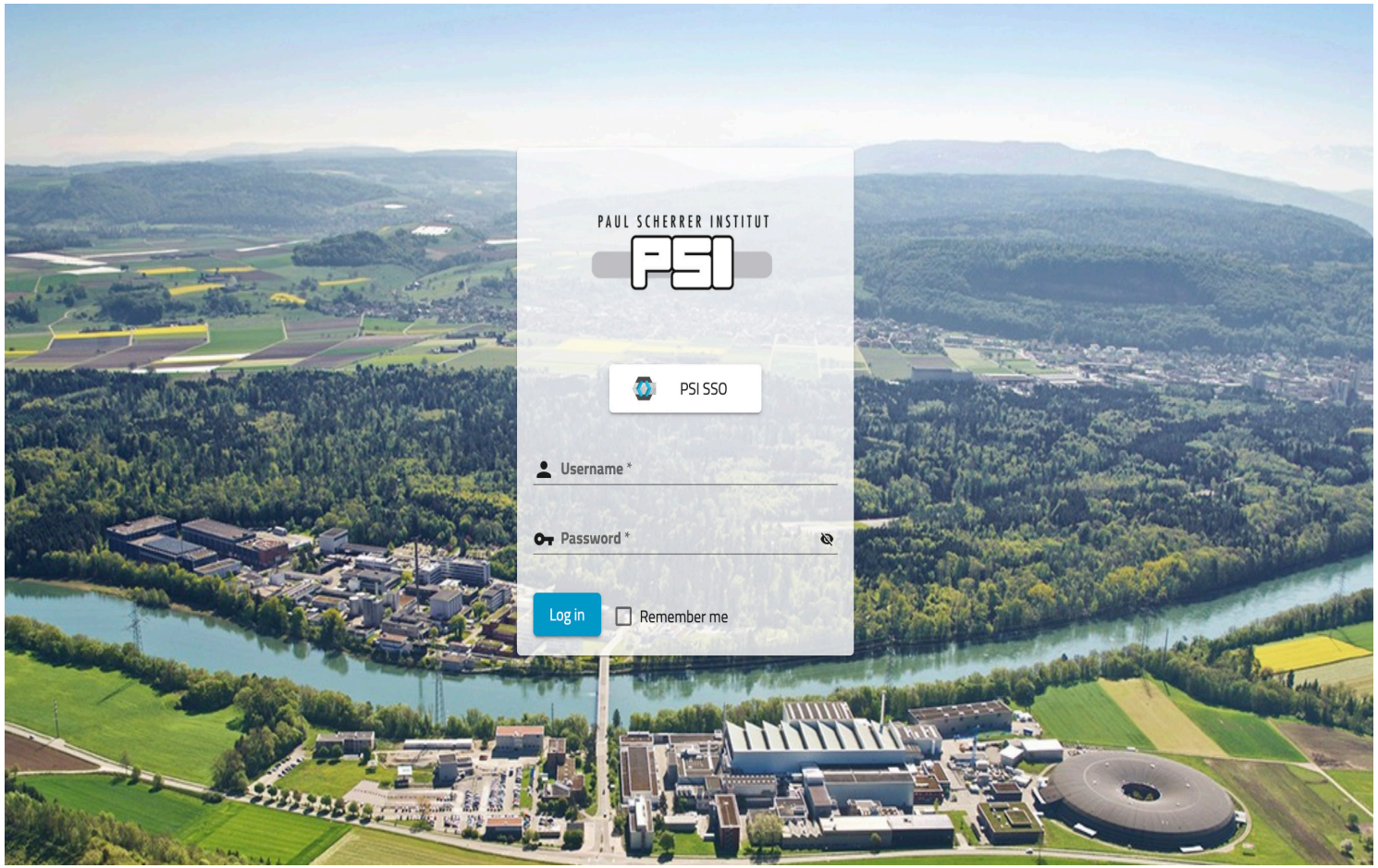
Interactions with the data catalogue



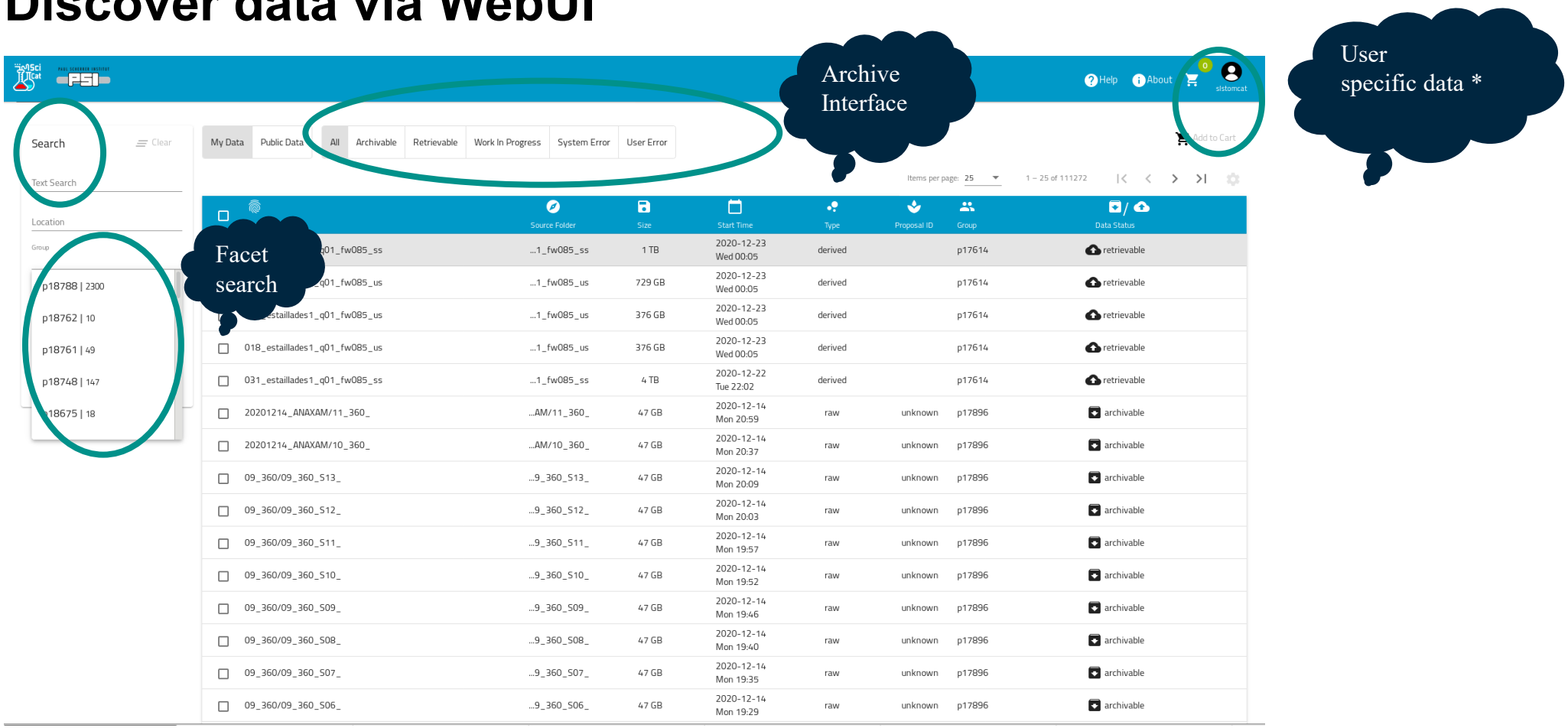
Where does SciCat help the Scientists?

- **Organize** the scientific data into datasets
- Annotate the Datasets with **administrative** and **flexible scientific metadata**
- Make the data **searchable/discoverable**
- Provides the infrastructure for **publishing** the data, DOI generation
- Can be used as frontend for **longterm** storage (Archive) solutions of mass data (PB regime)
- Supports both **open access** and **embargoed** data

User authentication

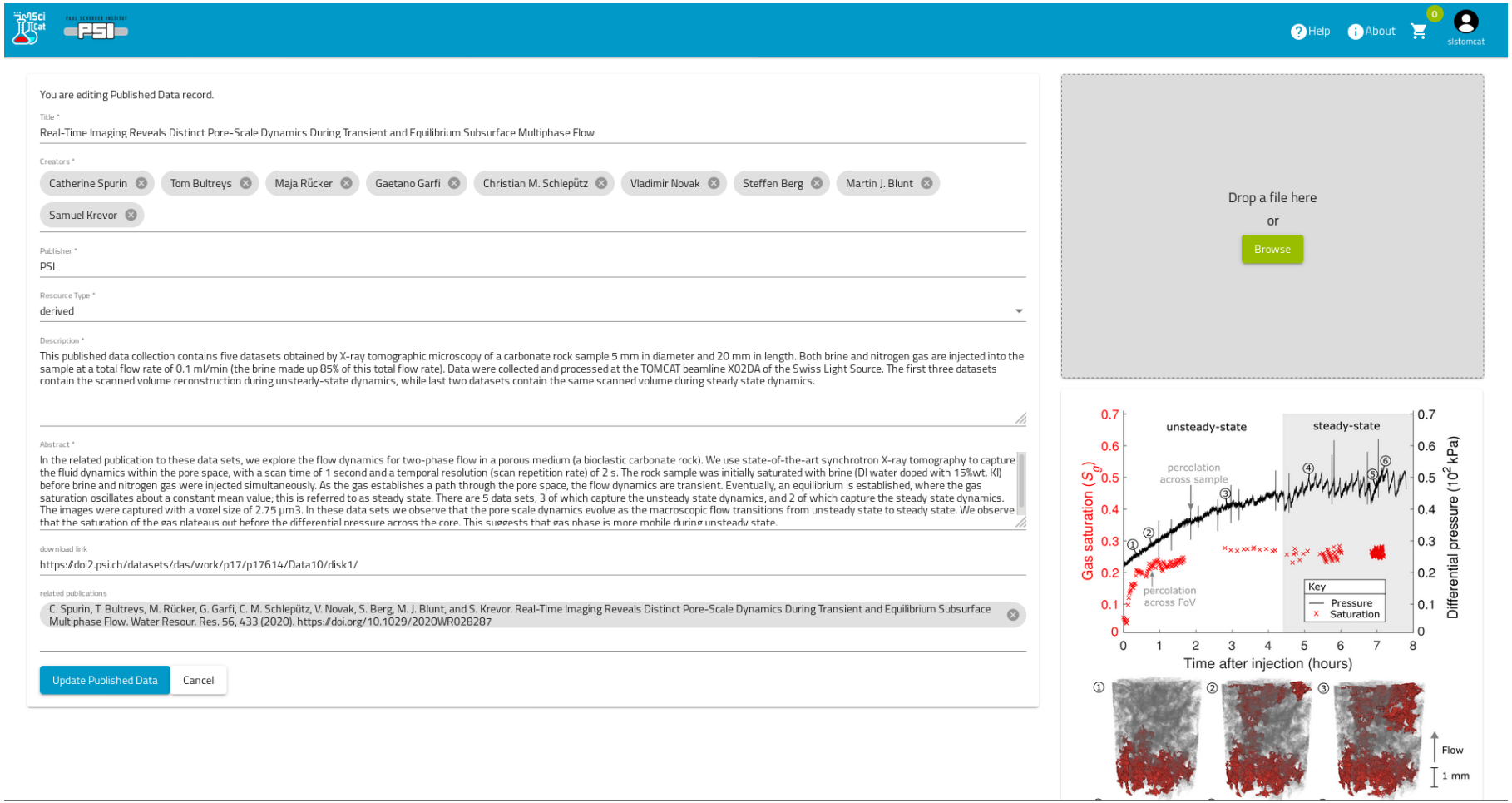


Discover data via WebUI

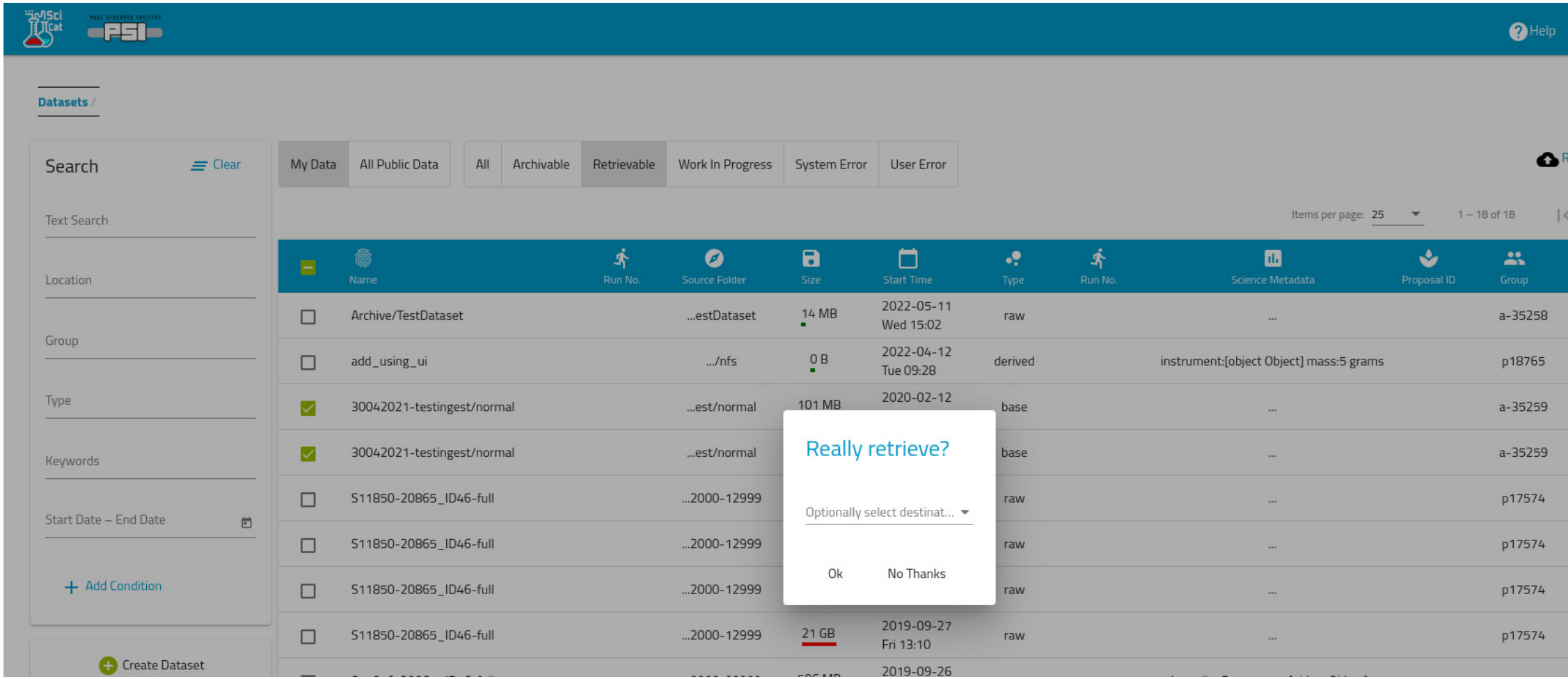


* User authorisation is handled based on group membership which is checked against the ownership of datasets. Group membership can come from external systems (e.g. DUO).

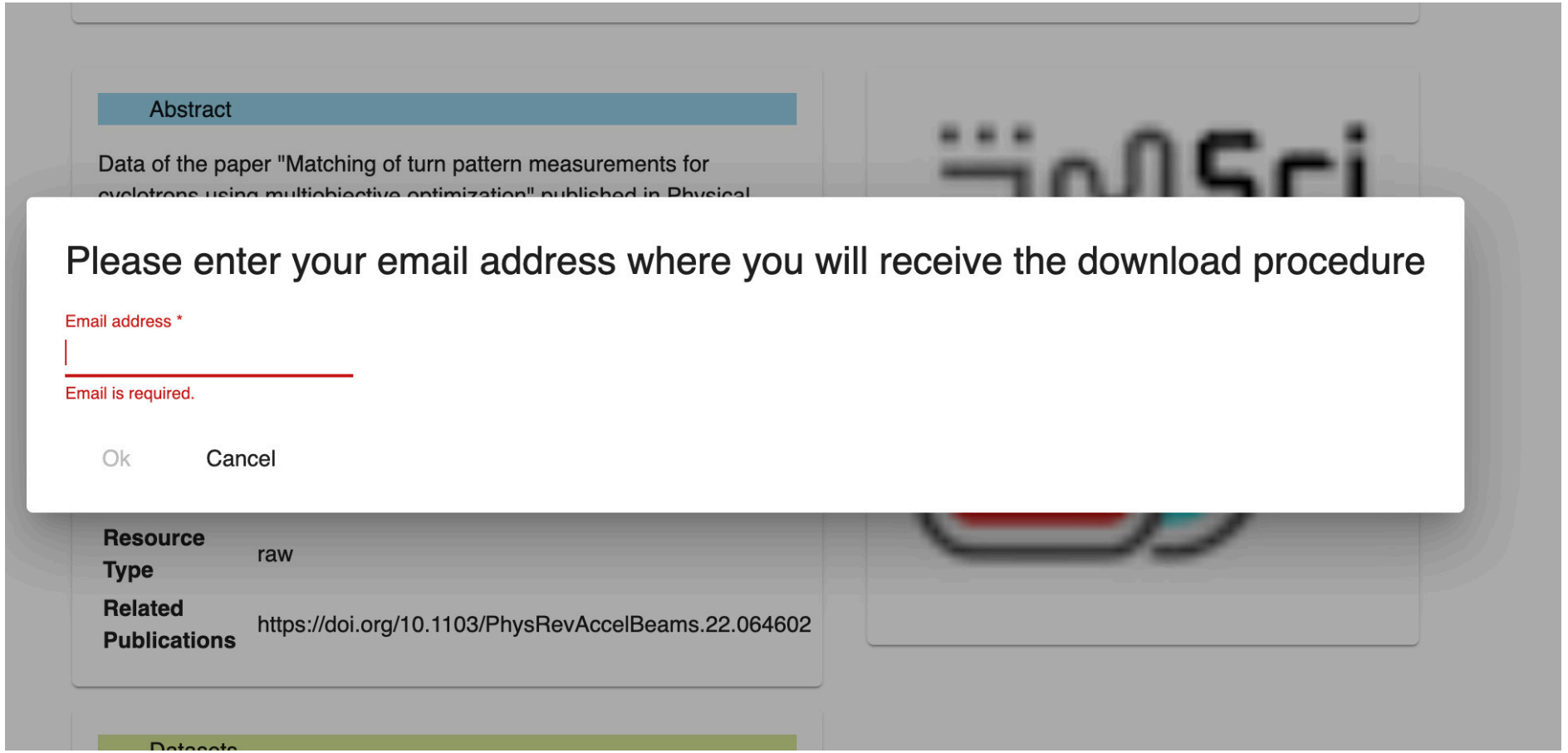
Editing of Metadata



Retrieving data from tape



Retrieving public data from tape



Published Data = List of Datasets + Metadata + DOI

Real-Time Imaging Reveals Distinct Pore-Scale Dynamics During Transient and Equilibrium Subsurface Multiphase Flow

Catherine Spurin,Tom Bultreys,Maja Rücker,Gaetano Garfi,Christian M. Schlepütz,Vladimir Novak,Steffen Berg,Martin J. Blunt,Samuel Krevor; PSI (2021)

Abstract

In the related publication to these data sets, we explore the flow dynamics for two-phase flow in a porous medium (a bioclastic carbonate rock). We use state-of-the-art synchrotron X-ray tomography to capture the fluid dynamics within the pore space, with a scan time of 1 second and a temporal resolution (scan repetition rate) of 2 s. The rock sample was initially saturated with brine (DI water doped with 15%wt. KI) before brine and nitrogen gas were injected simultaneously. As the gas establishes a path through the pore space, the flow dynamics are transient. Eventually, an equilibrium is established, where the gas saturation oscillates about a constant mean value; this is referred to as steady state. There are 5 data sets, 3 of which capture the unsteady state dynamics, and 2 of which capture the steady state dynamics. The images were captured with a voxel size of 2.75 µm3. In these data sets we observe that the pore scale dynamics evolve as the macroscopic flow transitions from unsteady state to steady state. We observe that the saturation of the gas plateaus out before the differential pressure across the core. This suggests that gas phase is more mobile during unsteady state.

Publication details

DOI https://doi.org/10.16907/46a4d882-4dec-4097-8289-8f6311a4aa36

Resource Type derived

Related Publications C. Spurin, T. Bultreys, M. Rücker, G. Garfi, C. M. Schlepütz, V. Novak, S. Berg, M. J. Blunt, and S. Krevor. Real-Time Imaging Reveals Distinct Pore-Scale Dynamics During Transient and Equilibrium Subsurface Multiphase Flow. Water Resour. Res. 56, 433 (2020). https://doi.org/10.1029/2020WR028287

Datasets

Description This published data collection contains five datasets obtained by X-ray tomographic microscopy of a carbonate rock sample 5 mm in diameter and 20 mm in length. Both brine and nitrogen gas are injected into the sample at a total flow rate of 0.1 ml/min (the brine made up 85% of this total flow rate). Data were collected and processed at the TOMCAT beamline X02DA of the Swiss Light Source. The first three datasets contain the scanned volume reconstruction during unsteady-state dynamics, while last two datasets contain the same scanned volume during steady state dynamics.

20.500.11935/64af1e80-c539-4a90-a051-b7db5e6e714d

20.500.11935/e151f4d6-198a-47e7-ac63-0b258ef36ed3

20.500.11935/441fdcd9-fa0c-491c-b102-d114cc841609

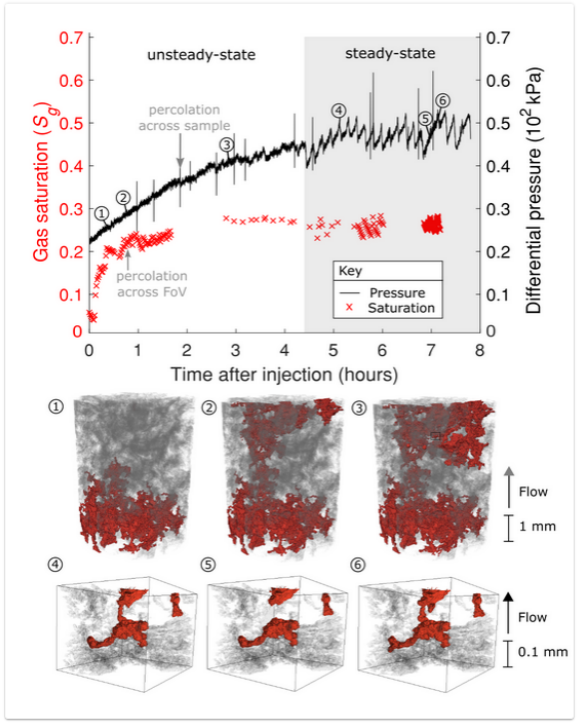
20.500.11935/b9782901-be3b-40fe-91d0-3e0a784337c4

20.500.11935/5899a0eb-7e3b-451f-b01e-17ddfc0d0938

Actions

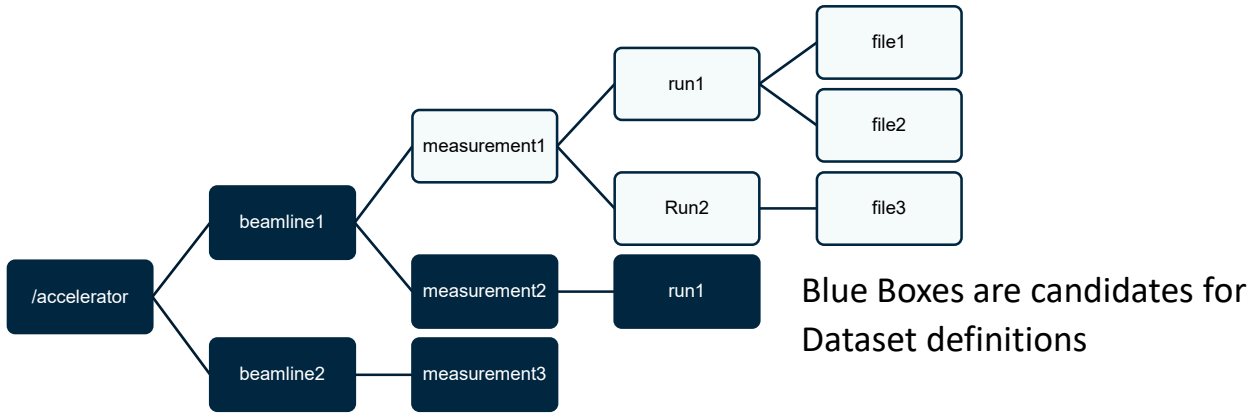
To access the data associated with this DOI click below and follow the instructions

Access Data



Metadata ingestion: 1. start e.g. from existing folder structure to define Datasets

- Datasets are the smallest unit for archiving, retrieving and publication
- Create them by defining a list of files, e.g. for raw data list all the files that logically belong to a measurement/data taking run, or any other criteria. For example: define all the files in the same directory (e.g. measurement1) as part of one dataset.



- In addition to “raw” Datasets you can create “derived” datasets containing the results of your analysis derived from the raw data. This ingest step is usually done by the user pursuing the analysis

Metadata ingestion: 2. Define Scientific Metadata

The definition of scientific meta data is fully flexible.

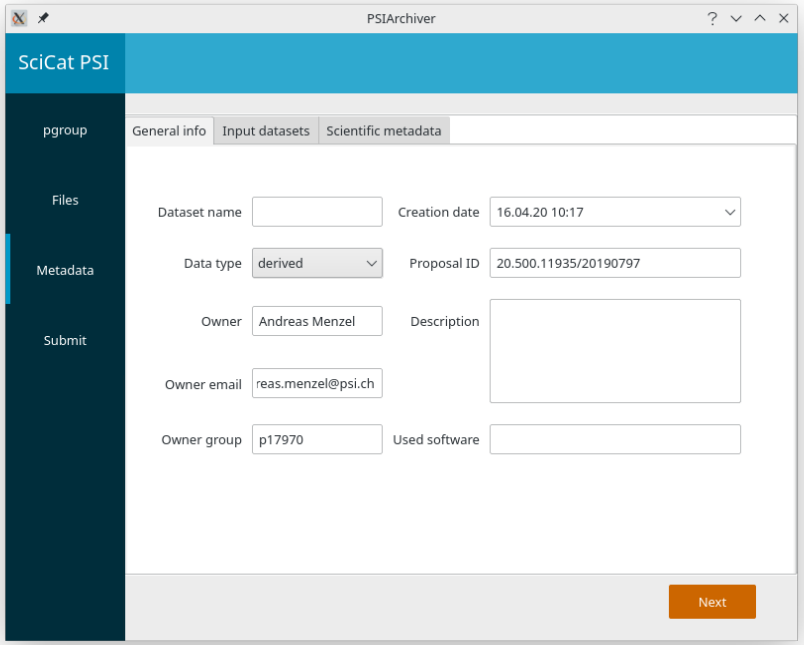
Ideally following a standard if it exists, e.g. NeXus based HDF5 files, extracted from instrument.

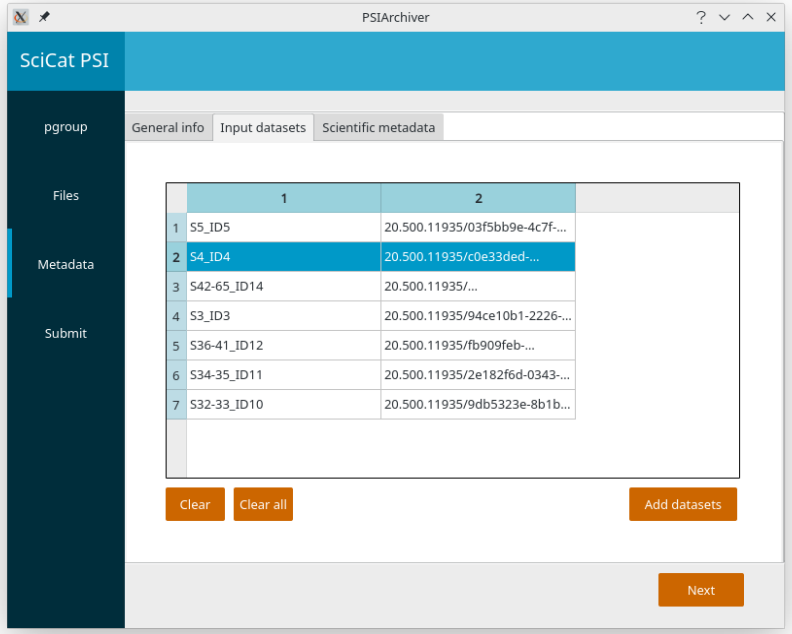
Example:

```
    "scientificMetadata": {
    "beamlineParameters": {
        "monostripe": "Ru/C",
        "ring_current": {
            "value": 0.402246,
            "units": "A"
        },
        "beam_energy": {
            "value": 22595,
            "units": "eV"
        }
    },
    "detectorParameters": {
        "objective": 20,
        "scintillator": "LAG 20um",
        "exposure_time": {
            "value": 0.4,
            "units": "s"
        }
    }…
}
```

Manual Ingests via Qt GUI tool at PSI

Especially for derived data:





Manual Ingests via CLI tool at PSI

Linux and Windows command line tool (datasetIngestor example):

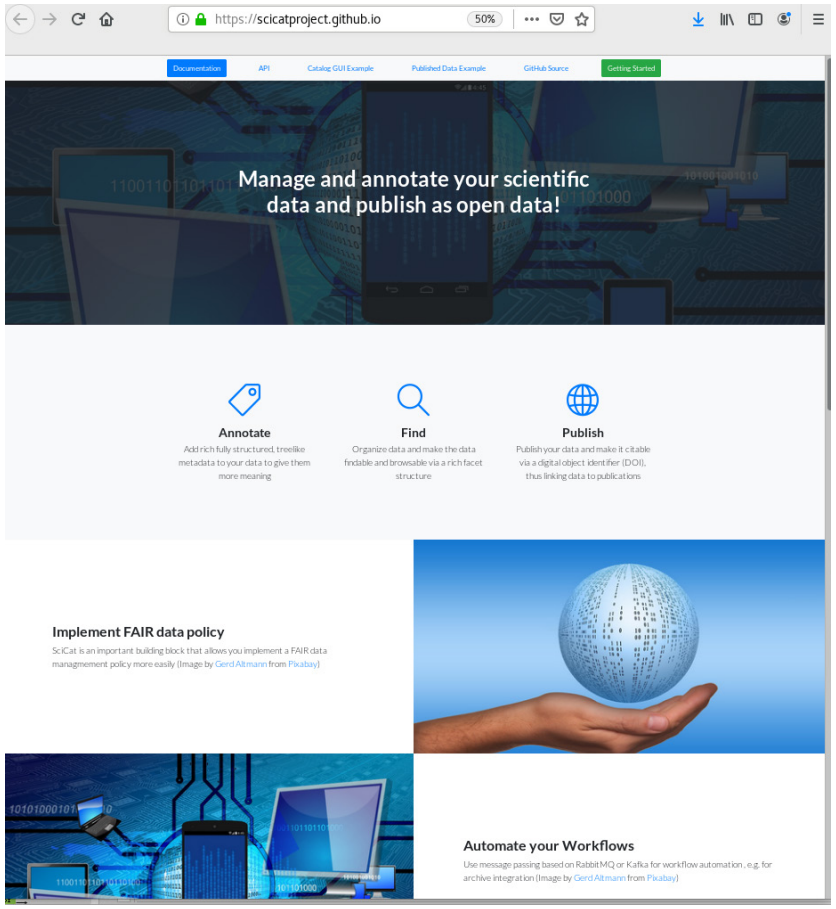
```
datasetIngestor [options] metadata-file [filelisting-file|'folderlisting.txt']

  -allowexistingsource
        Defines if existing sourceFolders can be reused
  -autoarchive
        Option to create archive job automatically after ingestion
  -copy
        Defines if files should be copied from your local system to a central server before ingest.
  -devenv
        Use development environment instead of production environment (developers only)
  -ingest
        Defines if this command is meant to actually ingest data
  -linkfiles string
        Define what to do with symbolic links: (keep|delete|keepInternalOnly) (default "keepInternalOnly")
  -noninteractive
        If set no questions will be asked and the default settings for all undefined flags will be assumed
  -tapecopies int
        Number of tapecopies to be used for archiving (default 1)
  -testenv
        Use test environment (qa) instead of production environment
  -user string
        Defines optional username:password string
```

PSI guide:
http://melanie.gitpages.psi.ch/SciCatPages/

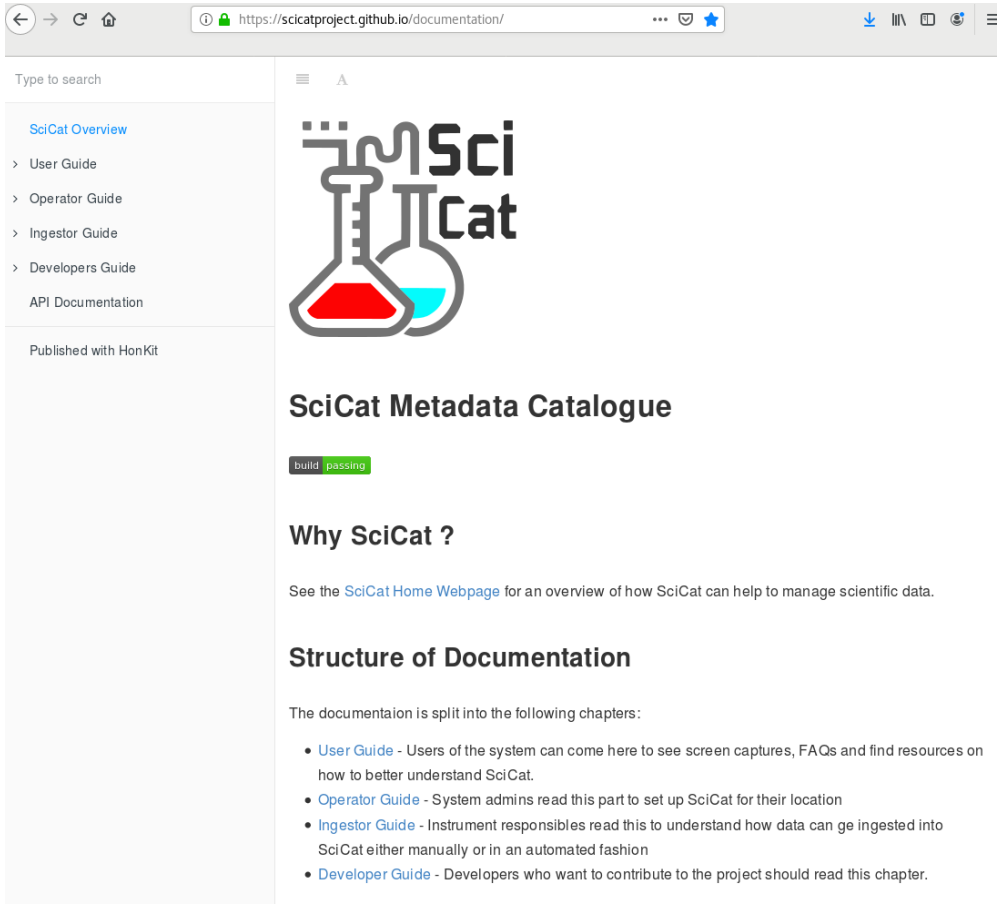
datasetIngestor usage example:

```
datasetIngestor metadata.json [filelisting.txt | 'folderlisting.txt']
```

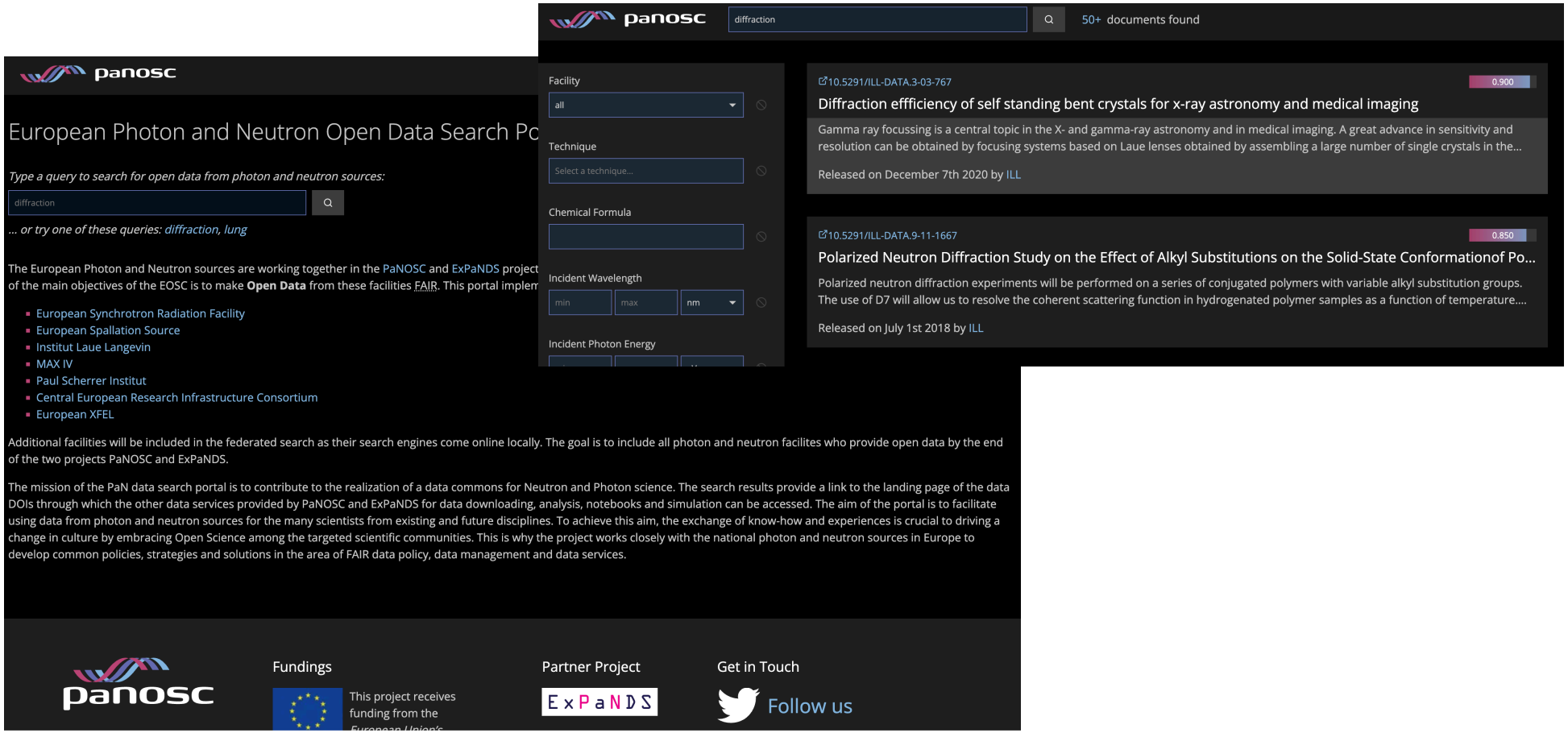
Documentation: scicatproject.github.io



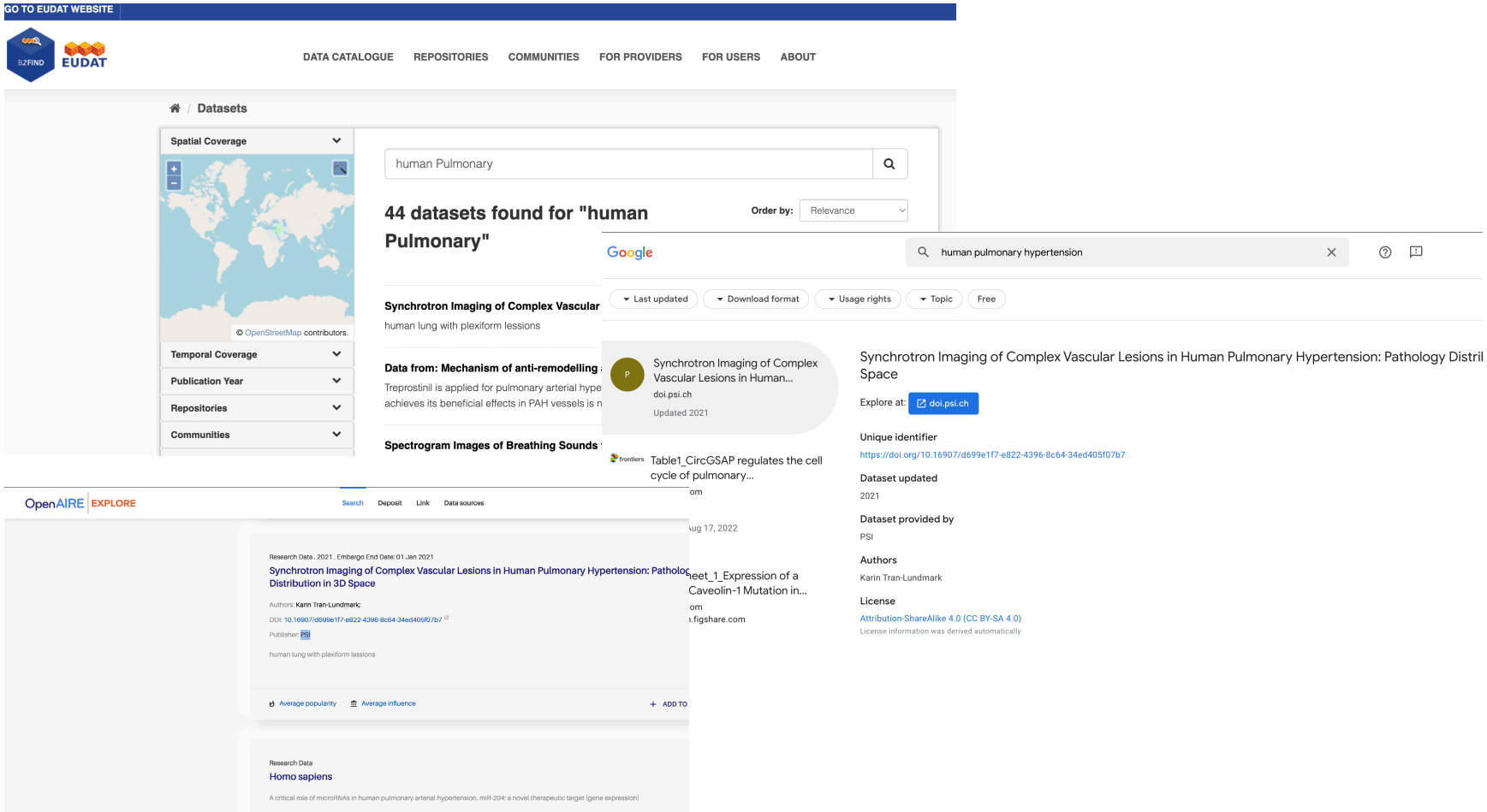
Full Documentation for Users and Operators



PaNOSC and ExPaNDS Open Data Search Portal



Data repositories



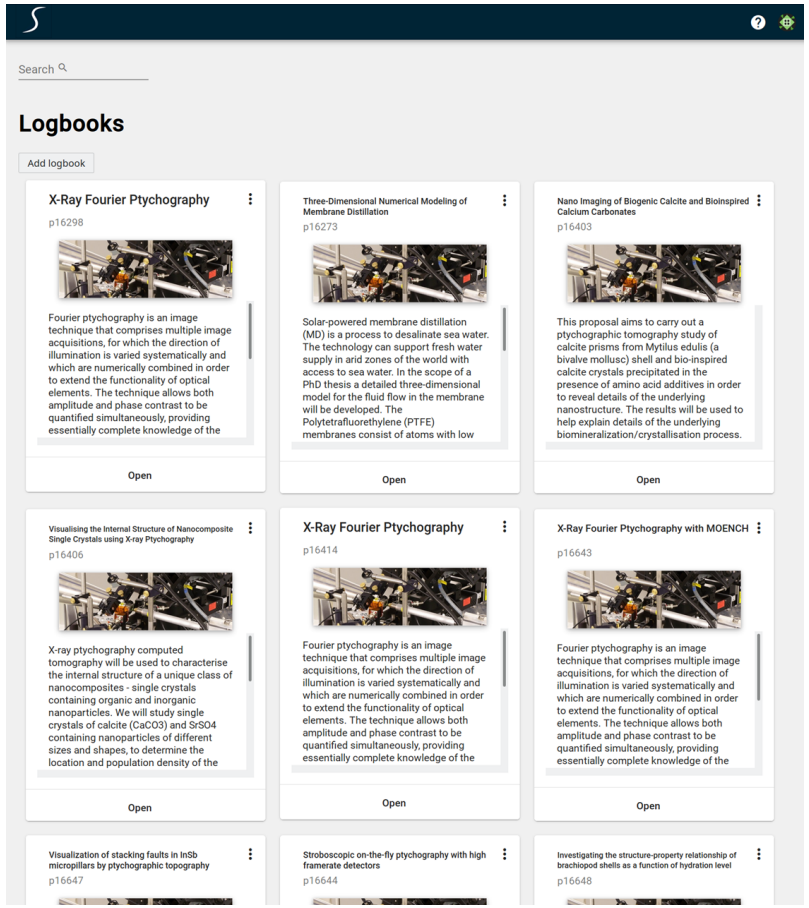
Thanks to all contributors!

SciLog electronic logbook

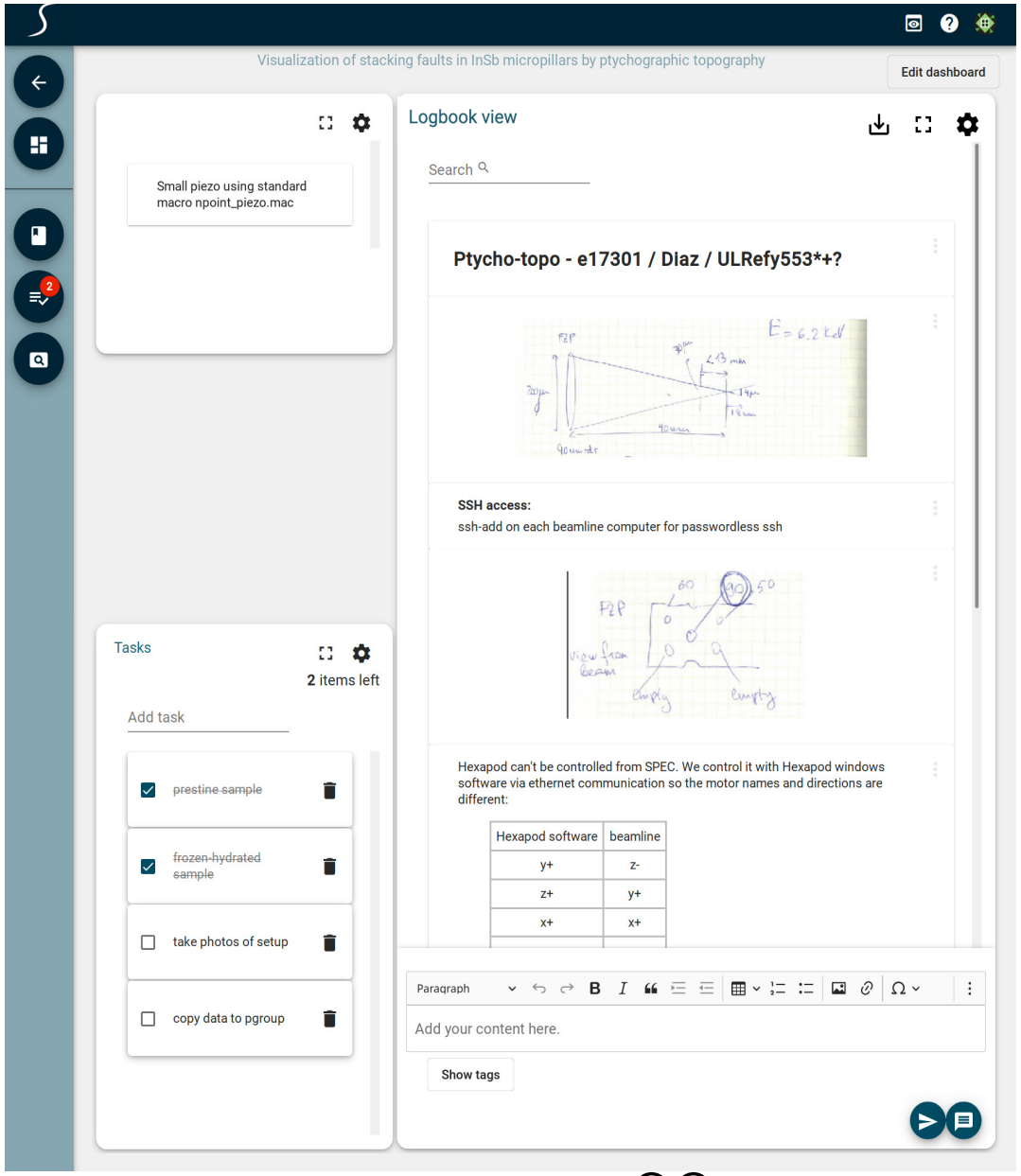
- Started as development effort of **Klaus Wakonig** and Stephan Egli within PSD department
- Requests for state-of-the art electronic logbook which is easy to use, can be reached from anywhere, can be integrated into existing environments (automation) and has fine grained access control.
- Source hosted on https://github.com/paulscherrerinstitute/scilog
- Deployment code at https://github.com/paulscherrerinstitute/scilog-ci
- Production instance at https://scilog.psi.ch

User authentication

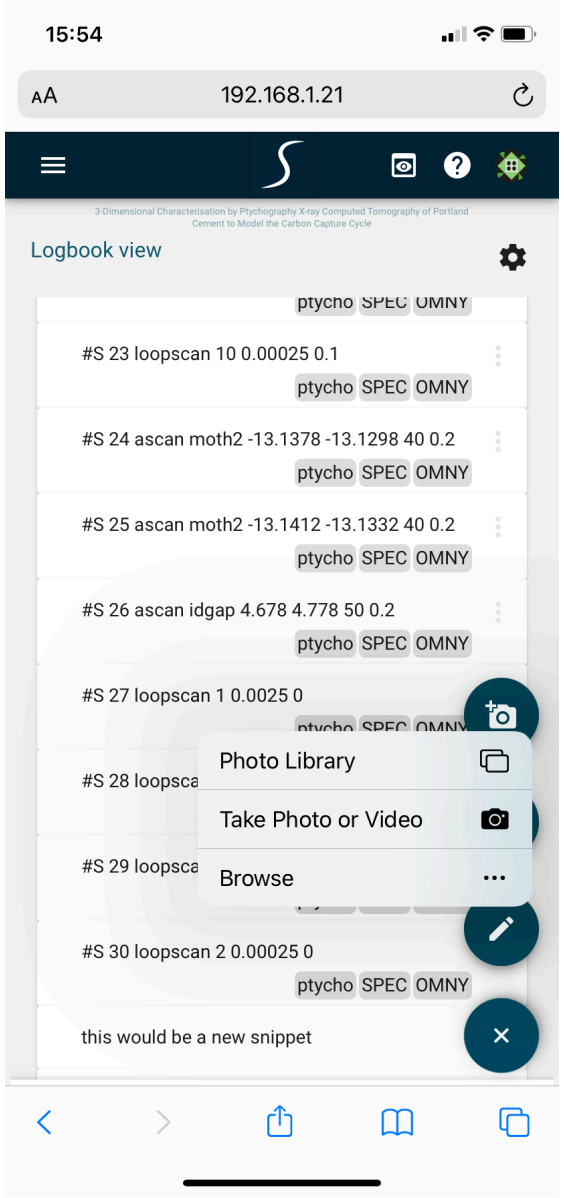
Viewing, searching, adding and editing

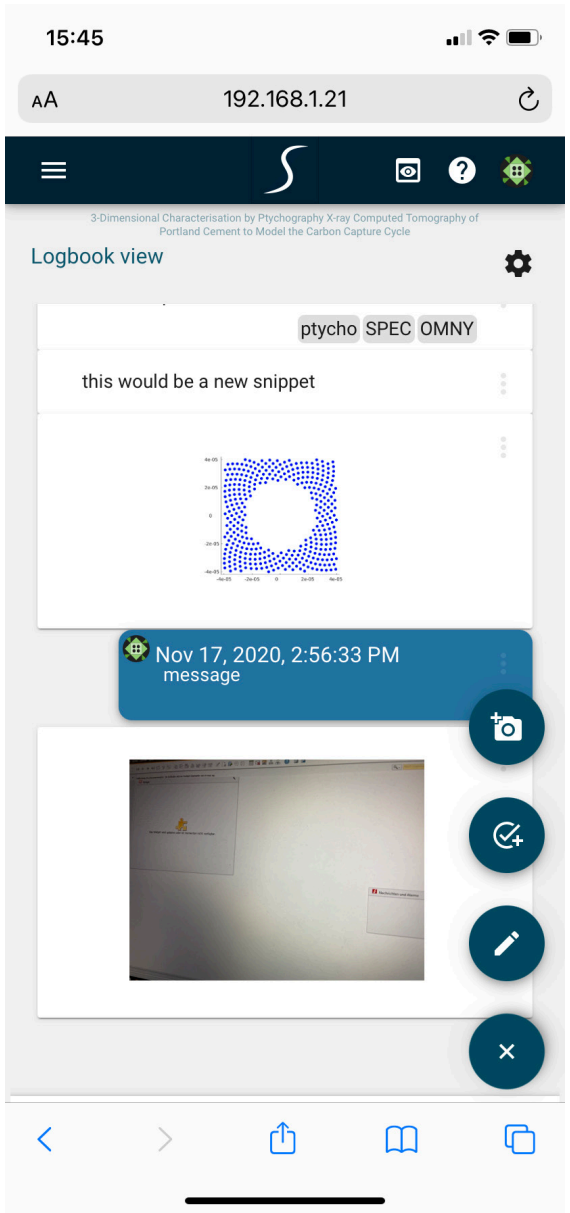


Inside a logbook



Mobile experience



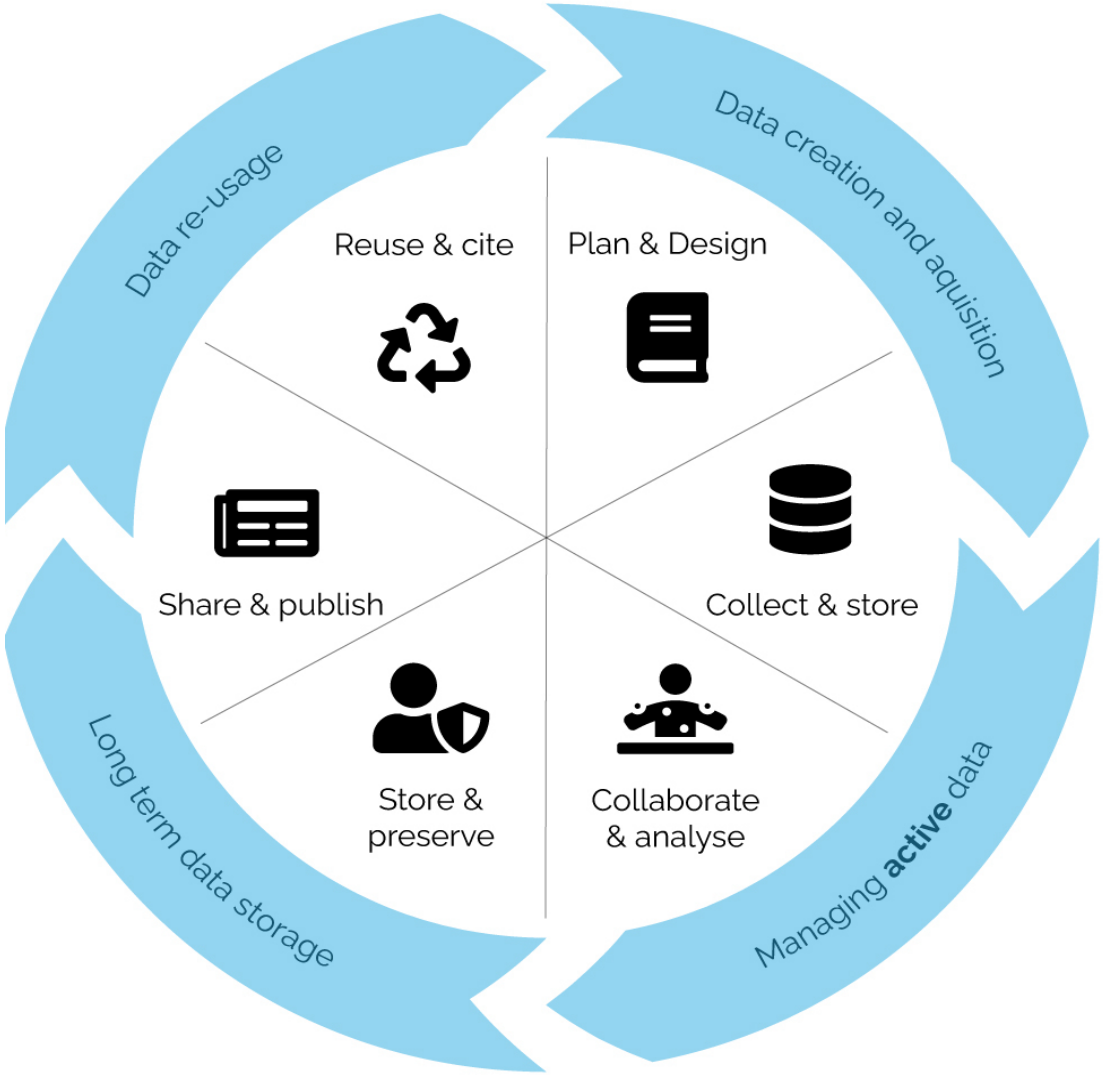


Plan & Design:

Data Management Plan (DMP)

Plan & Design: Why?

Bibliothèqe de l'EPFL, «RDM Horror stories | Episode 2 – Stranger Data Things», 11th February 2020. https://bit.ly/3qPWMlS



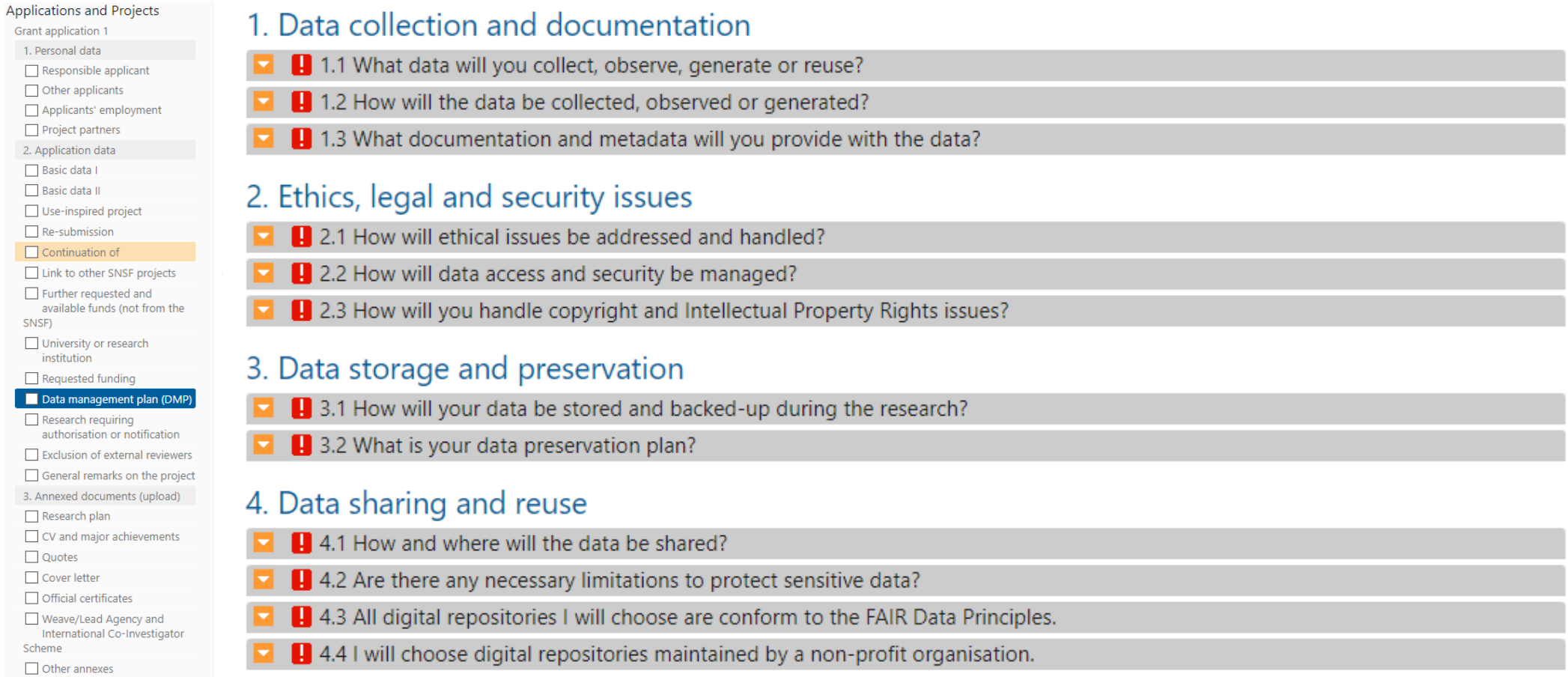
DMP

Covers the whole Research Data Life Cycle

Plan & Design: DMP

- What types of data will be collected and which code (incl. software) will be created or used?
- How will you document the data used and code programmed?
- Where will data and code be stored?
- Who owns the data and code is responsible for security and backup?
- Which data and code will be shared and preserved?
- How will data be shared and with whom?

Plan & Design: DMP



Plan & Design: DMP

- Keep it short and simple
- Be stingy with words
- Have one idea per sentence
- Use the active form — «we used the method» not «the method was used»
- Use positive phrases — «the results are different» not «the results are not the same»
- Use concrete terms — «it will be published in Nature» not «it will be published in a reputable journal»

- Don’t write in «sophisticated style»
- Save on adjectives and adverbs
- Avoid unnecessary constructions — e.g. «It is clear that», «the fact is that», «in an attempt to», «in order to»
- Don’t nominalise — «reduce» not «achieve a reduction in length»
- Don’t use empty modifiers — e.g. «basically», «indeed», «quite», «actually»
- Don’t use tautologous modifiers — e.g. «completely finish», «may potentially», «ultimate result», «blue in colour»

Plan & Design: DMP

1. **Organize yourselves in 4 groups (5 minutes)**

2. **Each group will engage with one of the four sections of the SNSF DMP (20 minutes)**
   - Read requirements
   - Write answers and questions
   - Discuss with other group members
   - Designate presenter

3. **Presentation and discussion of findings (20 minutes)**

Plan & Design: DMP - Data Collection and Documentation

1.1 What data will you collect, observe, generate or reuse?

- Type, format (NEAD), content, volume of data, reference to data (if reused)

1.2 How will the data be collected, observed, generated?

- Standards methodology, quality assurance
- File organisation and versioning (folder structures, git, ELN/LIMS, etc.)

1.3 What documentation and metadata will you provide?

- Scientific Metadata (README, metadata standards)
- General Metadata (Depending on choice of data repository)

Plan & Design: DMP - Ethics, Legal and security issues

2.1 How will ethical issues be addressed and handled?

- Information and consent to using personal data, location of critical infrastructure ase well as rare and protected species
- Requirements for assessments by ethical review boards, premission by third parties
- Description of Pseudonymisation or Anonymisation Methods

2.2 How will the data access and security be managed?

- Distinguish datasets according to the level of risk (cf. §2.1) and use an adverb to describe the level of risk («high», «medium», «low»)
- State Storage Location, secure transmission, access restruction, IT infrastructure

2.3 How will you handle copyright and Intellectual Property Rights Issues?

- Consider non-dislosure agreements, potential patents, research collaborations accross institutions
- Recommendation to use CC0 where possible

Plan & Design: DMP - Data Storage and Preservation

3.1 How will your data be stored and backed-up during the research?

- Backup strategy for work at all stages of research (amount of storage needed, frequency of updates, responsibilities, security measures)

3.2 What is your data preservation plan?

- Data formats
- Selection mode for data to be preserved (all relevant data related to reported results, long term preservation of unique datasets)

Plan & Design: DMP - Data Sharing and Reuse

4.1 How and where will the data be shared?

- Repository of choice (non-commercial preferred and required for contribution of up to 10’000 CHF for storage)
- Metadata Policy of said repository

4.2 Are there necessary limitations to protect sensitive data?

- Reasons data cannot be published at certain times (Section §2.1)

4.3 All Digital Repositories I will choose conform to FAIR Data?

- Check box

4.4 All Digital Repository I will choos are mainained by a non-profit oranisation?

- If no, provide justification (costs will not be covered)

Thank you for your attention!

Feedback!

Please give us a short feedback

Questions?

Presentation slides: lib4ri.ch > Learn > Trainings

Appendix

Appendix: PSI

- https://intranet.psi.ch/en/ord
- https://intranet.psi.ch/en/ord/data-management-tools

Appendix: File Formats EPFL

Bibliothèque de l'EPFL, Research Data, fast guide #4», 2019, https://bit.ly/3NFloYx

TYPE OF DATA	APPROPRIATE	ACCEPTABLE	DEPRECATED
Tabular (extensive metadata)	CSV – HDF5	TXT – HTML – TEX – FASTQ [3] – POR	
Tabular (minimal metadata)	CSV – TAB – ODS – SQL – TSV	XML (if appropriate DTD) – XLSX	XLS – XLSB
Textual / Presentation	TXT – PDF – ODT – ODM – TEX – MD – HTM – XML – EXTXYZ [4] – ODF	PPTX – RTF – DOCX – PDF (with embedded forms) – EPS – IPF	DOC – PPT – DVI – PS
Code / Computation	M – R – PY – IYPNB – RSTUDIO – RMD – NETCDF – AIML	SDD	MAT – RDATA
Image & Spectroscopy	TIF – PNG – SVG – JPEG – FITS	JCAMP – JPG – JP2 – TIF – TIFF – PDF – GIF – BMP – DM3 – OIR – LSM [5]	INDD – AIT – PSD – SPC
Audio	FLAC – WAV – OGG – MXL – MIDI – MEI – HUMDRUM	MP3 – AIF	
Video	MP4 – MJ2 – AVI – MKV	OGM – MP4 – WEBM	WMV – MOV – QT
Geospatial	NETCDF – tabular GIS attribute data – SHP – SHX – DBF – PRJ – SBX – SBN – POSTGIS – TIF – TFW – GEOJSON	MDB – MIF	
3D structures & images	X3D – X3DV – X3DB – PDF3D – POV – PDBML	DWG – DXF – PDB	PXP
Generic	XML – JSON – RDF		

Appendix: File Formats ETH Zürich

ETH-Library, File formats for archiving, 2022, https://bit.ly/3DBqXmb

Assessment of various file formats

Table 1: Our assessment of future readability of some common file formats. (For more detailed information we refer to the recommendations of the Bundesarchiv (German), the KOST (German or French), the Memoriav, the Forschungsdatenzentrums Archäologie & Altertumswissenschaften IANUS (Germany), the Library of Congress and the Harvard Library.)

File type	Recommended	Suitable to only a limited extent	Not suitable for archiving
Text	• PDF/A (*.pdf, preferred subtypes 2b and 2u) • Plain Text (*.txt, *.asc, *.c, *.h, *.cpp, *.m, *.py, *.r etc.) coded as ASCII, UTF-8, or UTF-16 using byte order mark • XML (inclusive XSD/XSL/XHTML etc.; with included or accessible schema and character encode explicitly specified)	• PDF (*.pdf) with embedded fonts • Plain text (*.txt, *.asc, *.c, *.h, *.cpp, *.m, *.py, *.r etc.) (ISO 8859-1 coded) • Rich Text Format (*.rtf) • HTML and XML (The ASCII text is readable over long term; try to avoid external links.) Not accepted for publication, OK for supplementary materials: • Word *.docx • PowerPoint *.pptx • LaTeX, TeX (The ASCII text is readable over long term; open source software required for formatting and the resulting PDF should be included.) • OpenDocument formats (*.odm, *.odt, *.odg, *.odc, *.odf)	• Word *.doc • PowerPoint *.ppt
Spreadsheet or table	• Comma- or tab delimited text files (*.csv)	• Excel *.xlsx (container format) • OpenDocument spreadsheets (*.ods)	• Excel *.xls, *.xlsb (binary formats)
Raw data and workspace		• ASCII Text is suitable for long-term use, but the data import may be time-consuming. • S-Plus files (*.sdd) may be saved as text files. • Matlab *.mat files may be saved in HDF Format. Saving nontrivial ASCII Matlab *.mat files should be avoided because they are not readable with the Matlab load command (see table 2). • Network Common Data Format or NetCDF (*.nc, *.cdf) • Hierarchical Data Format (HDF5) (*.h5, *.hdf5, *.he5)	• Binary files such as the standard Matlab files *.mat or the R files *.RData
Raster image (bitmap)	• TIFF (*.tif) (uncompressed, preferentially TIFF 6.0, Part 1: baseline TIFF). TIFF is preferred as compared to PNG or JPEG2000. • Portable Network Graphics (*.png, uncompressed) • JPEG2000 (*.jp2, lossless compression) • Digital-Negative-Format (*.dng) to keep raw data of digital fotos in addition to an second copy in TIFF format	• TIFF (*.tif) (compressed) • GIF (*.gif) • BMP (*.bmp) • JPEG/JFIF (*.jpg) • JPEG2000 (lossy compression) (*.jp2)	
Vector graphics	• SVG without JavaScript binding (*.svg)		• Graphics InDesign (*.indd), Illustrator (*.ait) • Encapsulated Postscript (*.eps) • Photoshop (*.psd)
CAD	• AutoCAD Drawing (*.dwg) • Drawing Interchange Format, AutoCAD (*.dxf) • Extensible 3D, X3D (*.x3d, *.x3dv, *.x3db)		
Audio	• WAV (*.wav) (uncompressed, pulse-code modulated)	• Advanced Audio Coding (*.mp4) • MP3 (*.mp3)	
Video [1]	• FFV1 codec (version 3 or later) in Matroska container (*.mkv)	• MPEG-2 (*.mpg,*.mpeg) • MP4, which is also called MPEG-4 Part 14 (*.mp4) • QuickTime Movie (*.mov) [2] • Audio Video Interleave (*.avi) • Motion JPEG 2000 (*.mj2, *.mjp2)	• Windows Media Video (*.wmv)

Footnotes

[1] In addition to the file format (or container format), also the codec and the compression method are important. See Ianus, Memoriav and KOST for further information.

[2] In the Version of Nov 21, 2018 of the current document, the format QuickTime Movie was downgraded from „Recommended" to „Suitable to only a limited extent". Apple discontinued the support of Windows QuickTime Player in the year 2016. Windows Media Player thus only supports file format versions 2.0, or earlier, of QuickTime Movie files.

Appendix: References (Slide 18)

[1] SPARC Europe, «The Open Data Citation Advantage», 2017, https://sparceurope.org/open-data-citation-advantage/.

[2] Digital Science, «The state of Open Data Report», 2019, https://digitalscience.figshare.com/articles/report/The_State_of_Open_Data_Report_2019/9980783/2

[3] European Commission and PwC, «Cost-Benefit analysis fro FAIR research Data», 2019.

https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1

[4] Baker, M., “1,500 scientists lift the lid on reproducibility”. *Nature* 533, 452–454 (2016). https://doi.org/10.1038/533452a

Appendix: Icon References

Slide 17:

- Le Moign, Vincent, «Lab Scientist Icon», https://icon-icons.com/icon/lab-scientist/101049, free for commercial use.
- Flaticon, «Checkliste», https://www.flaticon.com/de/kostenloses-icon/checkliste_2666469, free for personal and commercial use.
- PLoS, «Open Access logo», https://de.wikipedia.org/wiki/Datei:Open_Access_logo_PLoS_white.svg, CC-0.
- «Databases and People», https://freesvg.org/databases-and-people, CC-0.

Slide 18

- Felixmh, «Krischen-Früchte-Natur-Symbol», free commercial use.