Scopes of SciCat at DESY

Towards FAIR, useful and open data

Regina Hinzmann
22. Jan. 2024

HELMHOLTZ

DESY.

Scientific Metadata Management at DESY

SciCat in operation

01 General Strategy for SciCat @Beamlines

- Reasons to use SciCat
- Common language: meta data standards

02 FS/IT activities for public data access

- Workflow for DOI minting, status, next steps
- Activities around public-data.desy.de

03 What's next?!

SciCat activities at DESY

Different groups with the same goal: make SciCat useful.

Deployment of several SciCat instances on basis of OpenStack and Kubernetes infrastructure, supporting multiple concurrent R&D investigations:

1. Test instances at the beamline: **demonstrator beamlines**

- **P08, a Petra III beamline,**
- **A FLASH beamline.**
- More are worked on, FS-EC aims at **a setup common for all beamlines.**

for internal purposes

2. FS test instances for more general tasks

- **DOI minting**: Different DESY groups (IT, FS and L- library) **collaborated to agreed** on a service to gain experience in the workflow
- **Public data** at DESY (RIC): a **pilot service** is being set up that
  - combines DESY and Helmholtz products (dCache HIFIS storage) with SciCat.

for the outside world

Single SciCat instances, internal use.

Demonstrator beamlines

Towards a useful SciCat

Useful for

- Beamline staff/„user“ (=scientist/physicist):
  - What are the key words you search for? Would you search for *undulator gap* ? What else? Would you analyse data to reproduce experiment conditions?
    - *try to limit the number of entries to less 10 in* **`dataset.scientificMetada.xyz`**
  - Common language: Do we mean the same thing if we use *instruments* or *energy*?
  - I learned SciCat has many pre-defined fields, for free-text entries and data quality assessment. Eg a dataset has attributes like:
    - **`dataset.comment`**
    - **`dataset.dataQualityMetrics`**
    - **`dataset.description`**

DOI minting and open data

Different groups with the same goal: make SciCat useful.

FS/IT activities for tasks:

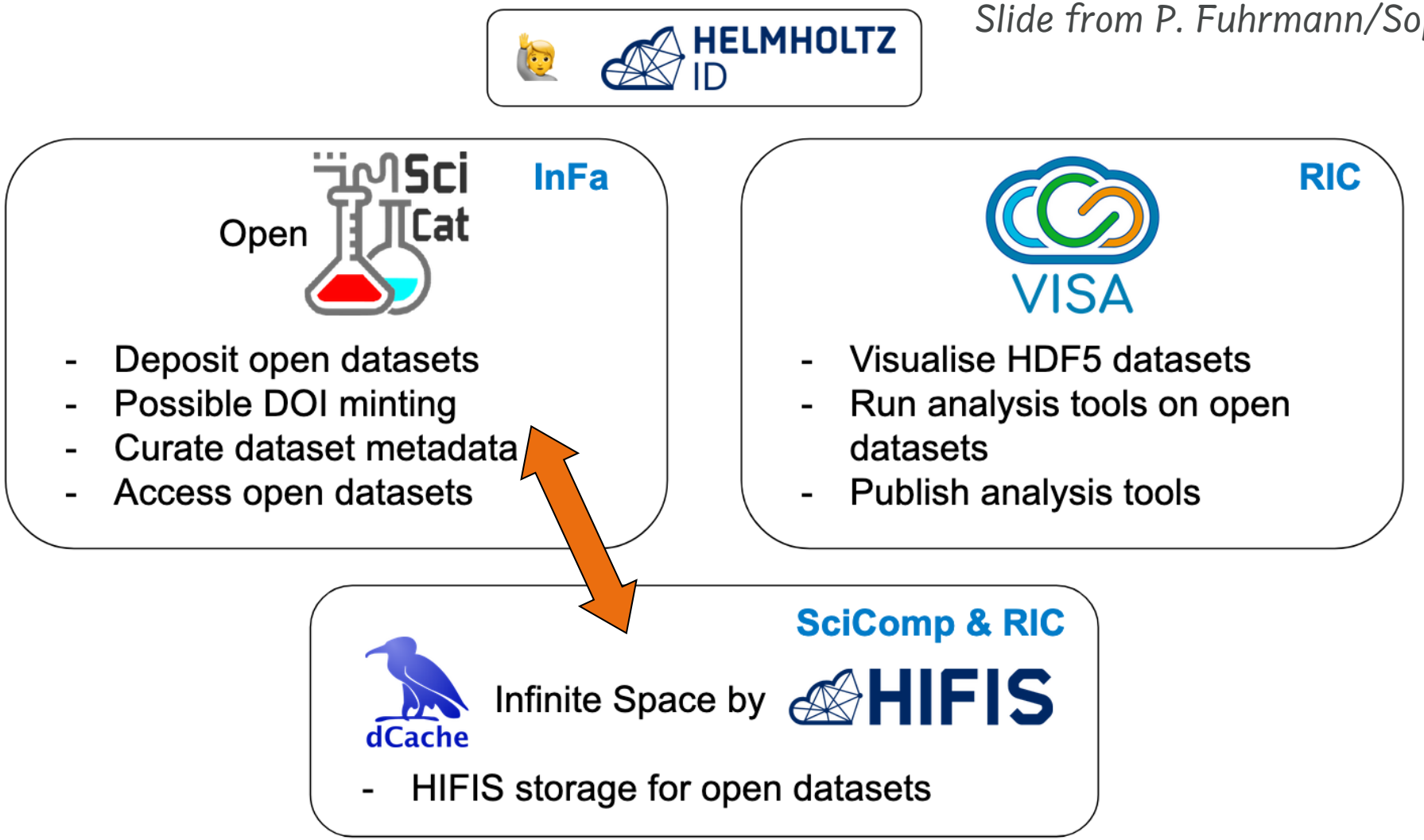
1. **DOIs provision for DESY data**: Different DESY groups (IT, FS and L- library) **agreed** at a first step on how a DOI can be provided:
   - manual process, user awareness
   - granularity of a datasets
   - access to *raw* datasets first, what is the status?

2. **Public data** at DESY (RIC): a **pilot service** is being set up
   - combines DESY and Helmholtz products (dCache HIFIS storage) with SciCat.

FAIR data and open science at DESY

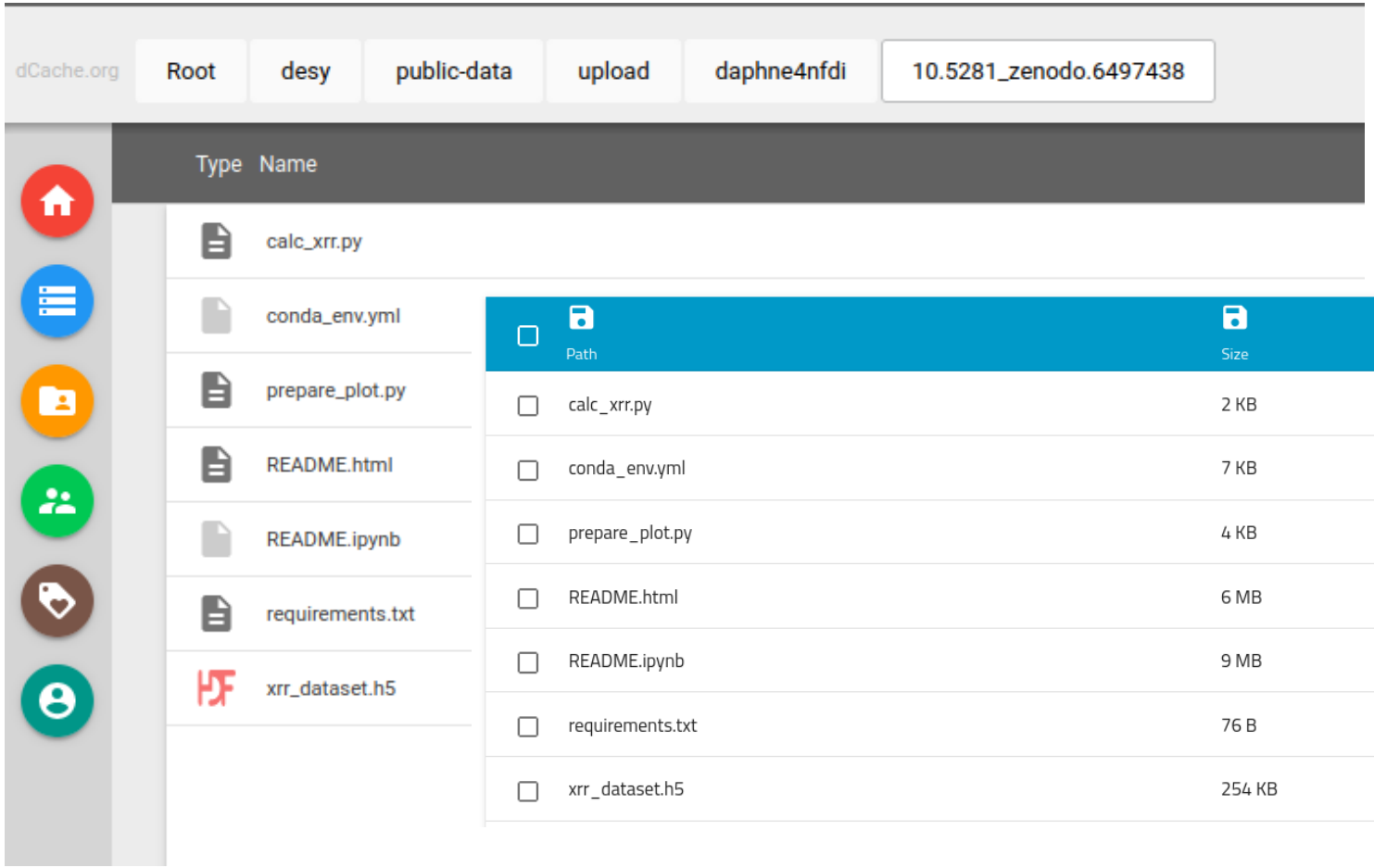
Composing existing components into a future repository service

Slide from P. Fuhrmann/Sophie Servan (IT-RIC)



Pilot Service

hifis-storage.desy.de/



Thanks to RIC group, Tim Wetzel!

public-data.desy.de/

Path	Size	Time
calc_xrr.py	2 KB	2024-01-16 14:11
conda_env.yml	7 KB	2024-01-16 14:11
prepare_plot.py	4 KB	2024-01-16 14:11
README.html	6 MB	2024-01-16 14:11
README.ipynb	9 MB	2024-01-16 14:11
requirements.txt	76 B	2024-01-16 14:11
xrr_dataset.h5	254 KB	2024-01-16 14:11

Pilot Service: public-data.desy.de

General Information

Name	Reflectometry curves (XRR and NR) and corresponding fits for machine learning
Description	This is a compiled dataset of raw X-ray reflectivity (XRR, reflectometry) measurements together with corresponding fit parameters, intentionally published to use as training or test data for machine learning models. (The authors aim to include NR data in further versions of this dataset and plan to include other substrates and materials for XRR. Contributions welcome!)
PID	undefined/c5fabbc4-852b-4304-8c52-194af5cf94c5
Type	raw
Creation Time	2024-01-16 14:09
Keywords	

Creator Information

Owner	Linus Pithan
Principal Investigator	linus.pithan@desy.de
Contact Email	linus.pithan@desy.de
Owner Group	fsec
Access Groups	

File Information

Source Folder	/desy/public-data/upload/daphne4nfdi/10.5281_z

Scientific Metadata

View | Edit

Search

- DIP_1
- DIP_2
- DIP_3
- DNTT_PDIF_1to2

Experimentalists	Rußegger, Nadine; Greco, Alessandro
Layer_material	DNTT PDIF (1:2)
instrument	DESY, P08
q_max_fit	0.15 (1/Ang)
year_experiment	2021

- DNTT_PDIF_2to1
- PDIC5

Are you still with me?

Many people work on various important issues.

- Please report and collaborate. Thank you.
- FS + IT: Find a strategy to set up a stable, reliable catalogue for all beamlines at DESY.
- Gain experience harvesting meta data and provide urgent “features” for DESY data.

Thank you