PAUL SCHERRER INSTITUT

PSI

Carlo Minotti:: Paul Scherrer Institut

SciCat Present and Future

NOBUGS 2022

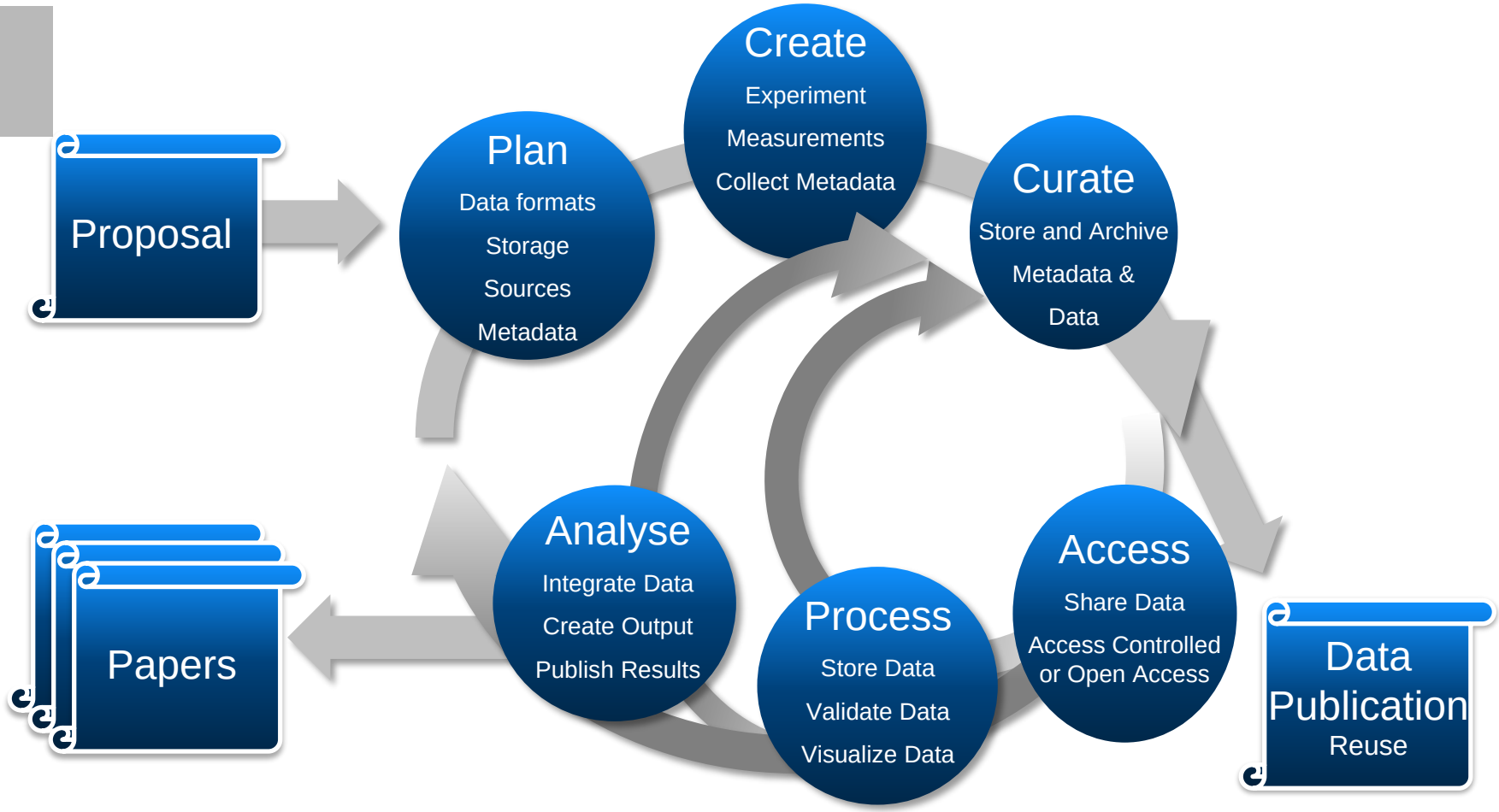
Paul Scherer Institute, Villigen, Switzerland

PAUL SCHERRER INSTITUT

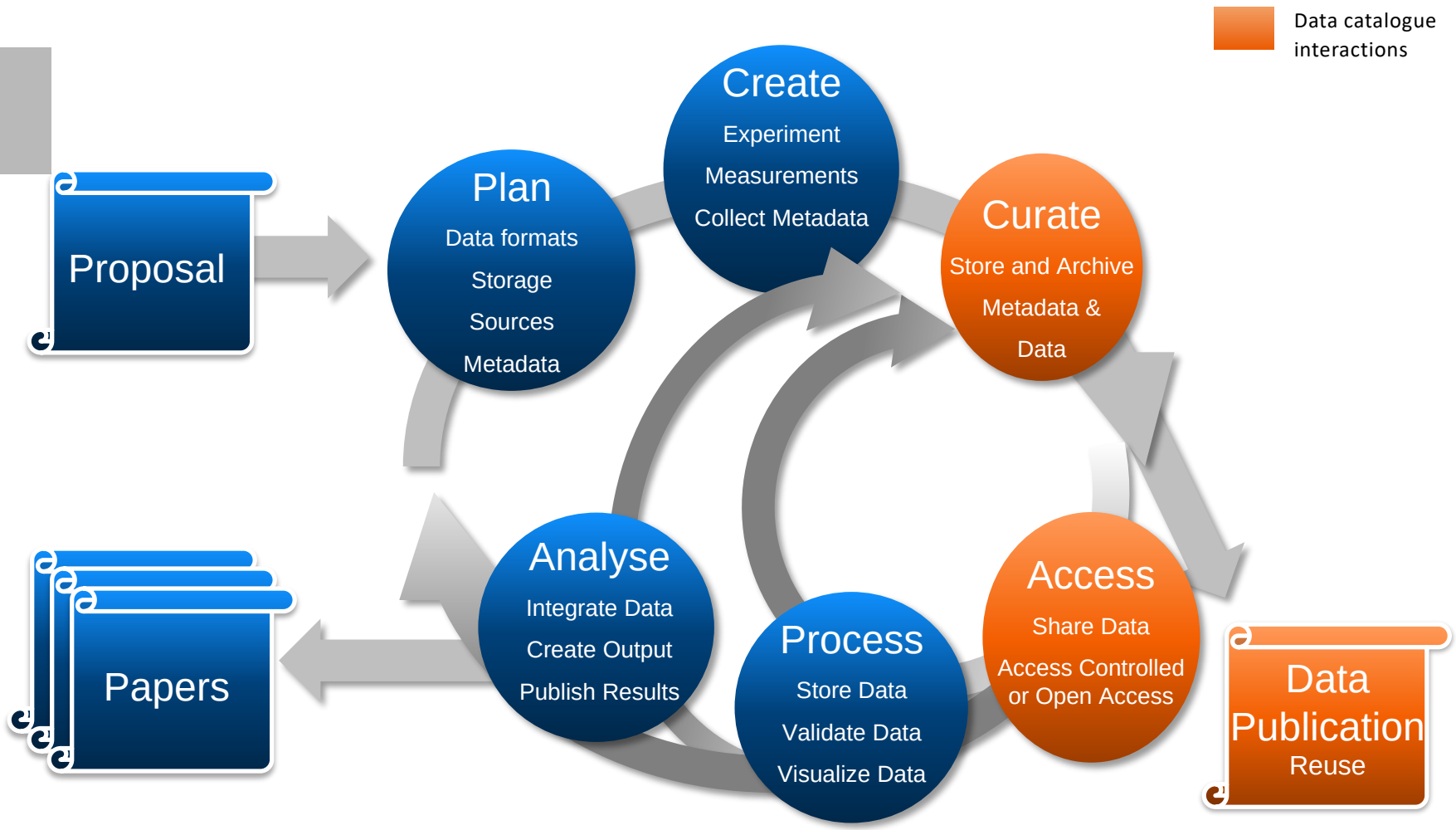
Outline

- The role of SciCat
- The user‘s perspective
- The developer/maintainer’s perspective
- The beamline scientist/ingestor’s perspective

A typical user journey



Interactions with the data catalogue



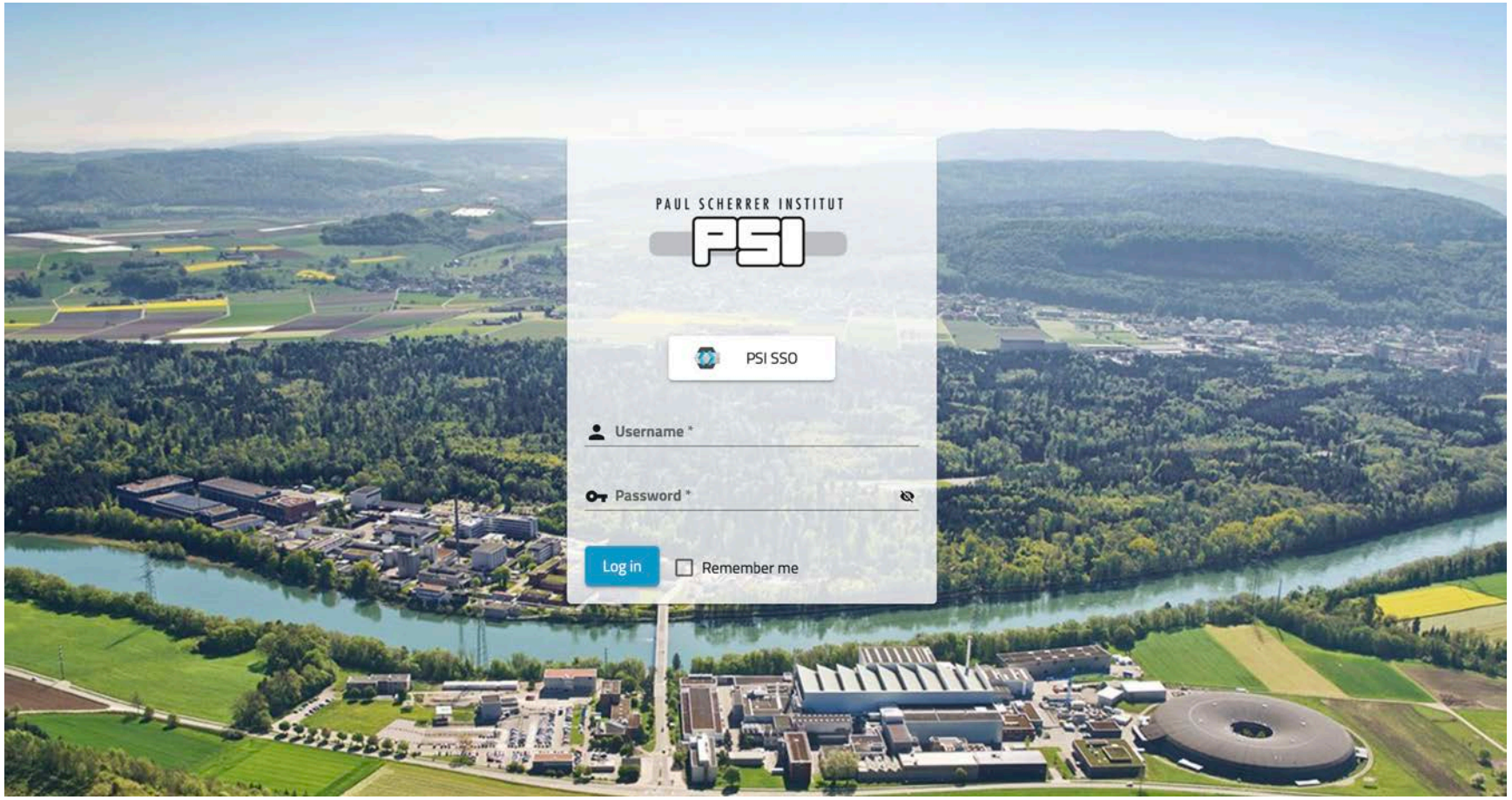
PAUL SCHERRER INSTITUT

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

- Supported authentication methods:
  - Default accounts
  - LDAP
  - ODIC (e.g. Keycloak/Sign in with Google…)

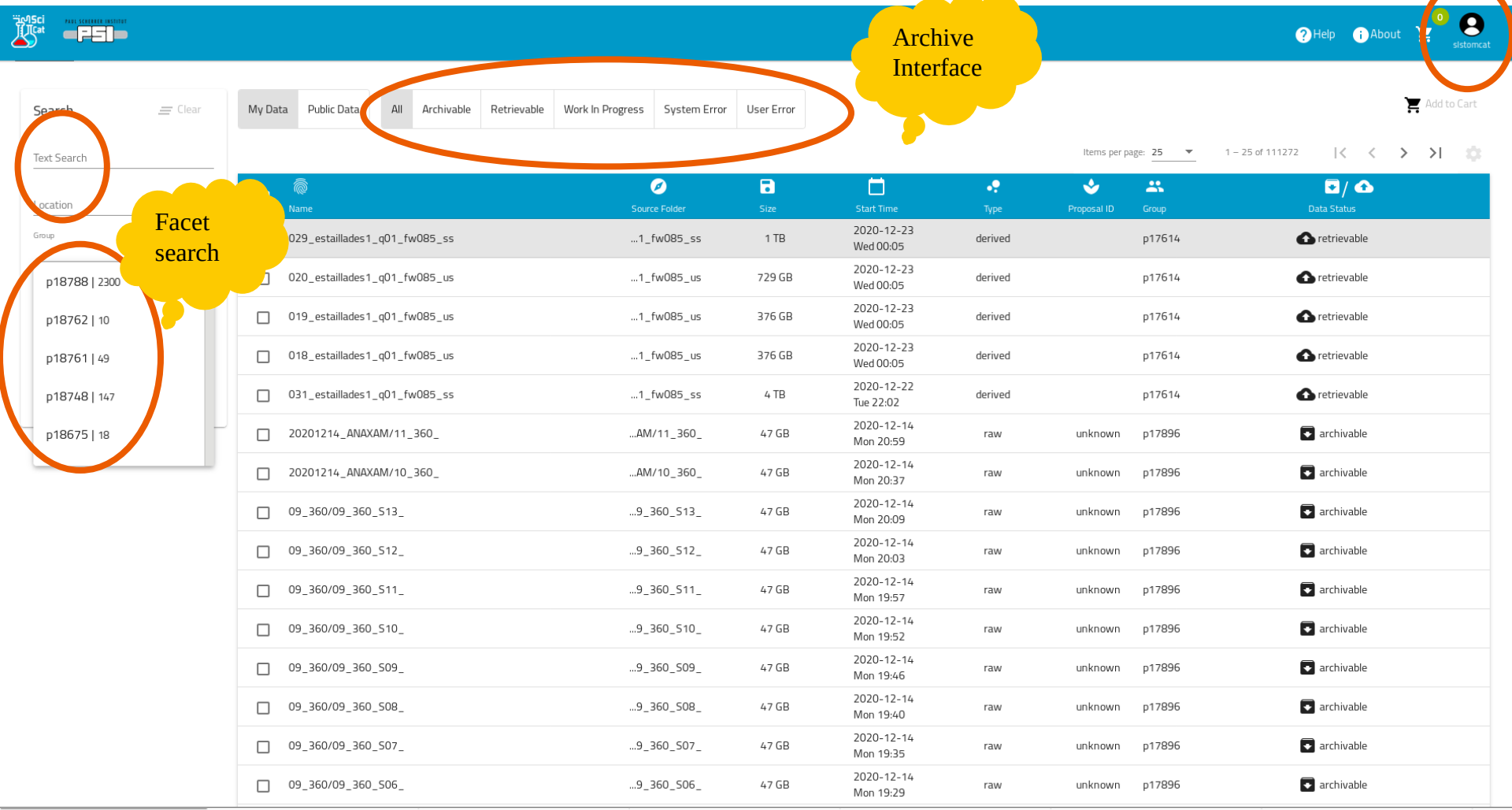


Discover data via WebUI

User specific data *

Archive Interface

Facet search



* 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

Help

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

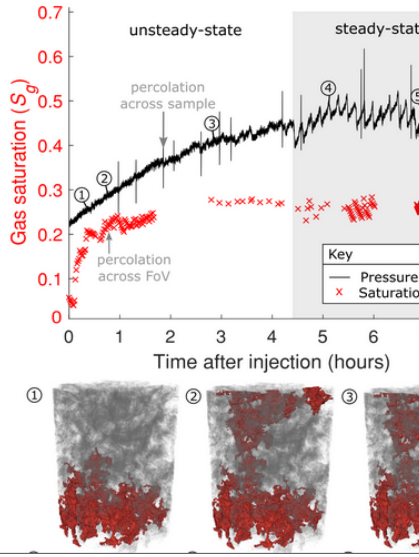
C. Spurin, T. Bultreys, M. Rücker, G. Garfi, C. M. Schlepütz, V. Novak, S. Berg, M. J. Blunt, and S. Krevor. Real-Time Imaging Reveals Distinct Pore-Scale Dynamics During Transient and Equilibrium Subsurface Multiphase Flow. Water Resour. Res. 56, 433 (2020). https://doi.org/10.1029/2020WR028287

Update Published Data | Cancel

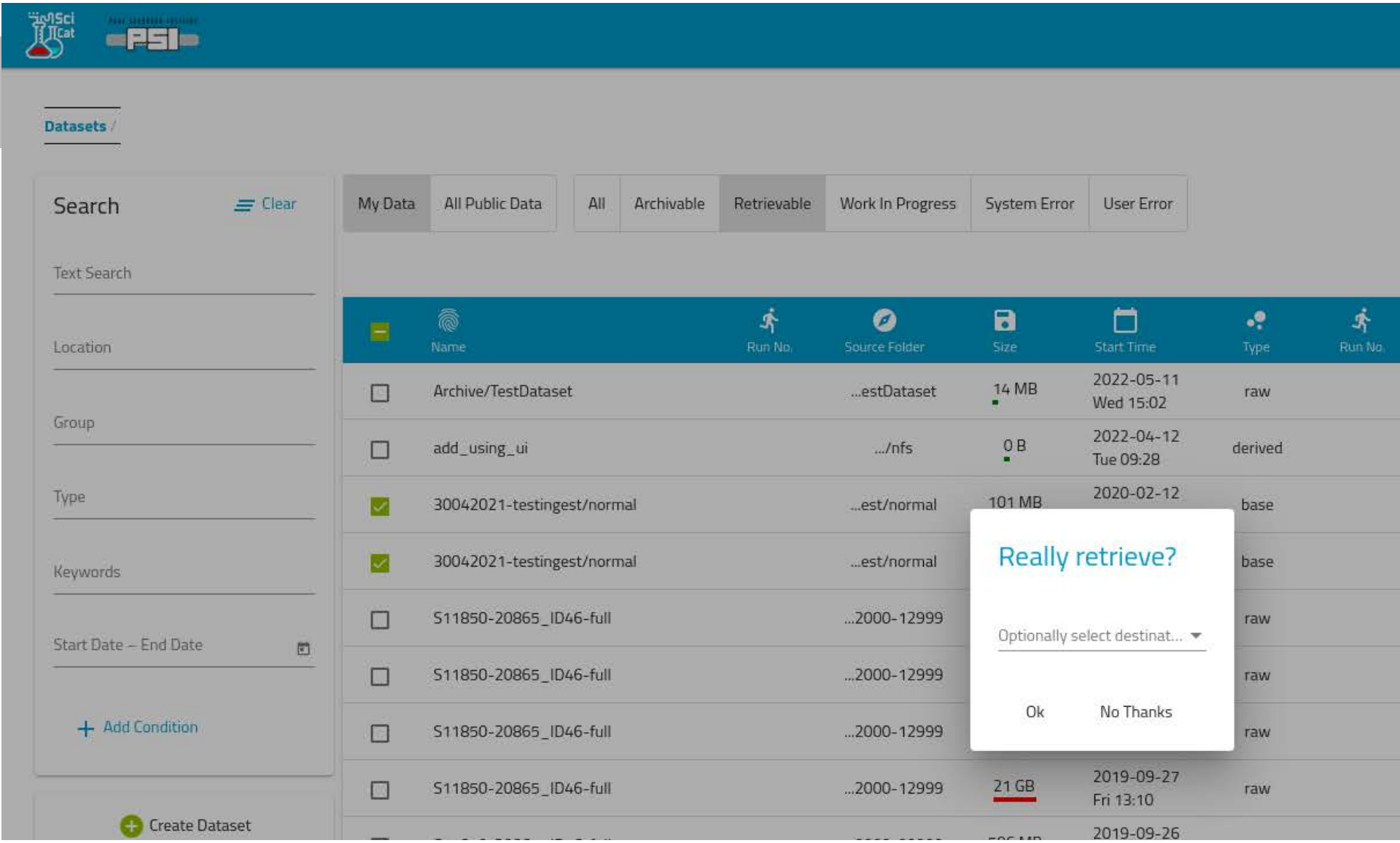
Drop a file here

or

Browse



Retrieving data from tape



Published Data = List of Datasets + Metadata + DOI

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

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

Abstract

In the related publication to these data sets, we explore the flow dynamics for two-phase flow in a porous medium (a bioclastic carbonate rock). We use state-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

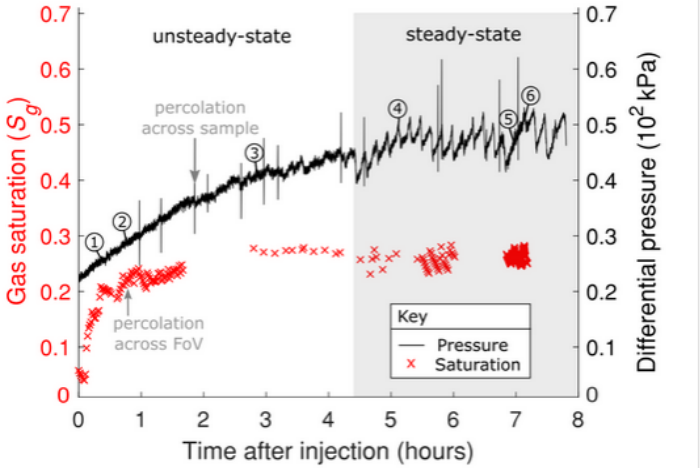
20.500.11935/5899a0eb-7e3b-451f-b01e-17ddfc0d0938

Actions

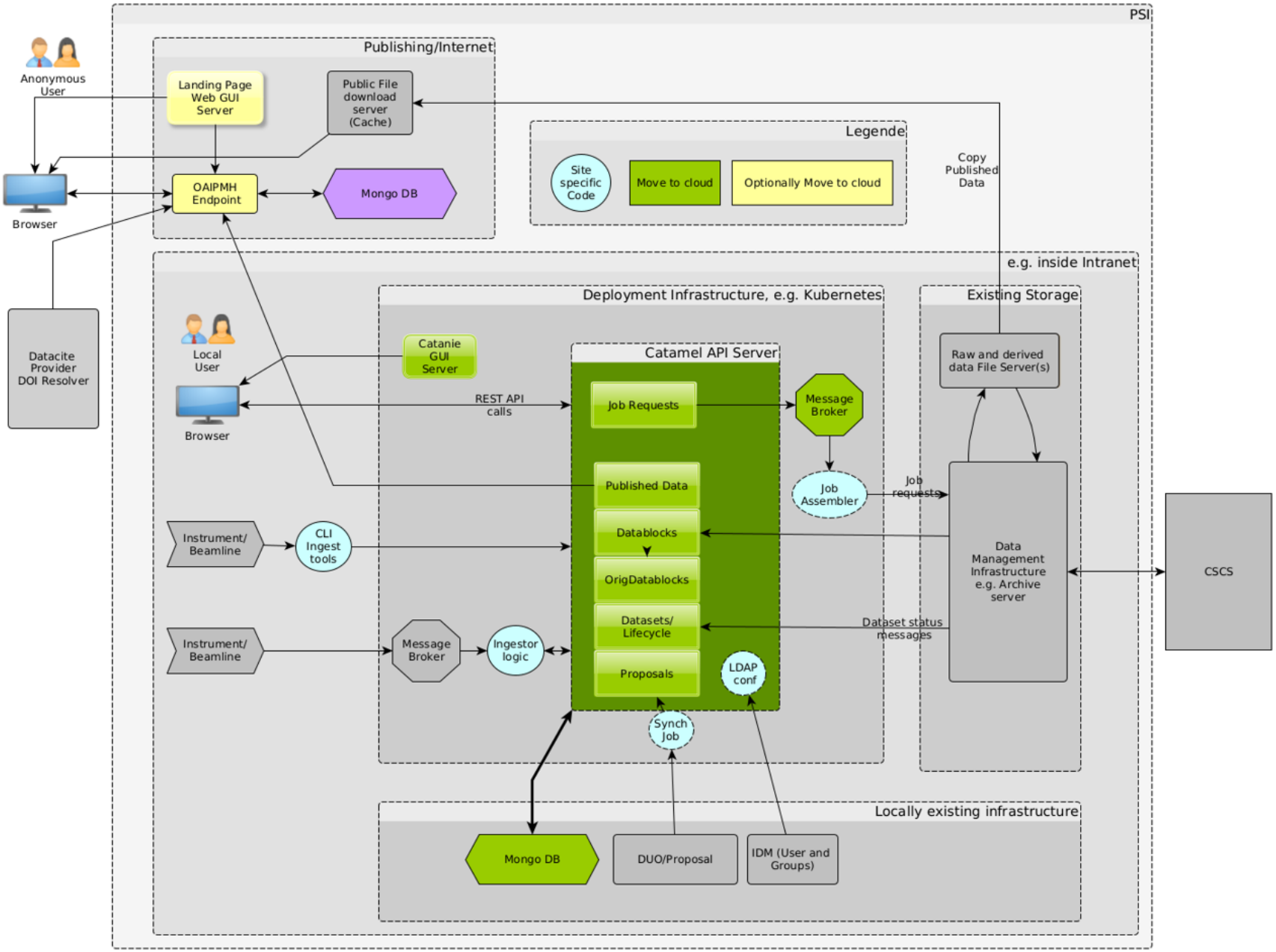
To access the data associated with this DOI click below and follow the instructions

Access Data

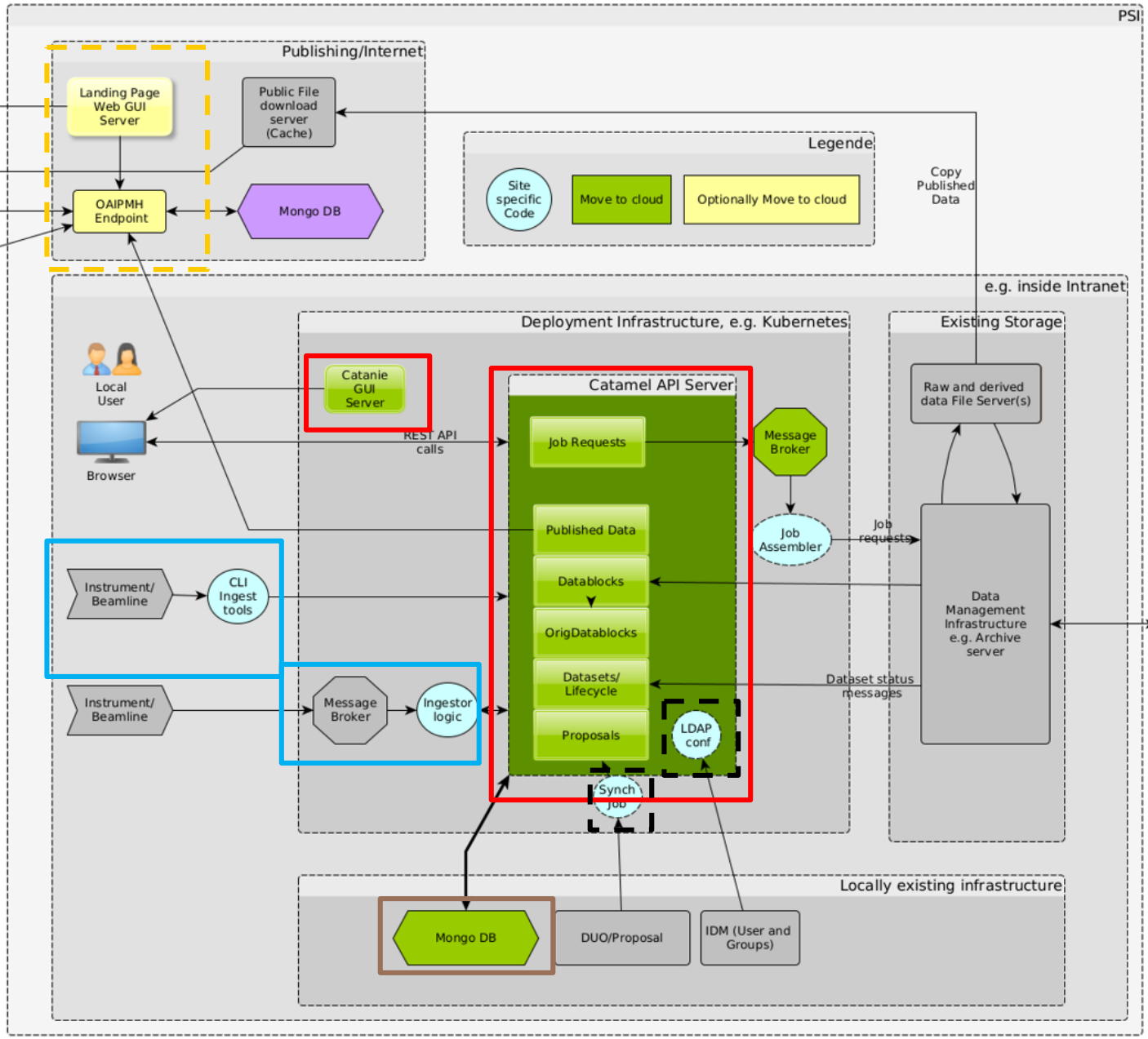
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License



Data Catalog Architecture and Integration



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

One-click Reference Installation

README.md

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
```

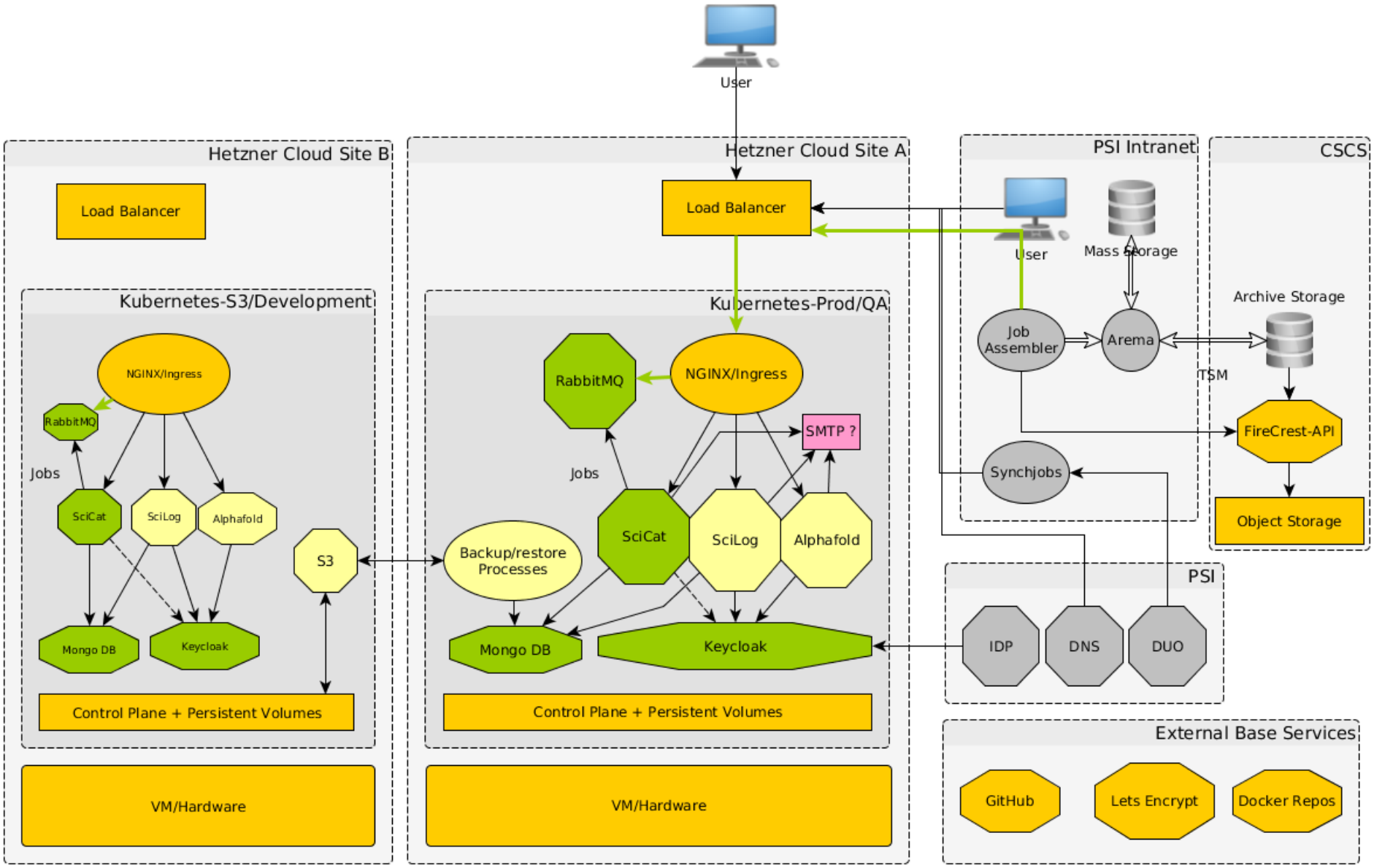
3. SciCat will now be available on http://localhost. The Loopback API explorer of catamel is available at http://localhost/explorer/, the one for the search-api at http://localhost/panosc-explorer/.

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

Cloud architecture and connections to PSI



CI: deployment of SciCat microservices

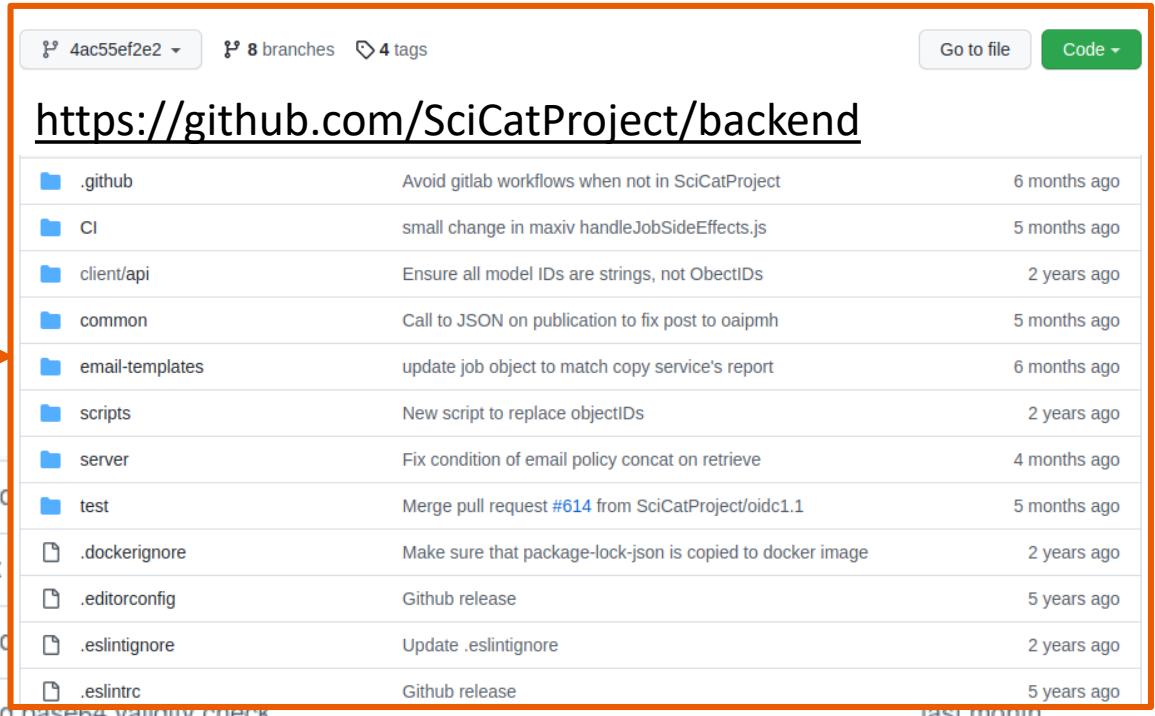
https://github.com/paulscherrerinstitute/scicat-ci

.github/workflows	Add oaipmh su… 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

PAUL SCHERRER INSTITUT PSI

CI

https://github.com/SciCatProject/backend



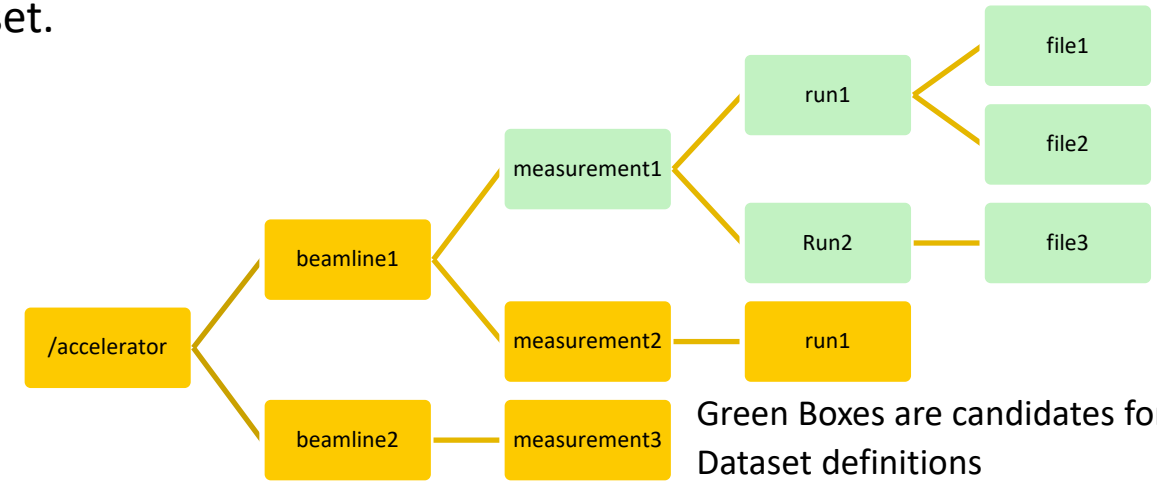
4ac55ef2e2 | 8 branches | 4 tags | Go to file | Code

.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 ObectIDs	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

.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

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

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



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

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"
        }
    }…
}
```

PAUL SCHERRER INSTITUT

Future work

- Migration to a new backend technology
- More granular permission model
- Increased configurability in tables in UI
- Kubernetes deployment manifests

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

- **Fortnightly developers'/operators' meetings + internal chat channel**
- **Please get in contact for questions about SciCat onboarding at this 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