23.11.2023

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

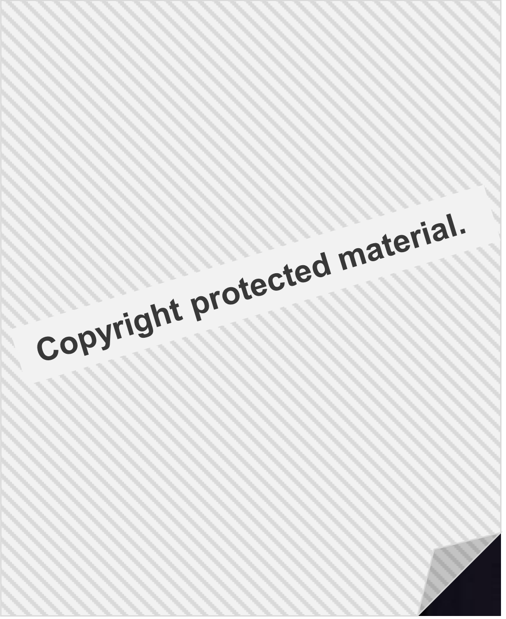
- Life cycle of research data
- Adequate metadata documentation for your code and data
- Storing and publishing data
- Using OpenBIS (ELN) and writing Data Management Plans (DMP)

Program

Topic	Speaker	Time
Introduction	Fabian Felder	9.00 - 9.15
Policies, Incentives & the Research Data Life Cycle	Fabian Felder	9.15 - 9.30
Collect & Store	Federico Cantini	9.30 - 10.05
Evaluate & Archive Share & Disseminate	Fabian Felder	10.05 - 10.15
Break		10.20 - 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

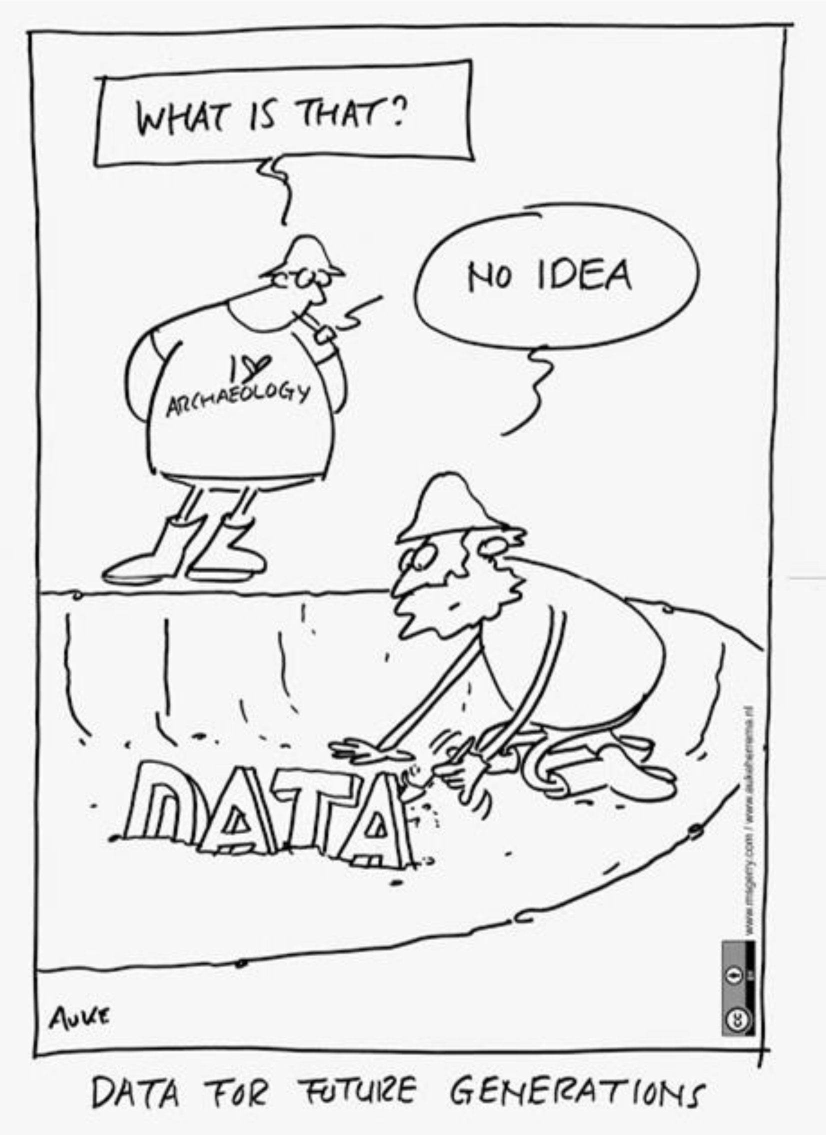
Cham, J. G., «Scratch: A context-changing framework for contextualizing nano informatic structures» (2014), International Journal of Temporal Deflective Behaviour, 4 (1689), p. 432.

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?

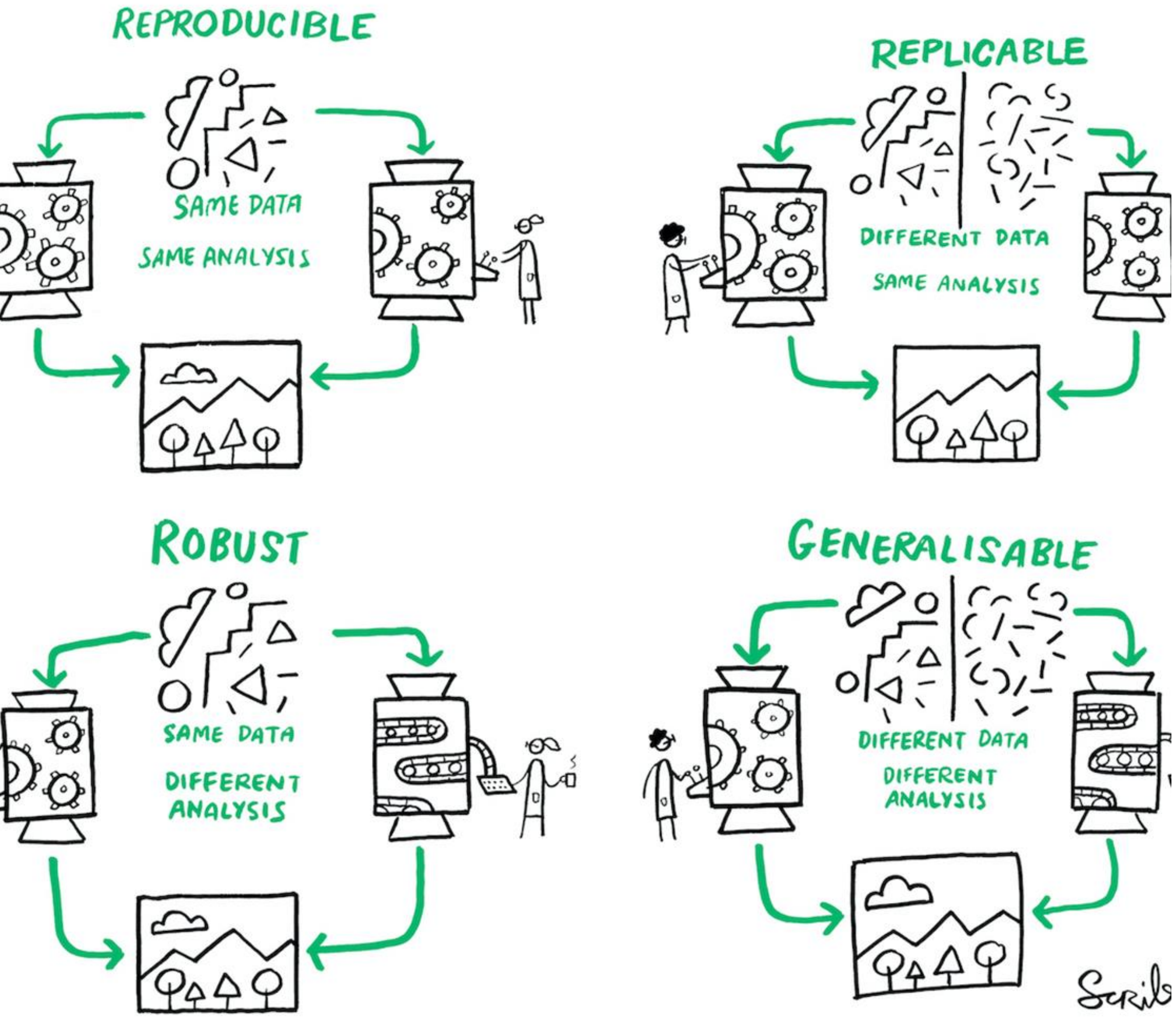
Proper metadata tagging

and

research description
is

time consuming

Reproducibility



Reproducibility

Scriberia, “Reproducible Research”, *The Turing Way*, CC-BY, DOI: [10.5281/zenodo.3332807](https://doi.org/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

What matters?

What will you need to record?

Is there a way to record it automatically?

Which structure do you use? Or do you rely on a narrative expression?

How do you describe your materials?

Which formats do you use?

Is there a standard?

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

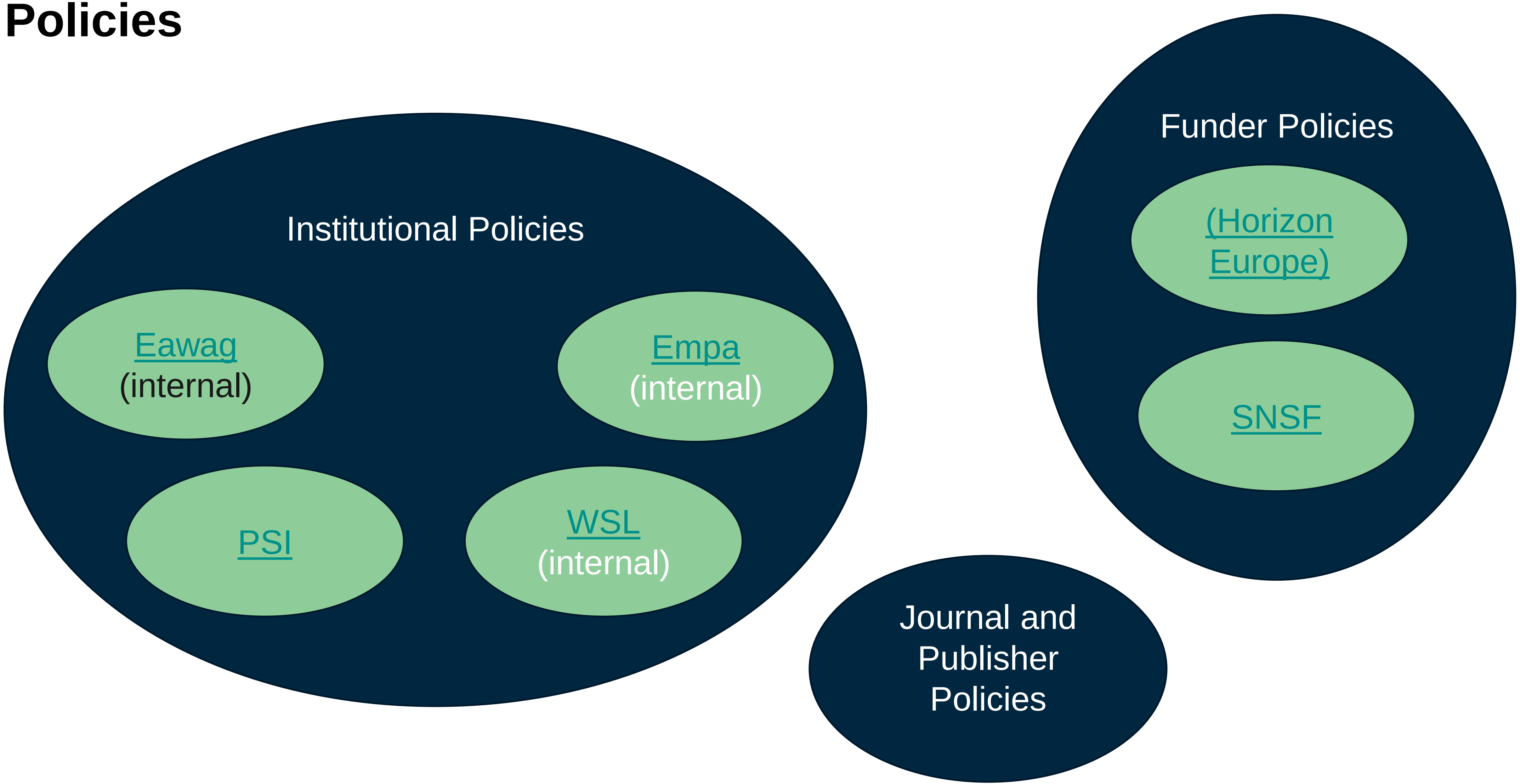
- I1 (Meta)data use formal, accessible, shared, and broadly applicable language for knowledge representation
- I2 (Meta)data use vocabularies that follow FAIR principles
- I3 (Meta)data indlude qualified references to other (meta)data

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

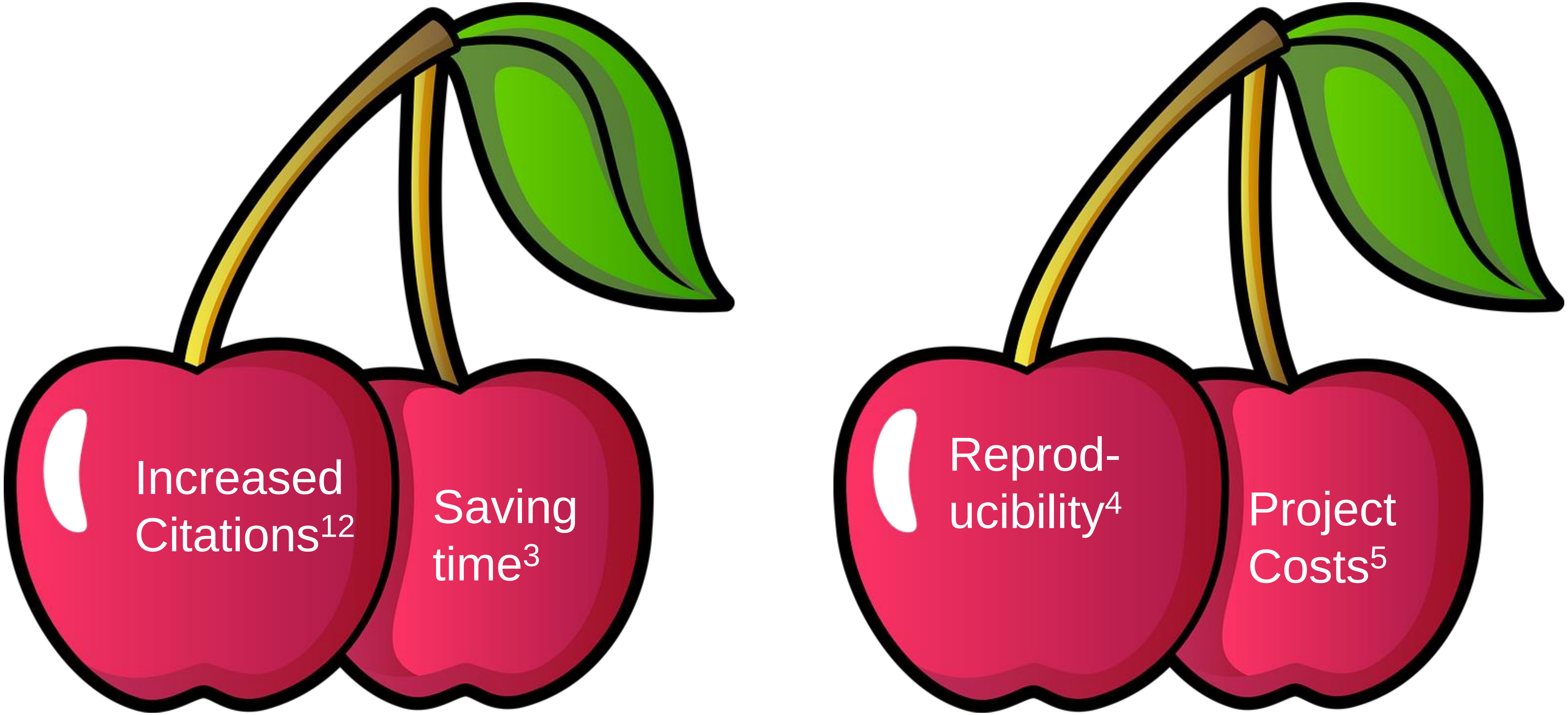
As open as possible,

as closed as necessary.

Repositories

FAIR and non-commercial

Policies



Increased Citations[12]

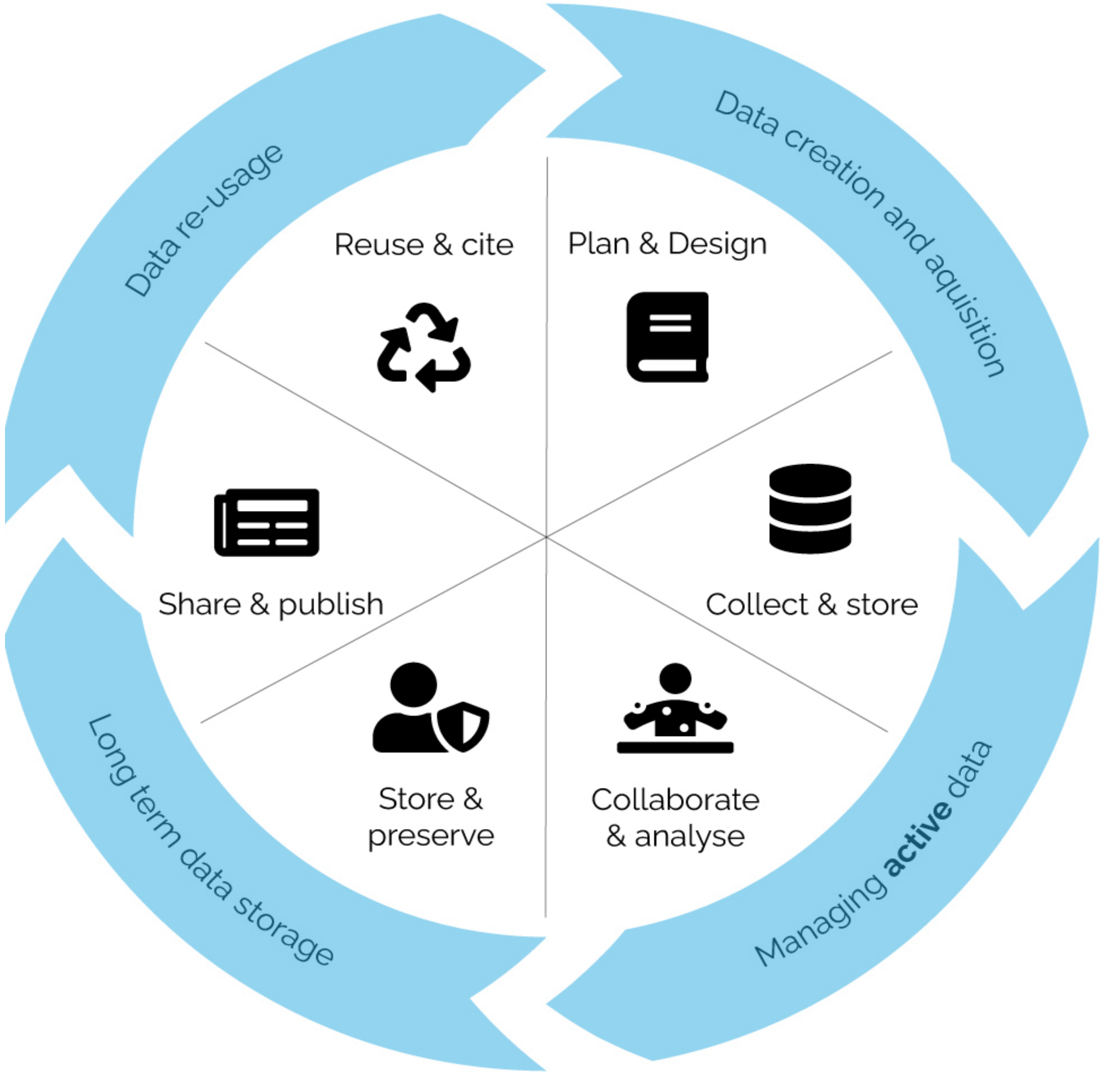
Saving time[3]

Reproducibility[4]

Project Costs[5]

Research Data Life Cycle

Research Data Life Cycle



Collect & Store

Collect & Store

```
01010100 01101000
01101001 01101110
01101011 00100000
01100100 01101001
01100110 01100110
01100101 01110010
01100101 01101110
01110100 00101110
```

Data

observational, experimental, simulation…

Code

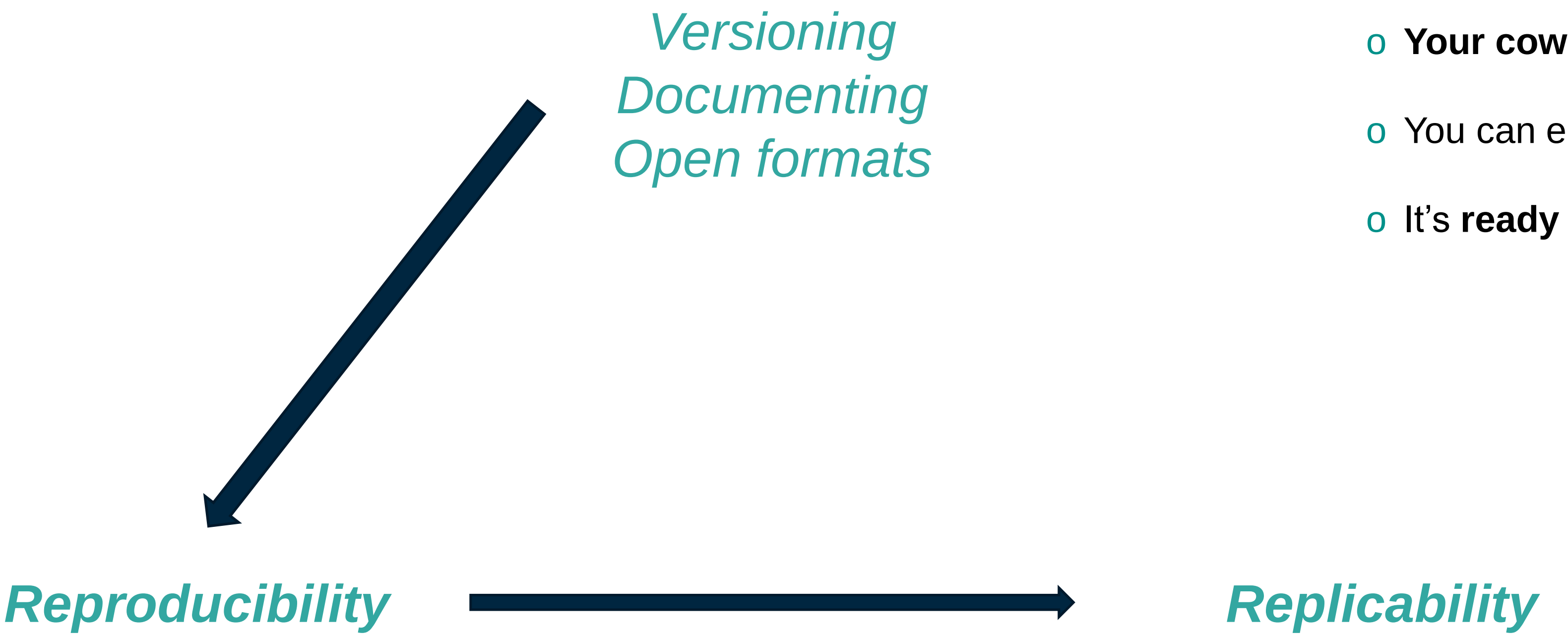
Applications, scripts…

Metadata

Structured information associated with data (and code)

The Who, What, Where, Why & How of data

Collect and Store



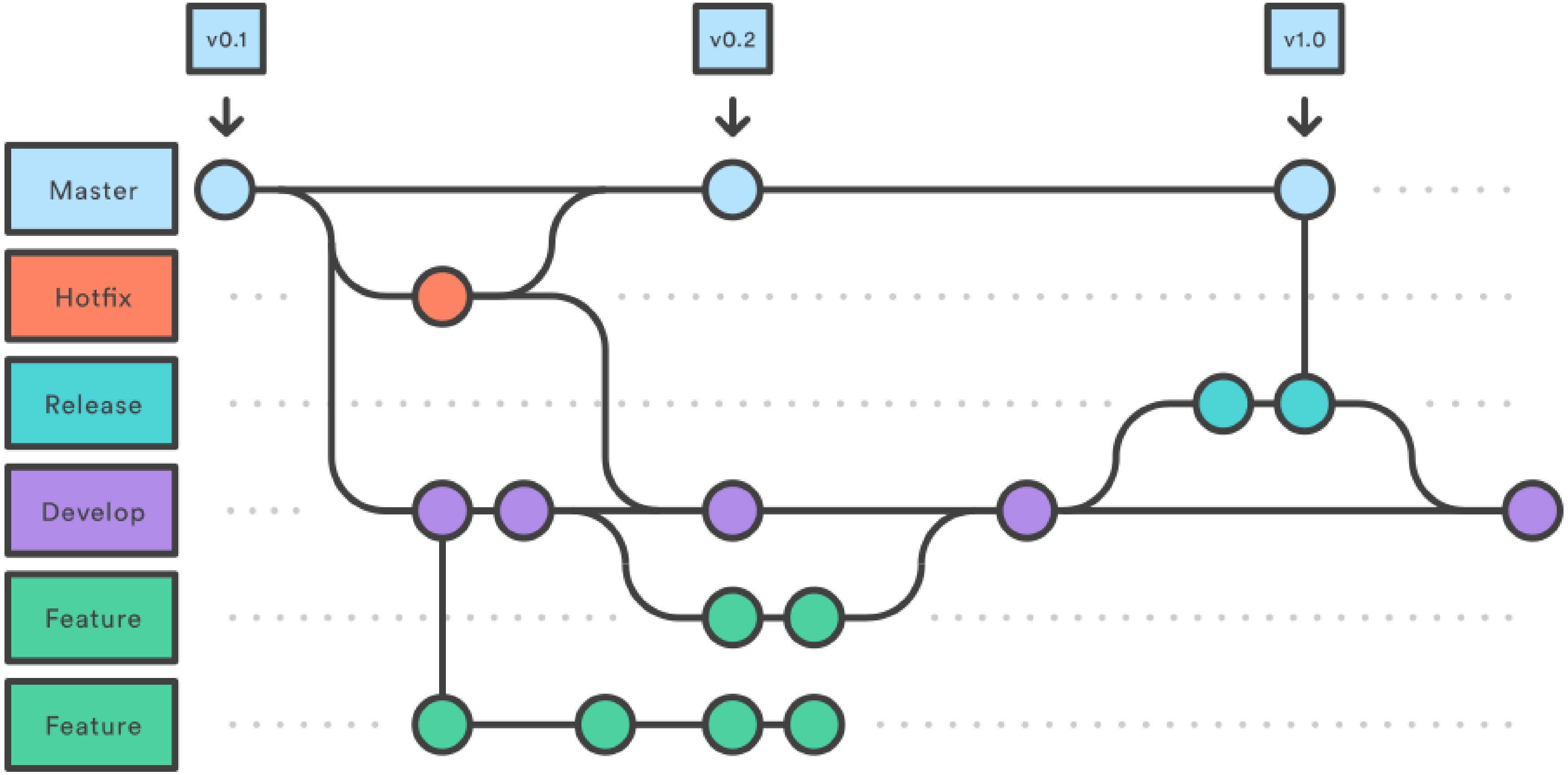
Versioning
Documenting
Open formats

Reproducibility

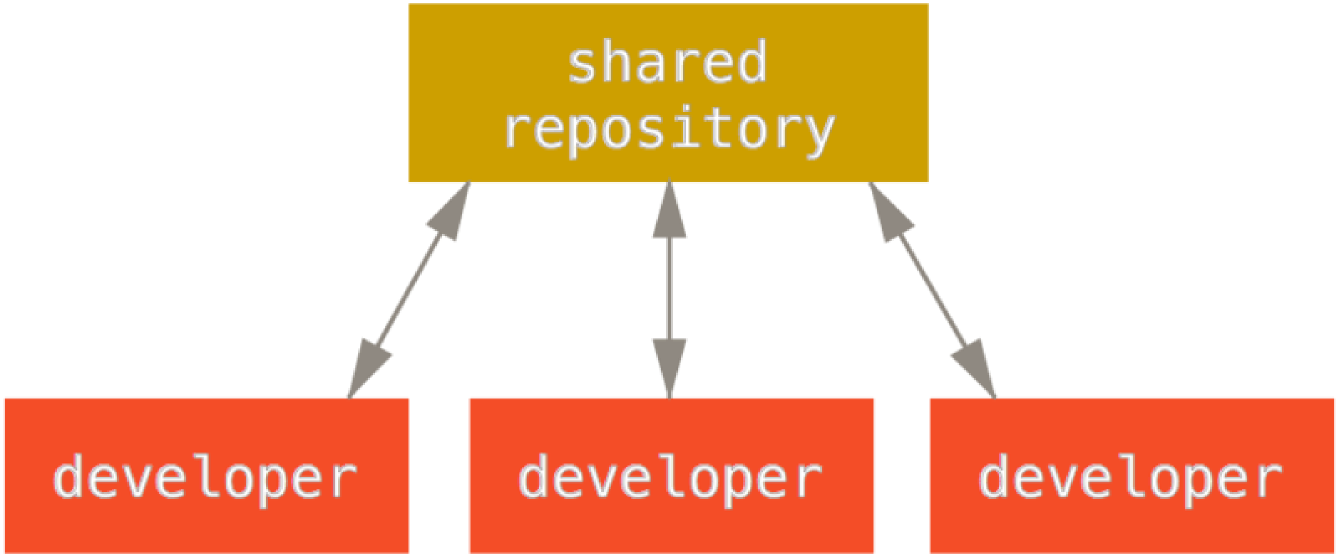
Replicability

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

Collect and Store: Software version control



Collect and Store: Software version control



Collect and Store: Software version control

git

https://git-scm.com/

- CLI (*Command Line interface)*
- GUIs (*Graphical User Interfaces)*
  https://git-scm.com/downloads/guis

Workstation

git

Your own server

git

Gitea

Forgejo

GitLab

Internet

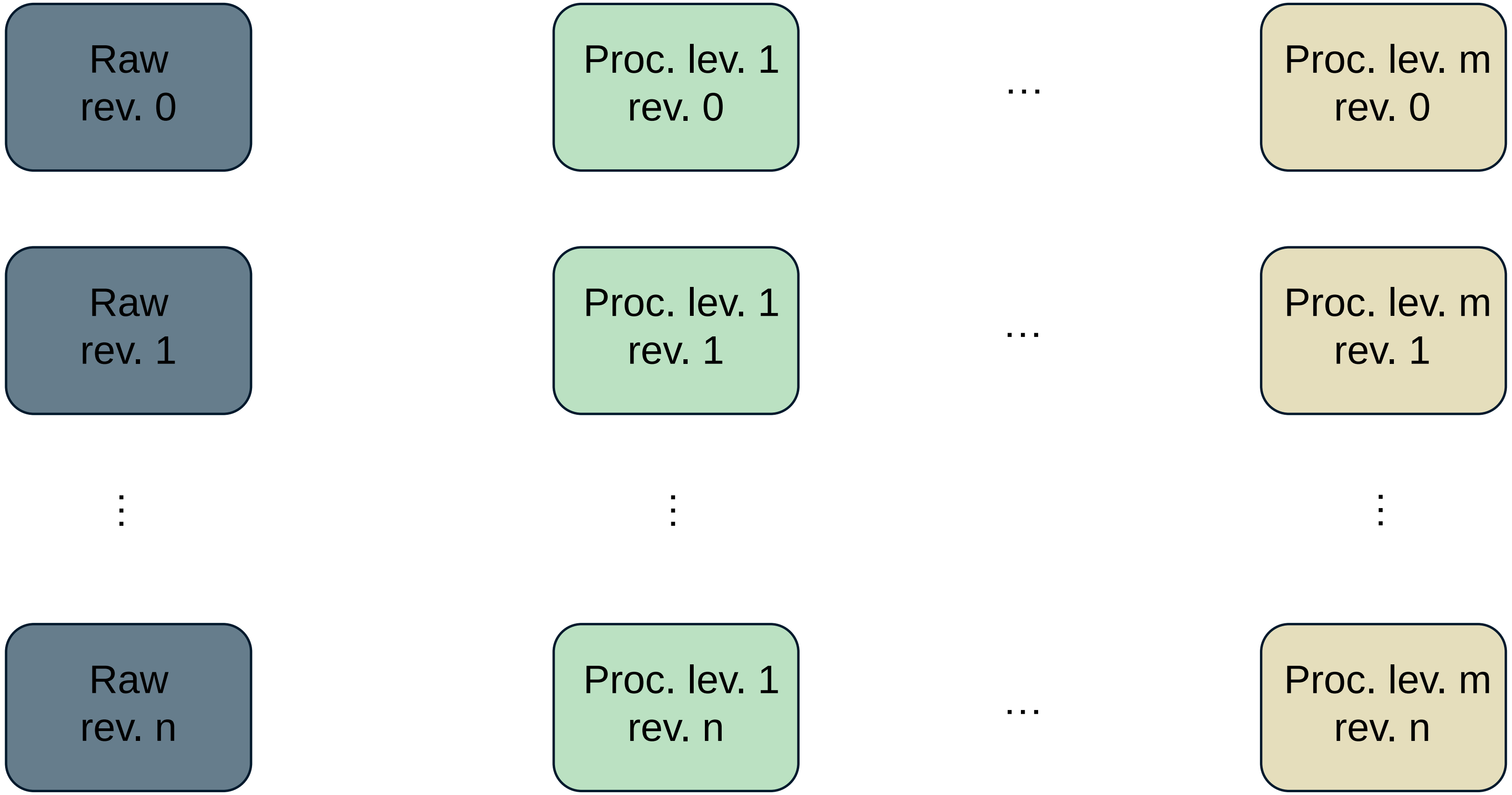
Codeberg

GitLab

GitHub

Bitbucket

Collect and Store: Data versioning



Raw rev. 0	Proc. lev. 1 rev. 0	…	Proc. lev. m rev. 0
Raw rev. 1	Proc. lev. 1 rev. 1	…	Proc. lev. m rev. 1
⋮	⋮		⋮
Raw rev. n	Proc. lev. 1 rev. n	…	Proc. lev. m rev. n

Collect and Store: Data versioning tools

Renku (*https://renku.readthedocs.io/en/stable/index.html*)

Data Version Control (*https://dvc.org*)

Git Large File Storage (*https://git-lfs.com*)

Lake FS(*https://docs.lakefs.io*)

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: File Formats (recommendation)

Data type	Recommended file formats
Text	• PDF/A • Plain Text coded as ACII. UTF-8 or UTF-16 • XML
Spreadsheet	• CSV (NEAD)
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 & 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

Evaluate & Archive

Evaluate & Archive: Data Protection

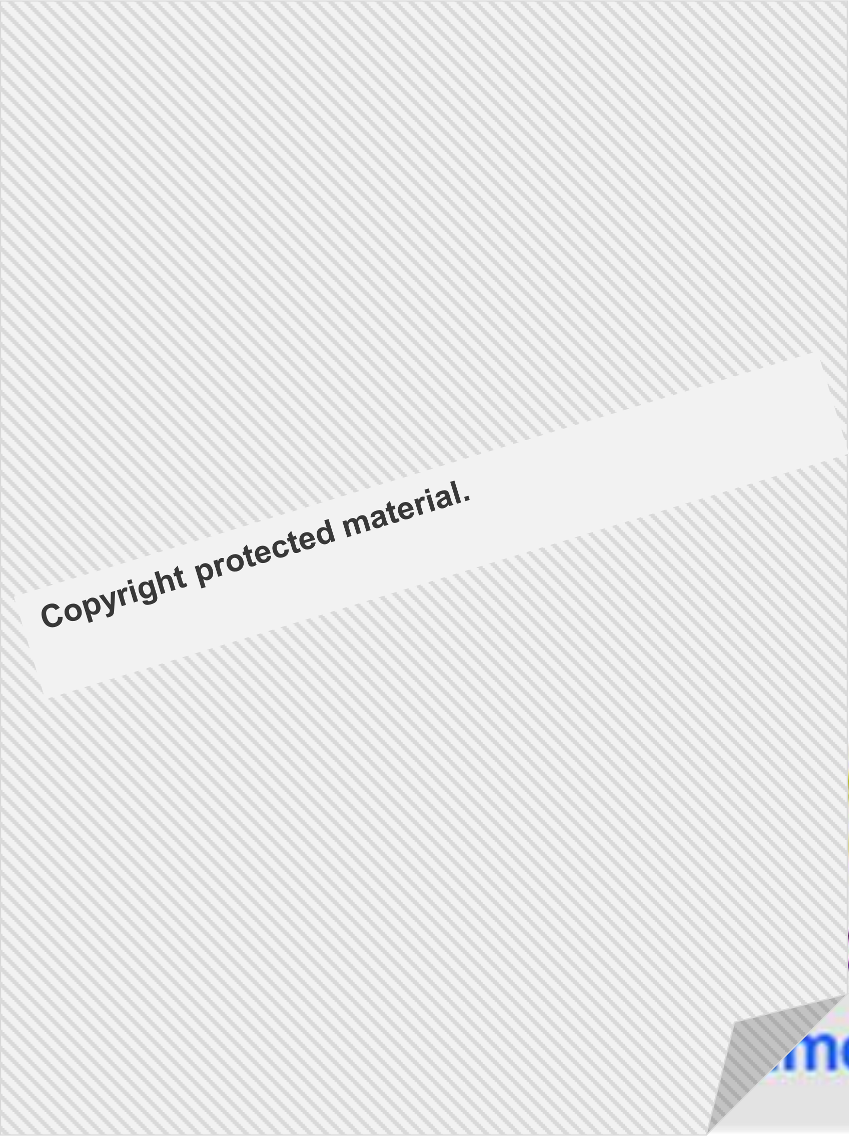
- **Relates to identified or identifiable person**
- **Solutions (https://dmlawtool.ccdigitallaw.ch/) :**
  - Identity irrelevant -> anonymisation
  - Identity relevant -> Ask for consent

    -> Pseudoanomization

    -> Manage access rights

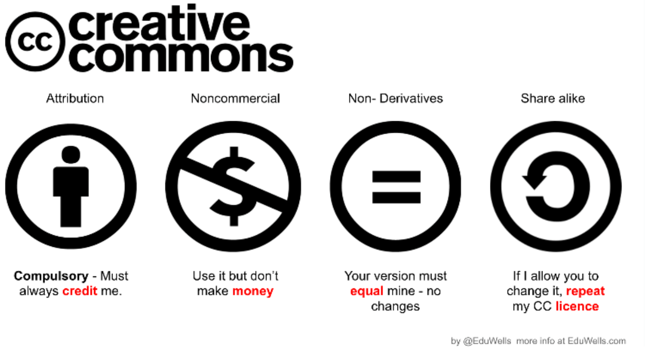
    -> Ability to address

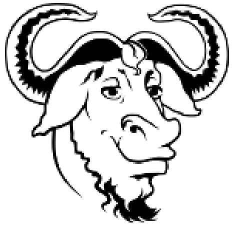
    subject’s rights
- **Always contact Data Protection Officers at your Research Institute if your research involves personal data**



Evaluate & Archive: Data Protection

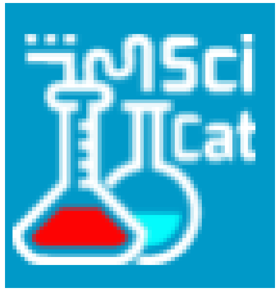
- Processed Data has copyright according to Swiss law
- Use CC licences when publishing factual data on data repositories (ideally CC 0)
- For software use licences specifically designed for software:
- Free Software (Open Source) licences like GPL, Apache, BSD and MIT.
- Exceptions! If you collaborated 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.


creative commons
Attribution
Noncommercial
Non- Derivatives
Share alike
Compulsory - Must always credit me.
Use it but don't make money
Your version must equal mine - no changes
If I allow you to change it, repeat my CC licence
by @EduWells more info at EduWells.com



Share & Disseminate

Share & Disseminate: The Choice of Data Repository

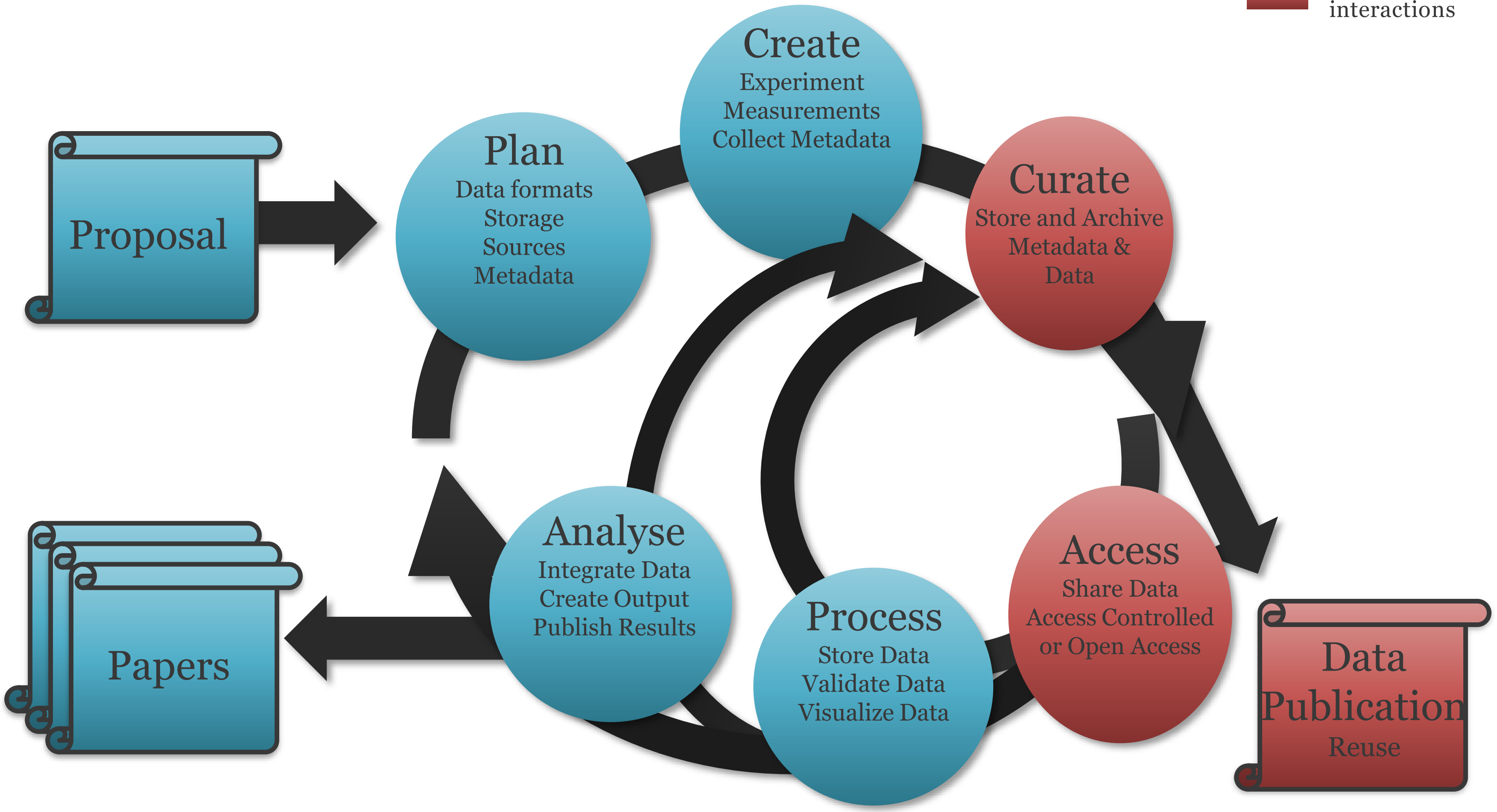


For alternatives: https://www.re3data.org/

RDM Services and Support at PSI

Interactions with the data catalogue

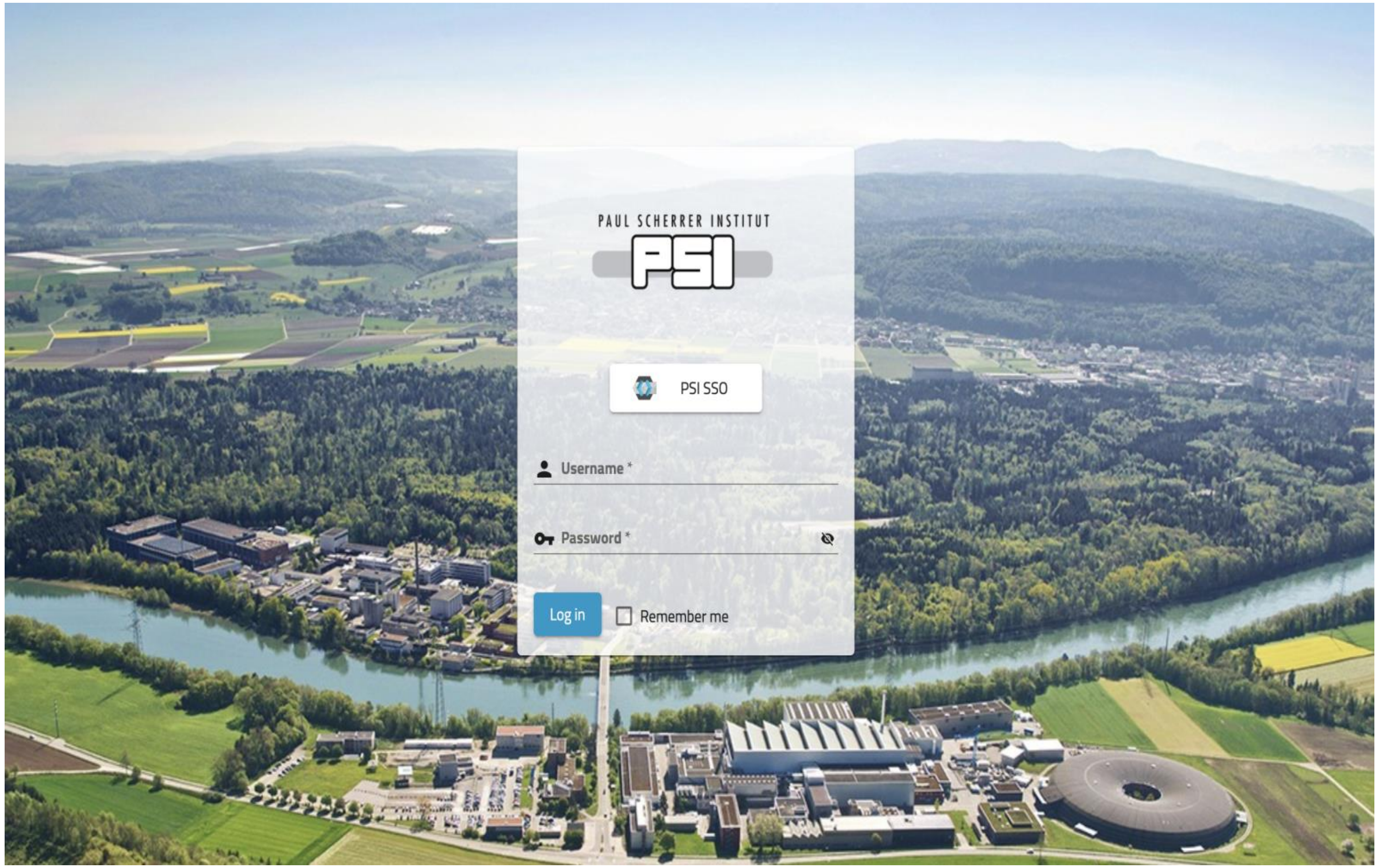
🟥 Data catalogue interactions



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



PAUL SCHERRER INSTITUT

PSI

PSI SSO

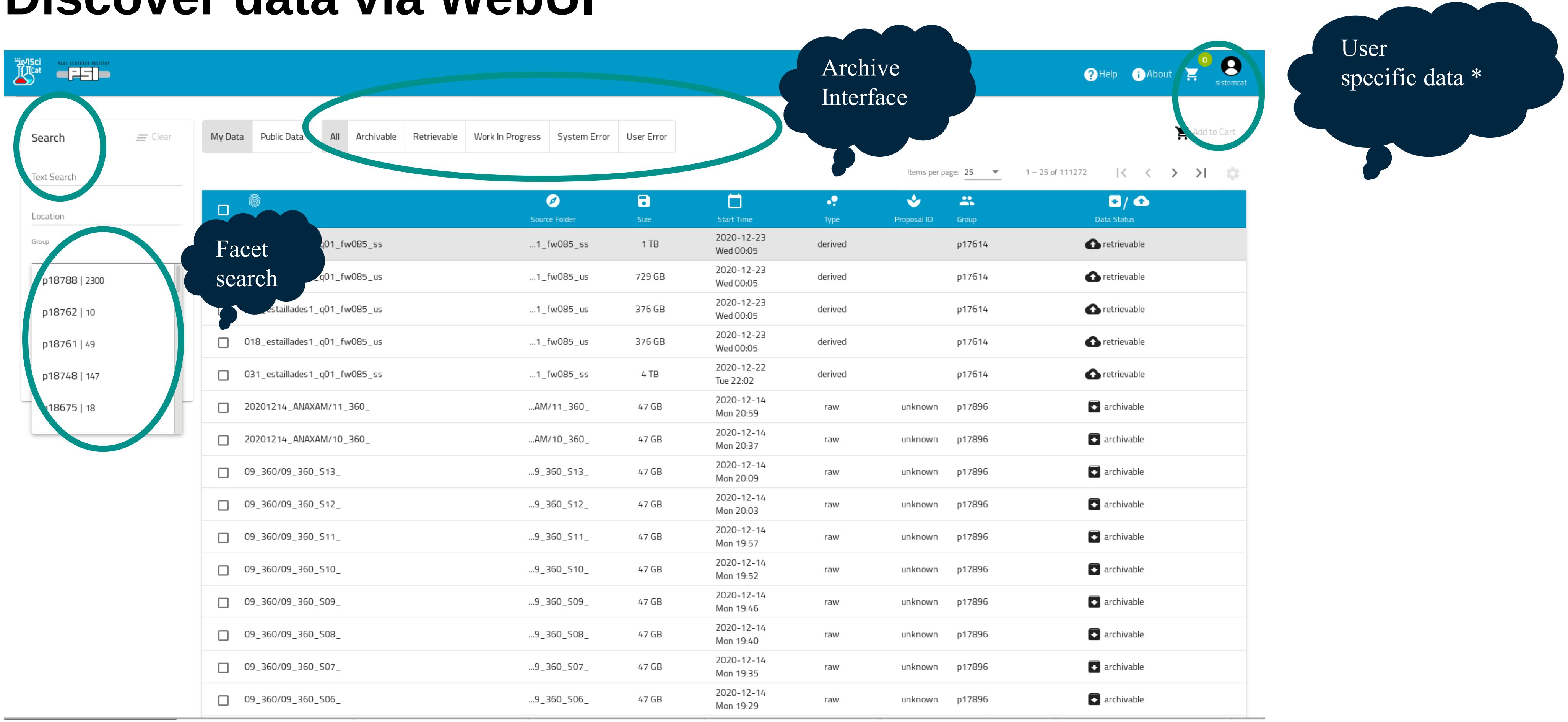
Username *

Password *

Log in

□ Remember me

Discover data via WebUI



Archive Interface

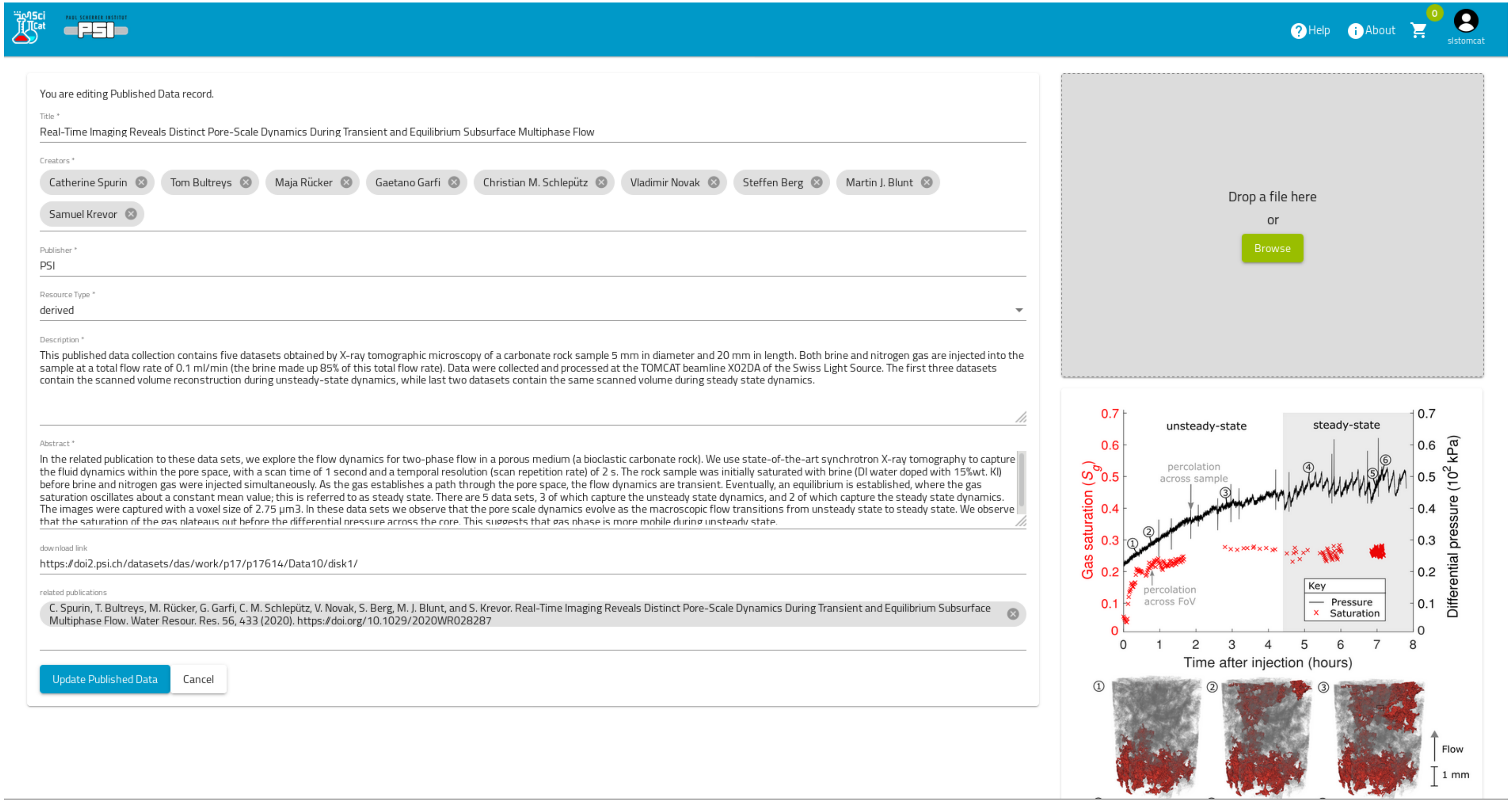
User specific data *

Facet search

□		Source Folder	Size	Start Time	Type	Proposal ID	Group	Data Status
	q01_fw085_ss	...1_fw085_ss	1 TB	2020-12-23 Wed 00:05	derived		p17614	retrievable
	_q01_fw085_us	...1_fw085_us	729 GB	2020-12-23 Wed 00:05	derived		p17614	retrievable
	estaillades1_q01_fw085_us	...1_fw085_us	376 GB	2020-12-23 Wed 00:05	derived		p17614	retrievable
□	018_estaillades1_q01_fw085_us	...1_fw085_us	376 GB	2020-12-23 Wed 00:05	derived		p17614	retrievable
□	031_estaillades1_q01_fw085_ss	...1_fw085_ss	4 TB	2020-12-22 Tue 22:02	derived		p17614	retrievable
□	20201214_ANAXAM/11_360_	...AM/11_360_	47 GB	2020-12-14 Mon 20:59	raw	unknown	p17896	archivable
□	20201214_ANAXAM/10_360_	...AM/10_360_	47 GB	2020-12-14 Mon 20:37	raw	unknown	p17896	archivable
□	09_360/09_360_S13_	...9_360_S13_	47 GB	2020-12-14 Mon 20:09	raw	unknown	p17896	archivable
□	09_360/09_360_S12_	...9_360_S12_	47 GB	2020-12-14 Mon 20:03	raw	unknown	p17896	archivable
□	09_360/09_360_S11_	...9_360_S11_	47 GB	2020-12-14 Mon 19:57	raw	unknown	p17896	archivable
□	09_360/09_360_S10_	...9_360_S10_	47 GB	2020-12-14 Mon 19:52	raw	unknown	p17896	archivable
□	09_360/09_360_S09_	...9_360_S09_	47 GB	2020-12-14 Mon 19:46	raw	unknown	p17896	archivable
□	09_360/09_360_S08_	...9_360_S08_	47 GB	2020-12-14 Mon 19:40	raw	unknown	p17896	archivable
□	09_360/09_360_S07_	...9_360_S07_	47 GB	2020-12-14 Mon 19:35	raw	unknown	p17896	archivable
□	09_360/09_360_S06_	...9_360_S06_	47 GB	2020-12-14 Mon 19:29	raw	unknown	p17896	archivable

* 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



You are editing Published Data record.

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

Creators *
Catherine Spurin ⊗ Tom Bultreys ⊗ Maja Rücker ⊗ Gaetano Garfi ⊗ Christian M. Schlepütz ⊗ Vladimir Novak ⊗ Steffen Berg ⊗ Martin J. Blunt ⊗ Samuel Krevor ⊗

Publisher *
PSI

Resource Type *
derived

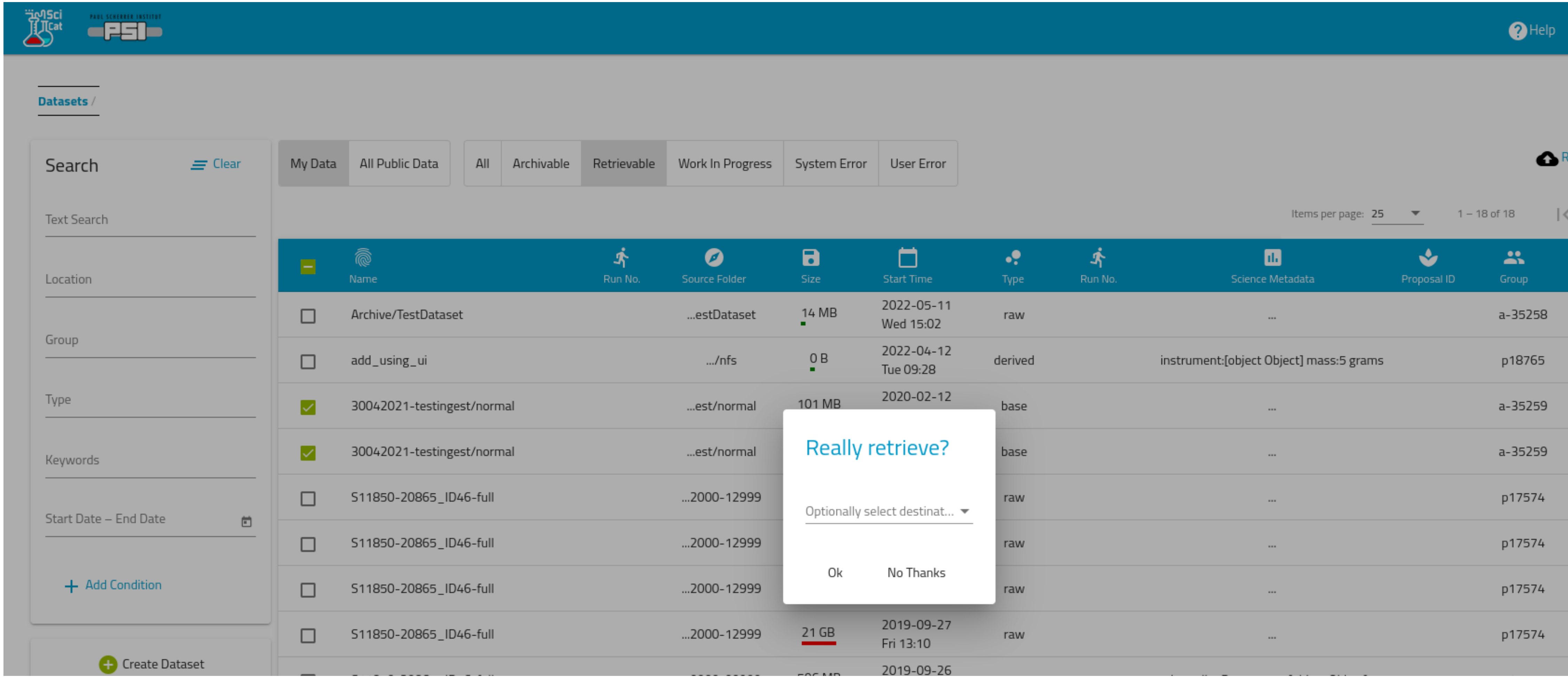
Description *
This published data collection contains five datasets obtained by X-ray tomographic microscopy of a carbonate rock sample 5 mm in diameter and 20 mm in length. Both brine and nitrogen gas are injected into the sample at a total flow rate of 0.1 ml/min (the brine made up 85% of this total 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.

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.

download link
https://doi2.psi.ch/datasets/das/work/p17/p17614/Data10/disk1/

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 ⊗

Retrieving data from tape



Items per page: 25   1 – 18 of 18

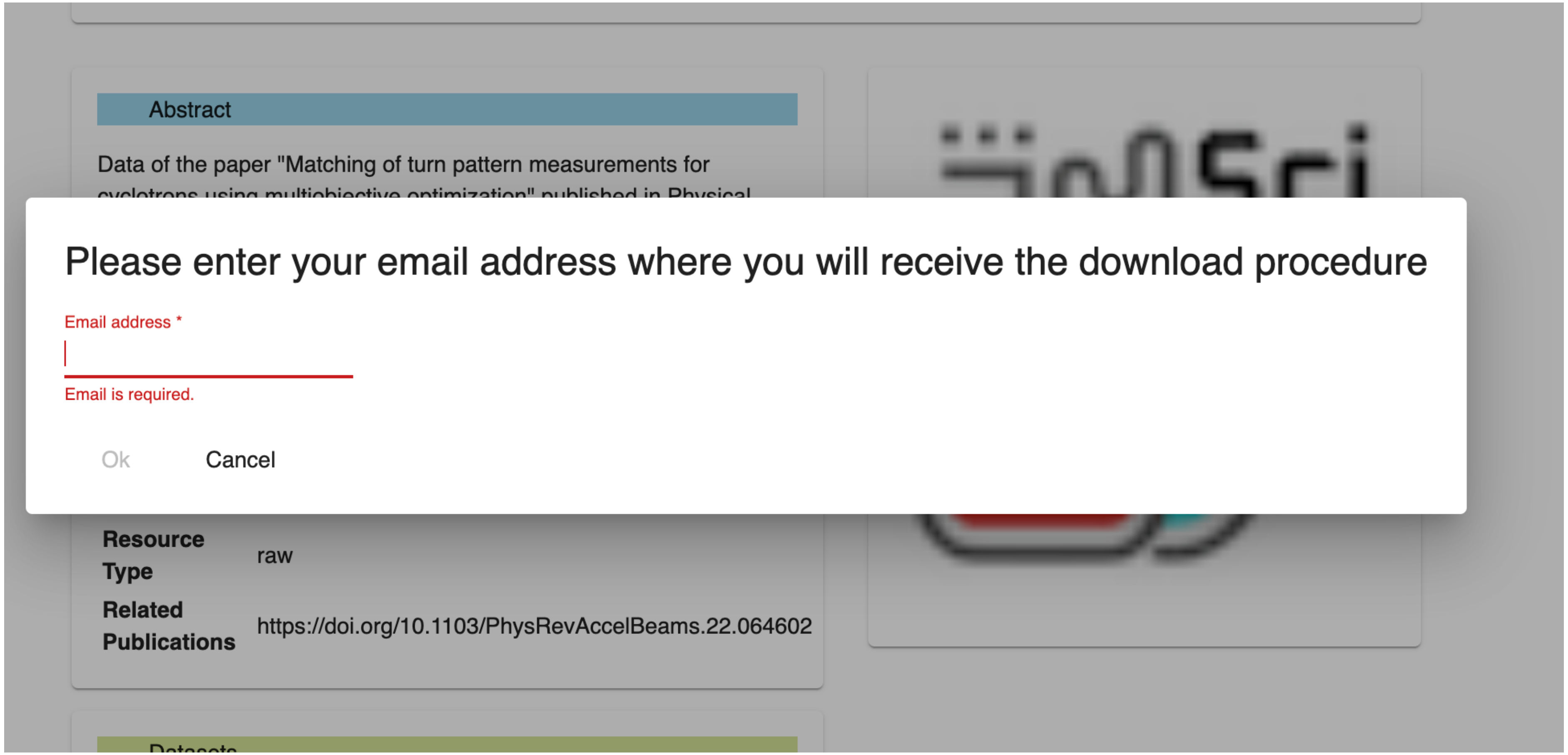
☐	Name	Run No.	Source Folder	Size	Start Time	Type	Run No.	Science Metadata	Proposal ID	Group
☐	Archive/TestDataset		...estDataset	14 MB	2022-05-11 Wed 15:02	raw		...		a-35258
☐	add_using_ui		.../nfs	0 B	2022-04-12 Tue 09:28	derived		instrument:[object Object] mass:5 grams		p18765
☑	30042021-testingest/normal		...est/normal	101 MB	2020-02-12	base		...		a-35259
☑	30042021-testingest/normal		...est/normal			base		...		a-35259
☐	S11850-20865_ID46-full		...2000-12999			raw		...		p17574
☐	S11850-20865_ID46-full		...2000-12999			raw		...		p17574
☐	S11850-20865_ID46-full		...2000-12999			raw		...		p17574
☐	S11850-20865_ID46-full		...2000-12999	21 GB	2019-09-27 Fri 13:10	raw		...		p17574
					2019-09-26					

Really retrieve?

Optionally select destinat...

Ok   No Thanks

Retrieving public data from tape



Abstract

Data of the paper "Matching of turn pattern measurements for cyclotrons using multiobjective optimization" published in Physical

Please enter your email address where you will receive the download procedure

Email address *

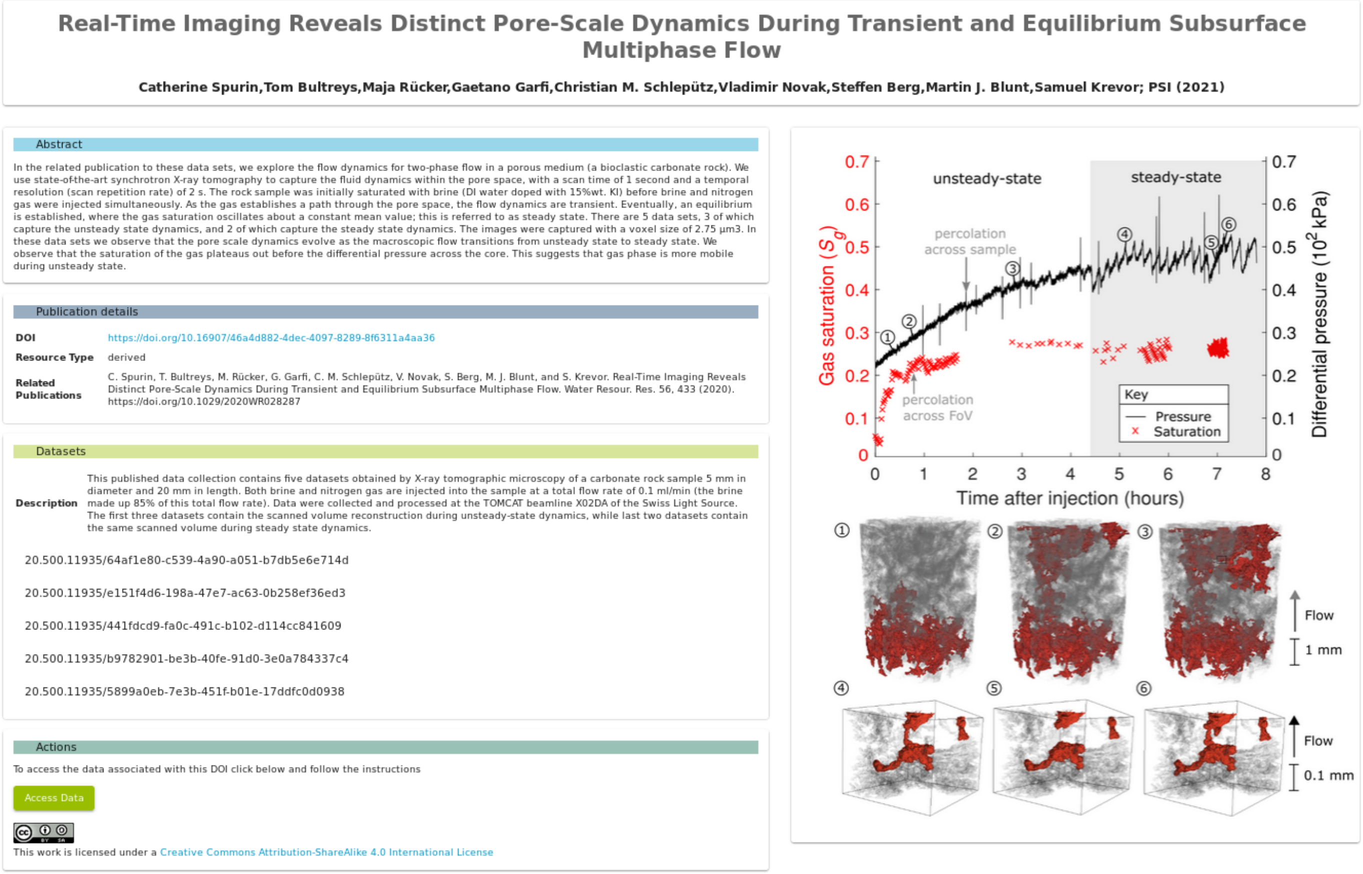
Email is required.

Resource Type: raw

Related Publications: https://doi.org/10.1103/PhysRevAccelBeams.22.064602

Datasets

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

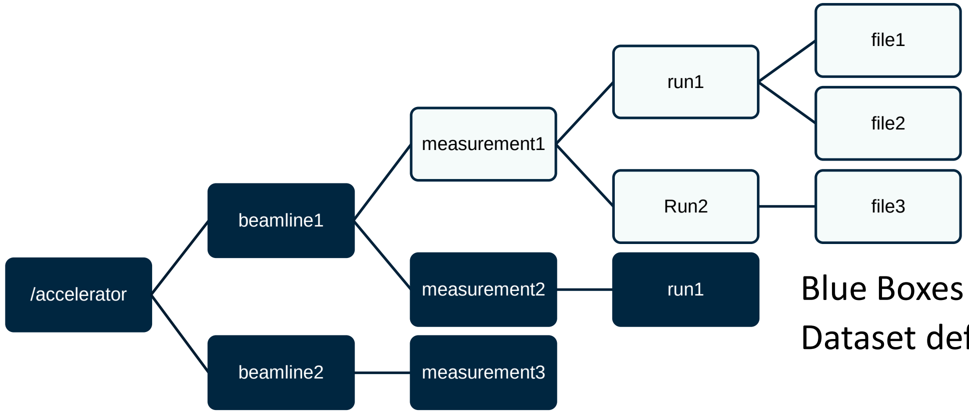
20.500.11935/5899a0eb-7e3b-451f-b01e-17ddfc0d0938

Actions

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

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.



Blue Boxes are candidates for Dataset definitions

- 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.

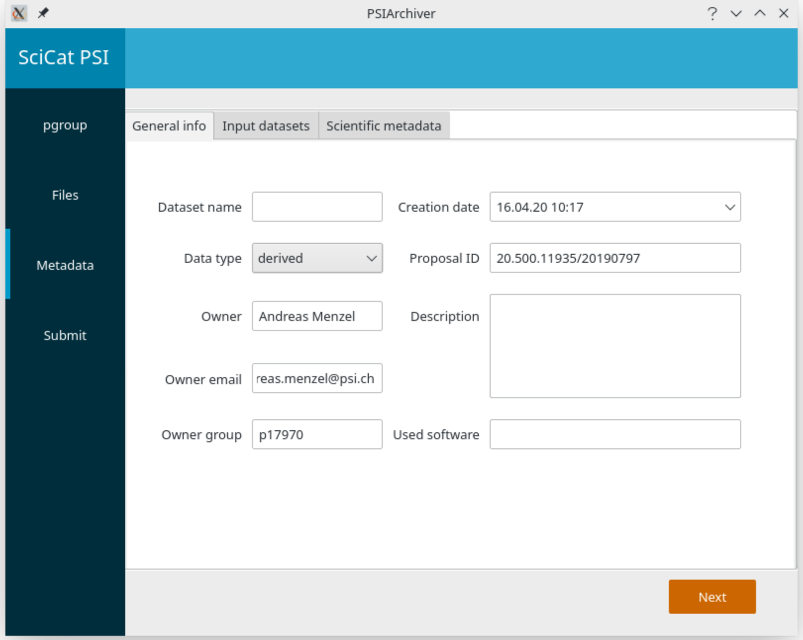
Ideally following a standard if it exists, e.g. NeXus based HDF5 files, extracted from instrument.

Example:

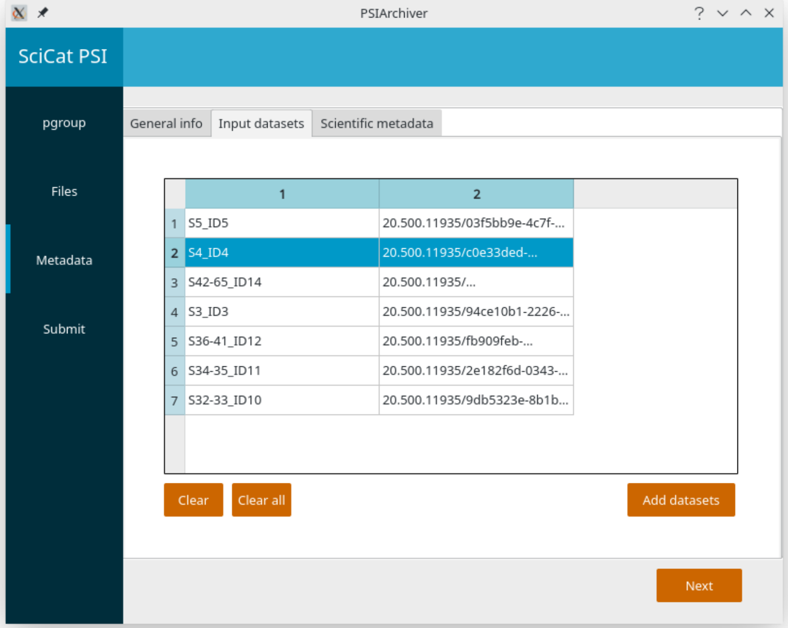
```
    “scientificMetadata": {
    "beamlineParameters": {
        ”monostripe": "Ru/C",
        ”ring_current": {
            "value": 0.402246,
            "units": "A"
        },
        ”beam_energy": {
            "value": 22595,
            "units": "eV"
        }
    },
    "detectorParameters":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

Manual Ingests via Qt GUI tool at PSI

Especially for derived data:



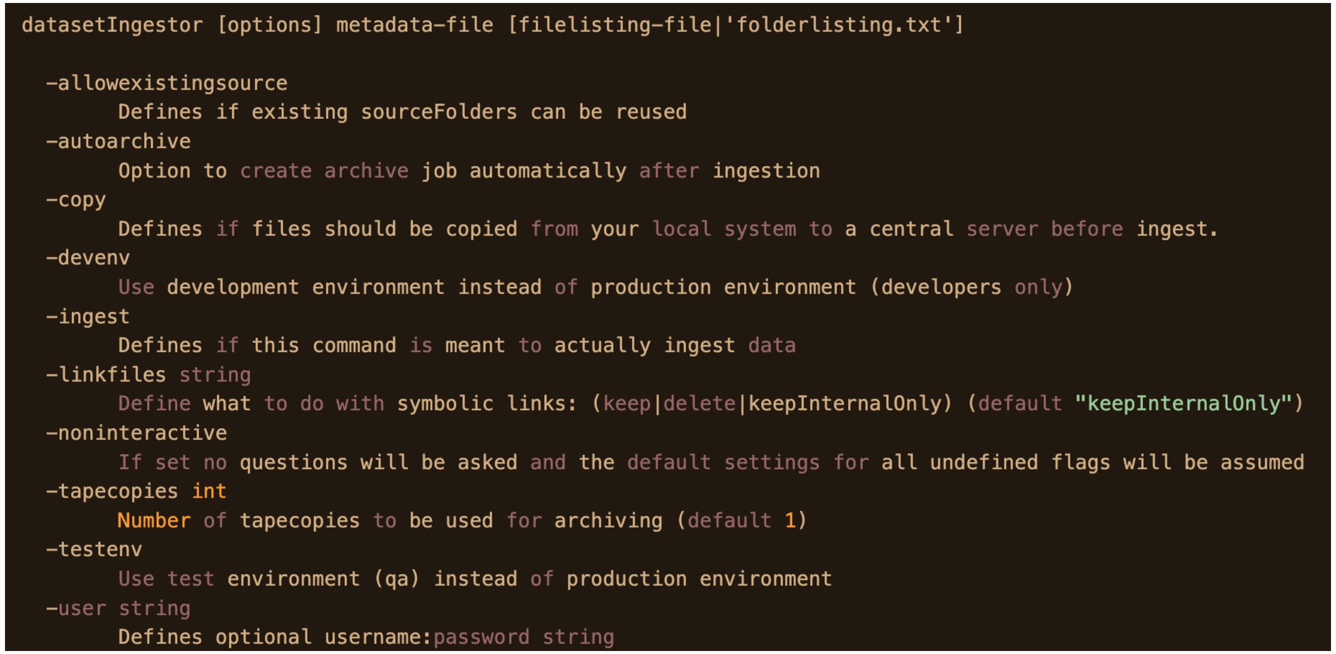
Dataset name		Creation date	16.04.20 10:17
Data type	derived	Proposal ID	20.500.11935/20190797
Owner	Andreas Menzel	Description	
Owner email	reas.menzel@psi.ch		
Owner group	p17970	Used software	



	1	2
1	S5_ID5	20.500.11935/03f5bb9e-4c7f-...
2	S4_ID4	20.500.11935/c0e33ded-...
3	S42-65_ID14	20.500.11935/...
4	S3_ID3	20.500.11935/94ce10b1-2226-...
5	S36-41_ID12	20.500.11935/fb909feb-...
6	S34-35_ID11	20.500.11935/2e182f6d-0343-...
7	S32-33_ID10	20.500.11935/9db5323e-8b1b...

Manual Ingests via CLI tool at PSI

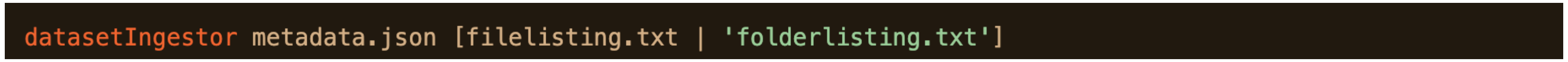
Linux and Windows command line tool (datasetIngestor example):



```
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
```

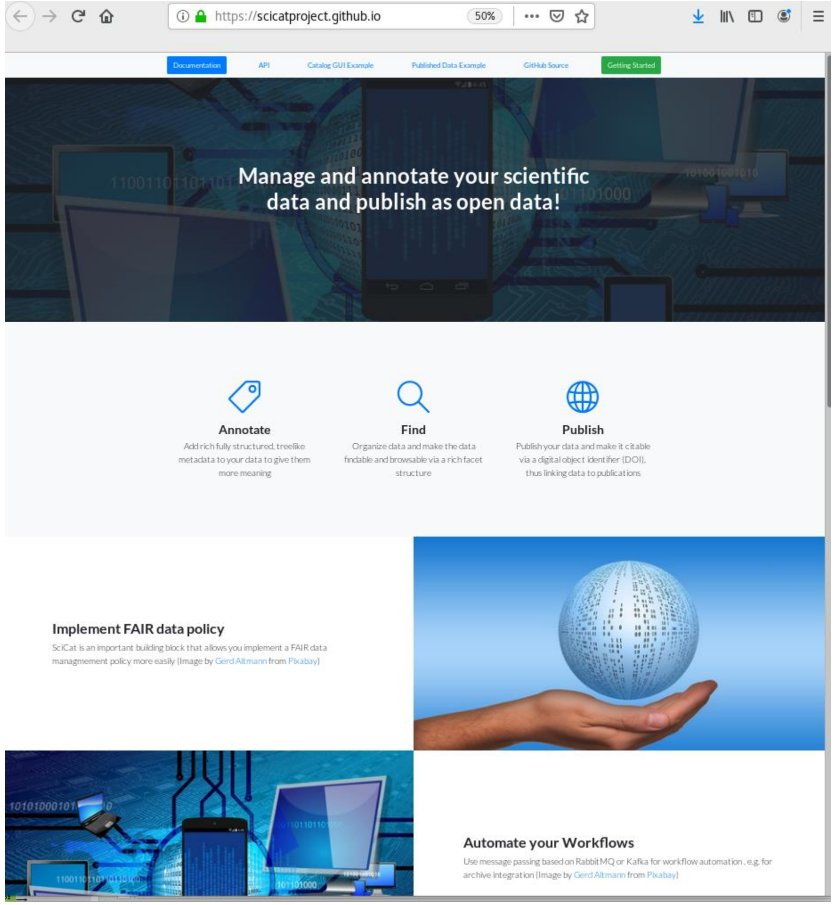
datasetIngestor usage example:



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

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

Documentation: scicatproject.github.io



Manage and annotate your scientific data and publish as open data!

Annotate
Add rich fully structured, treelike metadata to your data to give them more meaning

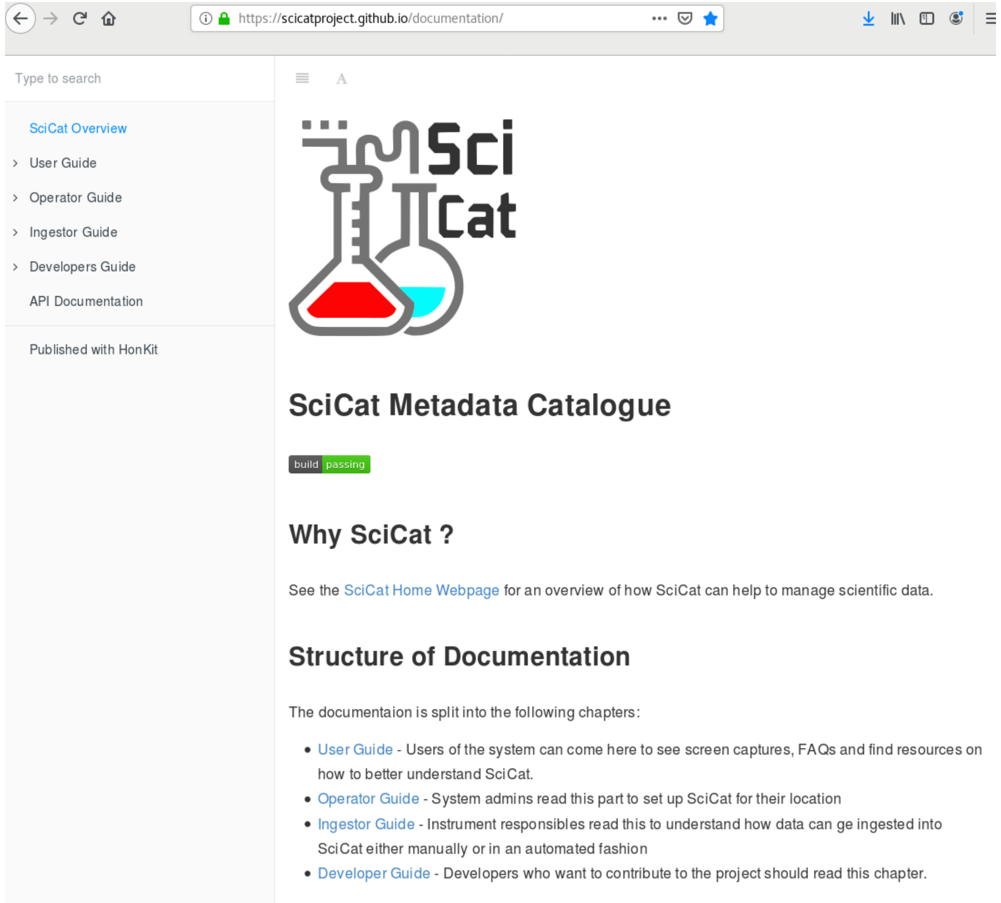
Find
Organize data and make the data findable and browsable via a rich facet structure

Publish
Publish your data and make it citable via a digital object identifier (DOI), thus linking data to publications

Implement FAIR data policy
SciCat is an important building block that allows you implement a FAIR data managmement policy more easily (Image by Gerd Altmann from Pixabay)

Automate your Workflows
Use message passing based on RabbitMQ or Kafka for workflow automation , e.g. for archive integration (Image by Gerd Altmann from Pixabay)

Full Documentation for Users and Operators



SciCat Metadata Catalogue

Why SciCat ?

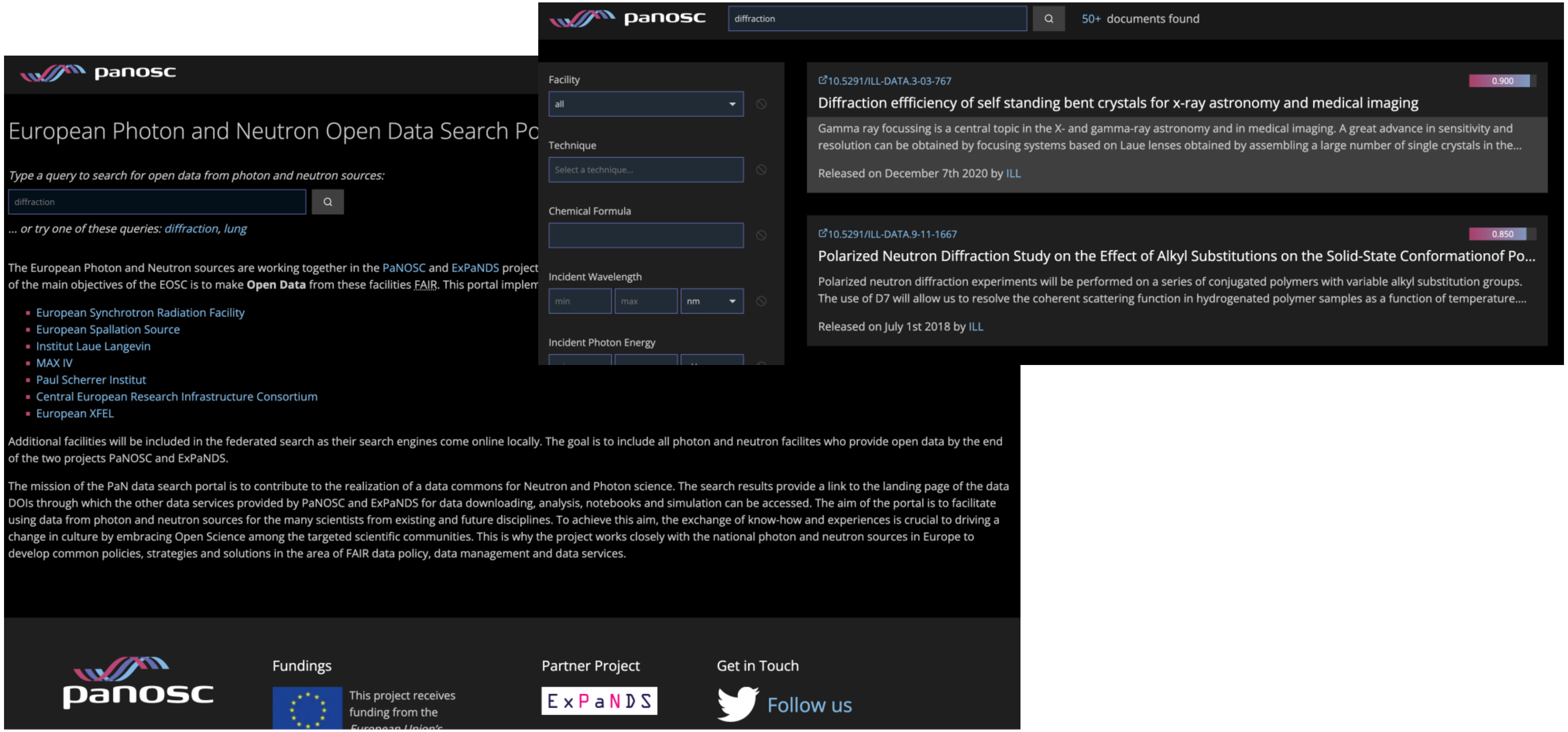
See the SciCat Home Webpage for an overview of how SciCat can help to manage scientific data.

Structure of Documentation

The documentaion is split into the following chapters:

- User Guide - Users of the system can come here to see screen captures, FAQs and find resources on how to better understand SciCat.
- Operator Guide - System admins read this part to set up SciCat for their location
- Ingestor Guide - Instrument responsibles read this to understand how data can ge ingested into SciCat either manually or in an automated fashion
- Developer Guide - Developers who want to contribute to the project should read this chapter.

PaNOSC and ExPaNDS Open Data Search Portal



European Photon and Neutron Open Data Search Po

Type a query to search for open data from photon and neutron sources:

diffraction

... or try one of these queries: diffraction, lung

The European Photon and Neutron sources are working together in the PaNOSC and ExPaNDS project
of the main objectives of the EOSC is to make **Open Data** from these facilities FAIR. This portal implem

- European Synchrotron Radiation Facility
- European Spallation Source
- Institut Laue Langevin
- MAX IV
- Paul Scherrer Institut
- Central European Research Infrastructure Consortium
- European XFEL

Additional facilities will be included in the federated search as their search engines come online locally. The goal is to include all photon and neutron facilites who provide open data by the end of the two projects PaNOSC and ExPaNDS.

The mission of the PaN data search portal is to contribute to the realization of a data commons for Neutron and Photon science. The search results provide a link to the landing page of the data DOIs through which the other data services provided by PaNOSC and ExPaNDS for data downloading, analysis, notebooks and simulation can be accessed. The aim of the portal is to facilitate using data from photon and neutron sources for the many scientists from existing and future disciplines. To achieve this aim, the exchange of know-how and experiences is crucial to driving a change in culture by embracing Open Science among the targeted scientific communities. This is why the project works closely with the national photon and neutron sources in Europe to develop common policies, strategies and solutions in the area of FAIR data policy, data management and data services.

diffraction

50+ documents found

Facility: all

Technique: Select a technique...

Chemical Formula:

Incident Wavelength: min max nm

Incident Photon Energy

10.5291/ILL-DATA.3-03-767 — 0.900

Diffraction effficiency of self standing bent crystals for x-ray astronomy and medical imaging

Gamma ray focussing is a central topic in the X- and gamma-ray astronomy and in medical imaging. A great advance in sensitivity and resolution can be obtained by focusing systems based on Laue lenses obtained by assembling a large number of single crystals in the...

Released on December 7th 2020 by ILL

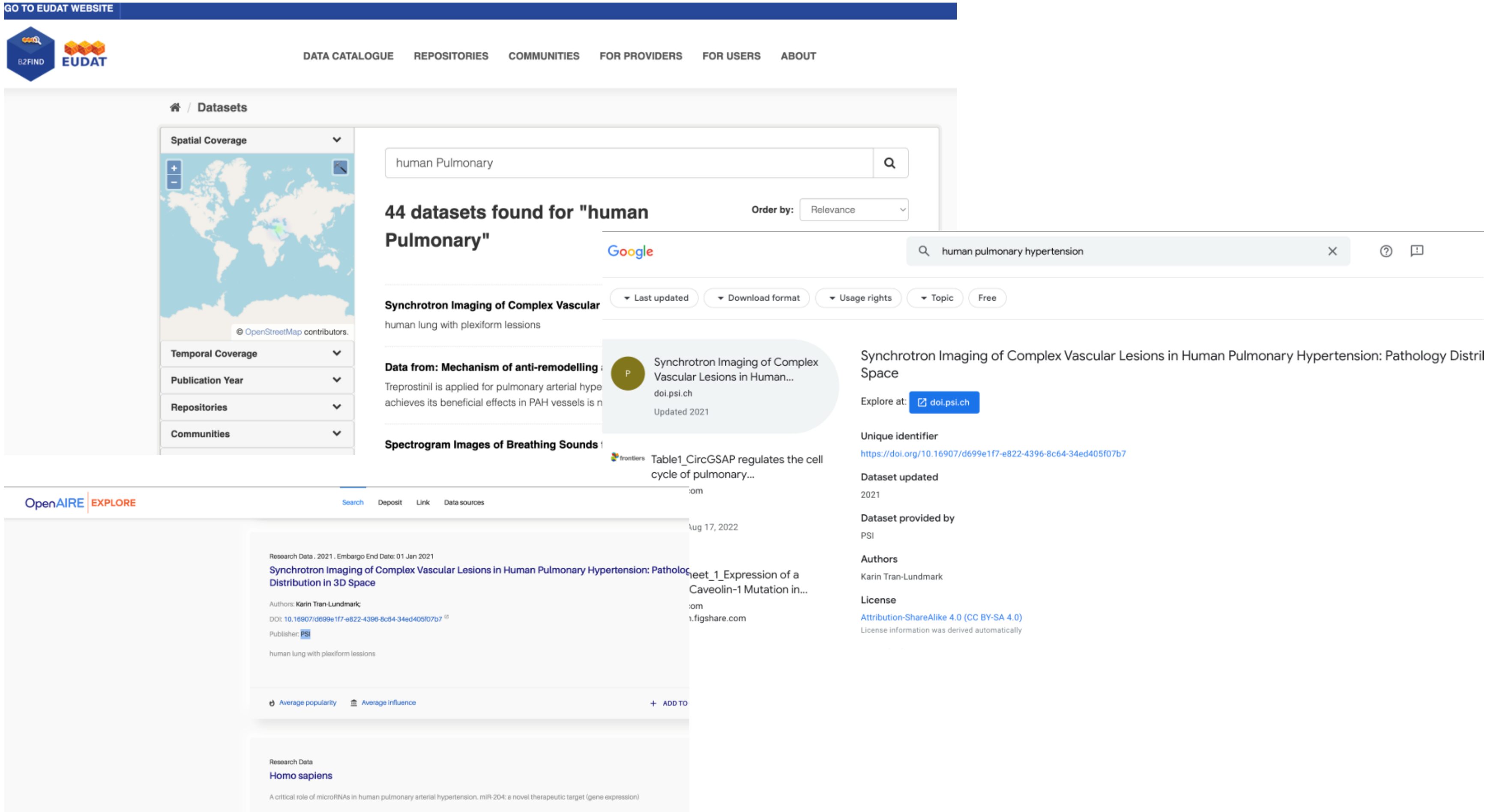
10.5291/ILL-DATA.9-11-1667 — 0.850

Polarized Neutron Diffraction Study on the Effect of Alkyl Substitutions on the Solid-State Conformationof Po...

Polarized neutron diffraction experiments will be performed on a series of conjugated polymers with variable alkyl substitution groups. The use of D7 will allow us to resolve the coherent scattering function in hydrogenated polymer samples as a function of temperature....

Released on July 1st 2018 by ILL

Data repositories



human Pulmonary

44 datasets found for "human Pulmonary"

Synchrotron Imaging of Complex Vascular
human lung with plexiform lessions

Data from: Mechanism of anti-remodelling
Treprostinil is applied for pulmonary arterial hype
achieves its beneficial effects in PAH vessels is n

Spectrogram Images of Breathing Sounds

human pulmonary hypertension

Synchrotron Imaging of Complex Vascular Lesions in Human...
doi.psi.ch
Updated 2021

Table1_CircGSAP regulates the cell cycle of pulmonary...
om
Aug 17, 2022

heet_1_Expression of a Caveolin-1 Mutation in...
om
.figshare.com

Synchrotron Imaging of Complex Vascular Lesions in Human Pulmonary Hypertension: Pathology Distril Space

Explore at: doi.psi.ch

Unique identifier
https://doi.org/10.16907/d699e1f7-e822-4396-8c64-34ed405f07b7

Dataset updated
2021

Dataset provided by
PSI

Authors
Karin Tran-Lundmark

License
Attribution-ShareAlike 4.0 (CC BY-SA 4.0)
License information was derived automatically

Research Data . 2021 . Embargo End Date: 01 Jan 2021
Synchrotron Imaging of Complex Vascular Lesions in Human Pulmonary Hypertension: Patholog Distribution in 3D Space
Authors: Karin Tran-Lundmark;
DOI: 10.16907/d699e1f7-e822-4396-8c64-34ed405f07b7
Publisher: PSI
human lung with plexiform lessions

Research Data
Homo sapiens
A critical role of microRNAs in human pulmonary arterial hypertension. miR-204: a novel therapeutic target (gene expression)

Thanks to all contributors!

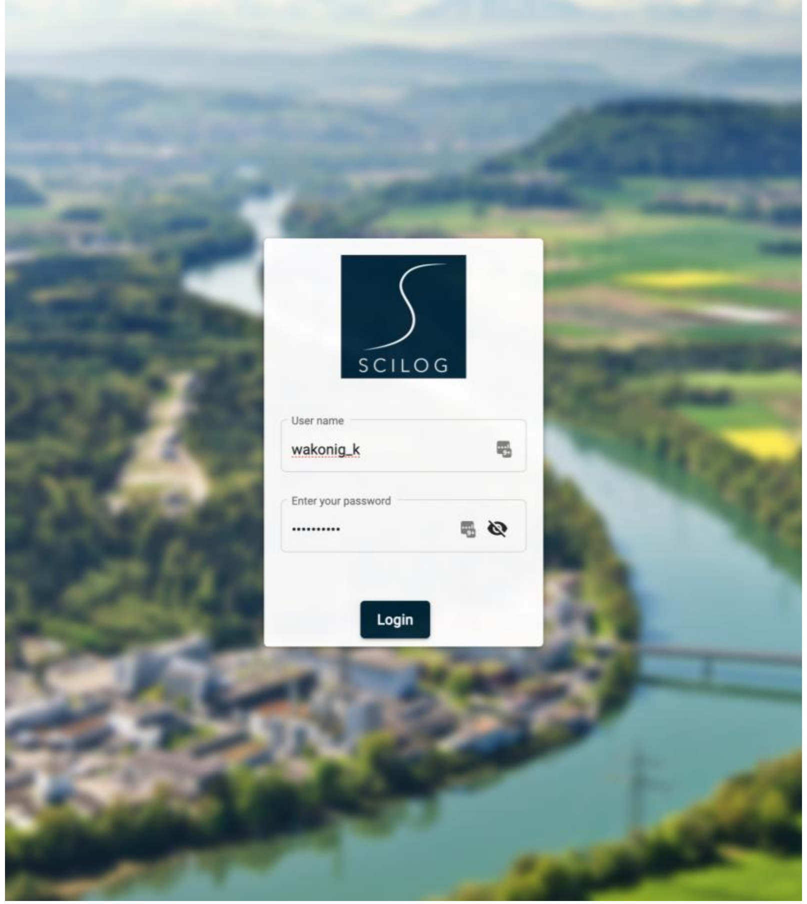
- Laura Shemilt
- Linus Pithan
- Dylan McReynolds
- Tobias Richter
- Chris Gwilliams
- Luke Gorman
- Hannes Petri
- Gareth Murphy

- Stephan Egli
- Frederik Bolmsten
- Carlo Minotti
- Max Novelli
- Henrik Johansson
- Marco Leorato
- Linh Nguyen
- Anastasiia Pylypenko

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

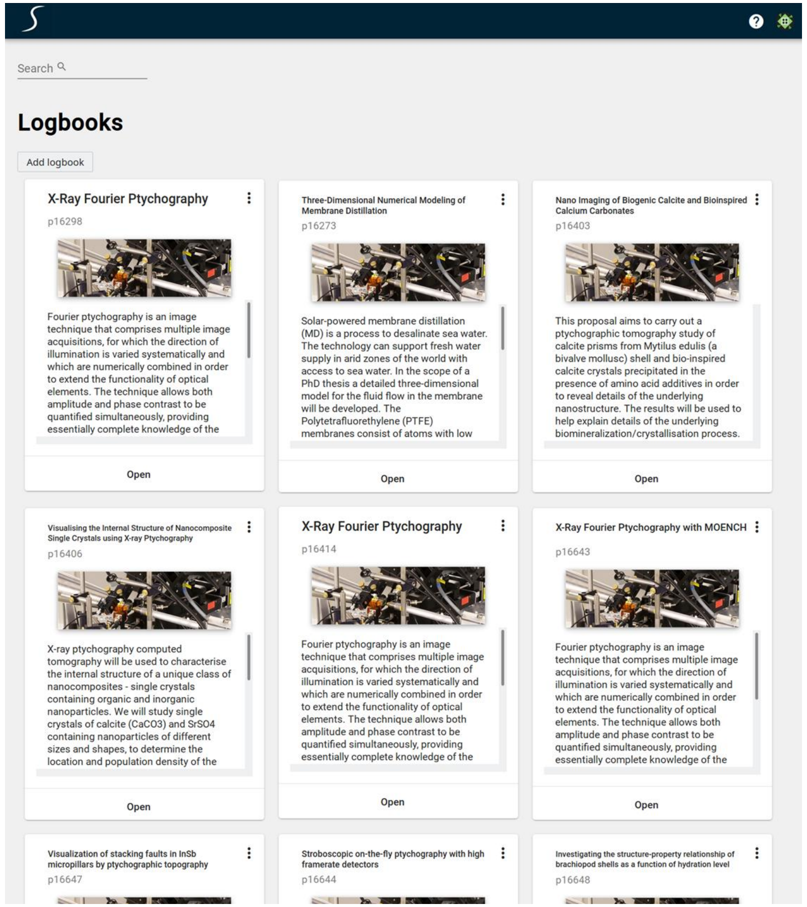
User authentication



User name
wakonig_k

Enter your password
••••••••••

Viewing, searching, adding and editing



Logbooks

X-Ray Fourier Ptychography
p16298

Fourier ptychography is an image technique that comprises multiple image acquisitions, for which the direction of illumination is varied systematically and which are numerically combined in order to extend the functionality of optical elements. The technique allows both amplitude and phase contrast to be quantified simultaneously, providing essentially complete knowledge of the

Three-Dimensional Numerical Modeling of Membrane Distillation
p16273

Solar-powered membrane distillation (MD) is a process to desalinate sea water. The technology can support fresh water supply in arid zones of the world with access to sea water. In the scope of a PhD thesis a detailed three-dimensional model for the fluid flow in the membrane will be developed. The Polytetrafluorethylene (PTFE) membranes consist of atoms with low

Nano Imaging of Biogenic Calcite and Bioinspired Calcium Carbonates
p16403

This proposal aims to carry out a ptychographic tomography study of calcite prisms from Mytilus edulis (a bivalve mollusc) shell and bio-inspired calcite crystals precipitated in the presence of amino acid additives in order to reveal details of the underlying nanostructure. The results will be used to help explain details of the underlying biomineralization/crystallisation process.

Visualising the Internal Structure of Nanocomposite Single Crystals using X-ray Ptychography
p16406

X-ray ptychography computed tomography will be used to characterise the internal structure of a unique class of nanocomposites - single crystals containing organic and inorganic nanoparticles. We will study single crystals of calcite (CaCO3) and SrSO4 containing nanoparticles of different sizes and shapes, to determine the location and population density of the

X-Ray Fourier Ptychography
p16414

Fourier ptychography is an image technique that comprises multiple image acquisitions, for which the direction of illumination is varied systematically and which are numerically combined in order to extend the functionality of optical elements. The technique allows both amplitude and phase contrast to be quantified simultaneously, providing essentially complete knowledge of the

X-Ray Fourier Ptychography with MOENCH
p16643

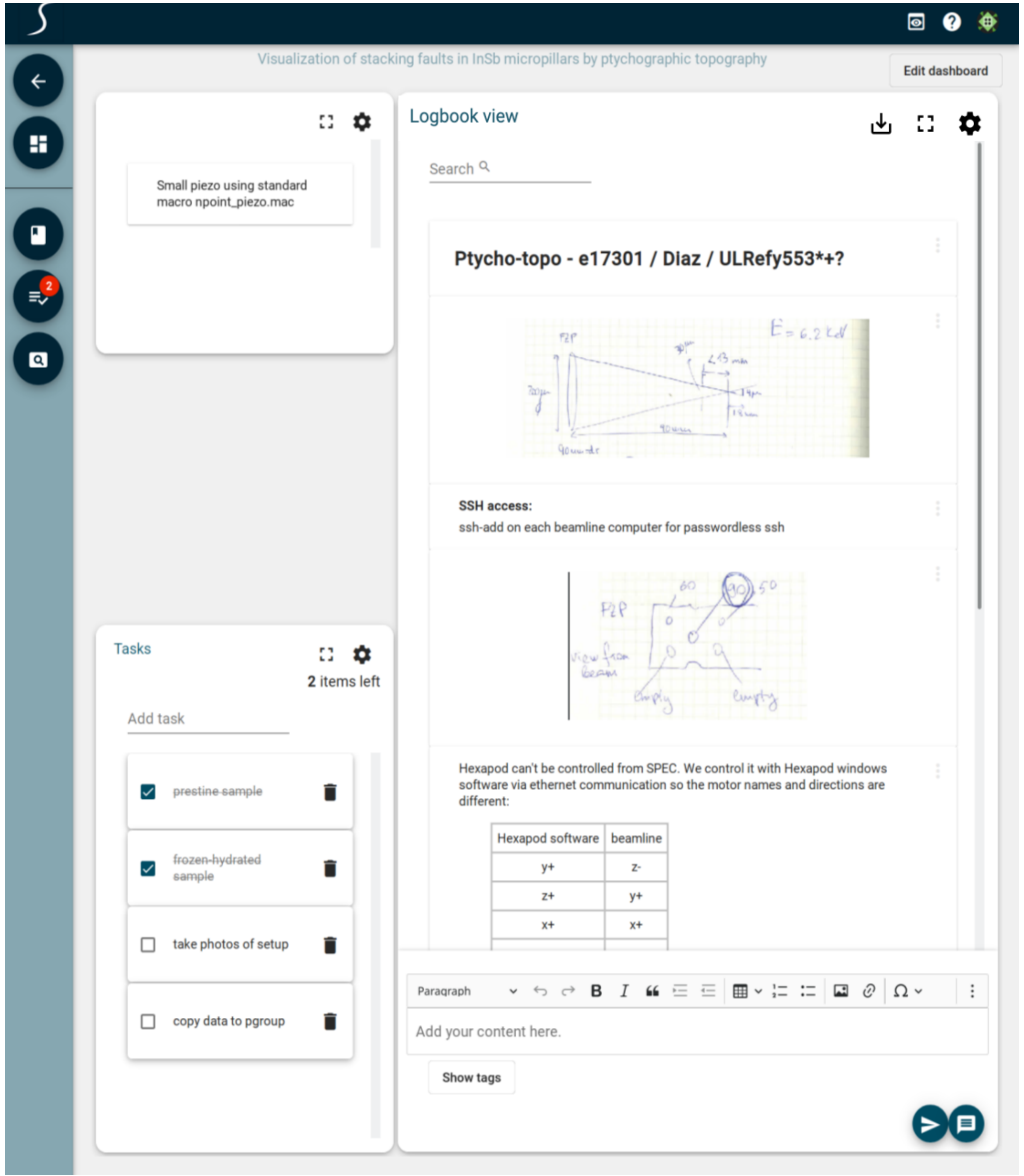
Fourier ptychography is an image technique that comprises multiple image acquisitions, for which the direction of illumination is varied systematically and which are numerically combined in order to extend the functionality of optical elements. The technique allows both amplitude and phase contrast to be quantified simultaneously, providing essentially complete knowledge of the

Visualization of stacking faults in InSb micropillars by ptychographic topography
p16647

Stroboscopic on-the-fly ptychography with high framerate detectors
p16644

Investigating the structure-property relationship of brachiopod shells as a function of hydration level
p16648

Inside a logbook



Visualization of stacking faults in InSb micropillars by ptychographic topography

Small piezo using standard macro npoint_piezo.mac

Tasks

2 items left

☑ ~~prestine sample~~

☑ ~~frozen-hydrated sample~~

☐ take photos of setup

☐ copy data to pgroup

Logbook view

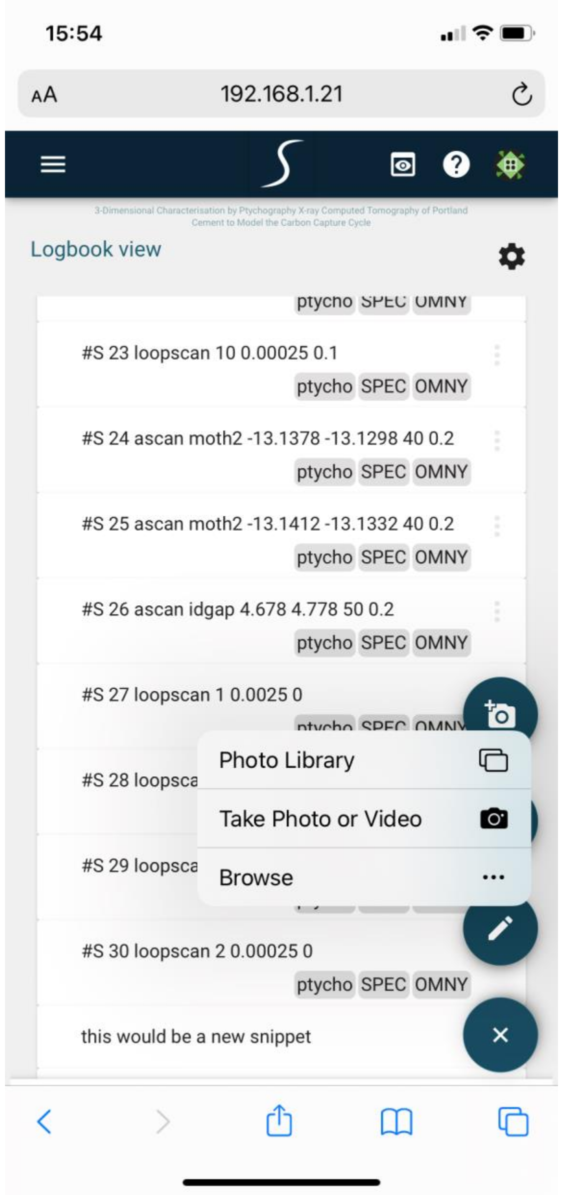
Ptycho-topo - e17301 / Diaz / ULRefy553*+?

SSH access:
ssh-add on each beamline computer for passwordless ssh

Hexapod can't be controlled from SPEC. We control it with Hexapod windows software via ethernet communication so the motor names and directions are different:

Hexapod software	beamline
y+	z-
z+	y+
x+	x+

Mobile experience



15:54

192.168.1.21

3-Dimensional Characterisation by Ptychography X-ray Computed Tomography of Portland Cement to Model the Carbon Capture Cycle

Logbook view

ptycho SPEC OMNY

#S 23 loopscan 10 0.00025 0.1

ptycho SPEC OMNY

#S 24 ascan moth2 -13.1378 -13.1298 40 0.2

ptycho SPEC OMNY

#S 25 ascan moth2 -13.1412 -13.1332 40 0.2

ptycho SPEC OMNY

#S 26 ascan idgap 4.678 4.778 50 0.2

ptycho SPEC OMNY

#S 27 loopscan 1 0.0025 0

ptycho SPEC OMNY

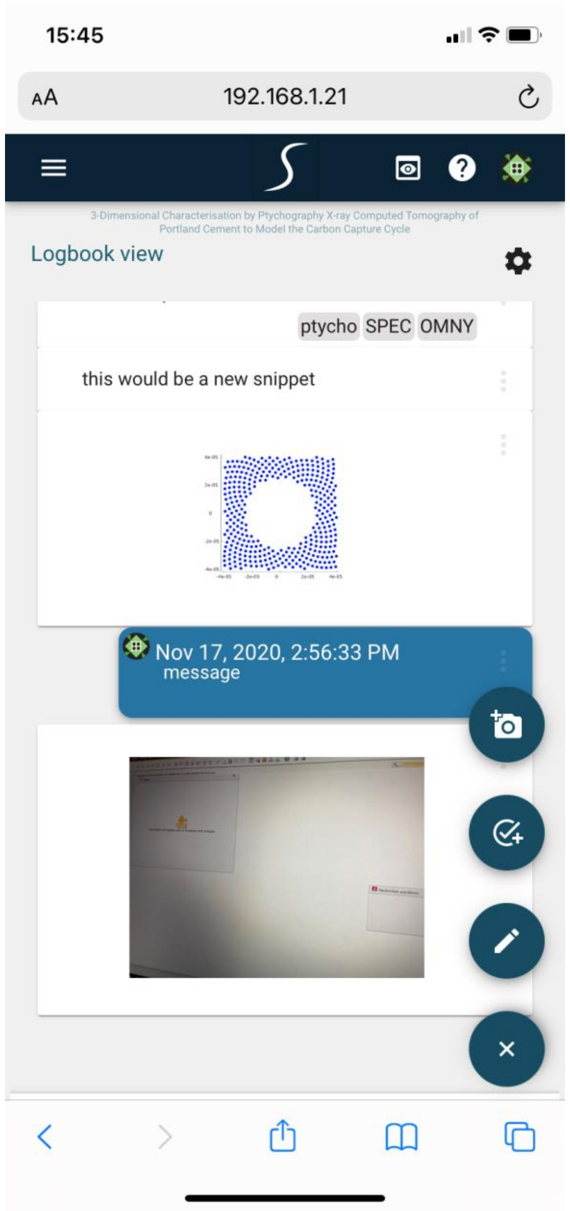
#S 28 loopsca

#S 29 loopsca

#S 30 loopscan 2 0.00025 0

ptycho SPEC OMNY

this would be a new snippet



15:45

192.168.1.21

3-Dimensional Characterisation by Ptychography X-ray Computed Tomography of Portland Cement to Model the Carbon Capture Cycle

Logbook view

ptycho SPEC OMNY

this would be a new snippet

Nov 17, 2020, 2:56:33 PM
message

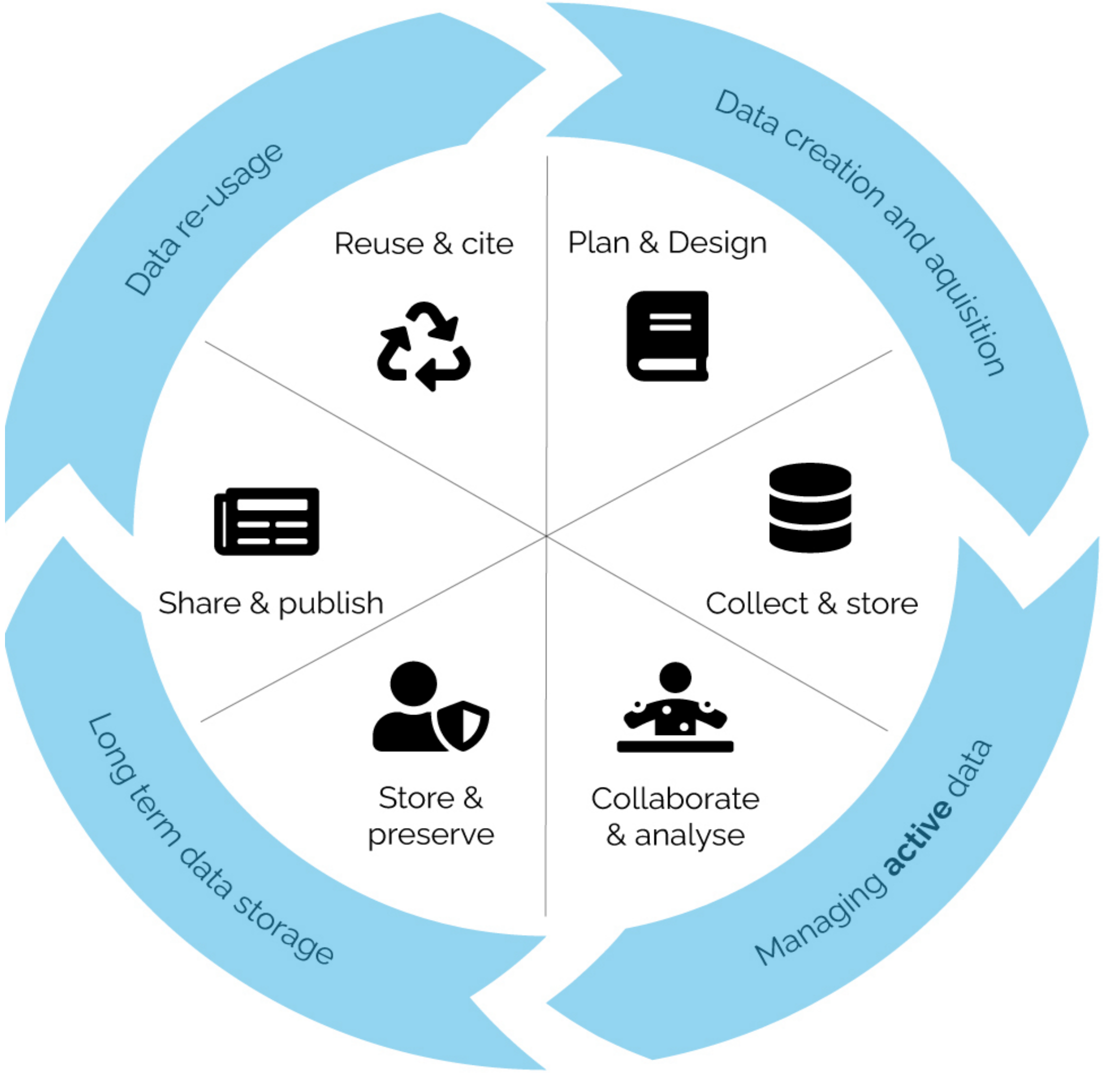
Plan & Design:

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

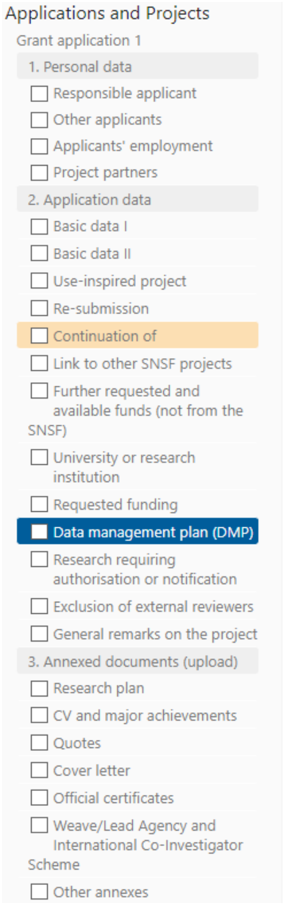


Covers the whole Research Data Life Cycle

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



Applications and Projects

Grant application 1

1. Personal data

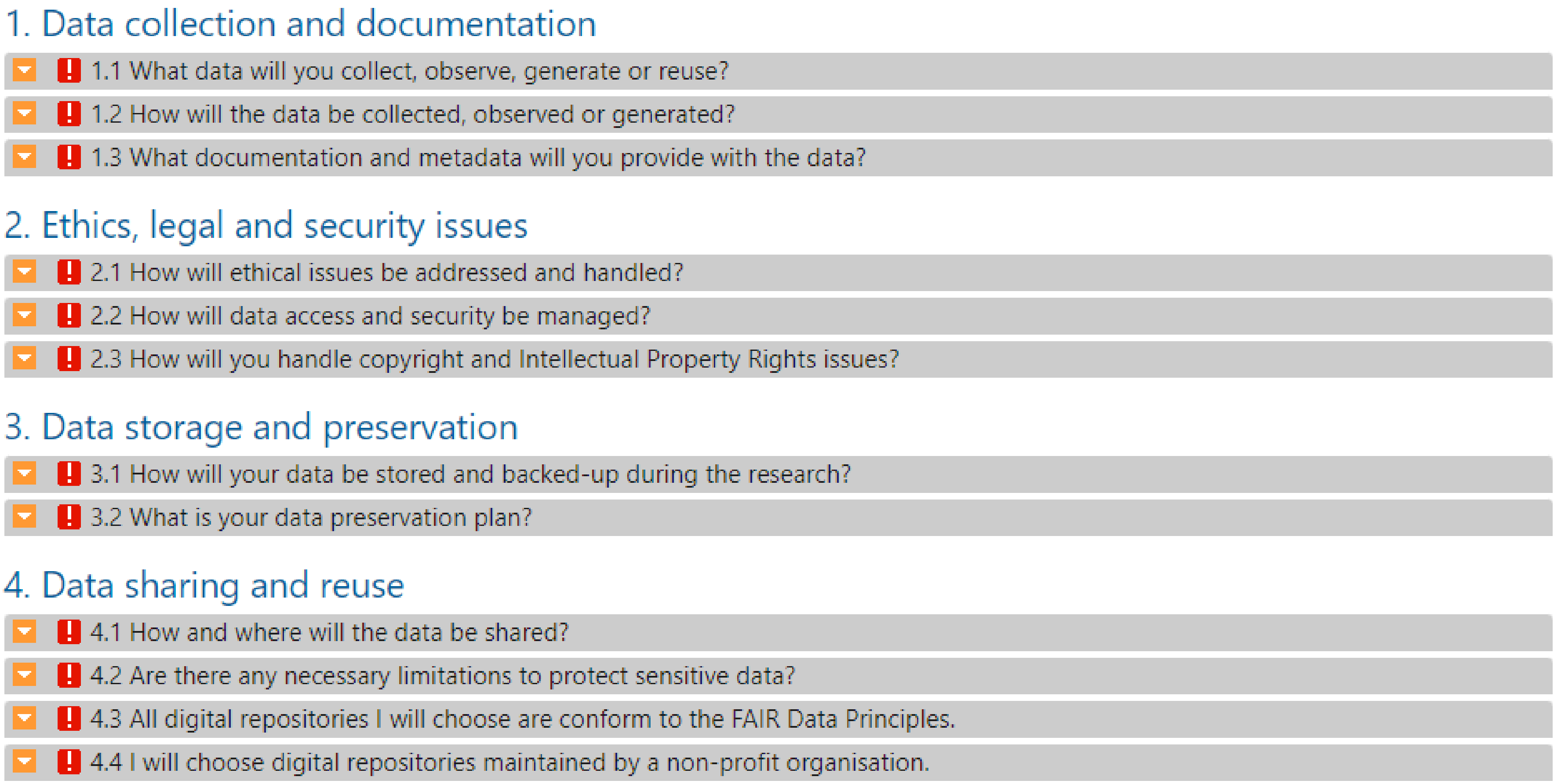
- □ Responsible applicant
- □ Other applicants
- □ Applicants' employment
- □ Project partners

2. Application data

- □ Basic data I
- □ Basic data II
- □ Use-inspired project
- □ Re-submission
- □ Continuation of
- □ Link to other SNSF projects
- □ Further requested and available funds (not from the SNSF)
- □ University or research institution
- □ Requested funding
- □ Data management plan (DMP)
- □ Research requiring authorisation or notification
- □ Exclusion of external reviewers
- □ General remarks on the project

3. Annexed documents (upload)

- □ Research plan
- □ CV and major achievements
- □ Quotes
- □ Cover letter
- □ Official certificates
- □ Weave/Lead Agency and International Co-Investigator Scheme
- □ Other annexes



1. Data collection and documentation

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

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

1.3 What documentation and metadata will you provide with the data?

2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

2.2 How will data access and security be managed?

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

3. Data storage and preservation

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

3.2 What is your data preservation plan?

4. Data sharing and reuse

4.1 How and where will the data be shared?

4.2 Are there any necessary limitations to protect sensitive data?

4.3 All digital repositories I will choose are conform to the FAIR Data Principles.

4.4 I will choose digital repositories maintained by a non-profit organisation.

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 groups of two (5 minutes)**

2. **Each group will engage with the first section 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

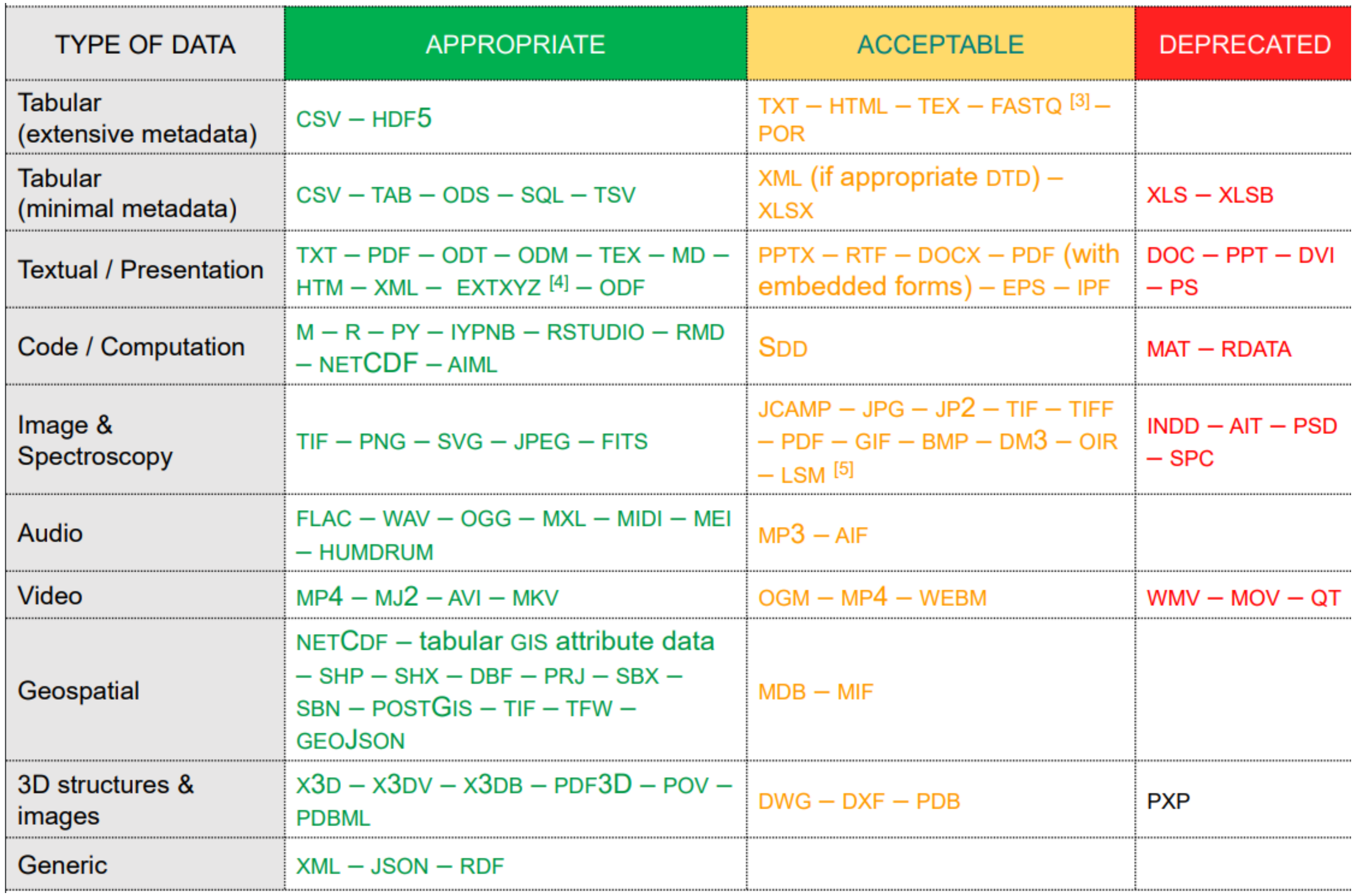
Appendix

Appendix: PSI

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

Appendix: File Formats EPFL

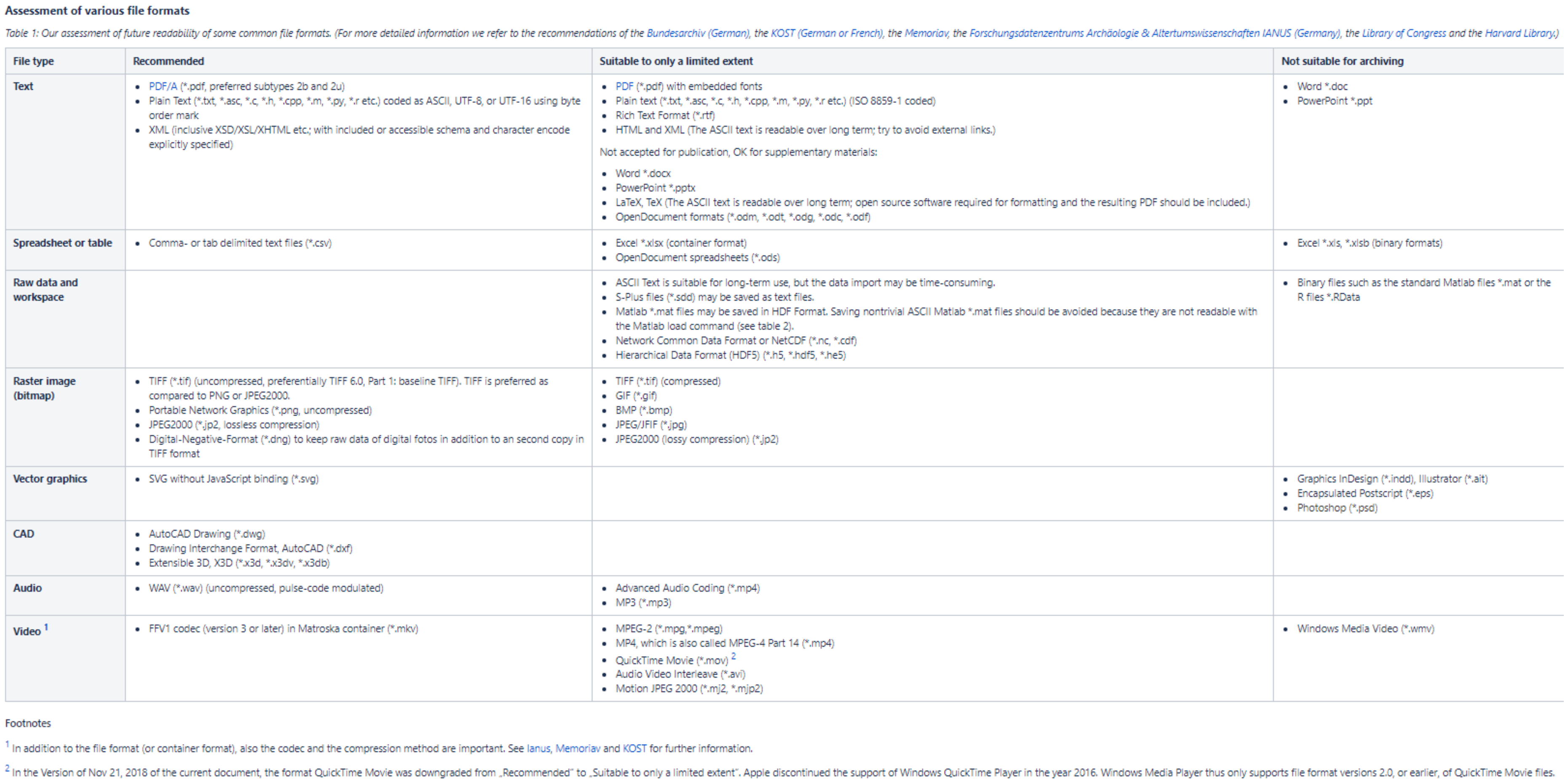
Bibliothèque de l’EPFL, Research Data, fast guide #4», 2019, https://bit.ly/3NFloYx



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		

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

Appendix: File Formats ETH Zürich



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 4:

- 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 8

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