PAUL SCHERRER INSTITUT

PSI

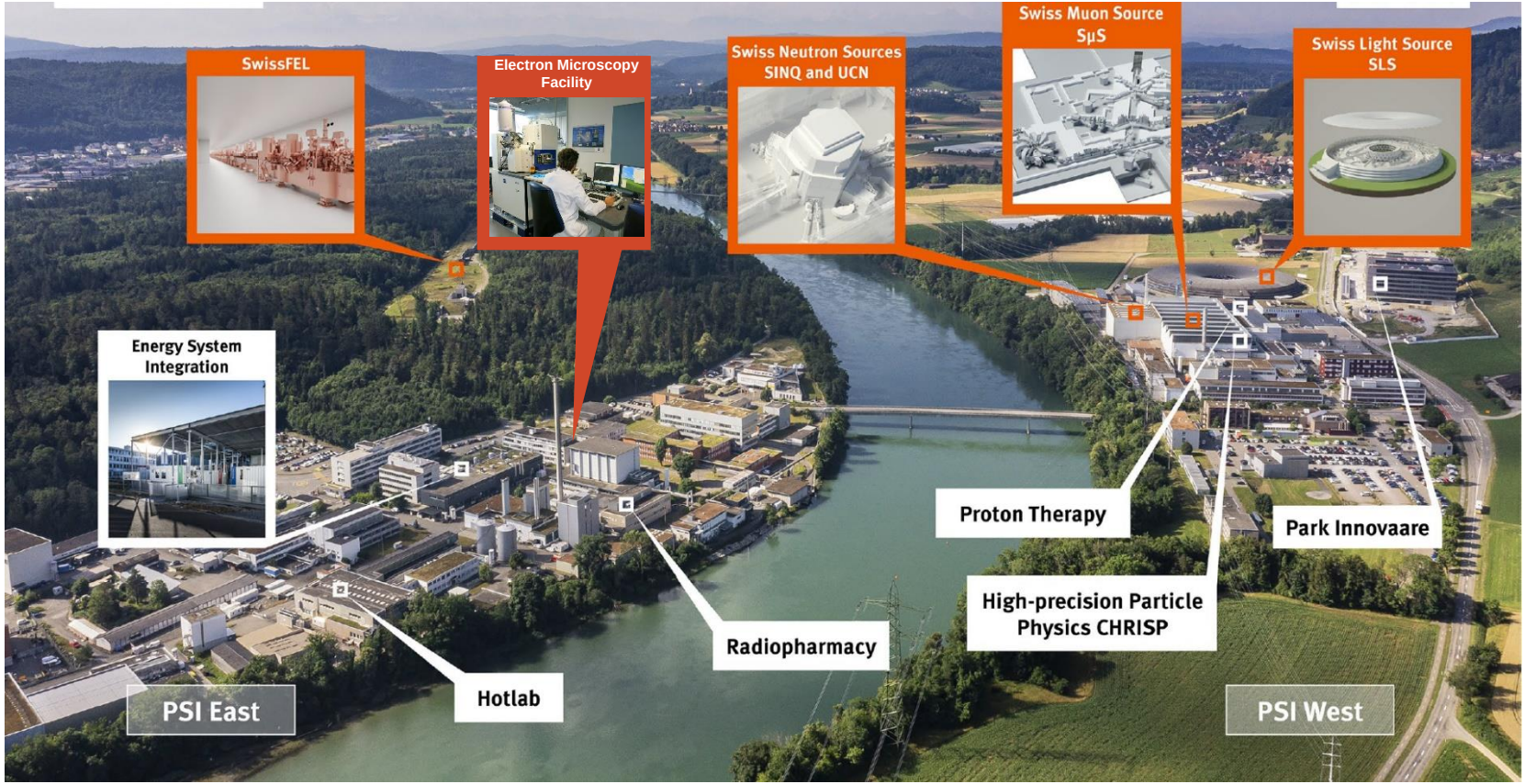
Alun Ashton :: Head of Science IT :: Paul Scherrer Institute

Scientific computing, data sharing and reuse at PSI

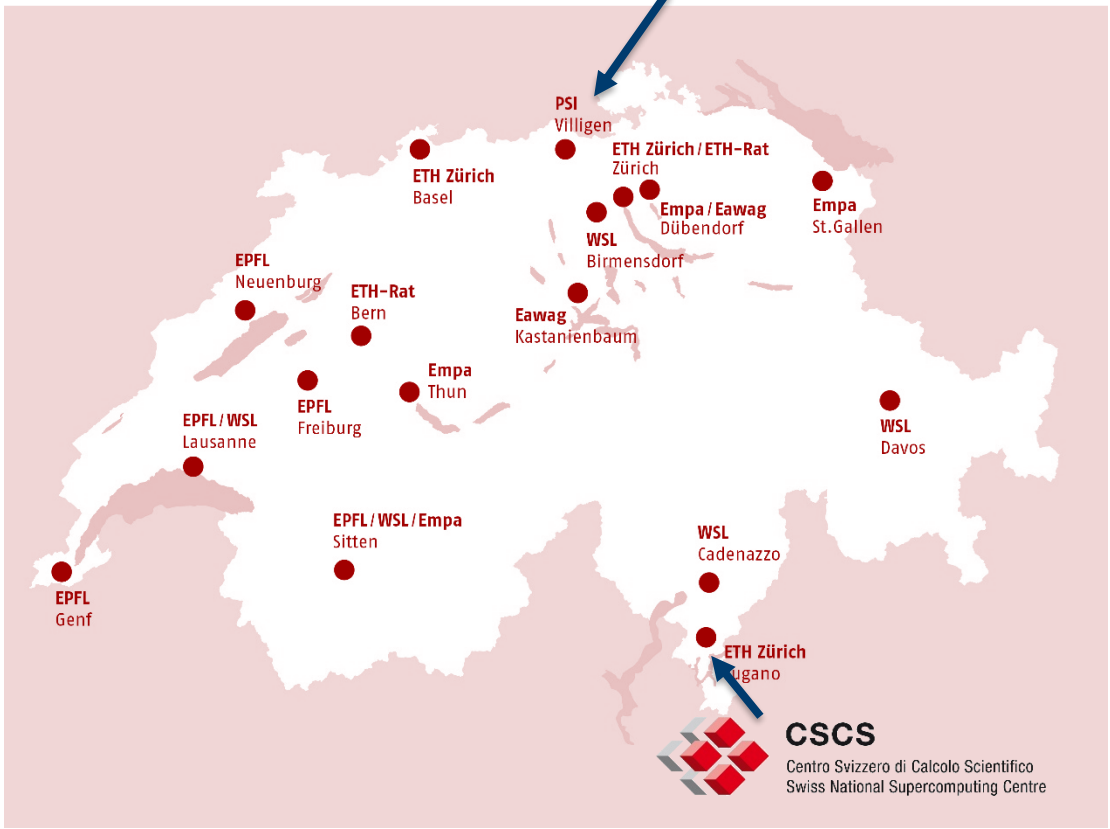
Raw diffraction data reuse: the good, the bad and the challenging: IUCR 2023

Melbourne 22nd August 2023

Facilities at the PSI Campus



The ETH Domain



PAUL SCHERRER INSTITUT

The Good

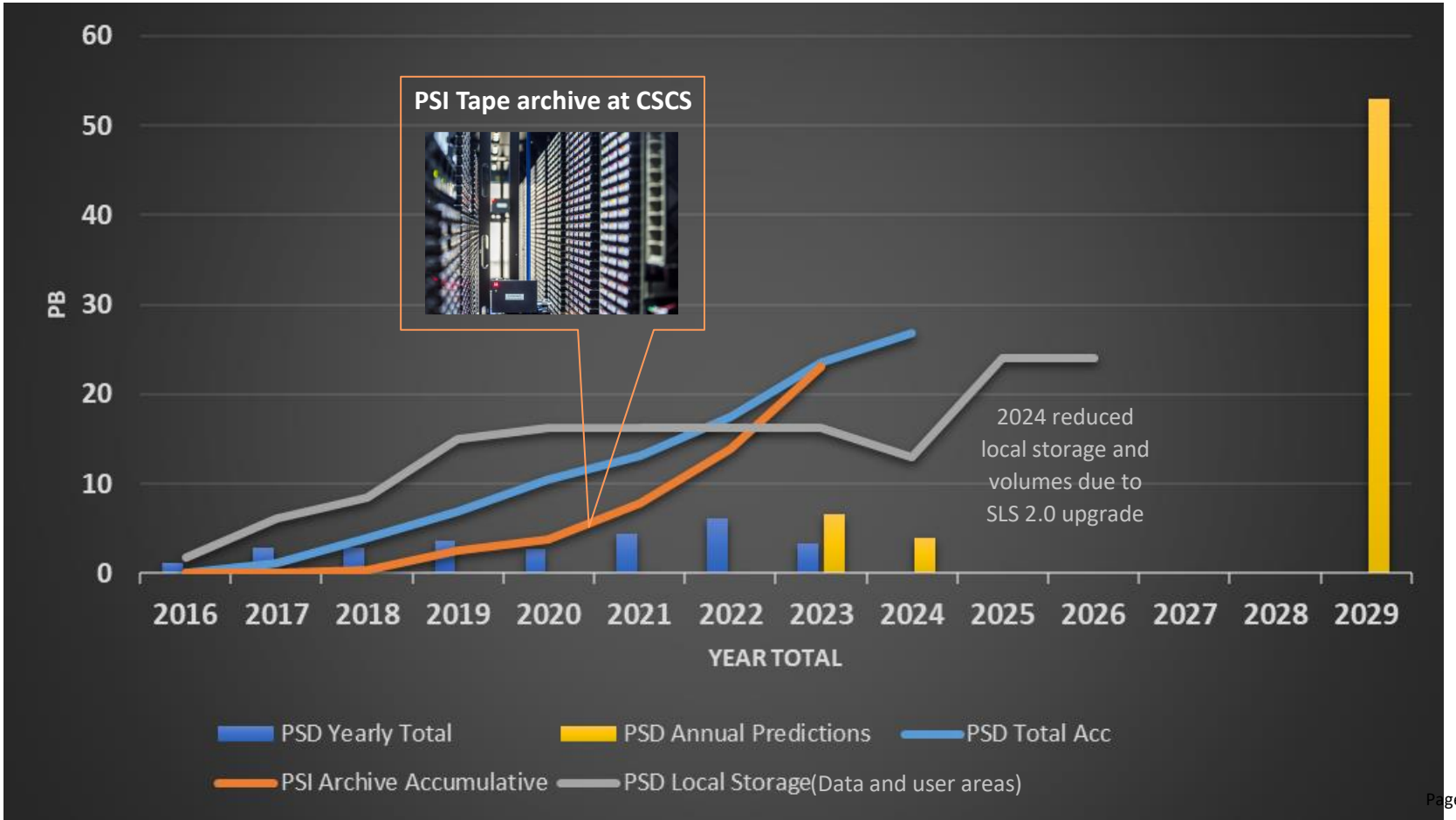
PSI Data Policy for the User Facilities

Data are a valuable and essential product and resource for research conducted at PSI. Especially the large research facilities operated at PSI are a source for large data sets relevant for national and international research teams. This updated document defines the general principles for research data management at PSI and its user facilities and has been approved in April 2022.

Excerpt

- This document pertains to the ownership, curation and access to research data and metadata generated, collected and/or processed and/or stored by PSI or at its facilities.
- It is binding for all PSI employees and external users of PSI facilities and the acceptance of this policy as far as applicable is a condition for the award of access to research infrastructures for internal and external users.
- All activities concerning data management and processing must be in accordance with the **PSI instruction on Research Integrity**.
- All Research Data and Metadata obtained as a result of Public Research will be Open Access after an initial embargo period during which access is restricted to the Experimental Team, represented by the PI or the main author of a proposal.
- The embargo period is three (3) years, starting at the end of data collection as agreed between the Data Steward and PI. Thereafter, the data will become openly accessible. On written request of the PI the embargo period can be shortened, omitted or extended.

Photon Science (PSD) and PSI Data Volumes

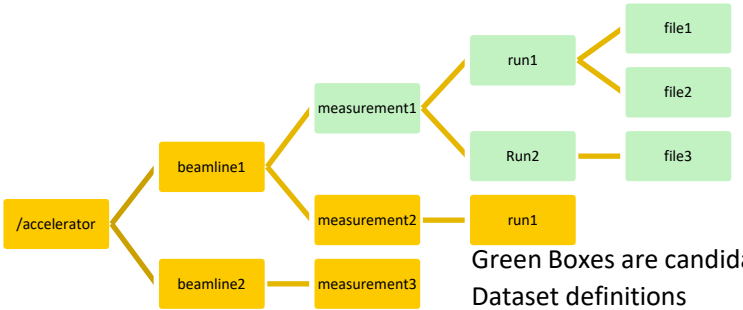


Data Catologue: Where does SciCat help?

- **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

Metadata ingestion: 1. 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"
        }
    }…
}
```

Editing of Metadata

Help About 0 sistomcat

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.

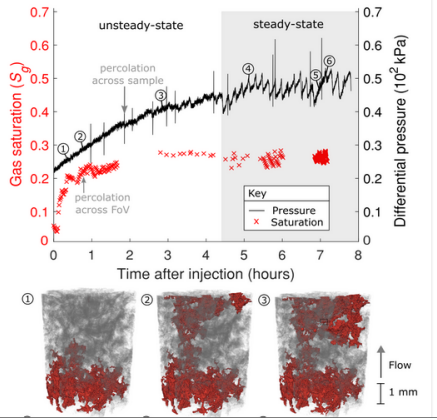
Abstract *
In the related publication to these data sets, we explore the flow dynamics for two-phase flow in a porous medium (a bioclastic carbonate rock). We use state-of-the-art synchrotron X-ray tomography to capture the fluid dynamics within the pore space, with a scan time of 1 second and a temporal 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

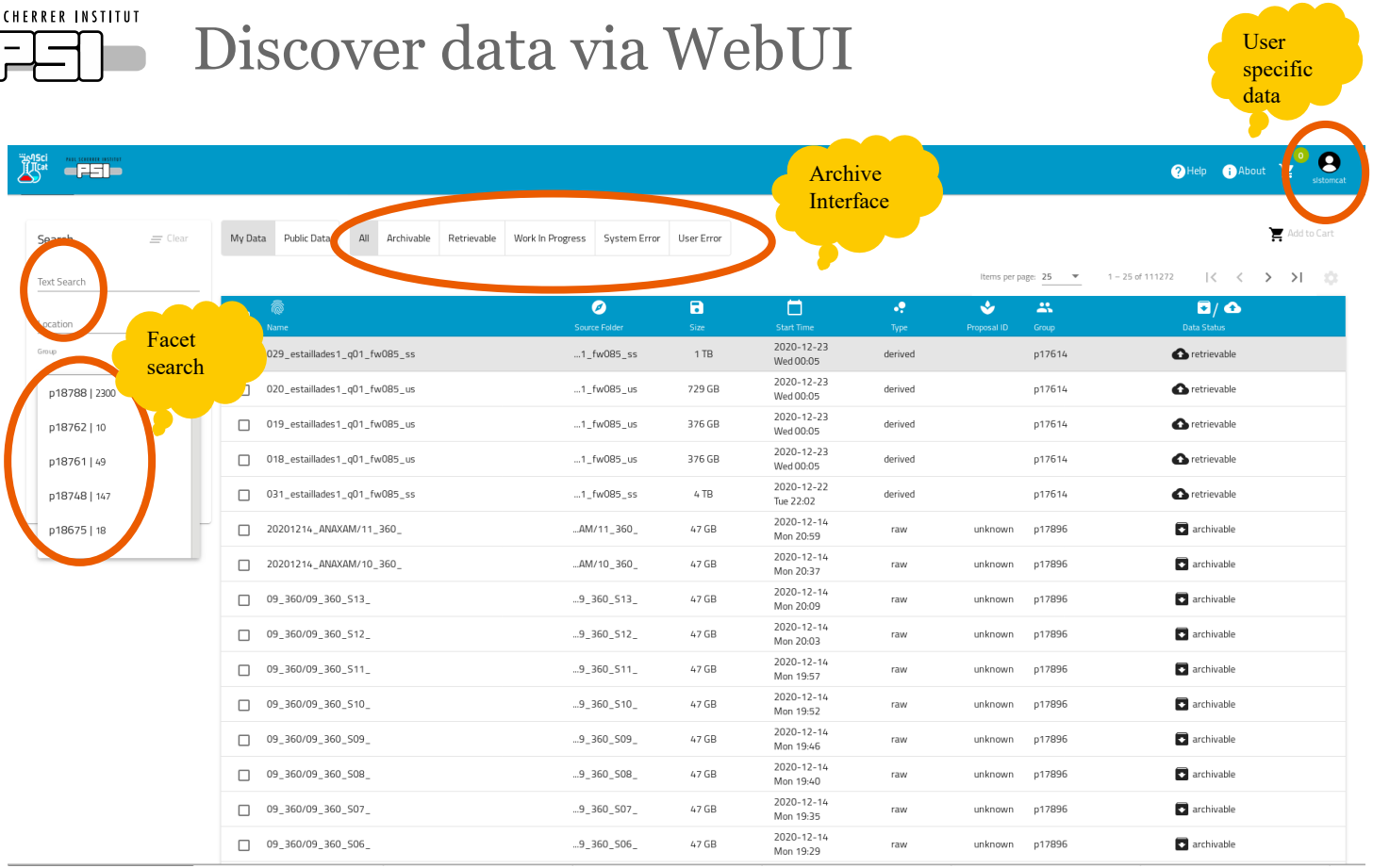
Update Published Data | Cancel

Drop a file here
or
Browse

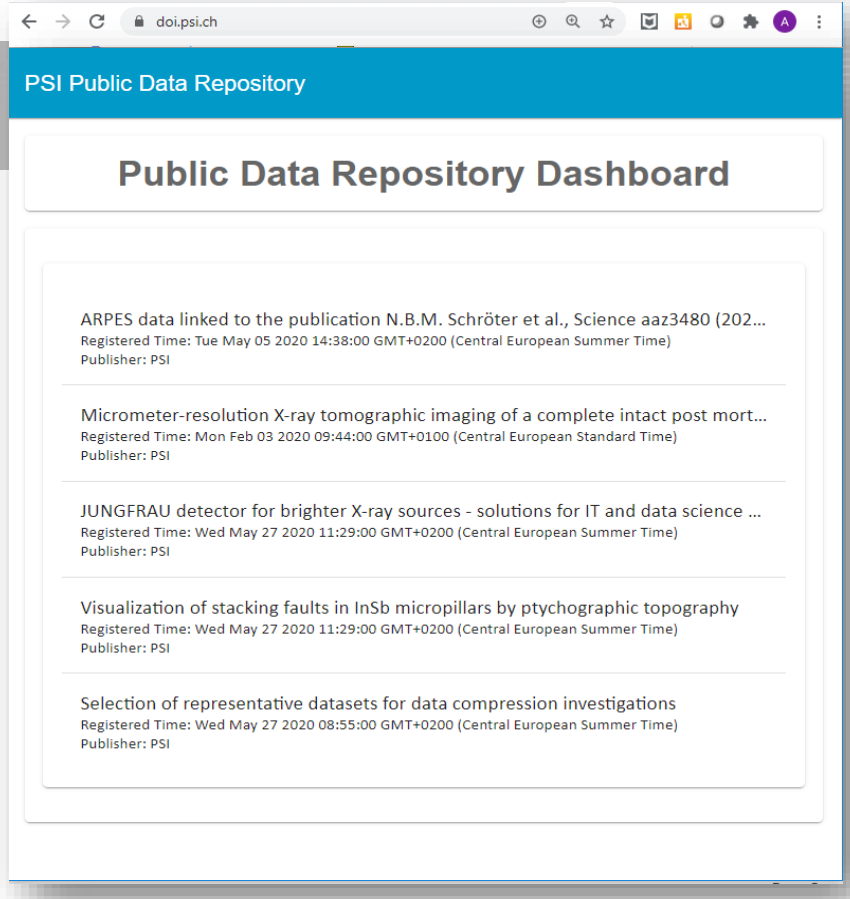


PAUL SCHERRER INSTITUT

Discover data via WebUI



Published Data = List of Datasets + Metadata + DOI



Published Data = List of Datasets + Metadata + DOI

doi.psi.ch

PSI Public Data Repository

Public Data Repos

ARPES data linked to the publication N.B.M
Registered Time: Tue May 05 2020 14:38:00 GMT+020
Publisher: PSI

Micrometer-resolution X-ray tomographic i
Registered Time: Mon Feb 03 2020 09:44:00 GMT+010
Publisher: PSI

JUNGFRAU detector for brighter X-ray sourc
Registered Time: Wed May 27 2020 11:29:00 GMT+02
Publisher: PSI

Visualization of stacking faults in InSb micro
Registered Time: Wed May 27 2020 11:29:00 GMT+02
Publisher: PSI

Selection of representative datasets for dat
Registered Time: Wed May 27 2020 08:55:00 GMT+02
Publisher: PSI

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-ofthe-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

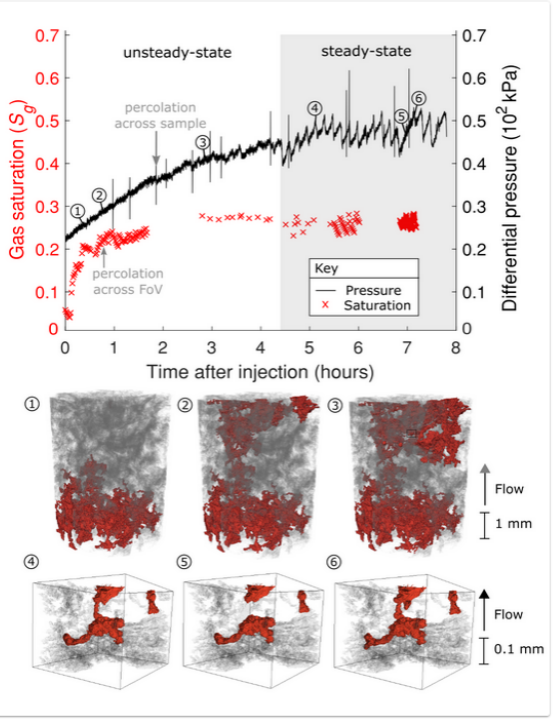
20.500.11935/b9782901-be3b-40fe-91d0-3e0a784337c4

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



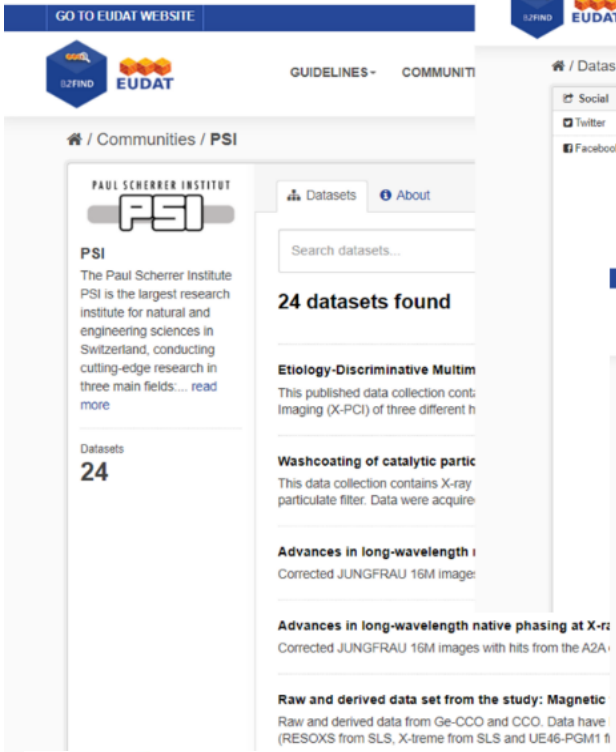
Data from published datasets On EOSC

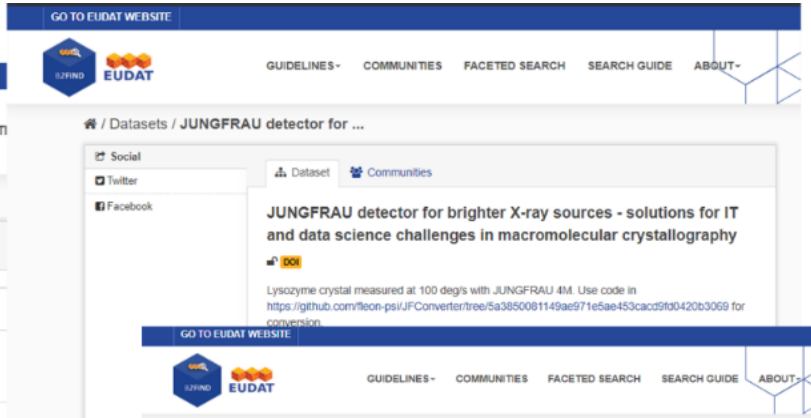
ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

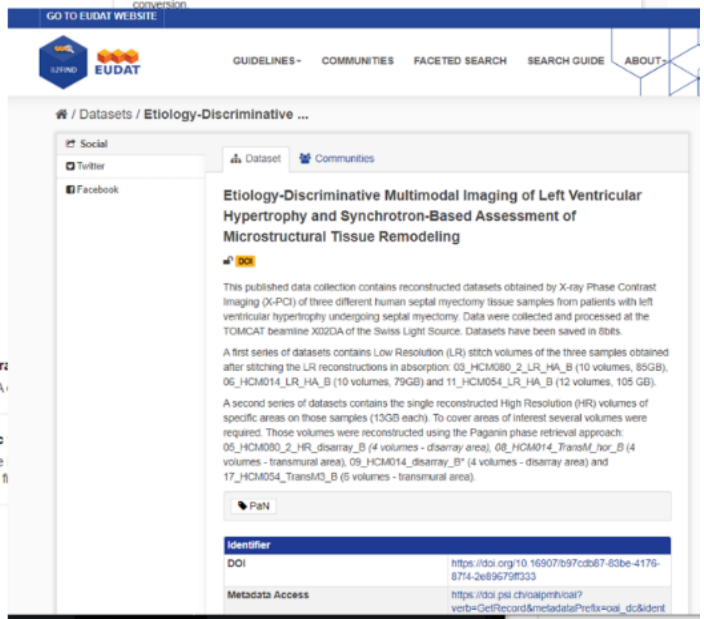
panosc
photon and neutron open science cloud

OpenAIRE

B2FIND





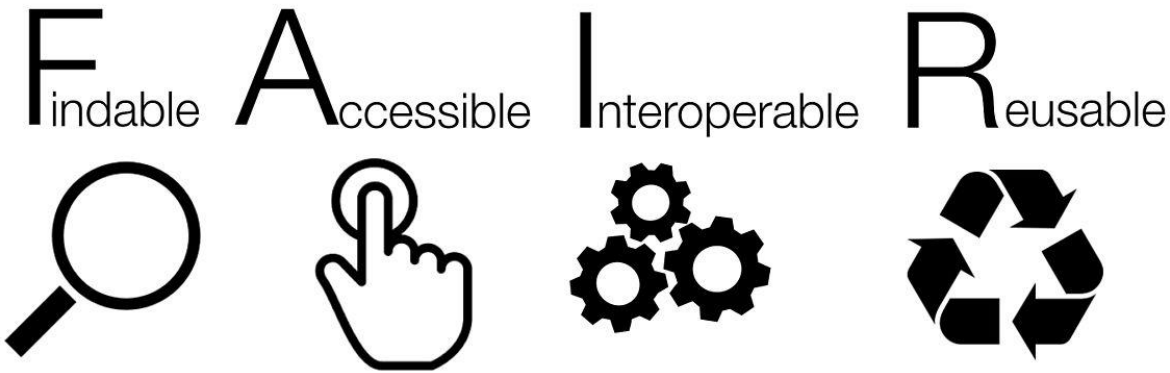


Google Dataset Search Beta

PAUL SCHERRER INSTITUT

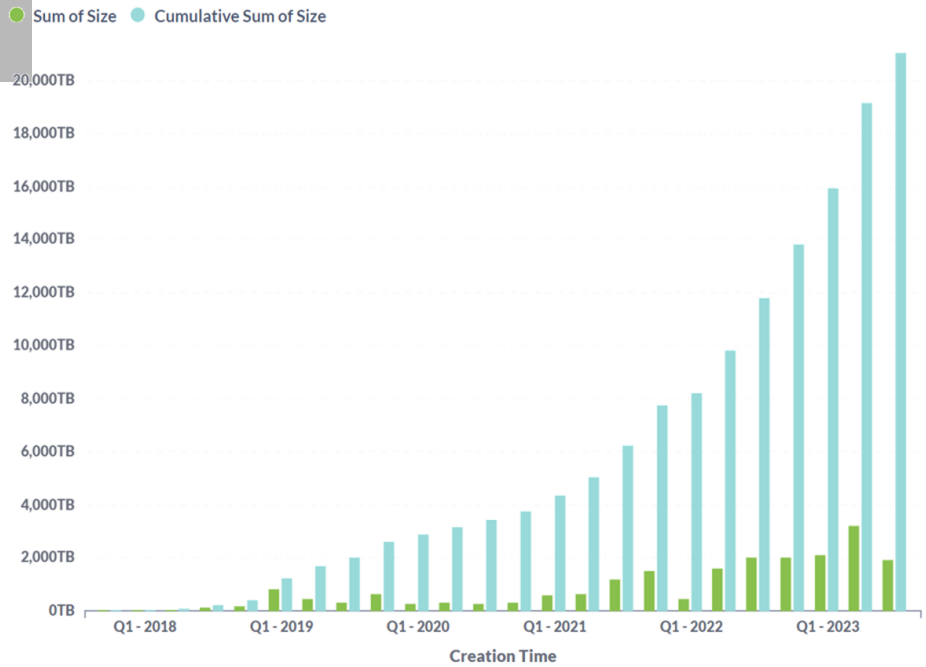
The Bad

Raw diffraction data reuse.....

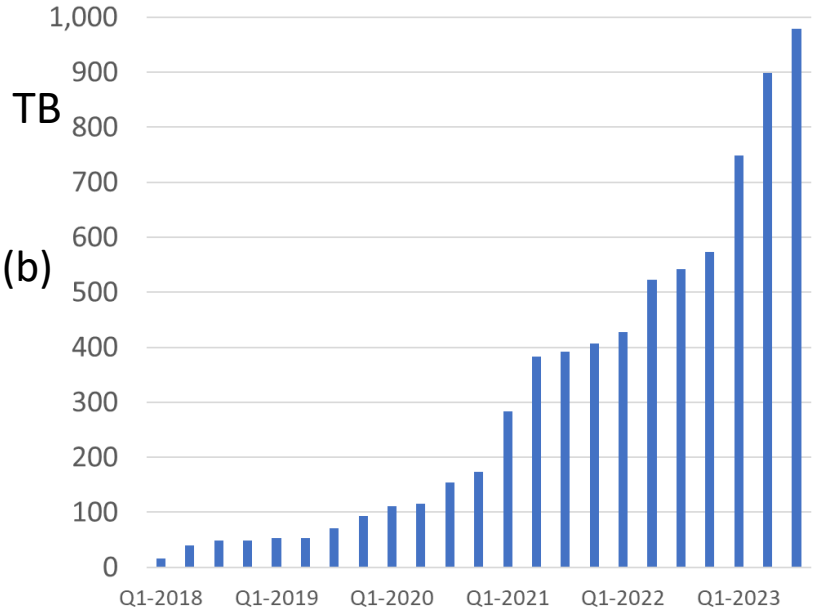


PSI Data Archive (a) and Retrieval* (b)

(a)



(b)



*Retrieval off Tape, not including 337TB of open data accessed off spinning disk

Metadata ingestion: 2. Define Scientific Metadata

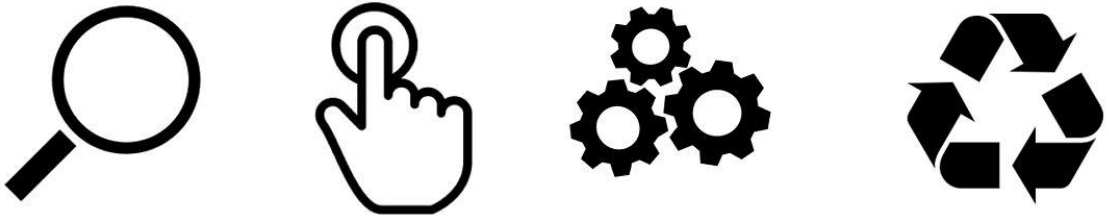
- The definition of scientific meta data is fully flexible.
- Ideally following a standard if it exists, e.g. NeXus based HDF5 files, extracted from instrument.

- 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"
        }
    }…
}
```

Acquiring For

Findable Accessible Interoperable Reusable



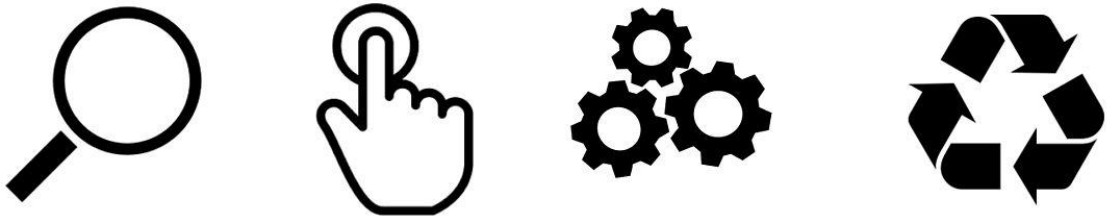
Scalably

PAUL SCHERRER INSTITUT

The Challenging

Acquiring For

Findable Accessible Interoperable Reusable



Scalably

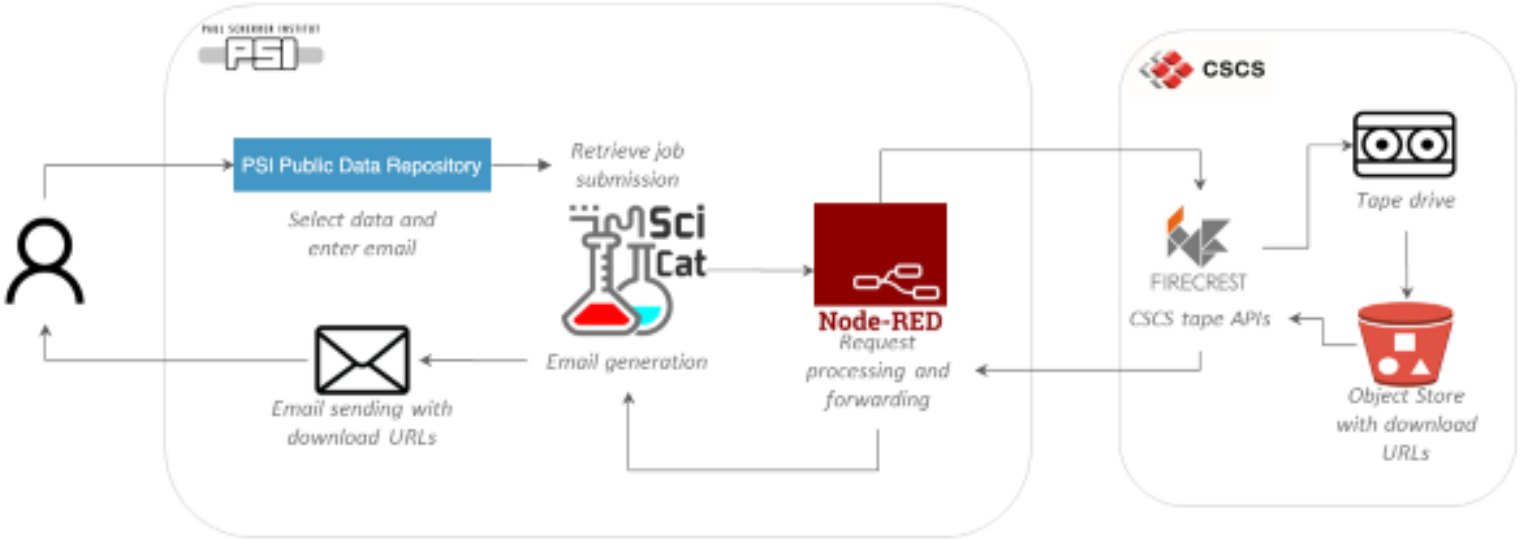
Scalably (and sustainably)

- Immediate access storage is expensive (short term solution) .
- Data on tape is cold.
- Tape technology can fail (currently only single copies to reduce costs).
- Recovery speeds are slower.

Current recovery of data from PSI



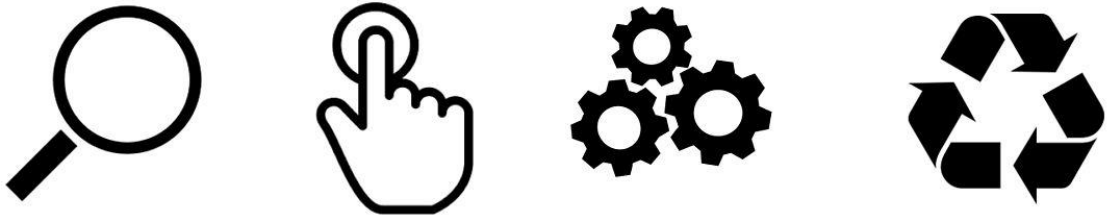
- Manual process to get data requested and staged at PSI for processing or at CSCS for download (and processing in the future)



- Data is not currently automatically made public at end of embargo period…..

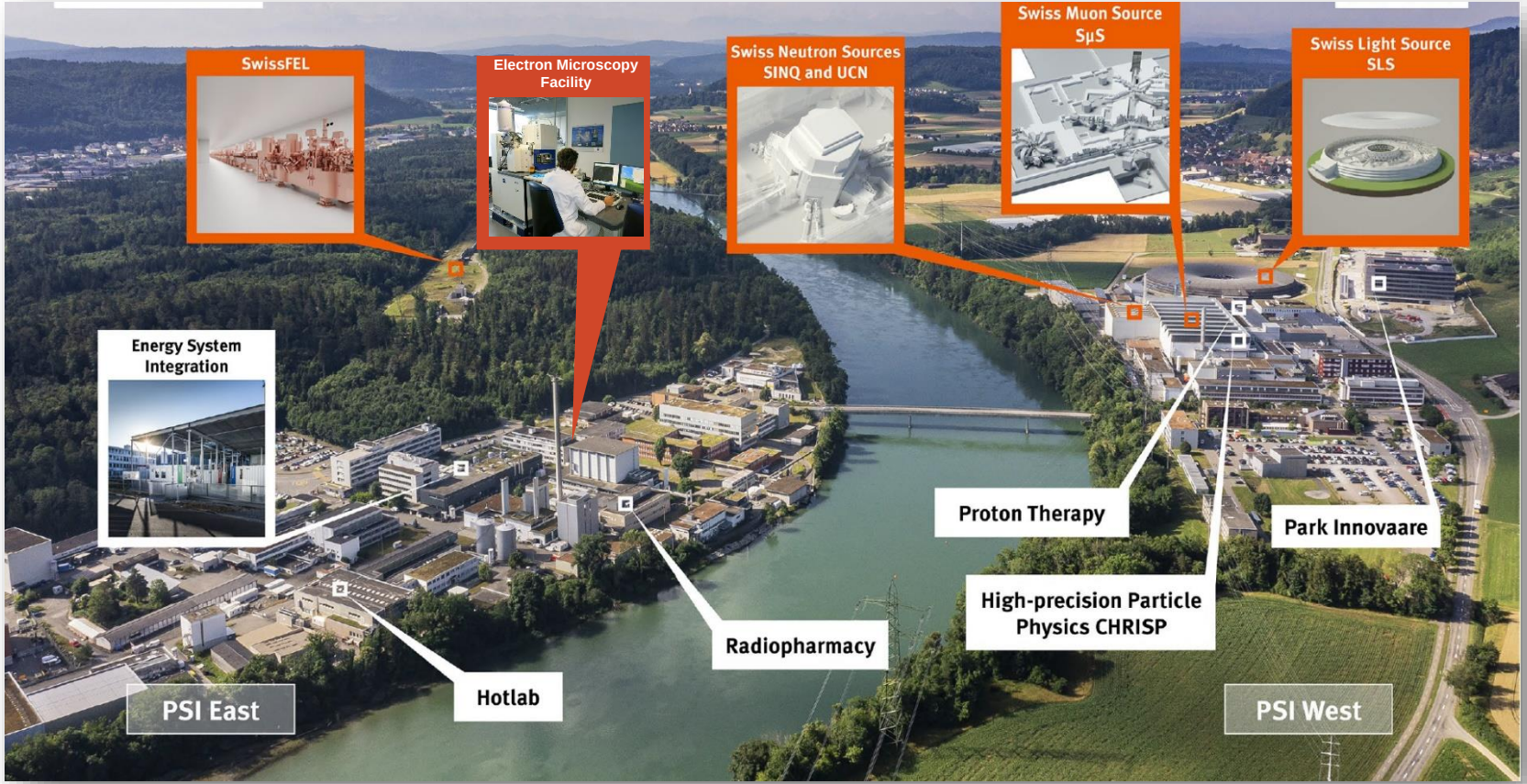
Acquiring For

Findable Accessible Interoperable Reusable



Scalably

Responsibilities at PSI Facilities



Acquiring For FAIR S

- Each facility and science domain will face different challenges
  - Data provenance
  - Data ‘format’
    - Community standards
    - Interdisciplinary standards
  - Data ingestion
  - Data quality

  - Responsibilities
    - Who should catalogue the data and when.

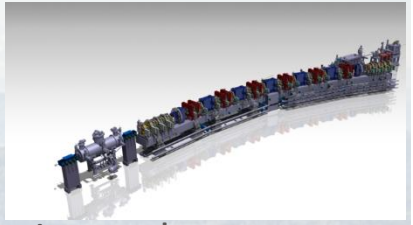
- Three PSI example activities:

1. SLS 2.0 upgrade

2. ETH Domain ORD
   - OpenEM

3. Materials cloud

1. SLS 2.0 Upgrade Project

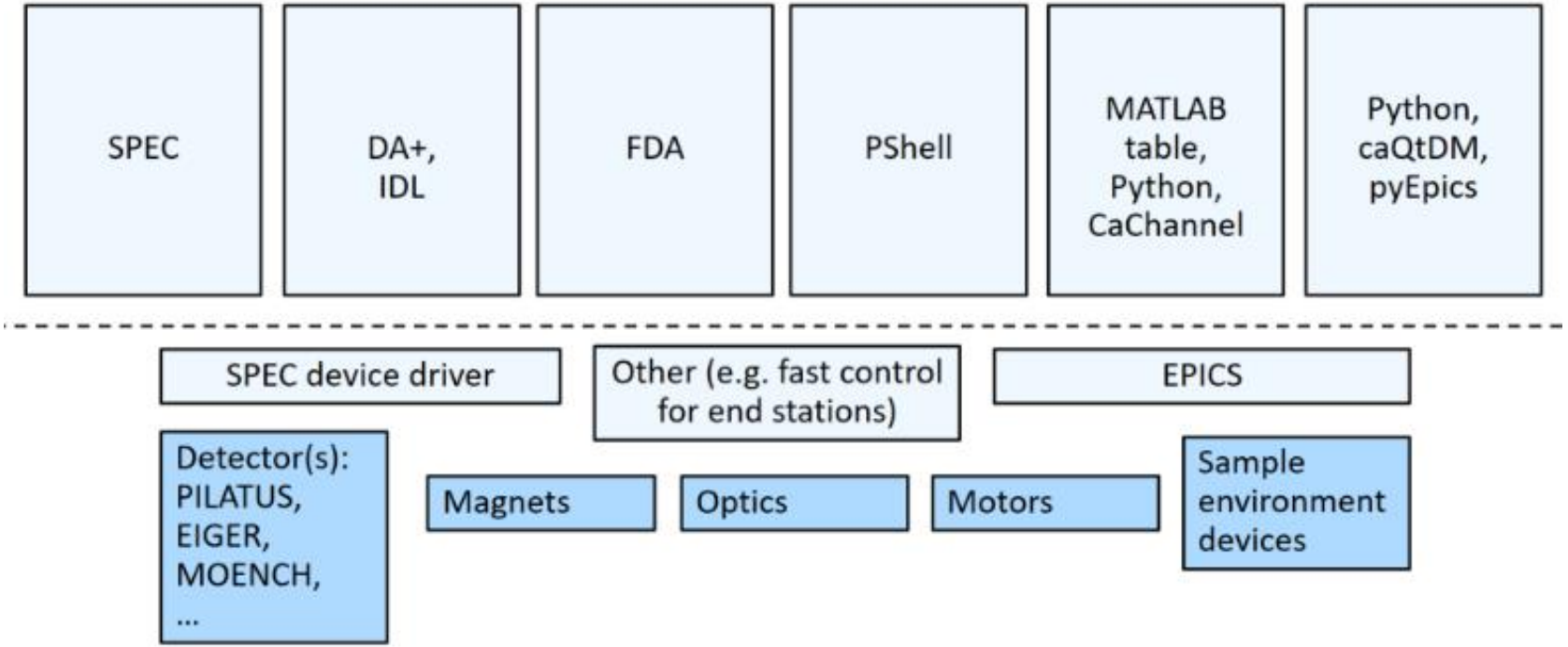


- The SLS 2.0 upgrade requires a comprehensive rebuild of the storage ring and magnet lattice, resulting in an improvement in emittance and associated increase in brightness by a factor of forty compared to the existing performance in the most commonly used hard x-ray regime.
- A phased program of upgrades of the beamlines will begin in parallel to optimize exploitation of the ring.

	2023												2024												2025												2026											
	Q1			Q2			Q3			Q4			Q1			Q2			Q3			Q4			Q1			Q2			Q3			Q4			Q1			Q2			Q3			Q4		
	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Overall	SLS user operation									Dark period																					SLS2 user operation with reduced number of beamlines						shutdown			com-mis-sioning			SLS2 User operation					
Machine	"									Dismantling SLS			Installation new ring												Beam commissioning and vacuum conditioning									user operation						ScSB & BI-Gr2 Id installation			SB com-mis-sioning			user oper-ation			user operation		
Beamlines Phase 1	"									installations, modifications and upgrades as 2nd priority															inst., modif. and upgr. 1st priority						commissioning									modifi-cations			com-mis-sioning								
Beamlines Phase 2	"									installations, modifications and upgrades as 3rd priority															inst., modif. and upgr. 2nd priority			installations, modifications and upgrades as 1st priority									front end completion			commissioning								

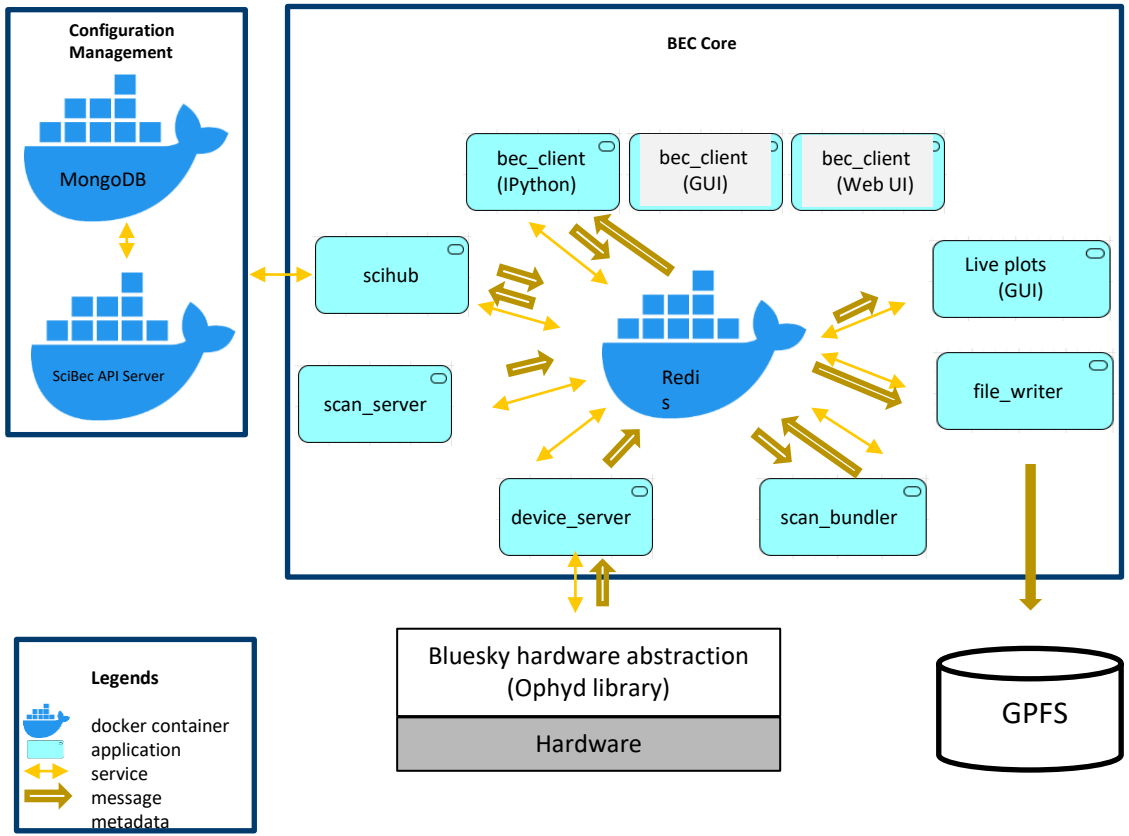
1. Opportunities from SLS 2.0 Shutdown

- Experiment Control on SLS



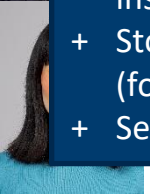
1. Beamline Experiment Control for post SLS 2.0



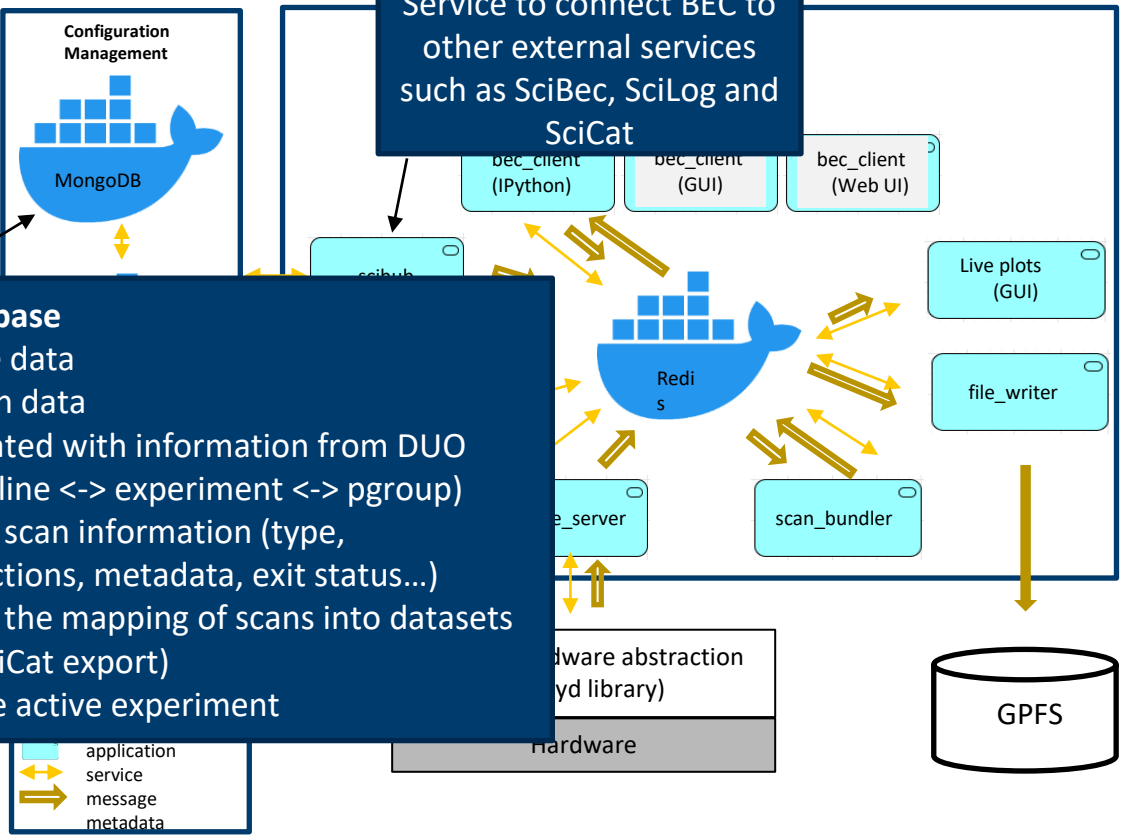



- Standard entry points for client interfaces or analysis
- Standardised file (and metadata) writing
- Hardware abstraction (Ophyd from BlueSky)
- Rollout onto beamlines underway.

1. Beamline Experiment Control for post SLS 2.0



scihub service
Service to connect BEC to other external services such as SciBec, SciLog and SciCat



BEC database
- Device data
- Session data

+ Populated with information from DUO (beamline <-> experiment <-> pgroup)
+ Stores scan information (type, instructions, metadata, exit status…)
+ Stores the mapping of scans into datasets (for SciCat export)
+ Set the active experiment

- Standard entry points for client interfaces or analysis
- Standardised file (and metadata) writing
- Hardware abstraction (Ophyd from BlueSky)
- Rollout onto beamlines underway.

2. ORD in ETH Domain, Switzerland

- “Science often benefits from multiple evaluations of the same data sets by researchers with different backgrounds, perceptions and ideas. ORD allows – and stimulates – new research and discoveries”, Position of the ETH Domain

- The ETH-board together with the ETH domain have committed 15 MCHF for the promotion of open research data (ORD) research practices and services.
  - Measure 1: Call for Field-Specific Actions
  - Measure 2: Coordination of Access to Research Data Management (RDM) Services & Infrastructures
  - Measure 3: Development of Online Course Material for RDM Training
  - Measure 4: Information on Legal Questions related to ORD
  - Measure 5: Career Paths for ORD Professionals

https://ethrat.ch/en/eth-domain/open-research-data/

2. The Open EM Data Network (OpEM)

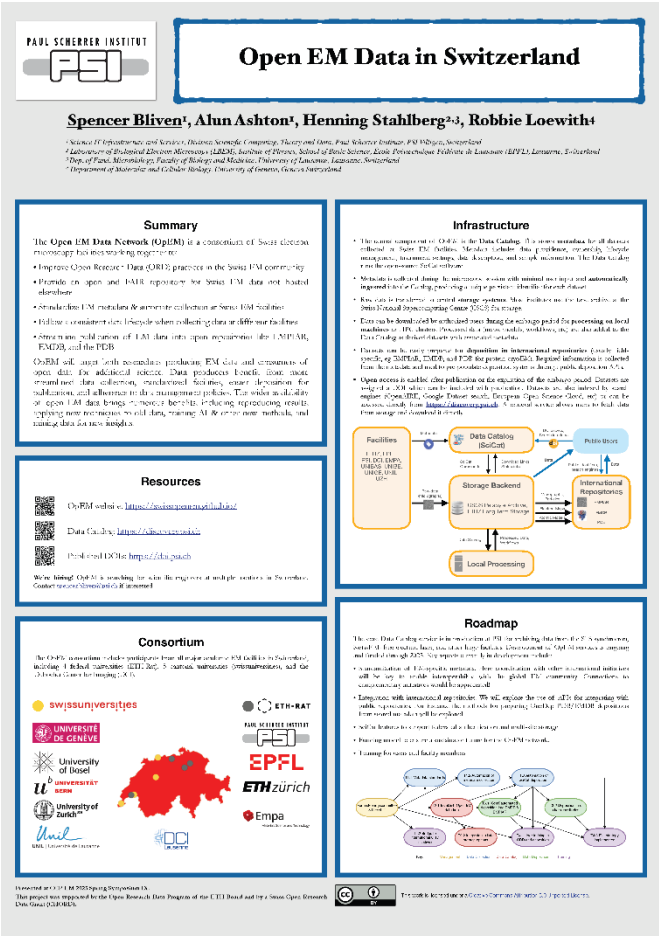
Funded as part of the ETH Domain ORD program:

OpEM is a consortium of Swiss electron microscopy facilities working together to:

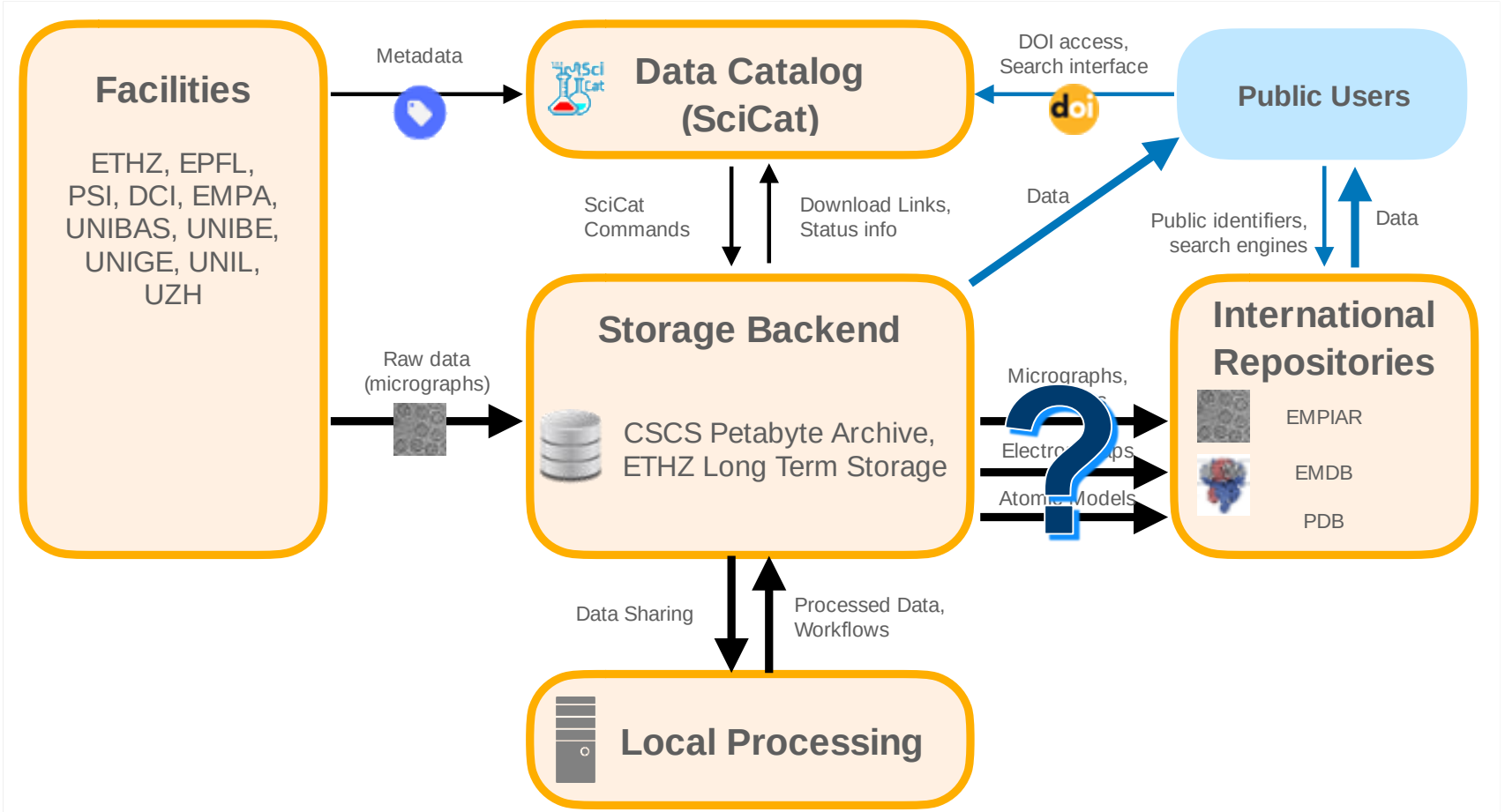
- Improve Open Research Data (ORD) practices in the Swiss EM community
- Provide an open and FAIR repository for Swiss EM data not hosted elsewhere
- Standardize EM metadata & automate collection at Swiss EM facilities
- Follow a consistent data lifecycle when collecting data at different facilities
- Streamline publication of EM data into open repositories

OpEM will target both researchers producing EM data and consumers of open data for additional science. Data producers benefit from more streamlined data collection, standardized facilities, easier deposition for publication, and adherence to data management policies. The wider availability of open EM data brings numerous benefits, including reproducing results, applying new techniques to old data, training AI & other new methods, and mining data for new insights.

https://ethrat.ch/en/eth-domain/open-research-data/



2. The Open EM Data Network



3. PREMISE goals

Open and Reproducible Materials Science Research

- Establish, promote and facilitate the **adoption of FAIR ORD practices** in Materials Science.
- Provide missing critical components to **enable open and reproducible research** (accessible, shareable)
- Address **interoperability between data from simulations and experiments** (currently: no established RDM practices)

- Key enabler of **emerging AI/ML-driven autonomous laboratories**, with native support for RDM and ORD practices

3. Example : Linking with Domain Tools

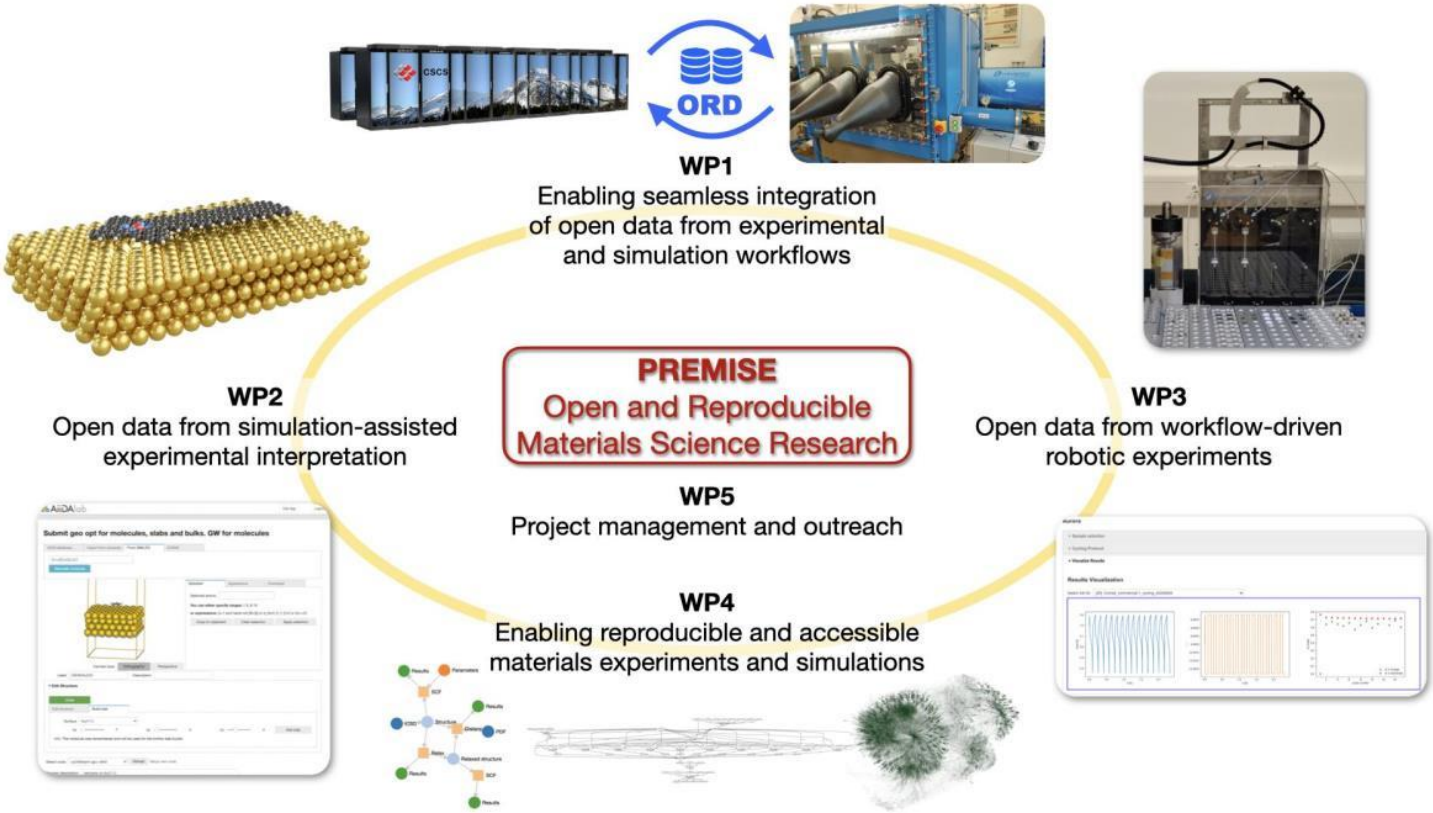
Materials Cloud is built to enable the seamless sharing and dissemination of resources in computational materials science, offering **educational, research, and archiving tools; simulation software and services; and curated and raw data.**

MATERIALSCLOUD

LEARN WORK DISCOVER EXPLORE ARCHIVE

https://www.materialscloud.org

3. PREMISE structure: structure and workpackages



PAUL SCHERRER INSTITUT

Summary

- The Good:
  - Facility tools and policies are now well advanced in Switzerland/Europe, benefiting greatly from the PANOSC and ExPaNDS initiatives.

- The Bad
  - The data you get out is only as good as the data you put in

- The Challenging
  - Ensuring Acquired For Findable Accessible Interoperable Reusable data Scalably (and Sustainably) needs further investment and engagement, not just top down

PAUL SCHERRER INSTITUT

PSI

Thanks to:

- All those pictured
- SciCat Collaborators
- PSI and ETH Domain colleagues
- ExPaNDS and PaNOSC colleagues