

WIR SCHAFFEN WISSEN – HEUTE FÜR MORGEN



Carlo Minotti and Stephan Egli :: Paul Scherrer Institut

Data Curation Services Overview

AWI Meeting Nov 4th 2022

Auditorium Paul Scherer Institute, Villigen, Switzerland

- Purpose/tasks of our data curation (DC) group
- Two main tools from Users Viewpoint:
 - SciCat
 - and SciLog
 - including a live demo
- Technologies used
 - Hosting, Container tools, CI/CD etc

Interactive style, please interrupt and ask questions

Data Curation Group Responsibilities

- Operate the **SciCat central data catalog** to allow to annotate, store, publish and archive data
- Operate the interface to the petabyte archive
- Manage the processes of data ingestion
- Interface to related systems, in particular to the digital user office DUO and the Identity management system for account and group handling.
- Operate, deploy and develop the **SciLog electronic logbook services (in preparation)**
- **Workpackage lead in ExPANDS** EU FP7 program
- Consultation in SciCat related question
- Collaboration in international developer teams for the SciCat project

AWI organization

26/10/22




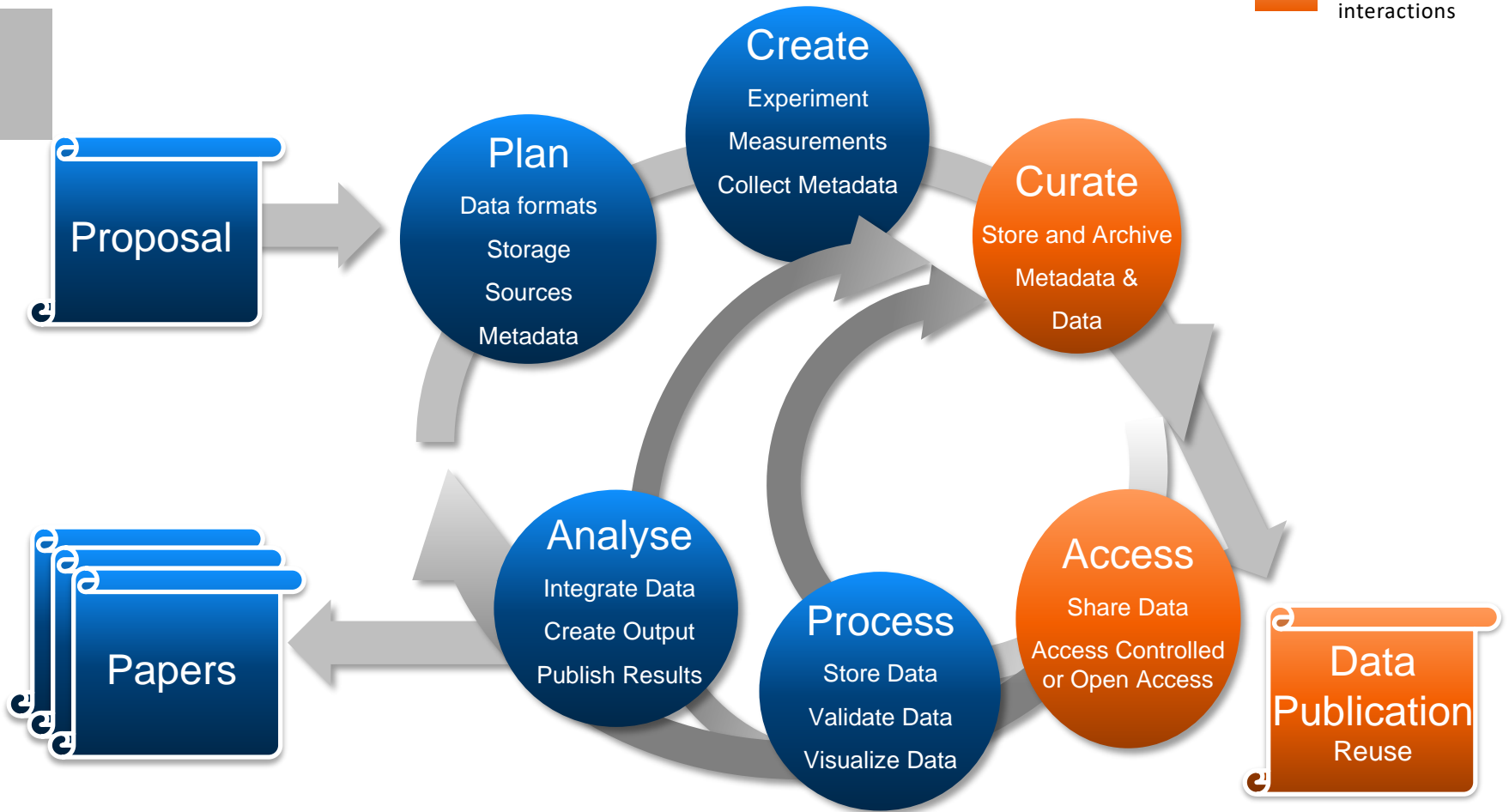
Key to Fixed Term Roles

- EU Funded
- PSI Funded fixed term
- Swiss Funded initiatives

- **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

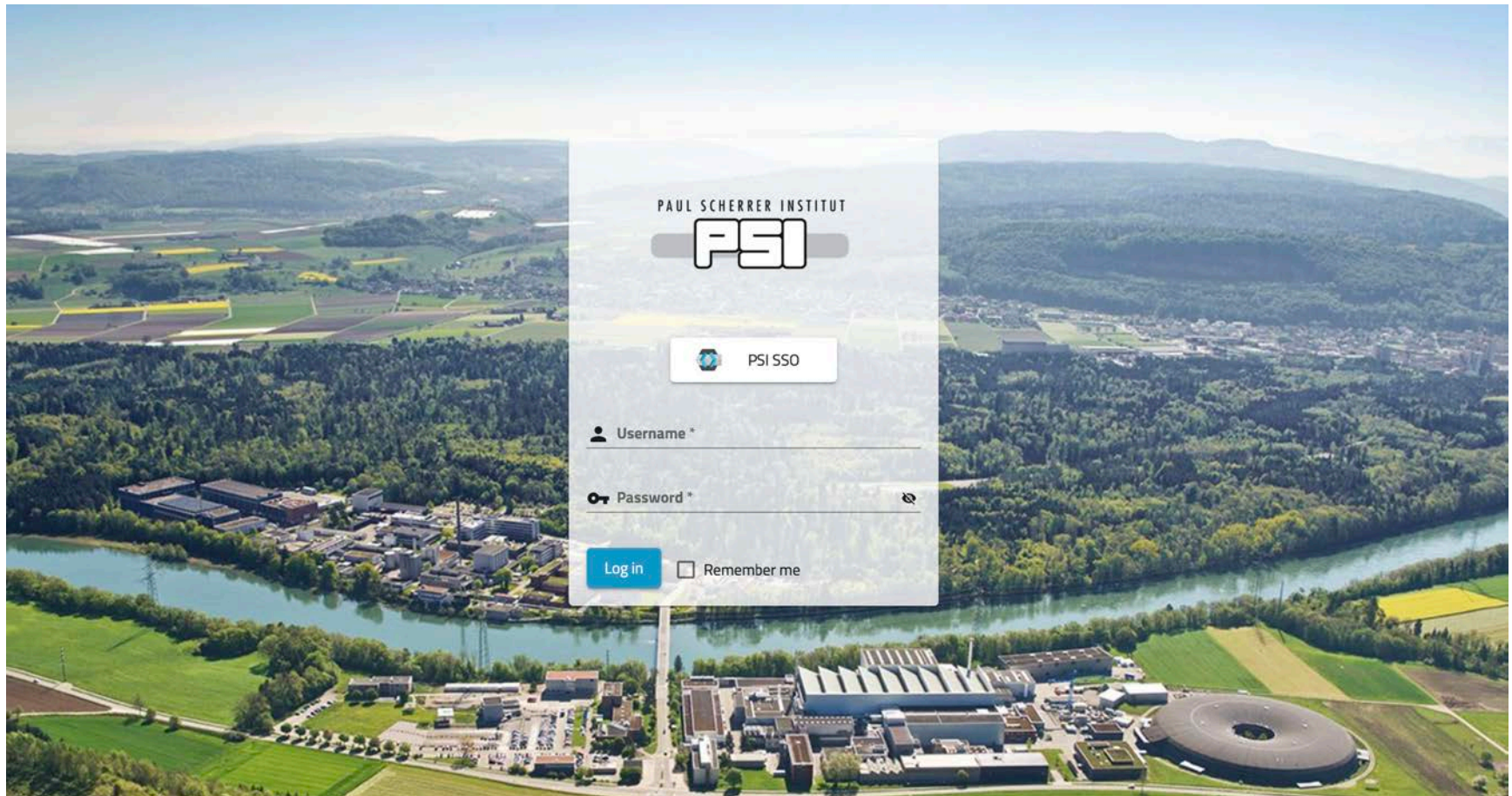
Interactions With the Data Catalogue

 Data catalogue interactions



User Authentication and Authorisation

- Connected to AIT AD via the PSI Identity Provider and Keycloak using OIDC protocol
- Authorisation is controlled by group membership inside AD
- In addition Functional accounts e.g. for Beamline scientists



User specific data *

Archive Interface

Facet search

My Data Public Data All Archivable Retrievable Work In Progress System Error User Error

Name	Source Folder	Size	Start Time	Type	Proposal ID	Group	Data Status
029_estaillades1_q01_fw085_ss	...1_fw085_ss	1 TB	2020-12-23 Wed 00:05	derived		p17614	retrievable
020_estaillades1_q01_fw085_us	...1_fw085_us	729 GB	2020-12-23 Wed 00:05	derived		p17614	retrievable
019_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).

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 Rucker ✕ Gaetano Garfi ✕ Christian M. Schlepütz ✕ Vladimir Novak ✕ Steffen Berg ✕ Martin J. Blunt ✕
 Samuel Krevor ✕

Publisher *
PSI

Resource Type *
derived

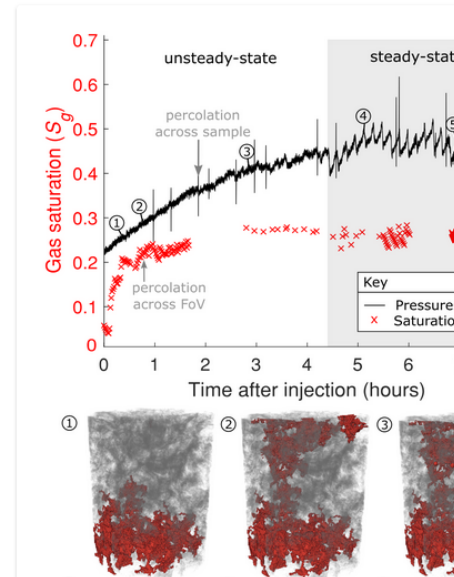
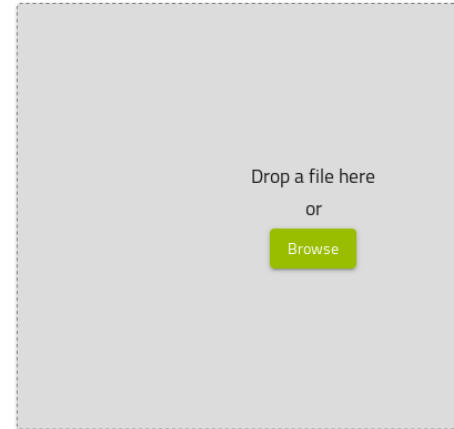
Description *
This published data collection contains five datasets obtained by X-ray tomographic microscopy of a carbonate rock sample 5 mm in diameter and 20 mm in length. Both brine and nitrogen gas are injected into the sample at a total flow rate of 0.1 ml/min (the brine made up 85% of this total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



Retrieving data from tape



Datasets /

Search

Clear

Text Search

Location

Group

Type

Keywords

Start Date – End Date

+ Add Condition

+ Create Dataset

My Data All Public Data All Archivable **Retrievable** Work In Progress System Error User Error

	Name	Run No.	Source Folder	Size	Start Time	Type	Run No.
<input type="checkbox"/>	Archive/TestDataset		...estDataset	14 MB	2022-05-11 Wed 15:02	raw	
<input type="checkbox"/>	add_using_ui		.../nfs	0 B	2022-04-12 Tue 09:28	derived	
<input checked="" type="checkbox"/>	30042021-testingest/normal		...est/normal	101 MB	2020-02-12	base	
<input checked="" type="checkbox"/>	30042021-testingest/normal		...est/normal			base	
<input type="checkbox"/>	S11850-20865_ID46-full		...2000-12999			raw	
<input type="checkbox"/>	S11850-20865_ID46-full		...2000-12999			raw	
<input type="checkbox"/>	S11850-20865_ID46-full		...2000-12999			raw	
<input type="checkbox"/>	S11850-20865_ID46-full		...2000-12999	<u>21 GB</u>	2019-09-27 Fri 13:10	raw	
					2019-09-26		

Really retrieve?

Optionally select destinat... ▾

Ok No Thanks

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

Catherine Spurin, Tom Bultreys, Maja Rucker, 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 μm^3 . 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. Rucker, 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-4e7e-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

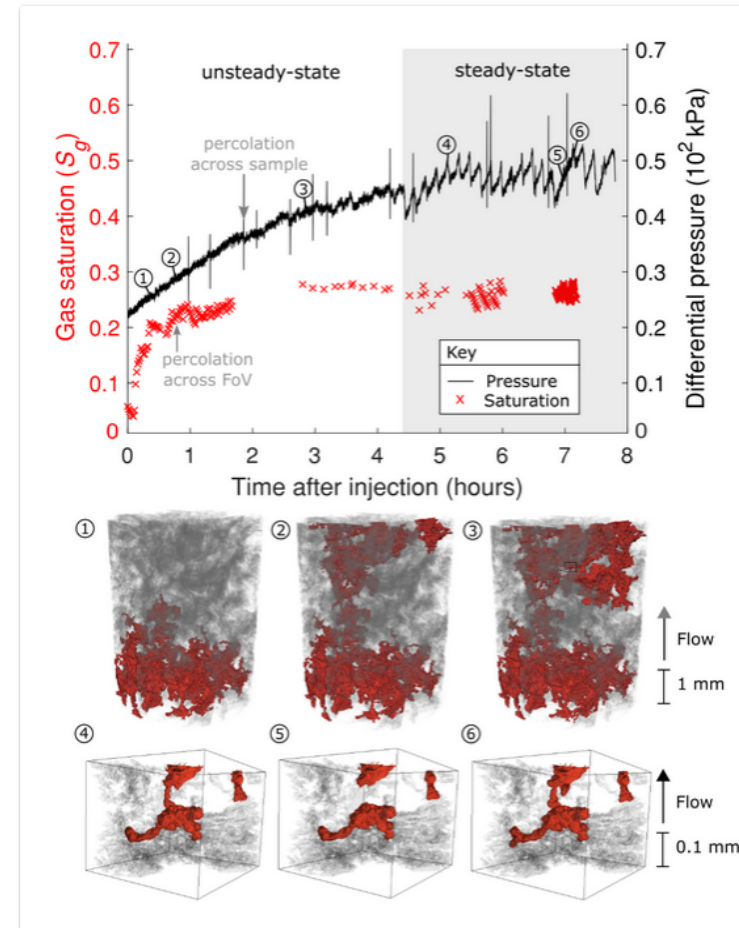
Actions

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

[Access Data](#)



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



Sites using or planning to use SciCat and contact information

- European Spallation Source (Sweden/Denmark)
- Paul Scherrer Institut (Switzerland)
- MaxIV Laboratory (Sweden)
- Advanced Light Source - micro-CT (USA)
- Bundesamt für Materialwirtschaft (Germany)
- Rosalind Franklin Institute (United Kingdom)
- Shanghai Facility SSRF (China)
- Beijing High Energy Photon Source (China)
- Deutsches Elektronen-Synchrotron - PETRA (Germany)
- SOLEIL (France)

- Several of the sites contribute actively to the SciCat codebase
- <https://github.com/SciCatProject>, <https://github.com/paulscherrerinstitut/scicat-ci>
- **Fortnightly developers'/operators' meetings + internal chat channel**
- **Get in touch email address: scicat-operator@lists.psi.ch**
- **Documentation for users and operators**

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

- 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>



User name

wakonig_k



Enter your password

.....



Login





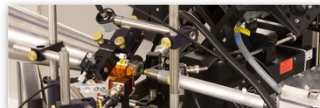
Search

Logbooks

[Add logbook](#)

X-Ray Fourier Ptychography

p16298

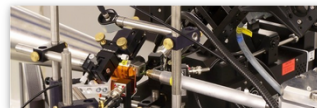


Fourier ptychography is an image technique that comprises multiple image acquisitions, for which the direction of illumination is varied systematically and which are numerically combined in order to extend the functionality of optical elements. The technique allows both amplitude and phase contrast to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

[Open](#)

Three-Dimensional Numerical Modeling of Membrane Distillation

p16273

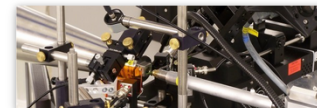


Solar-powered membrane distillation (MD) is a process to desalinate sea water. The technology can support fresh water supply in arid zones of the world with access to sea water. In the scope of a PhD thesis a detailed three-dimensional model for the fluid flow in the membrane will be developed. The Polytetrafluorethylene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

[Open](#)

Nano Imaging of Biogenic Calcite and Bioinspired Calcium Carbonates

p16403

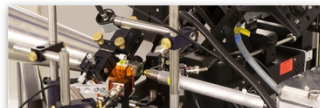


This proposal aims to carry out a ptychographic tomography study of calcite prisms from *Mytilus edulis* (a bivalve mollusc) shell and bio-inspired calcite crystals precipitated in the presence of amino acid additives in order to reveal details of the underlying nanostructure. The results will be used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

[Open](#)

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

X-Ray Fourier Ptychography

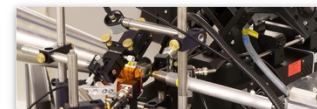
p16414



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

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

Visualization of stacking faults in InSb micropillars by ptychographic topography

Edit dashboard

Small piezo using standard macro npoint_piezo.mac

Tasks 2 items left

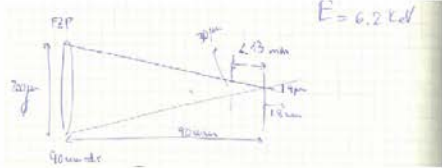
Add task

- prestine sample
- frozen-hydrated sample
- take photos of setup
- copy data to pgroup

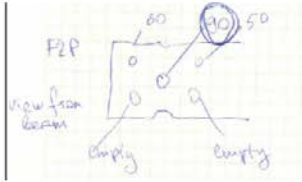
Logbook view

Search

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



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



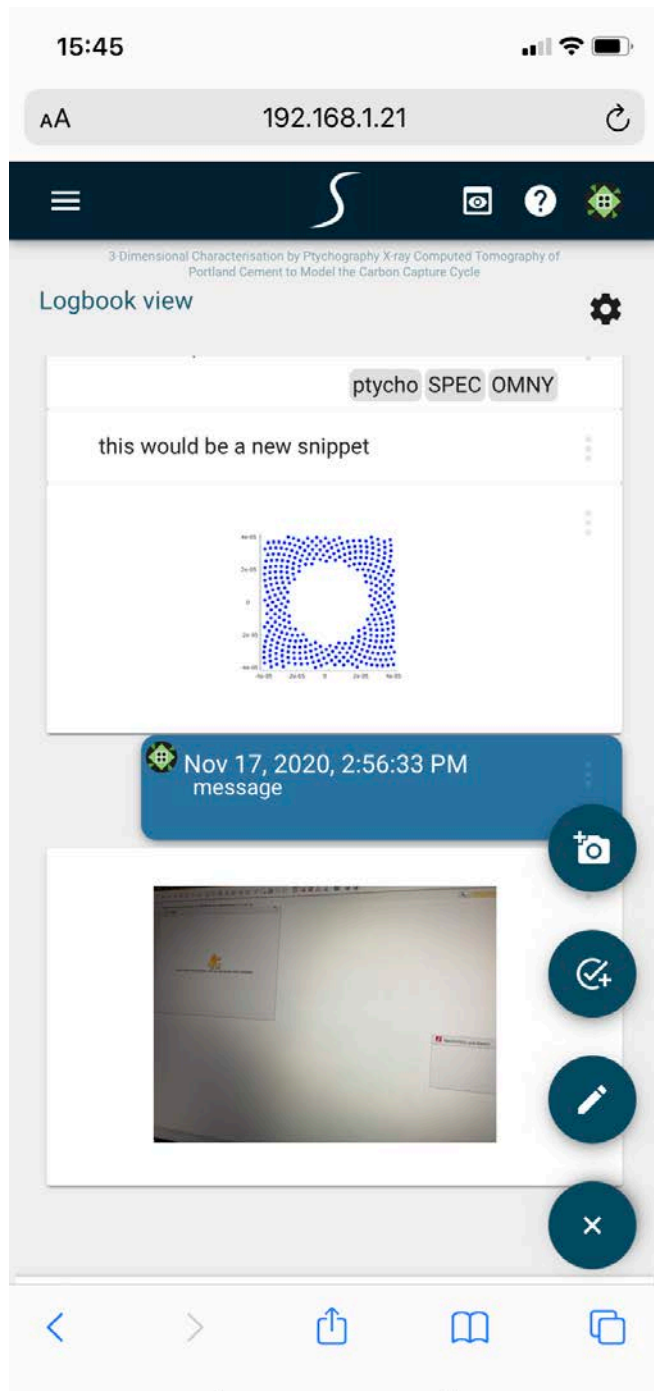
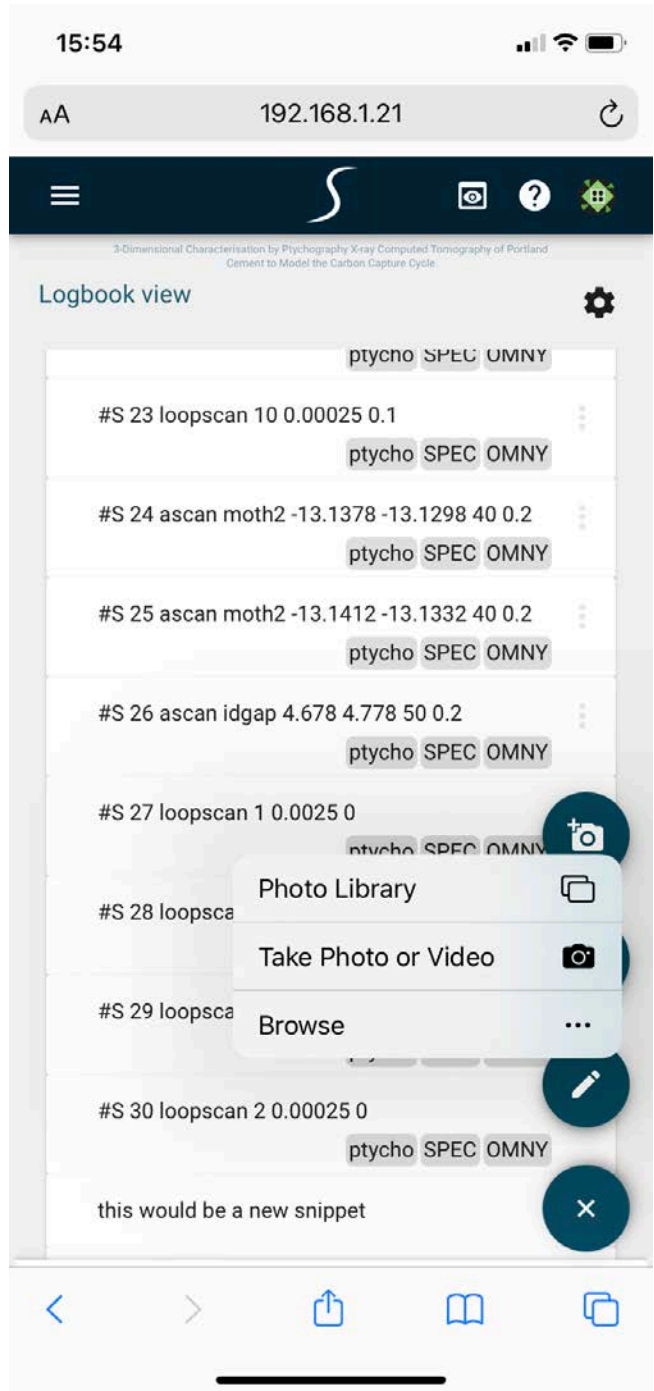
Hexapod can't be controlled from SPEC. We control it with Hexapod windows software via ethernet communication so the motor names and directions are different:

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

Paragraph

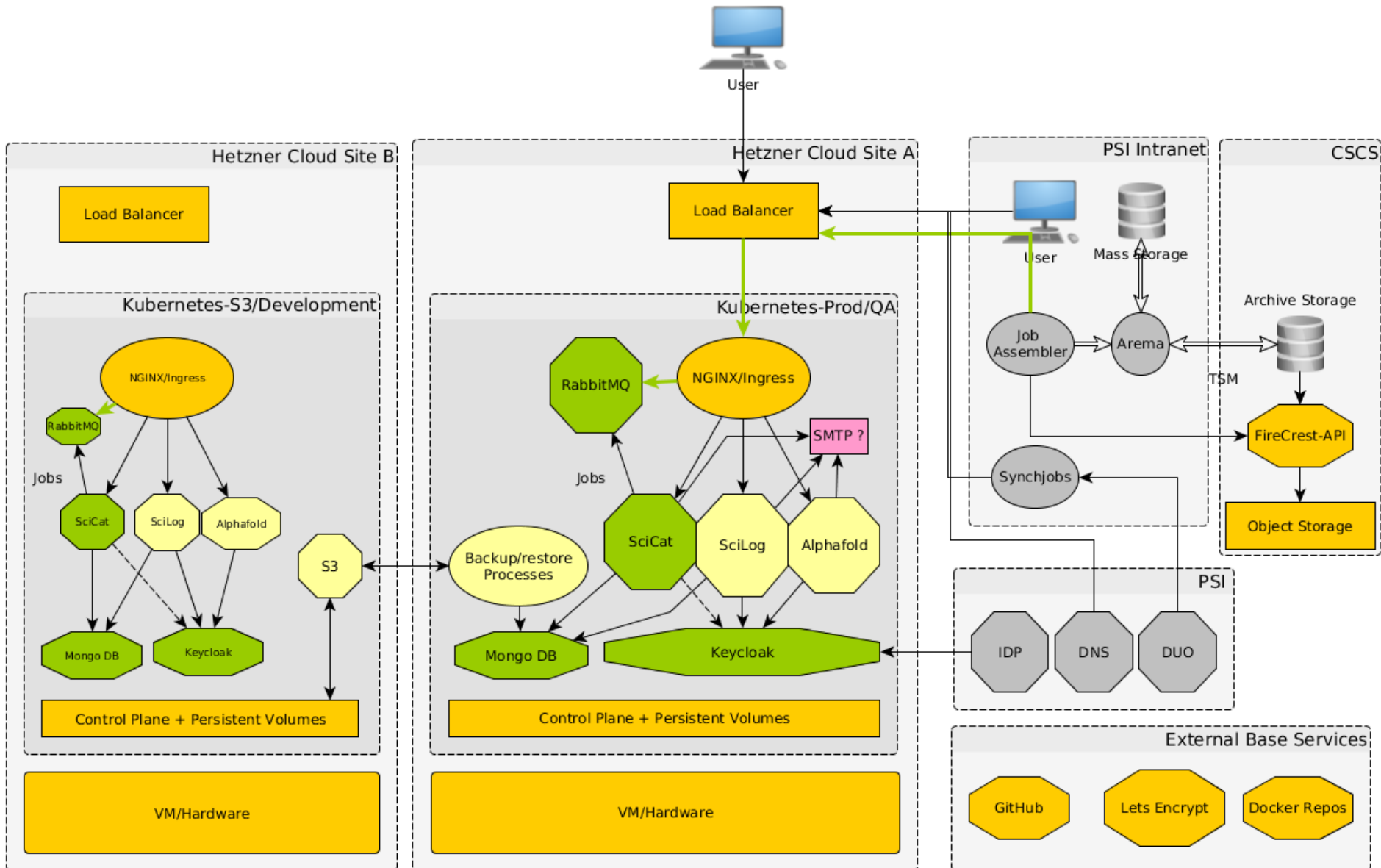
Add your content here.

Show tags



- **Both services rely on the same infrastructure and technologies**
- Key Technologies
 - GitHub and GitLab including their CICD automation tools, Helm charts
 - Hosting: Hetzner Cloud VMs , Network, Loadbalancer, Disks (Migration foreseen)
 - Taiga: Project Management, Shared ToDos
 - Deployment: Kubernetes based on KubeOne and terraform: (All the following layers are hosted inside Kubernetes)
<https://github.com/paulscherrerinstitut/cloudsetup>
 - Persistence: MongoDB , using the Percona distribution
 - Backup: Minio S3 compatible Object Store
 - API Server: Loopback 3 (NestJS in prep) for SciCat and Loopback 4 for SciLog
 - IDM: via Keycloak as IDP Gateway
 - Web certificates: auto deploy via letsencrypt
 - Message Broker: RabbitMQ
 - Archive Interface: NodeRed (Hosted in PSI, thanks Krisz!)
 - Logging and Monitoring: Grafana + Loki
- Main Languages: Nodejs, Typescript, Go, Python

Cloud architecture and connections to PSI



CI: deployment of SciCat microservices

<https://github.com/paulscherrerinstitute/scicat-ci>

📁	.github/workflows	Add oaipmh submodule and deploy Set runNumber enabled option to true	last month
➔	backend @ 4ac55ef	Fix condition of email policy concat on retrieve	2 months ago
➔	frontend @ 407e35f	Add frontend submodule	4 months ago
📁	generic_service_chart	Add base64 validity check	last month
📁	helm_configs	Set runNumber enabled option to true	19 hours ago
➔	oaipmh @ 24a3a49	Add oaipmh submodule and deploy	last month
➔	pan-ontologies-api @ f2b555d	Add pan-ontologies-api deploy	4 months ago
➔	search-api @ 7b2bb69	Add search-api deployment	4 months ago
📄	.gitmodules	Add oaipmh submodule and deploy	last month
📄	LICENSE	Initial commit	4 months ago
📄	README.md	Update README.md	2 days ago

<https://github.com/SciCatProject/backend>

📁 .github/workflows	Add		
📁 backend @ 4ac55ef	Fix		
📁 frontend @ 407e35f	Add		
📁 generic_service_chart	Add base64 validity check		last month
📁 helm_configs	Set runNumber enabled option to true		19 hours ago
📁 oaipmh @ 24a3a49	Add oaipmh submodule and deploy		last month
📁 pan-ontologies-api @ f2b555d	Add pan-ontologies-api deploy		4 months ago
📁 search-api @ 7b2bb69	Add search-api deployment		4 months ago
📁 .gitmodules	Add oaipmh submodule and deploy		last month
📄 LICENSE	Initial commit		4 months ago
📄 README.md	Update README.md		2 days ago

📁 .github	Avoid gitlab workflows when not in SciCatProject	6 months ago
📁 CI	small change in maxiv handleJobSideEffects.js	5 months ago
📁 client/api	Ensure all model IDs are strings, not ObjectIDs	2 years ago
📁 common	Call to JSON on publication to fix post to oaipmh	5 months ago
📁 email-templates	update job object to match copy service's report	6 months ago
📁 scripts	New script to replace objectIDs	2 years ago
📁 server	Fix condition of email policy concat on retrieve	4 months ago
📁 test	Merge pull request #614 from SciCatProject/oidc1.1	5 months ago
📄 .dockerignore	Make sure that package-lock-json is copied to docker image	2 years ago
📄 .editorconfig	Github release	5 years ago
📄 .eslintignore	Update .eslintignore	2 years ago
📄 .eslintrc	Github release	5 years ago

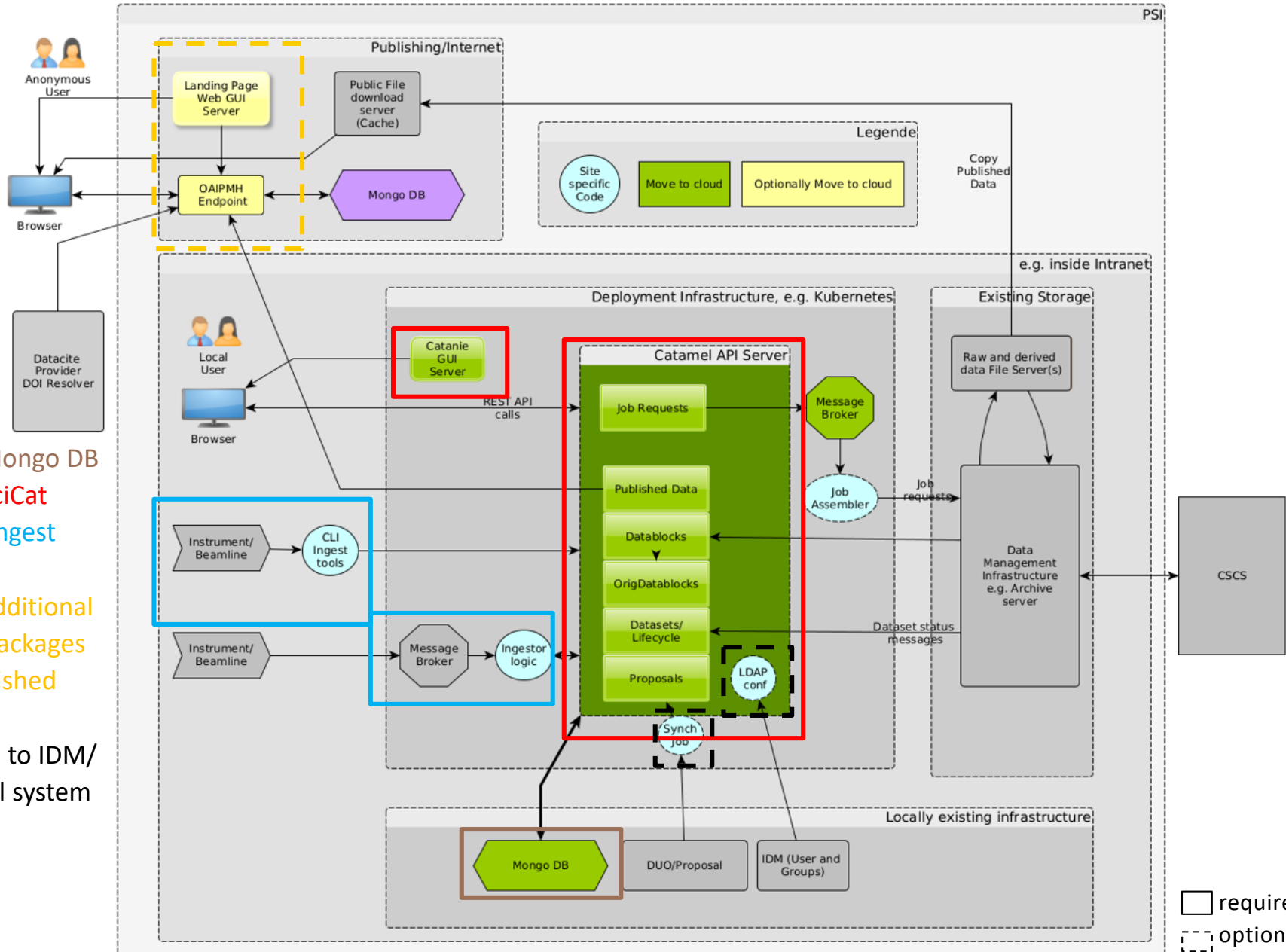
- SciCat: <https://discovery-qa.psi.ch>
- SciLog: <https://scilog.psi.ch>
- SciCat-ci: <https://github.com/paulscherrerinstitute/scicat-ci>
- Hetzner Dashboard: <https://accounts.hetzner.com/login>
- Kubernetes Dashboard: <https://k8sdashboard.psi.ch>
- Nodered: <http://nodered.psi.ch>
- Grafana: <https://awicloud.grafana.net>
- Minio: <https://minioconsole.development.psi.ch>
- Keycloak: <https://kc.psi.ch/auth>



Additional Material



Data Catalog Architecture and Integration



1. Setup Mongo DB
2. Setup SciCat
3. Define ingest process
4. Setup additional SciCat packages for published data
5. Connect to IDM/ proposal system

□ required
 □ optional

SciCat

Files for running SciCat with docker-compose.

Steps

1. Clone the repository

```
git clone https://github.com/SciCatProject/scicatlive.git
```

2. Run with the following command inside the directory

```
docker-compose up -d
```

3. SciCat will now be available on <http://localhost>. The Loopback API explorer of catamel is available at <http://localhost/explorer/>, the one for the search-api at <http://localhost/panosc-explorer/>.

Add Your Local Configuration

1. Add your local configuration to [config.local.js](#)
2. Uncomment the `volumes:` line and the line containing `config.local.js` in the catamel service section in [docker-compose.yaml](#) (if commented)
3. Restart the docker containers

<https://github.com/SciCatProject/scicatlive>

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"
    }
  }
}...
```