An Update from public-data.desy.de

Tim Wetzel, Patrick Fuhrmann, Armando Bermudez Martinez, Uwe Jandt, Paul Millar, Sophie Servan, Peter van der Reest, Regina Hinzmann, Johannes Reppin, Christian Voss, Linus Pithan, Anton Barty, … (DESY)
Julia Kobus, Philipp Jordt, Linus Liedtke, Bridget Murphy (CAU Kiel)

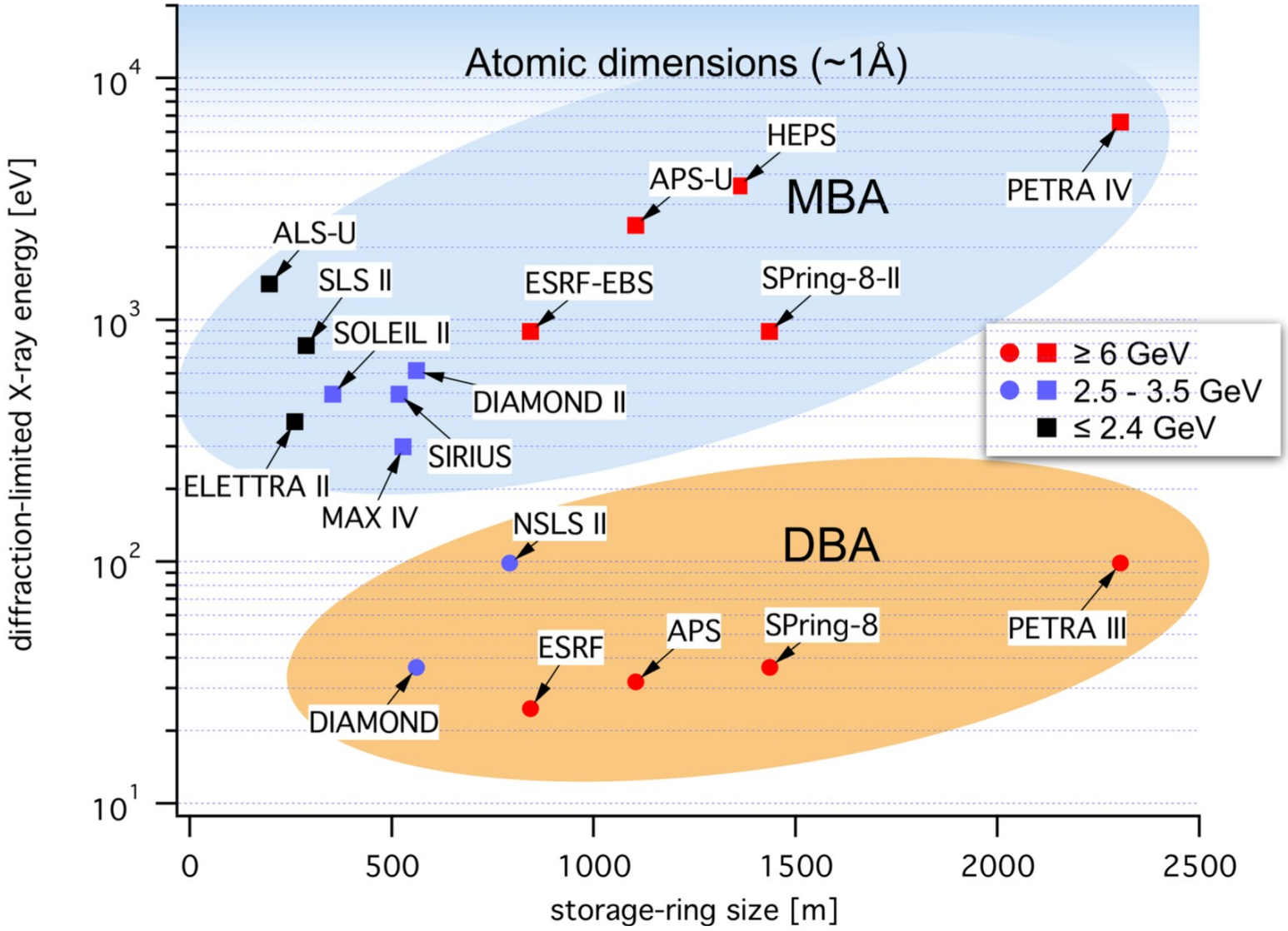
Slides partly with courtesy from Patrick Fuhrmann

This presentation is licensed under
https://creativecommons.org/licenses/by/4.0/

FH SciComp Workshop 2025

The Synchrotrons

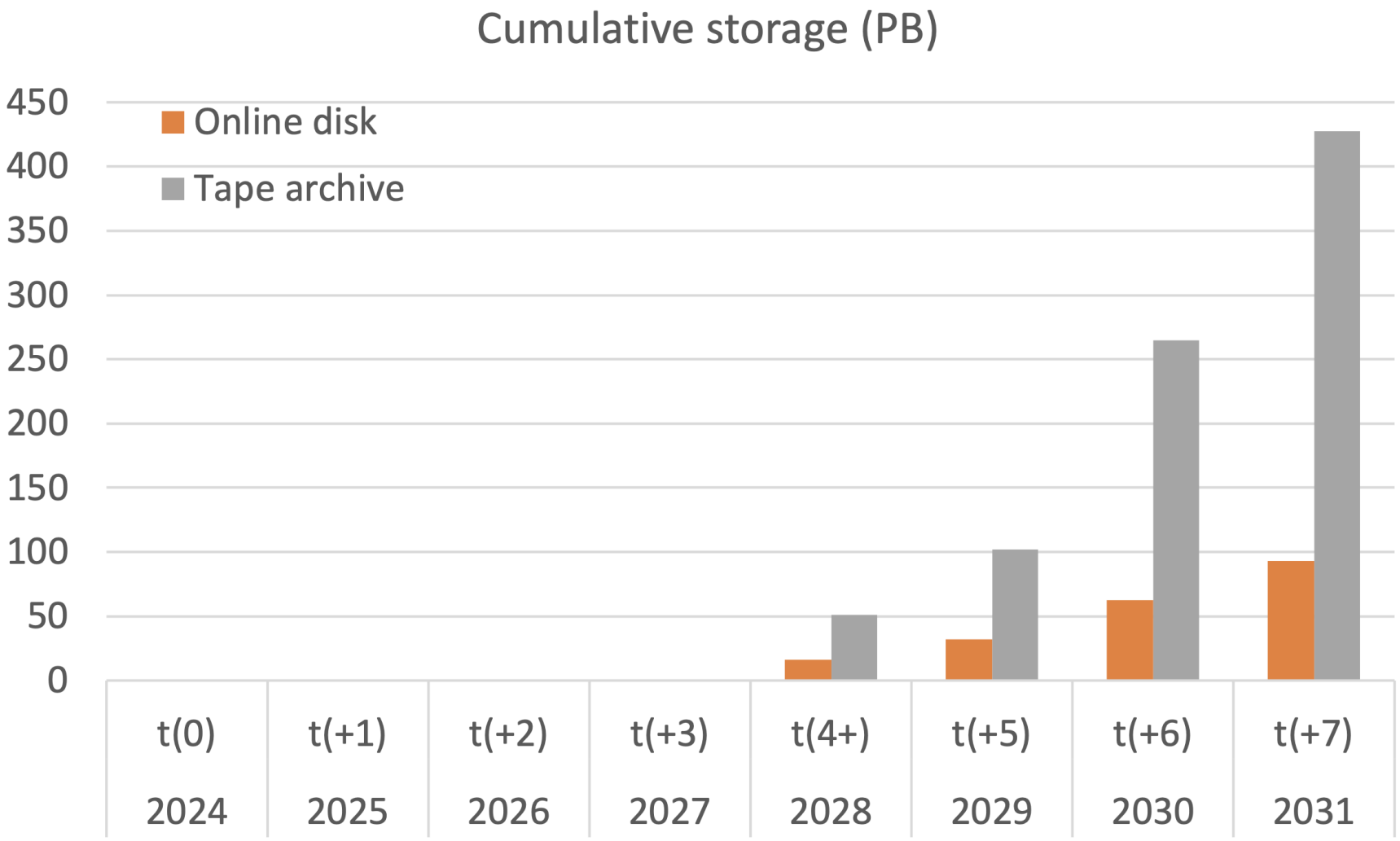
PETRA III/IV and Friends



Diffraction-limited photon energy for some synchrotron radiation sources and their future upgraders. PETRA IV will be the first source worldwide to reach the diffraction limit for hard X-ray energies.

The Consequence

Storage space for PETRA IV



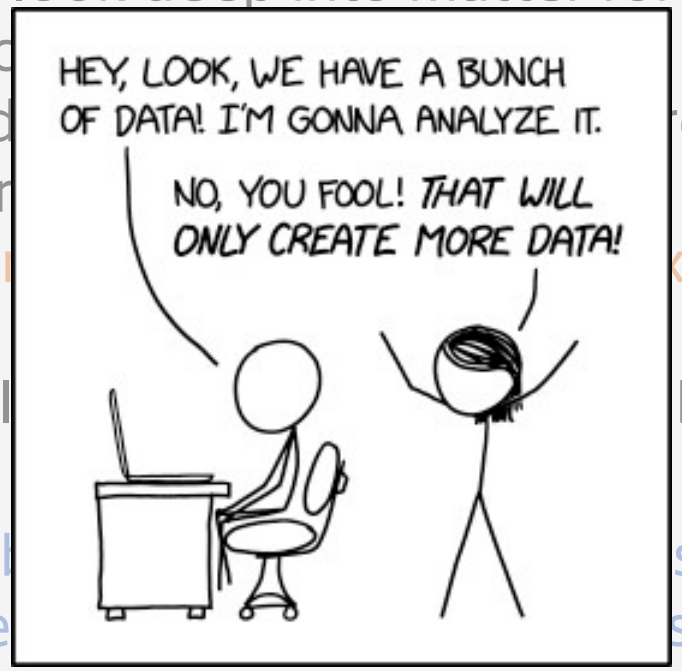
xkcd.com/1260

N.B.:
Timeline is shifted by one year!

The Data

Where and by whom is it produced?

- Light sources are huge infrastructures, providing extremely brilliant light, used with more than 200 techniques to look deep into matter for Biology, Material Science, Cultural Heritage and what nc
- In 2019 PETRA III at DESY had om 3150 national and international users at 21 bear
- Those beam times produce ar xtremely precious data, under the governance of the PI.
- After the 'embargo period' all become public 'Open Data'



xkcd.com/2582

- Data which wasn't used in pul st.
- After a publication, the future s on the PI, who will have moved on to his/her next project. The future of that data is undetermined.
- No one, except the original group, understands the details of those collected datasets.

The Solution!

SciCat

advance your data to advance your science

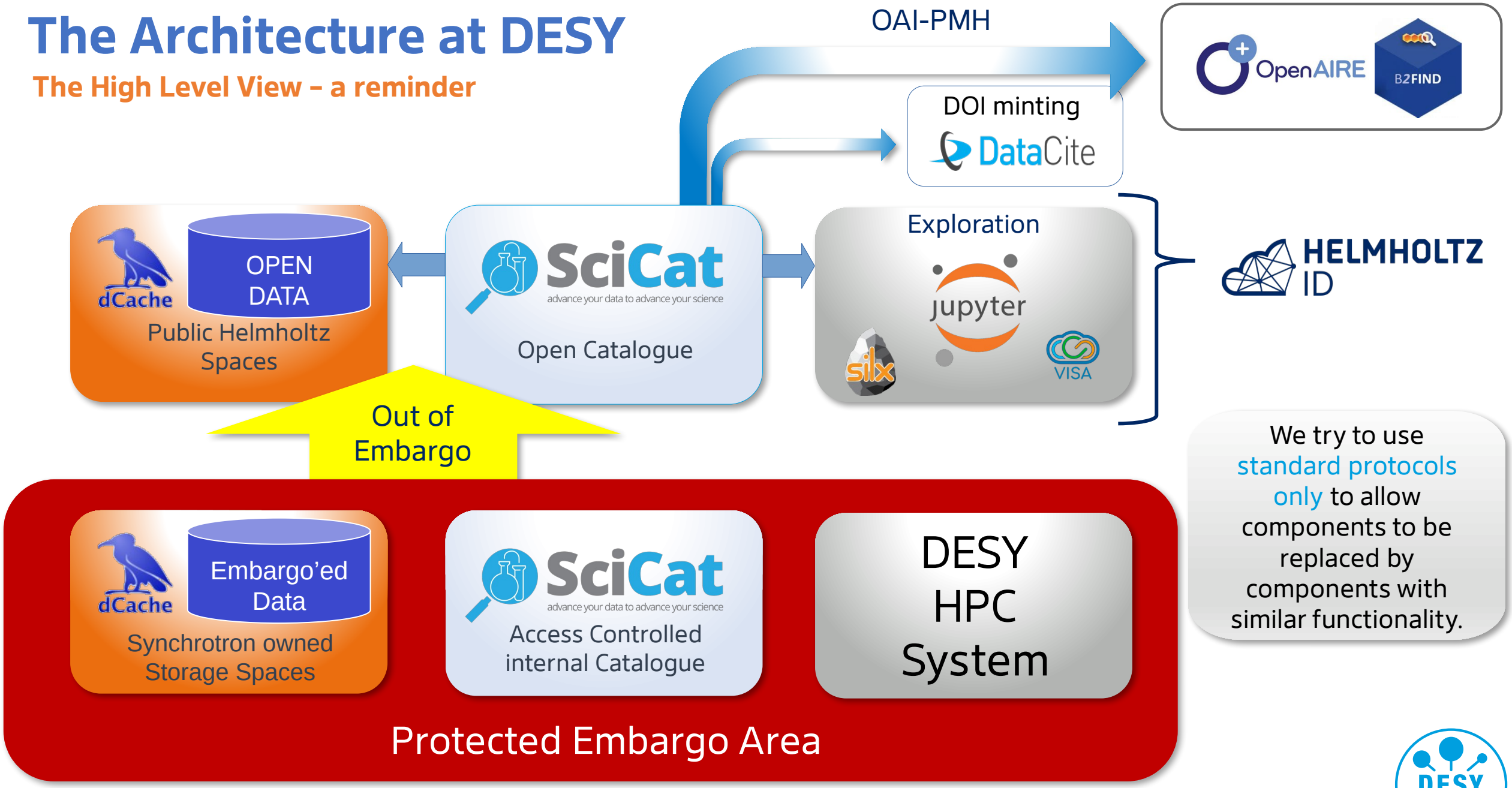
ESS EUROPEAN SPALLATION SOURCE *Max Novelli*	PSI *Carlo Minotti*	MAX IV *Carla Takahashi*
The Rosalind Franklin Institute *Laura Shemilt*	MLZ Heinz Maier-Leibnitz Zentrum *Björn Pedersen*	ALS ADVANCED LIGHT SOURCE *Dylan McReynolds*
DESY. *Regina Hinzmann*	SOLEIL SYNCHROTRON *Patrick Madela*	DAPHNE 4NFDI *Regina Hinzmann, Nicolas Hayen*
DECTRIS CLOUD	HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF *Oliver Knodel*	

Mission

SciCat is a community-driven scientific metadata catalogue designed to simplify metadata management, enabling data sharing, discovery, and collaboration. By providing a central hub for metadata, SciCat makes data easily findable and accessible for the entire community. With flexible integration into diverse infrastructures and a clear path to data publication, SciCat empowers organizations to foster an open and innovative data culture, driving high-quality science forward.

The Architecture at DESY

The High Level View – a reminder



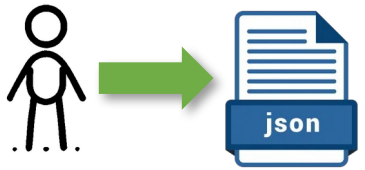
We try to use standard protocols only to allow components to be replaced by components with similar functionality.

The Publishing

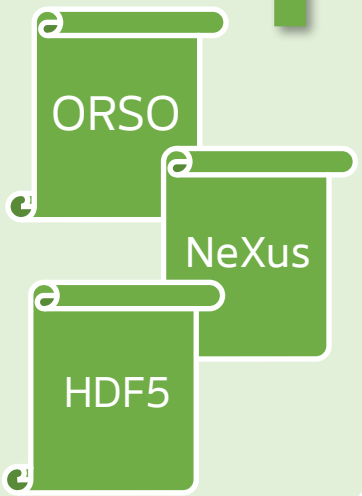
sisyphos.desy.de

User Space

Automatically or manually generated preliminary Metadata



Data



Daphne Sisyphos

PaN Reflectivity Database Upload Tool

Metadata validation against defined schemata enforced.

UI with field annotations

HIFIS Staging Space

Curation Process

- Metadata verification
- Licence/Legal checks
- Data Quality verification

Final Publishing

Open Data Portal

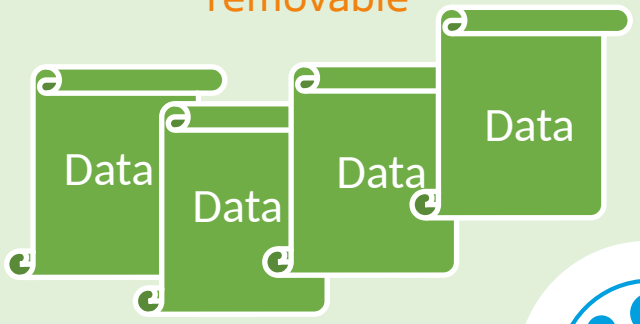
SciCat
advance your data to advance your science

DOI minting
DataCite
Landing Page

OAI-PMH
OpenAIRE
B2FIND

Persistent HIFIS Space

Data is immutable and no longer removable



The Schemata

Metadata Description Language

YAML

MDS Convertion

linkML

W3C OWL

* * *

{x} JSON Schema

Metadata Schema Registry

GitLab

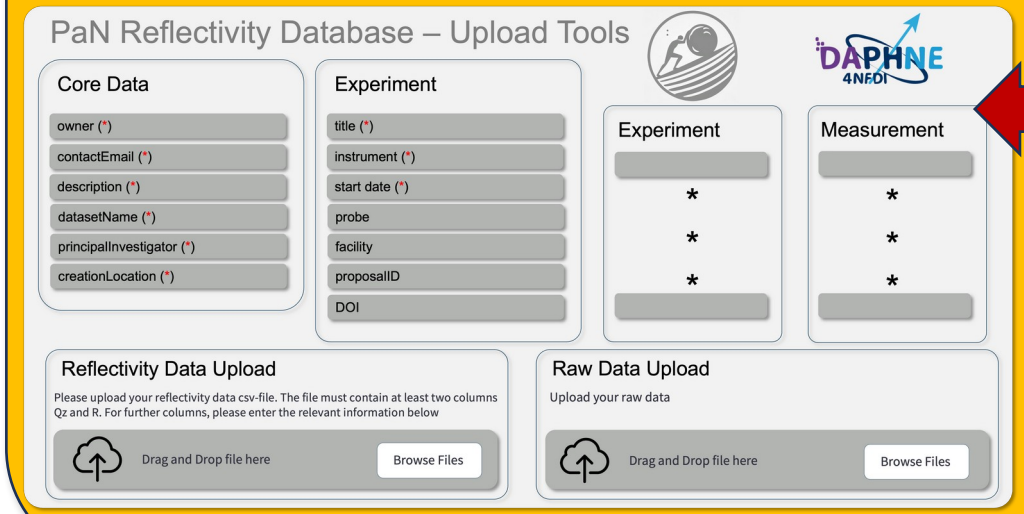
Reflectivity

Tomography

* * *

JSON

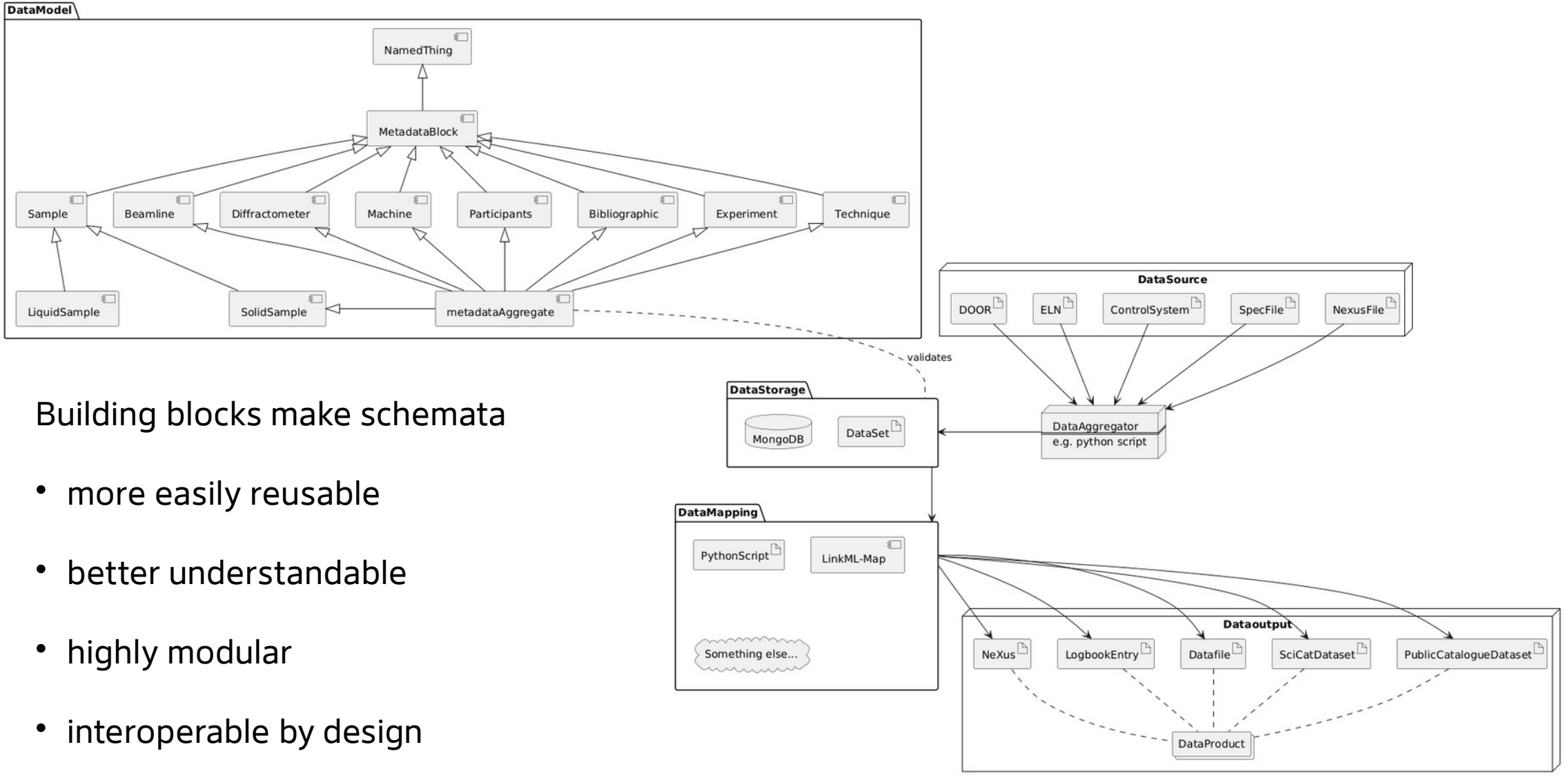
Domain Specific Meta Data Handling and Upload



Sisyphos

Meta Data Upload Form Generator

The building blocks



Building blocks make schemata

- more easily reusable
- better understandable
- highly modular
- interoperable by design

Source: P. Jordt, CAU, priv. comm.

The Form – sisyphos.desy.de

PaN Reflectivity Database – Upload Tools

Core Data

- owner (*)
- contactEmail (*)
- description (*)
- datasetName (*)
- principalInvestigator (*)
- creationLocation (*)

Experiment

- title (*)
- instrument (*)
- start date (*)
- probe
- facility
- proposalID
- DOI

Experiment

*
*
*

Measurement

*
*
*

Reflectivity Data Upload

Please upload your reflectivity data csv-file. The file must contain at least two columns Qz and R. For further columns, please enter the relevant information below

Drag and Drop file here

Browse Files

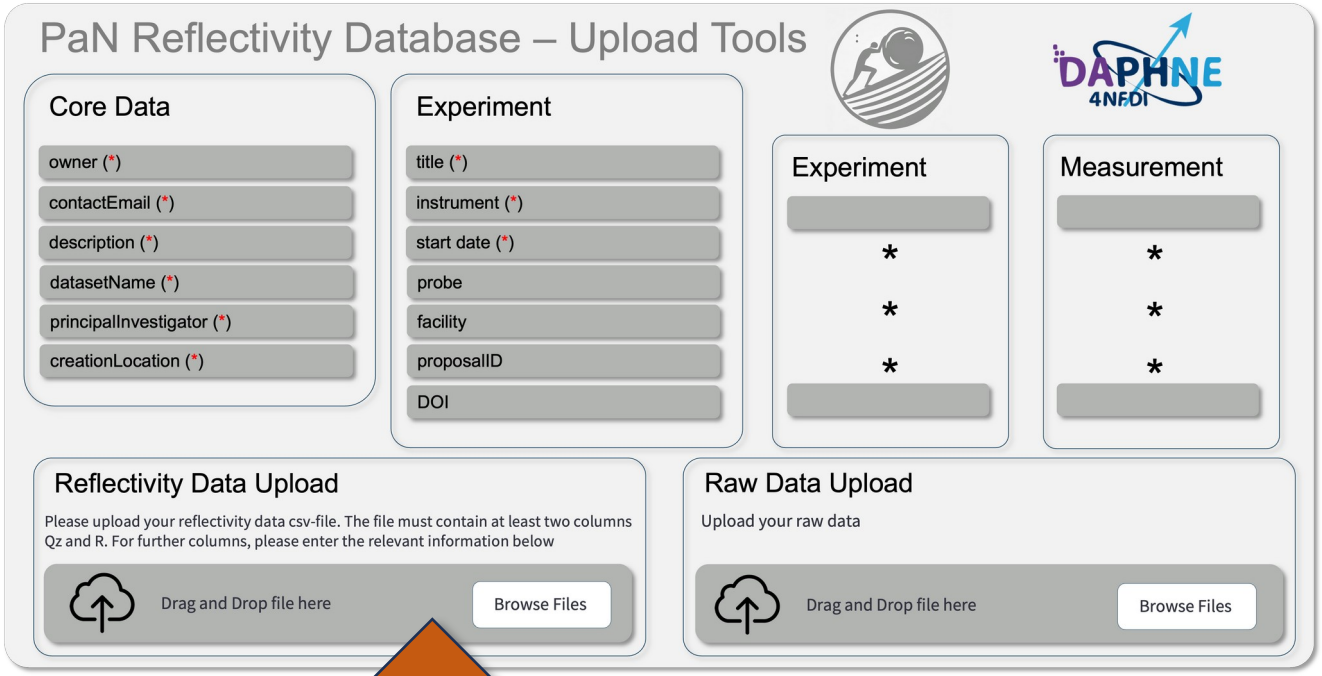
Raw Data Upload

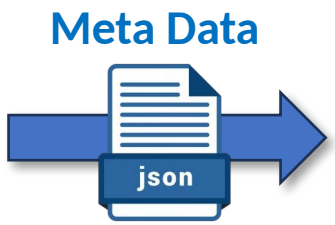
Upload your raw data

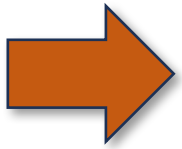
Drag and Drop file here

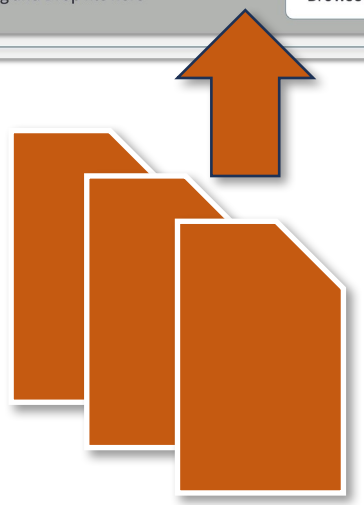
Browse Files

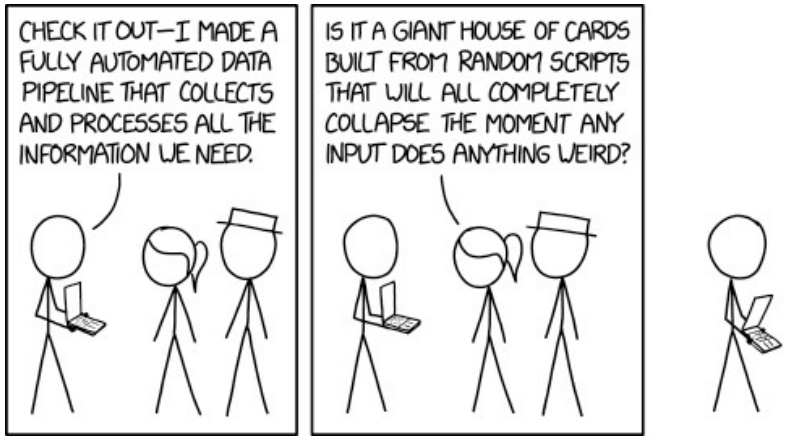
The Form – sisyphos.desy.de







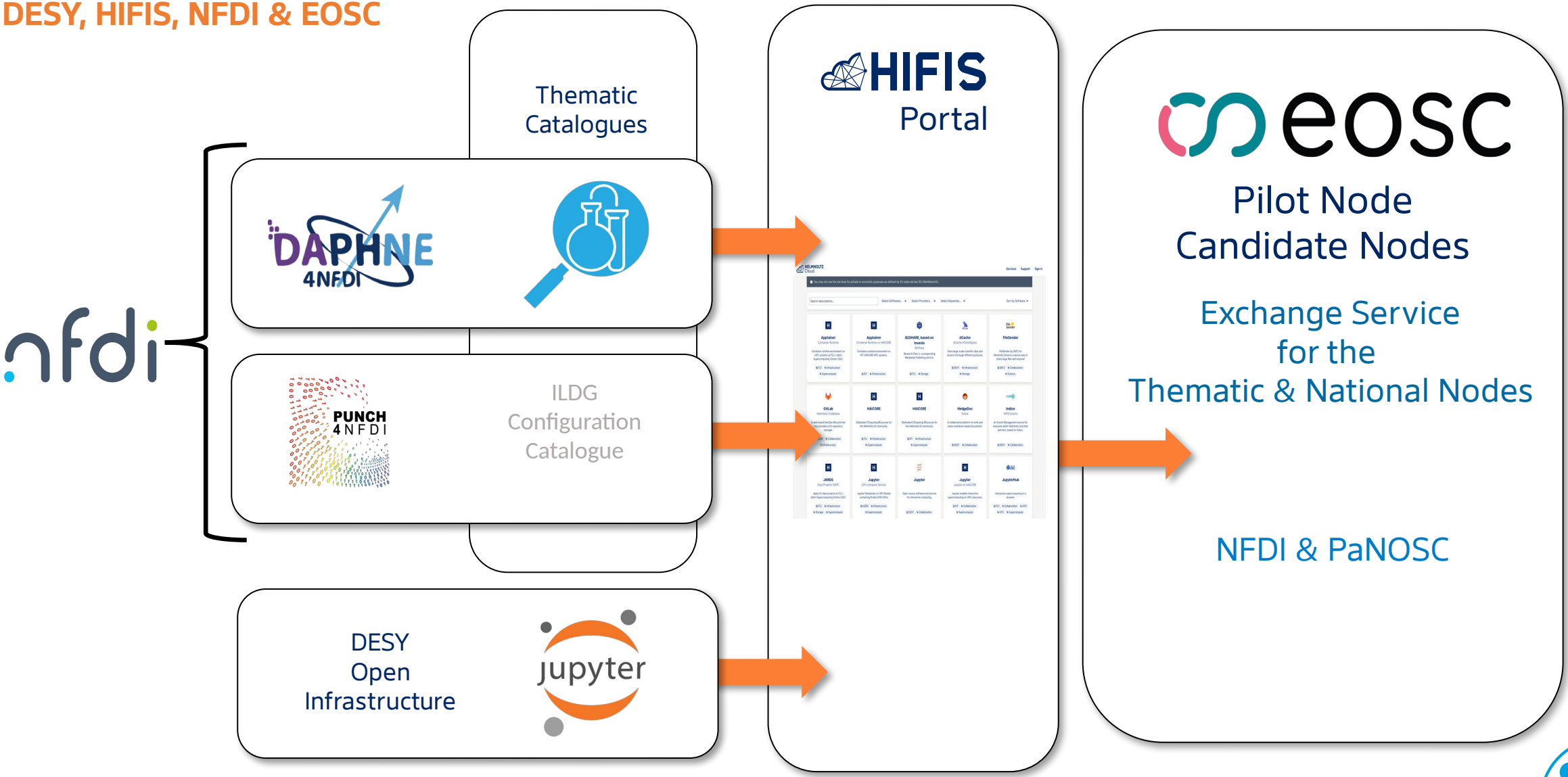




xkcd.com/2054

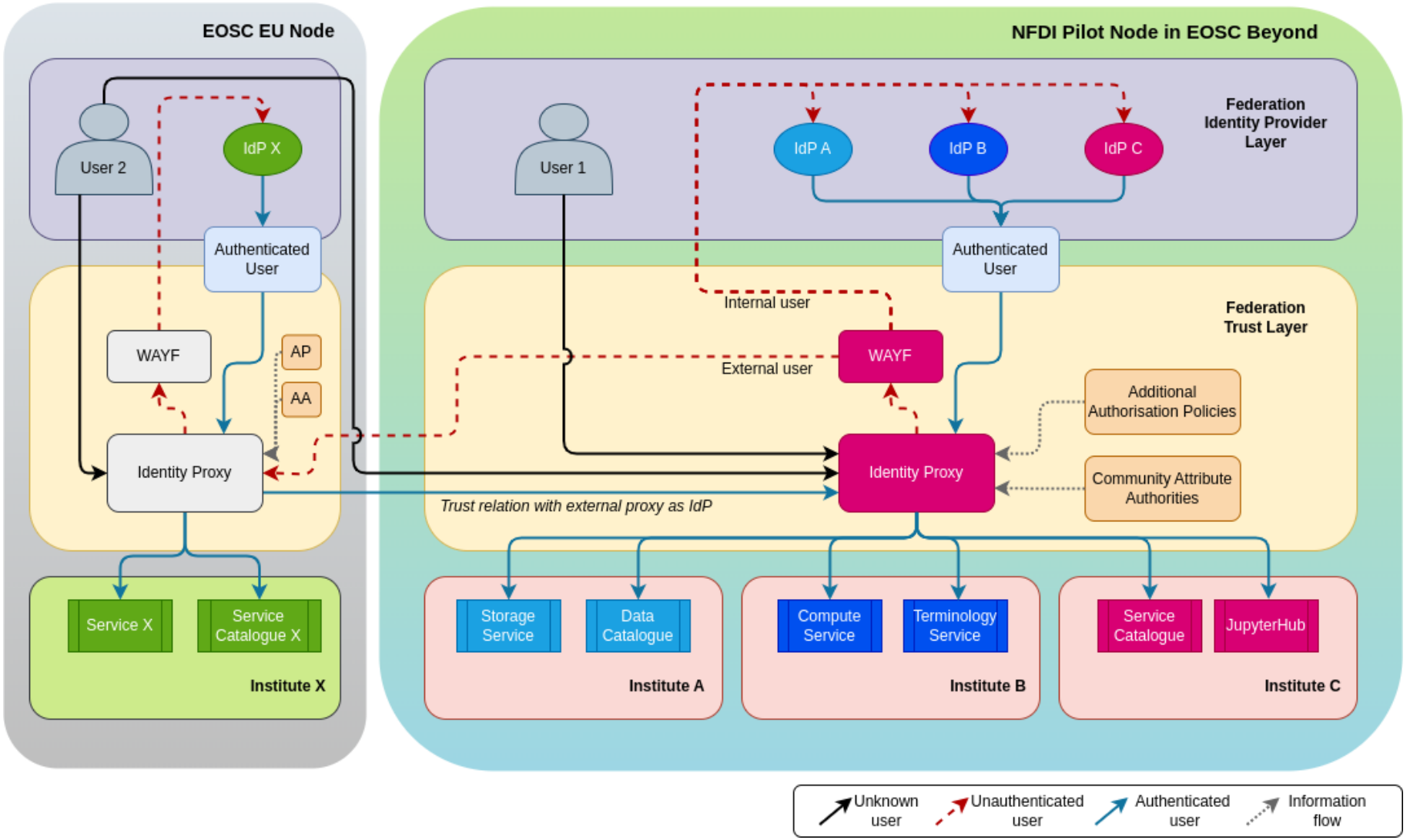
The further Landscapes

DESY, HIFIS, NFDI & EOSC



The Federations

Interfacing with other European communities and services - WIP



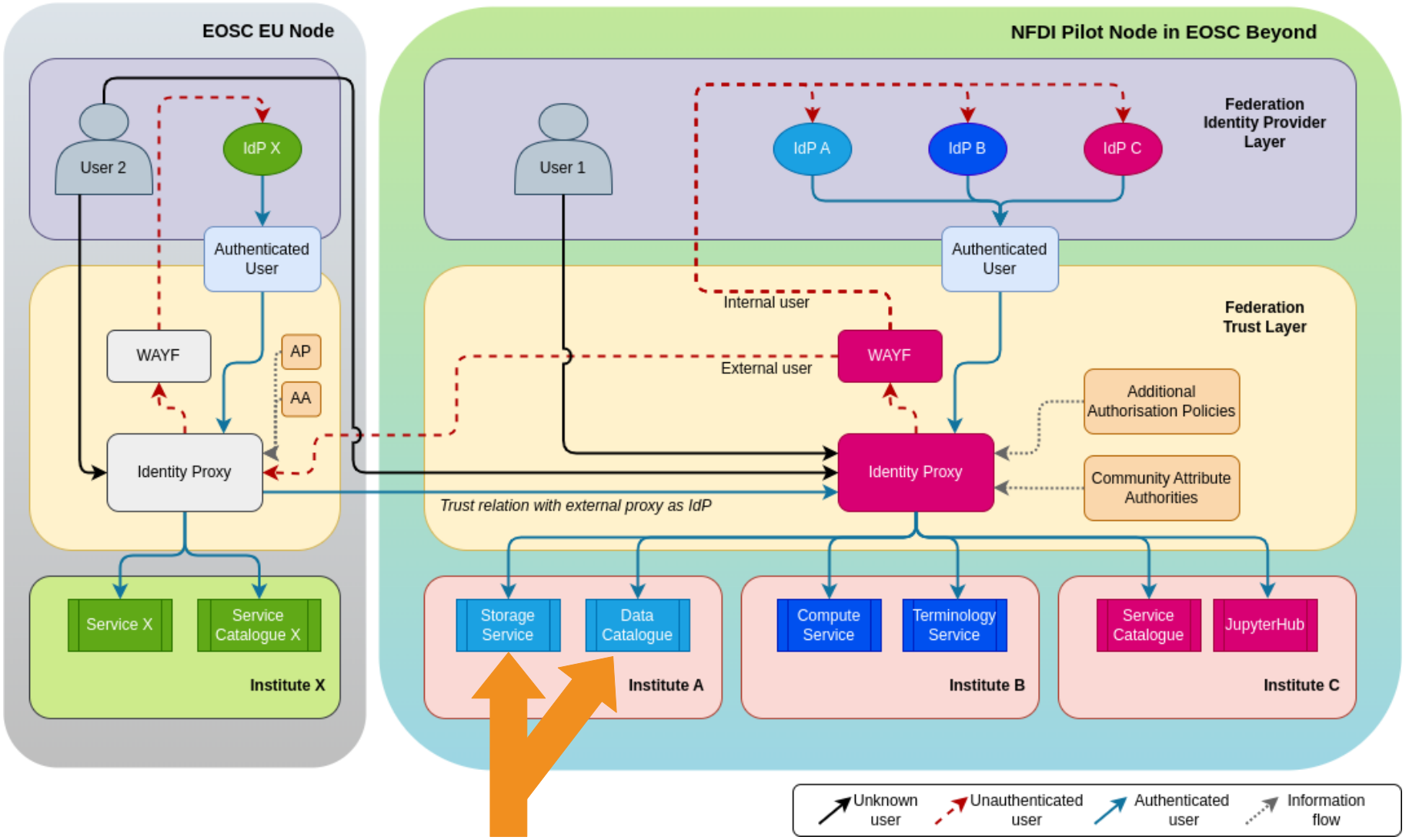
The European Open Science Cloud is currently being set up as a network of community-provided platforms ("Nodes")

Each node brings their own AAI and services behind it

The nodes are designed to be interoperable in order to bring synergies and interconnections for a more efficient scientific landscape

The Federations

Interfacing with other European communities and services - WIP



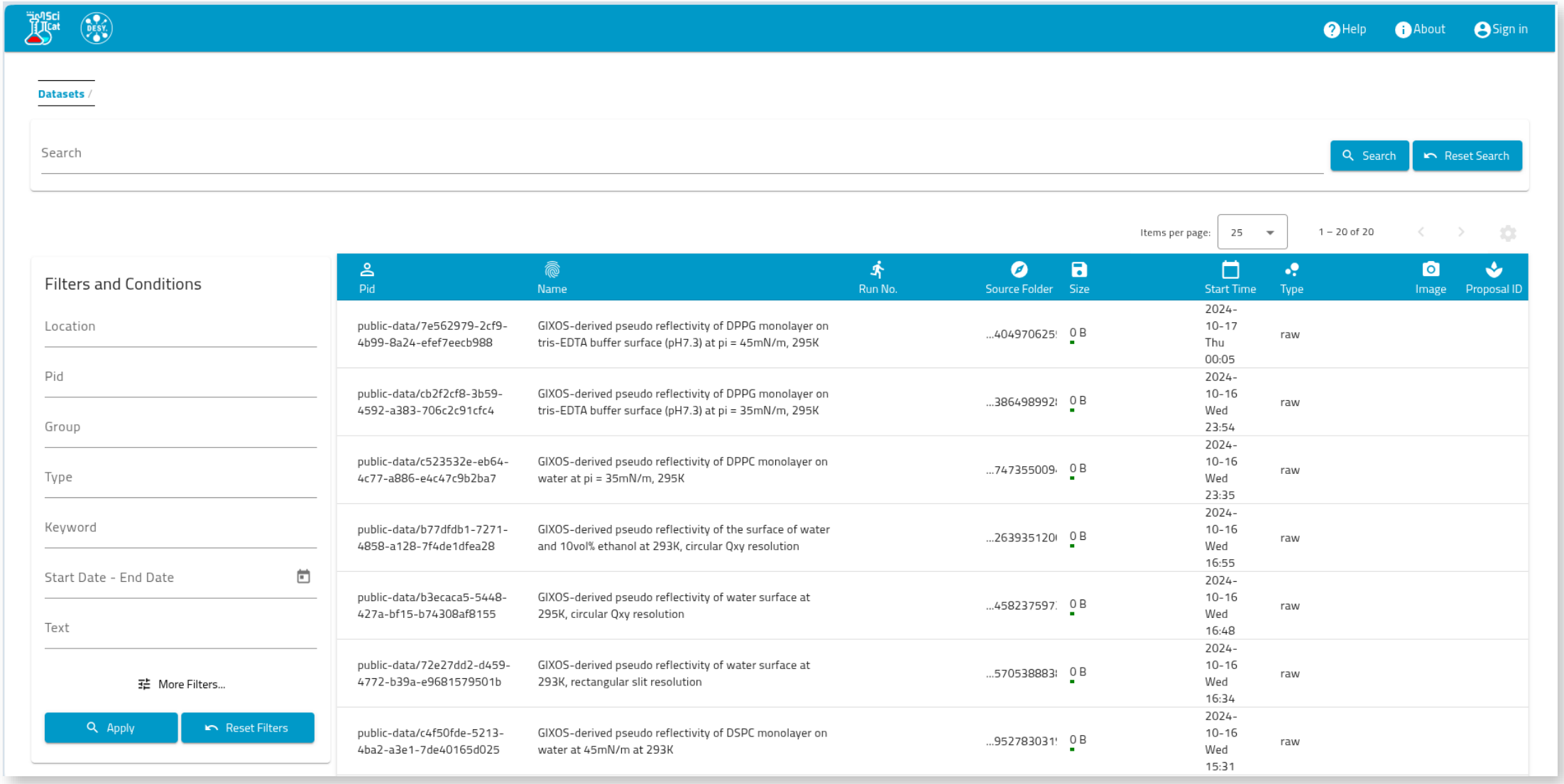
Mutual trust for authentication and authorization between communities allows for cross-community usage of

- Identities
- Services
- Data sources
- Infrastructure
- …

In a modular, unified and transparent way to make it useful and easy for the scientists

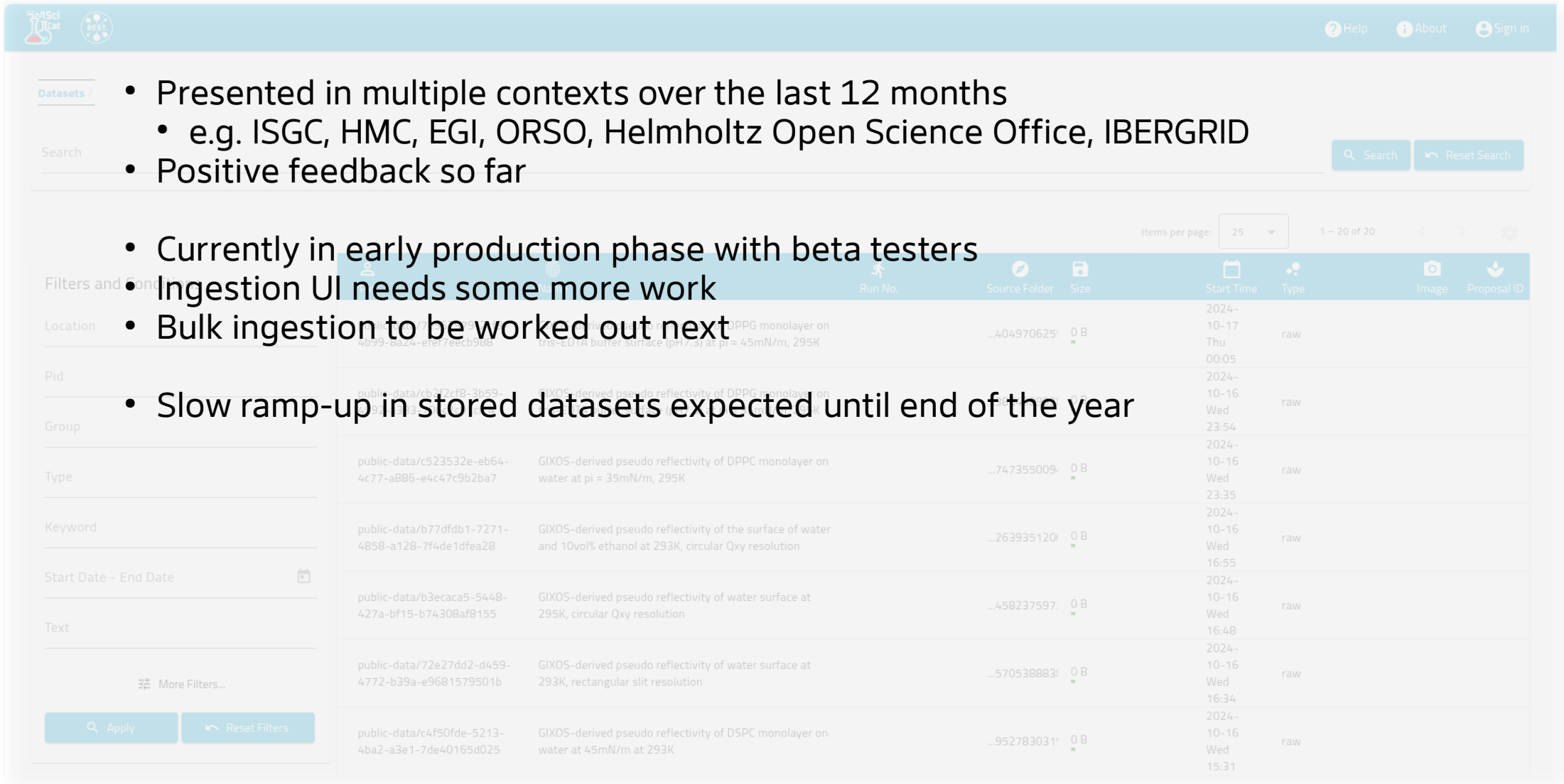
The Look

What does the catalogue look like?



The Outook

Where do we go from here?

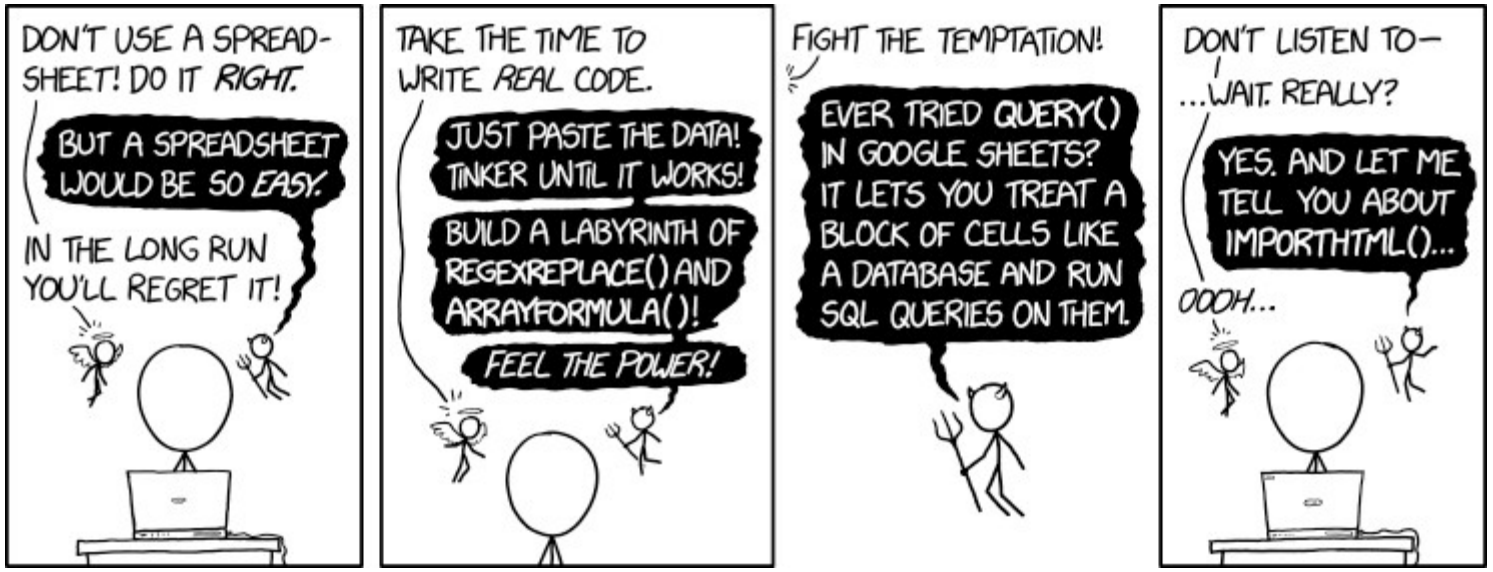


- Presented in multiple contexts over the last 12 months
  - e.g. ISGC, HMC, EGI, ORSO, Helmholtz Open Science Office, IBERGRID
- Positive feedback so far

- Currently in early production phase with beta testers
- Ingestion UI needs some more work
- Bulk ingestion to be worked out next

- Slow ramp-up in stored datasets expected until end of the year

Thank you



xkcd.com/2180

Patrick Fuhrmann, Tim Wetzel

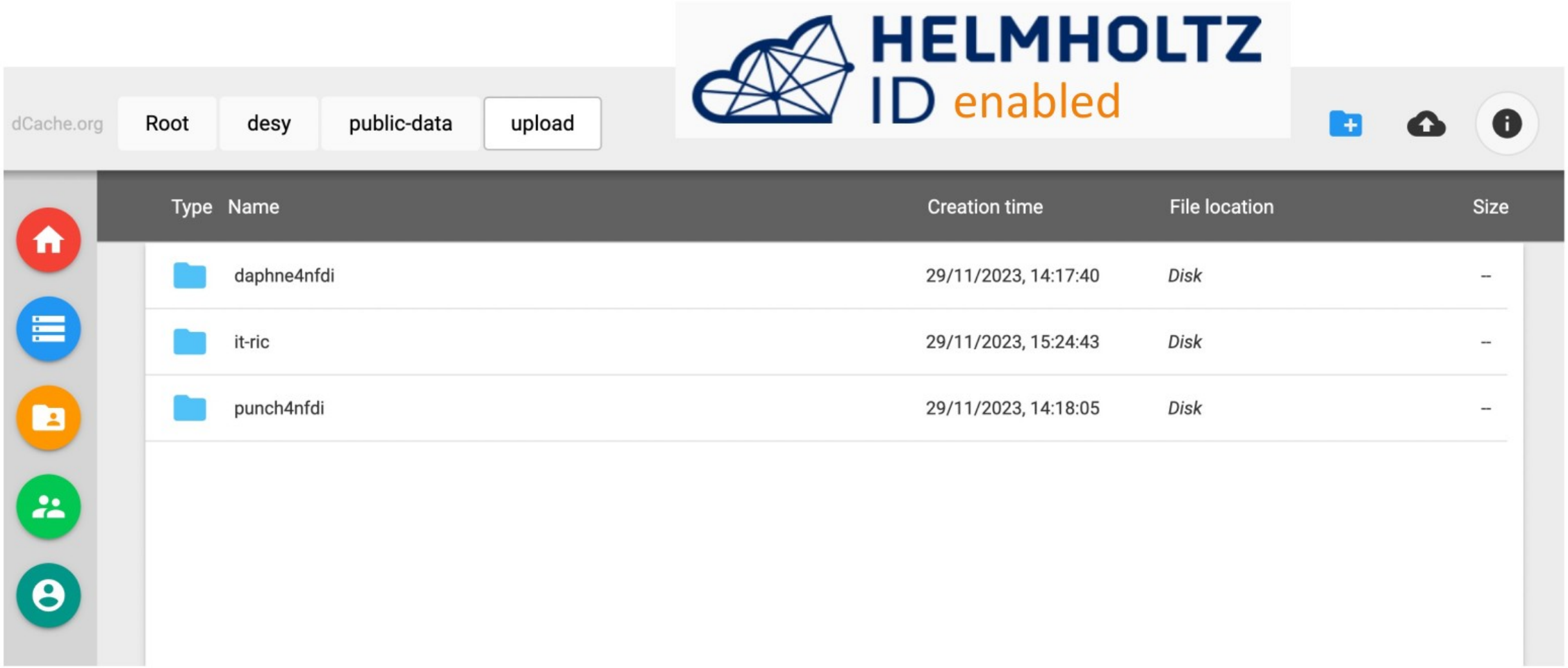
DESY IT

Patrick.Fuhrmann@desy.de, Tim.Wetzel@desy.de

Additional slides

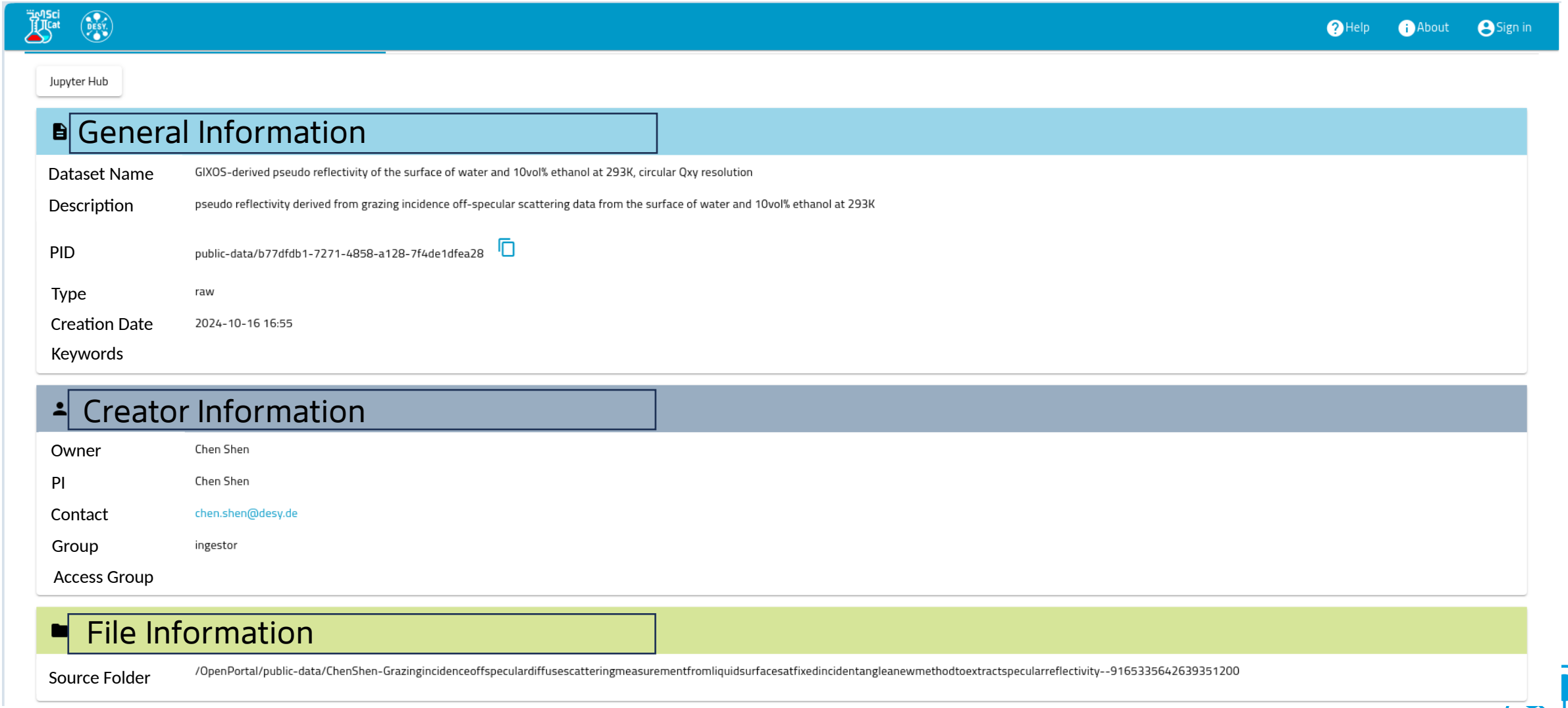
The Files

hifis-storage.desy.de



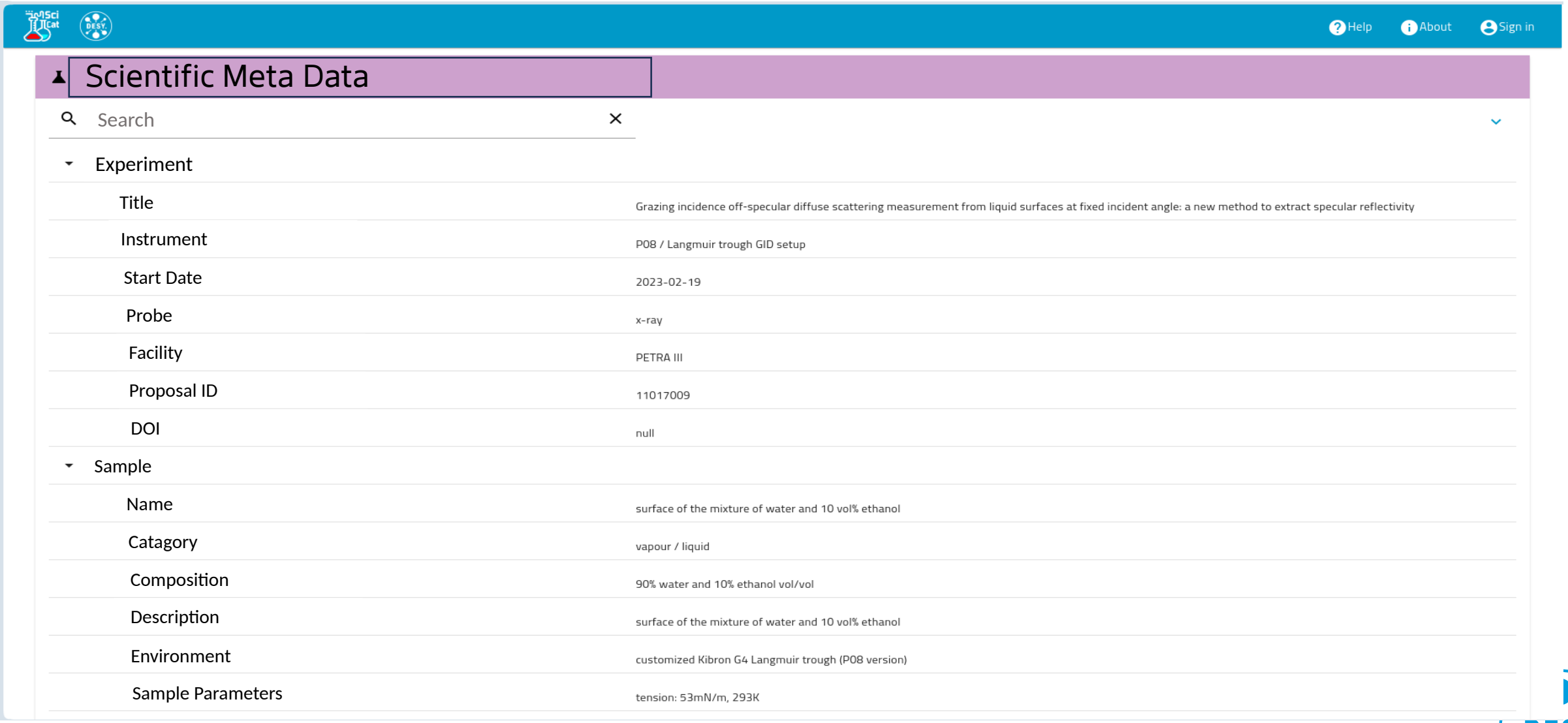
The Catalogue

public-data.desy.de



The Catalogue

public-data.desy.de



The Published Data

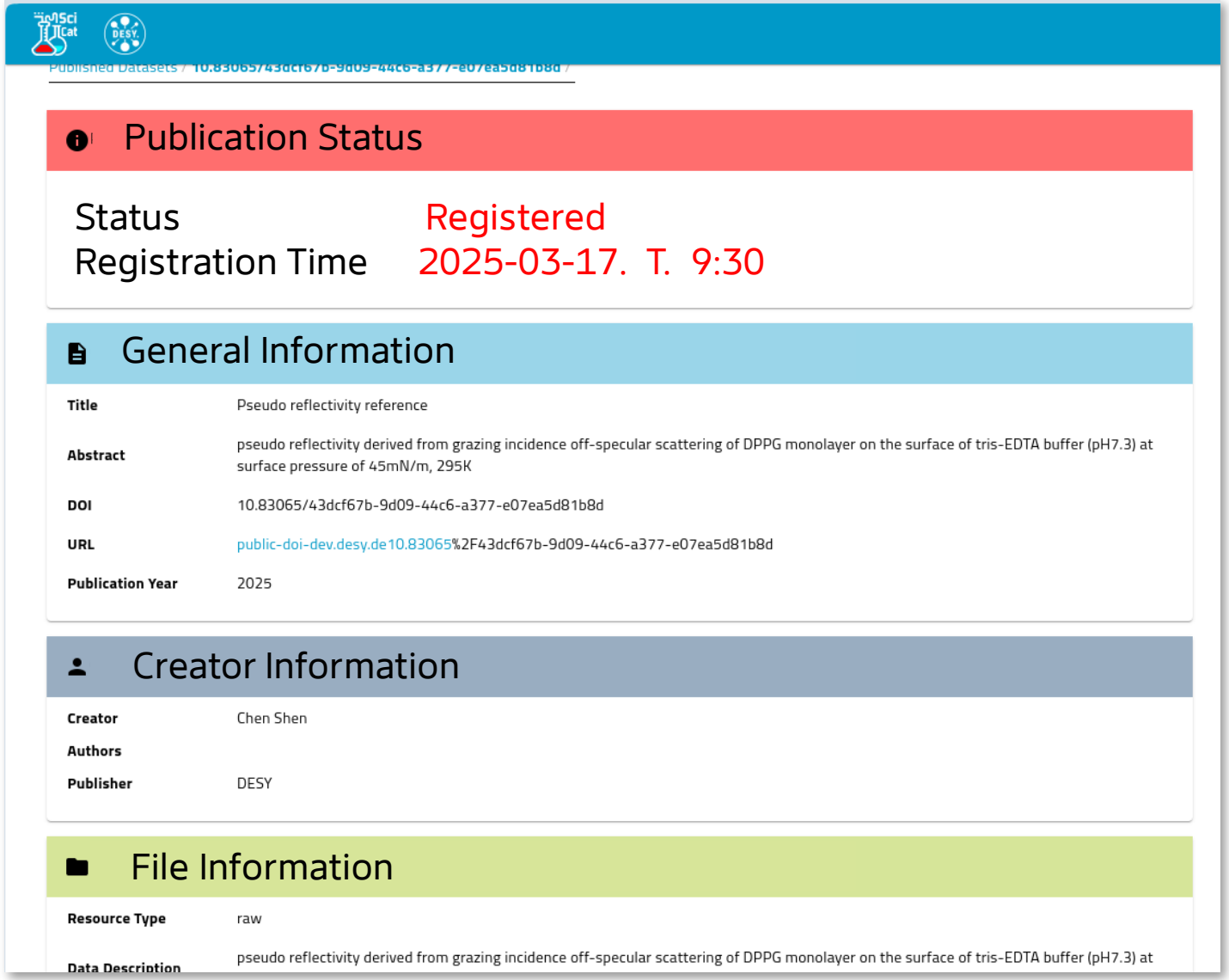
public-data.desy.de

The Minting

public-data.desy.de

The Minting

public-data.desy.de



Published Datasets / 10.83065/43dcf67b-9d09-44c6-a377-e07ea5d81b8d /

Publication Status

Status	Registered
Registration Time	2025-03-17. T. 9:30

General Information

Title	Pseudo reflectivity reference
Abstract	pseudo reflectivity derived from grazing incidence off-specular scattering of DPPG monolayer on the surface of tris-EDTA buffer (pH7.3) at surface pressure of 45mN/m, 295K
DOI	10.83065/43dcf67b-9d09-44c6-a377-e07ea5d81b8d
URL	public-doi-dev.desy.de10.83065%2F43dcf67b-9d09-44c6-a377-e07ea5d81b8d
Publication Year	2025

Creator Information

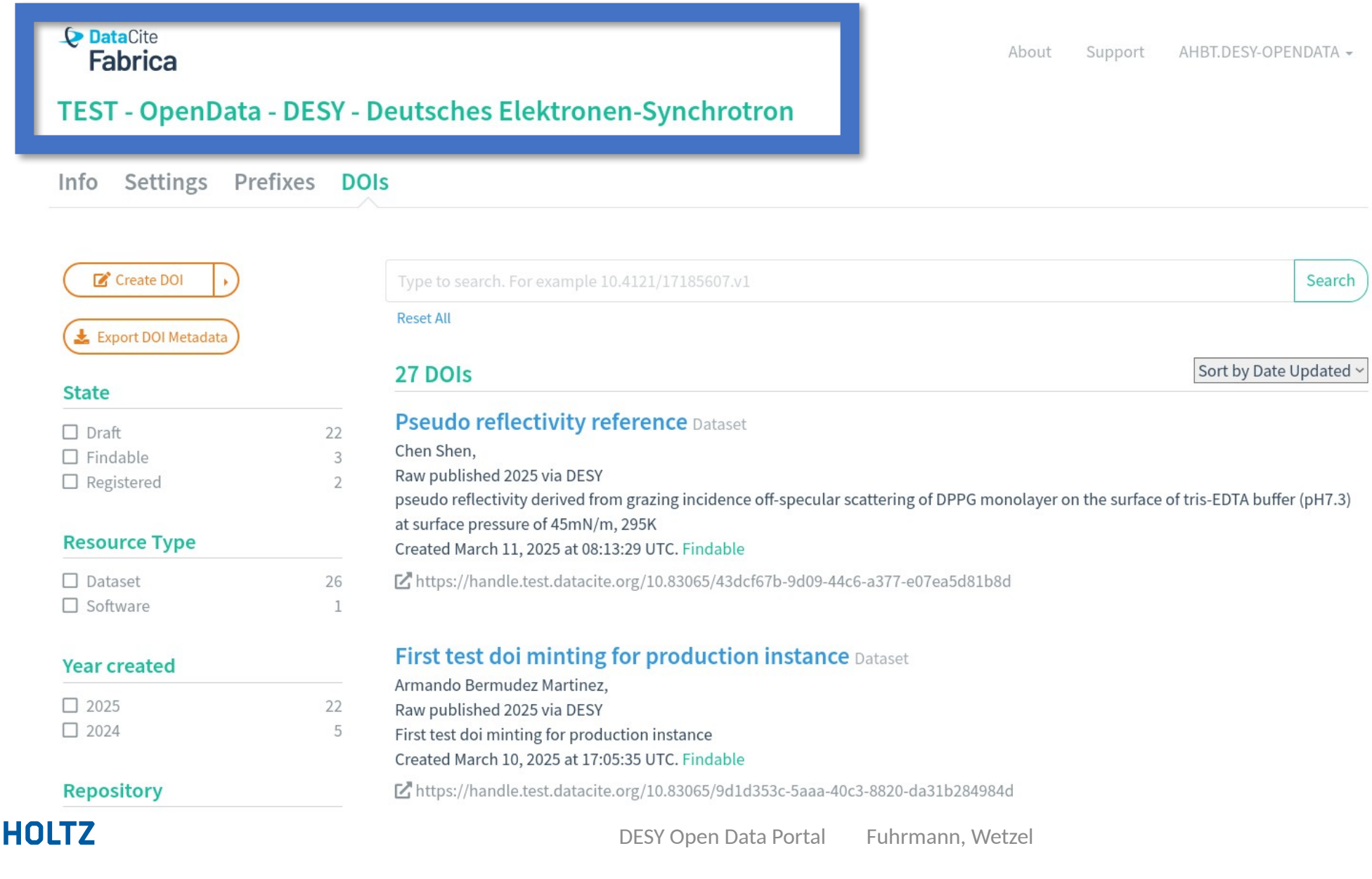
Creator	Chen Shen
Authors	
Publisher	DESY

File Information

Resource Type	raw
Data Description	pseudo reflectivity derived from grazing incidence off-specular scattering of DPPG monolayer on the surface of tris-EDTA buffer (pH7.3) at

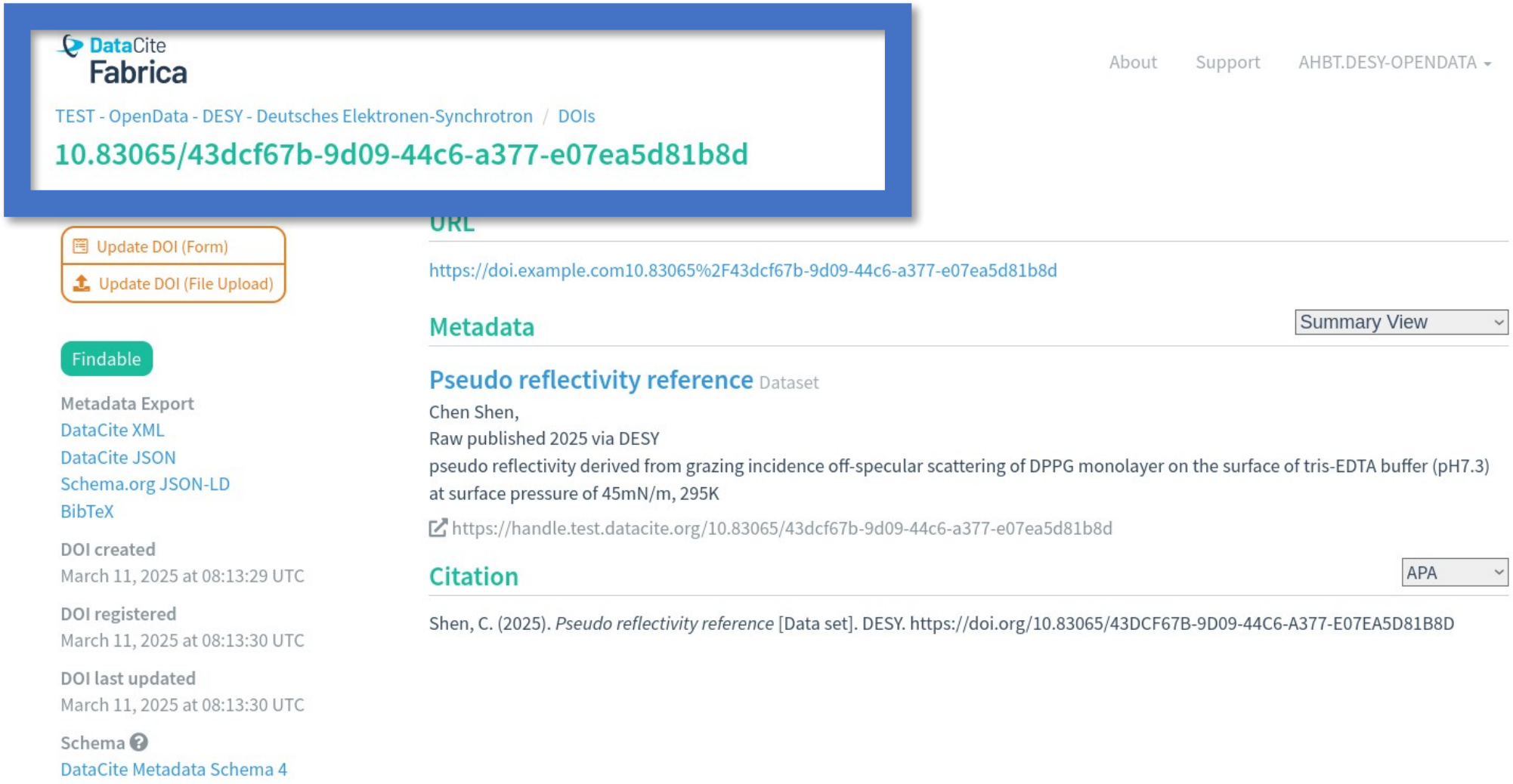
The Fabrica

DataCite Fabrica



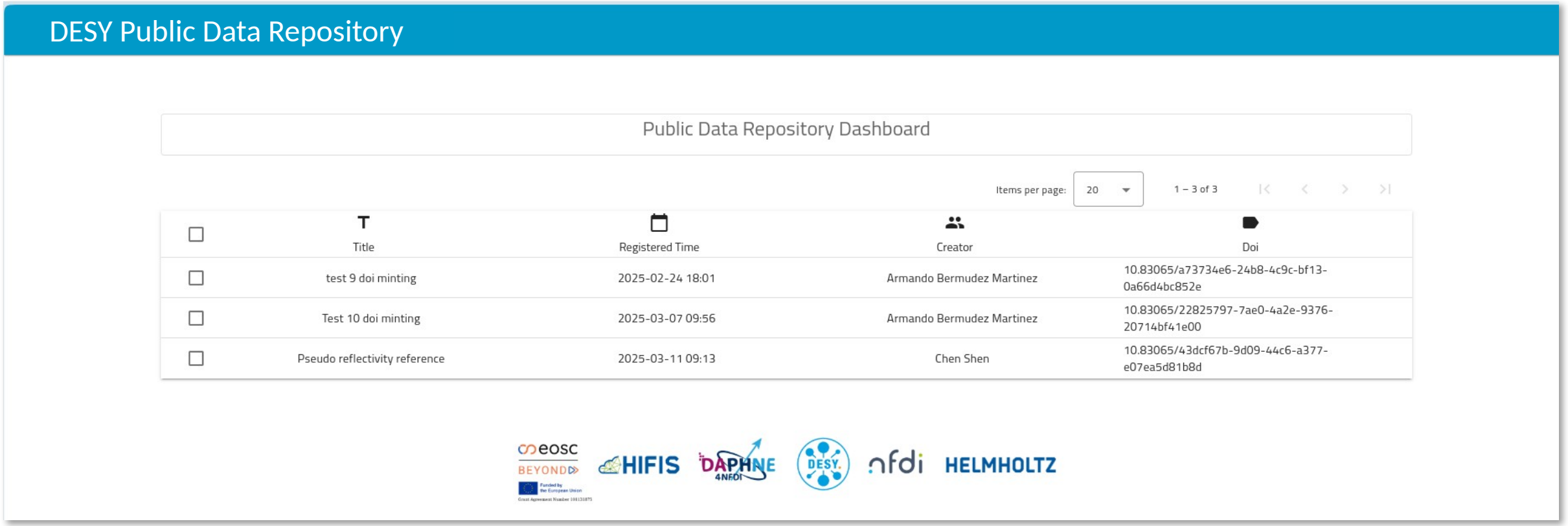
The Catch

DataCite Fabrica



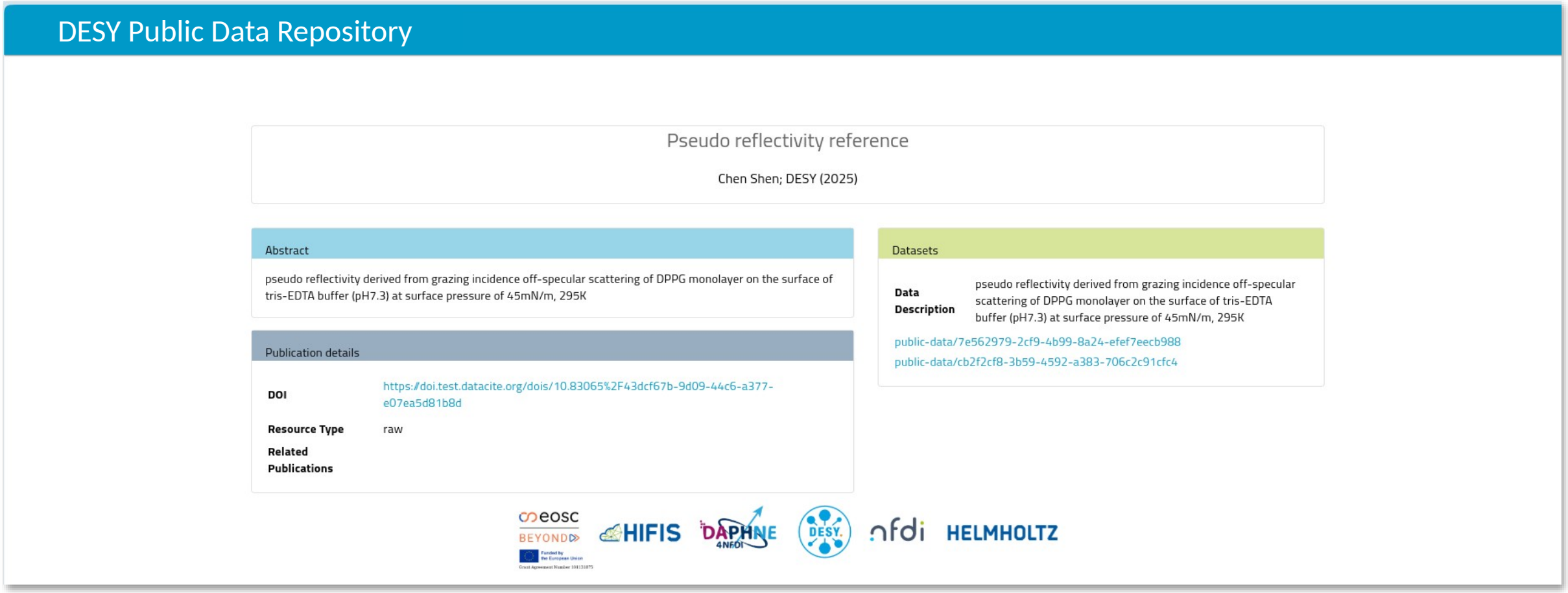
The Landing page

public-doi.desy.de



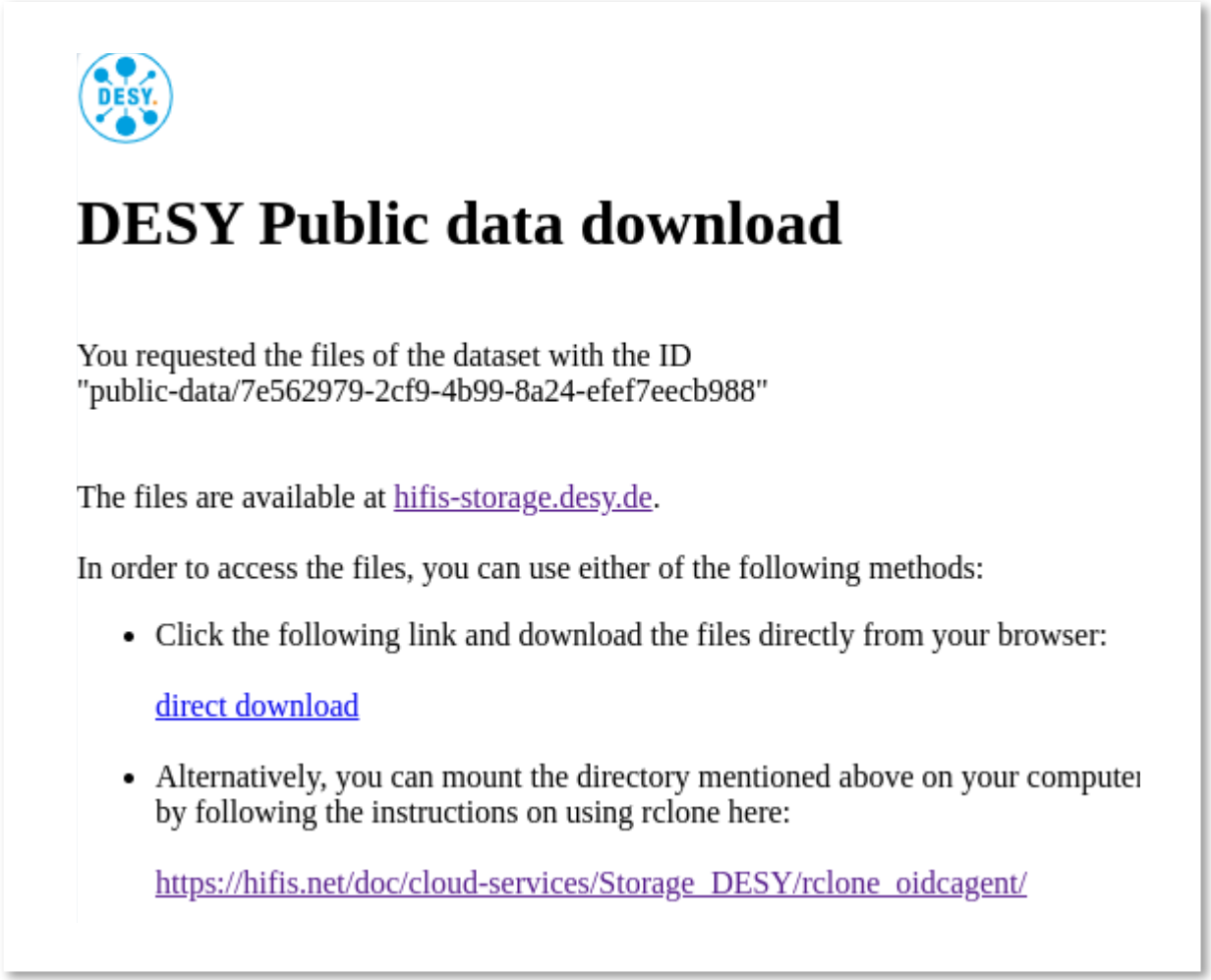
The Landing page

public-doi.desy.de



The Download

public-data.desy.de



DESY Public data download

You requested the files of the dataset with the ID
"public-data/7e562979-2cf9-4b99-8a24-efef7eecb988"

The files are available at hifis-storage.desy.de.

In order to access the files, you can use either of the following methods:

- Click the following link and download the files directly from your browser:

  direct download

- Alternatively, you can mount the directory mentioned above on your computer by following the instructions on using rclone here:

  https://hifis.net/doc/cloud-services/Storage_DESY/rclone_oidcagent/

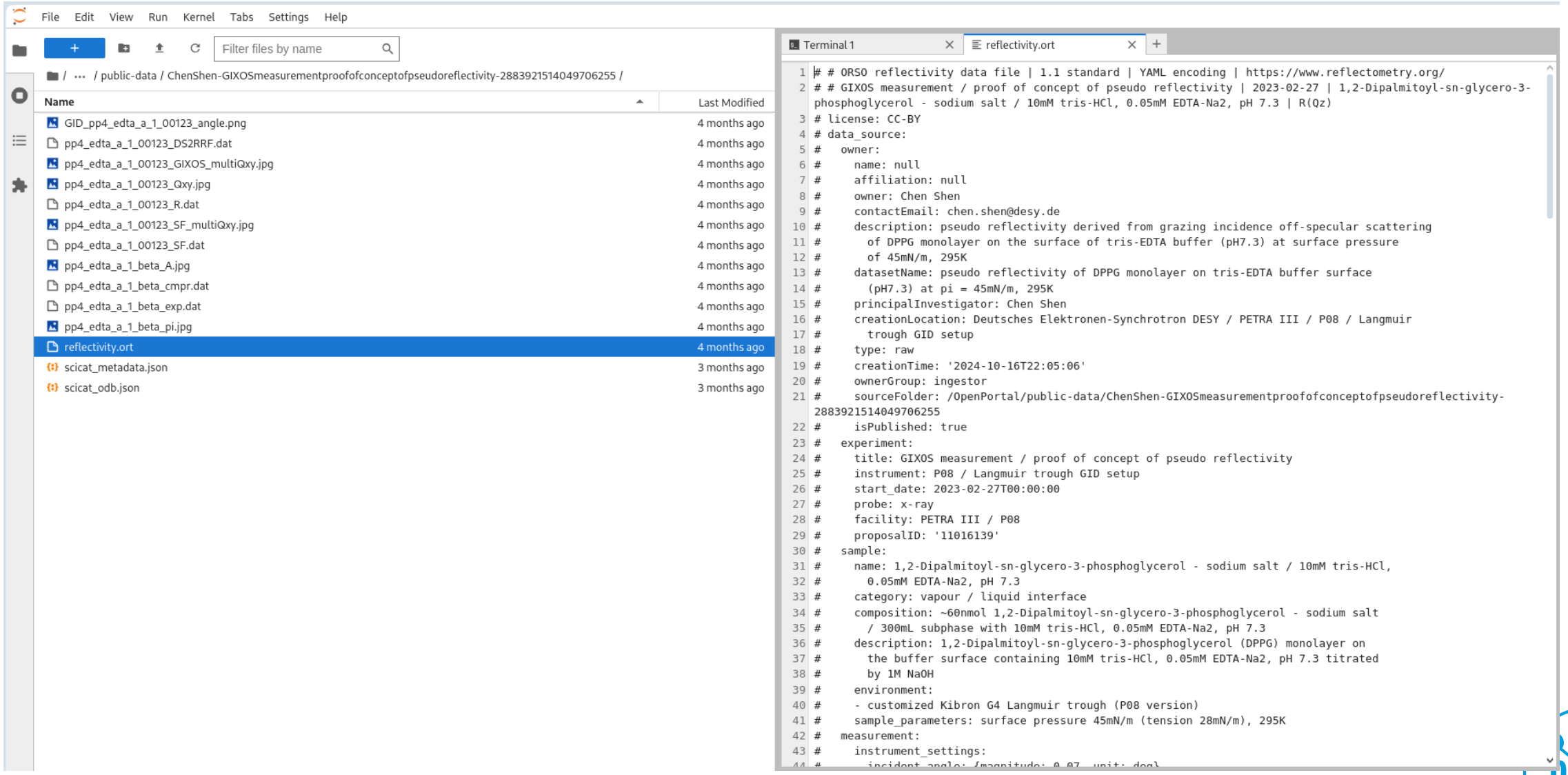
The Storage

public-data.desy.de

dCache

/ OpenPortal / public-data / ChenShen-GIXOSmeasurementproofofconceptofpseudoreflectivity-2883921514049706255

Name	Size	Last Modified
GID_pp4_edta_a_1_00123_angle.png	338512	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_00123_DS2RRF.dat	2169	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_00123_GIXOS_multiQxy.jpg	145231	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_00123_Qxy.jpg	97479	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_00123_R.dat	4163	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_00123_SF.dat	5034	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_00123_SF_multiQxy.jpg	147746	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_beta_A.jpg	430633	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_beta_cmpr.dat	26623	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_beta_exp.dat	25604	Thu Oct 17 11:34:44 CEST 2024
pp4_edta_a_1_beta_pi.jpg	479965	Thu Oct 17 11:34:44 CEST 2024
reflectivity.ort	12119	Thu Oct 17 11:34:44 CEST 2024
scicat_metadata.json	4880	Wed Nov 27 12:27:31 CET 2024
scicat_odb.json	3063	Fri Nov 22 09:13:50 CET 2024

The Exploration

jupyter.desy.de