

SciCat @ DESY

Where are we today and where are we going?

Regina Hinzmann
2024-12-09

2023 and 2024 summary activities

Let's hold on for a moment and look back

Stakeholders: definition by end of 2023

1. *All DESY beamlines: in-house research, DESY –user/scientists (e.g. universities), external users, industries.*
 - *FS-EC; FS-SC; PETRAIII, FLASH, RockIT Project, getting ready for PETRA IV*
 - *DESY-IT: RIC public data portal (HIFIS service, DAPHNE4NFDI, EOSC beyond); IT-InFa: setting up SciCat as service, joined forces with Benutzerservice and IT-SC.*
2. *DAPHNE4NFDI Project goals: FAIR data as “DESY does NOT want to stay on an island.”*
3. *Other German photon and neutron communities.*
4. *New since **2024**: SciCat Collaboration/Developers: contribution and sustainability of code development.*

System Requirements

What I expect from SciCat, what I want to use SciCat for!

- ❑ *Find my data.*
 - ❑ *Find my data: browse meta data from anywhere.*
 - ❑ *Find my data and show in a nice way.*
 - ❑ *Find my data, show in a nice way and understand it in a year's time after having had several similar beamtimes.*
 - ❑ *Find my data to optimize current beamtime.*
 - ❑ *Find my data to optimize my next beamtime.*
 - ❑ *Find my data to do first analysis.*
 - ❑ *Get a DOI.*
 - ❑ *Get a DOI and overcome Zenodo's limitations.*
 - ❑ *Get a DOI for data produced at DESY.*
 - ❑ *Get a DOI for my publication.*
- ❑ Finding data means having standardised, validated terms despite free format of scientific meta data
 - *PaNET integration, TA1 task (def.), IT and FS realise.*
 - ❑ Finding data means **having a nice frontend in place that enables a complex search.** e.g. made of a combination of queries of site specific terms.
 - *TA2 task, FS/IT collaborates.*
 - ❑ Finding data means **to receive a list first and not to be swamped by unwanted information (details at front)**
 - *TA2 task, IT and FS realise.*
 - ❑ DOI workflow initiated. *L, FS-EC and IT-RIC and IT-InFa.*

Progress

Step by step

- ✓ Proposal hierarchy now being implemented: Can define proposal types just as it was suggested in solution 2 of Max answer

- ✓ Parent proposal, e.g. DOOR proposal ID
- ✓ Child proposal (1st proposal): beamtime ID
- ✓ Grandchildren proposal (2nd proposal): any potential sub beamtime ID

General Information

Title

First hierarchical proposal

Abstract

This is the first hierarchical proposal

Identifier

60d3e77d080bb4817008cf38c56f52f4

Type

Default Proposal

Parent proposal

Hierarchical parent proposal

Creator Information

Main proposer

Fredrik Bolmsten

Principal investigator

Massimiliano Novelli

Metadata

View

Edit

Name

Value

Unit

Proposal I D

Title

First Name

Last Name

Start Date

End Date

contains

hierarchical

contains

contains

Start date – End date

Start date – End date

60d3e77d080bb4817008cf38c56f52f4

First hierarchical proposal

Fredrik

Bolmsten

Nov 20, 2024

Nov 20, 2025

121d427f84bd7bebe3809c391b983369

Second hierarchical proposal

Fredrik

Bolmsten

Nov 20, 2024

Nov 20, 2025

79e1b758d2f9fda13f2b1a4dc0bc0cf7

Hierarchical parent proposal

Fredrik

Bolmsten

Nov 20, 2024

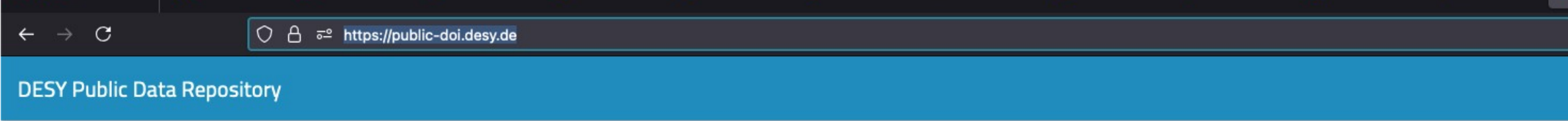
Nov 20, 2025

✗ missing tab that indicates relation of proposals, will be covered soon if not already done.

Progress

Step by step

✓ First prototyping of DOI minting: <https://public-doi.desy.de/>



Public Data Repository Dashboard

click: forward to Landing Page of DataCite

Items per page: 20 1 – 3 of 3

Title	Registered Time	Creator	Doi	
Hello		Chen Shen	10.83065/b0210f51-ac6f-45cf-8a66-5691f7b7bf38	
		Tim Wetzel	10.83065/add145a7-4543-4d41-b34c-710520505c8d	
		AArmando Bermudez Martinez	10.83065/94cdc0b3-913c-49d7-aed7-7cede52e43a1	

Hello

Chen Shen; DESY (2024)

Abstract

Hello abstract

Publication details

DOI <https://doi.test.datacite.org/doi/10.83065%2Fb0210f51-ac6f-45cf-8a66-5691f7b7bf38>

Resource Type raw

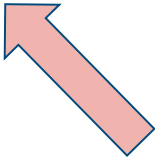
Related Publications

Datasets

Data Description 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

[public-data/7e562979-2cf9-4b99-8a24-efef7eecb988](#)

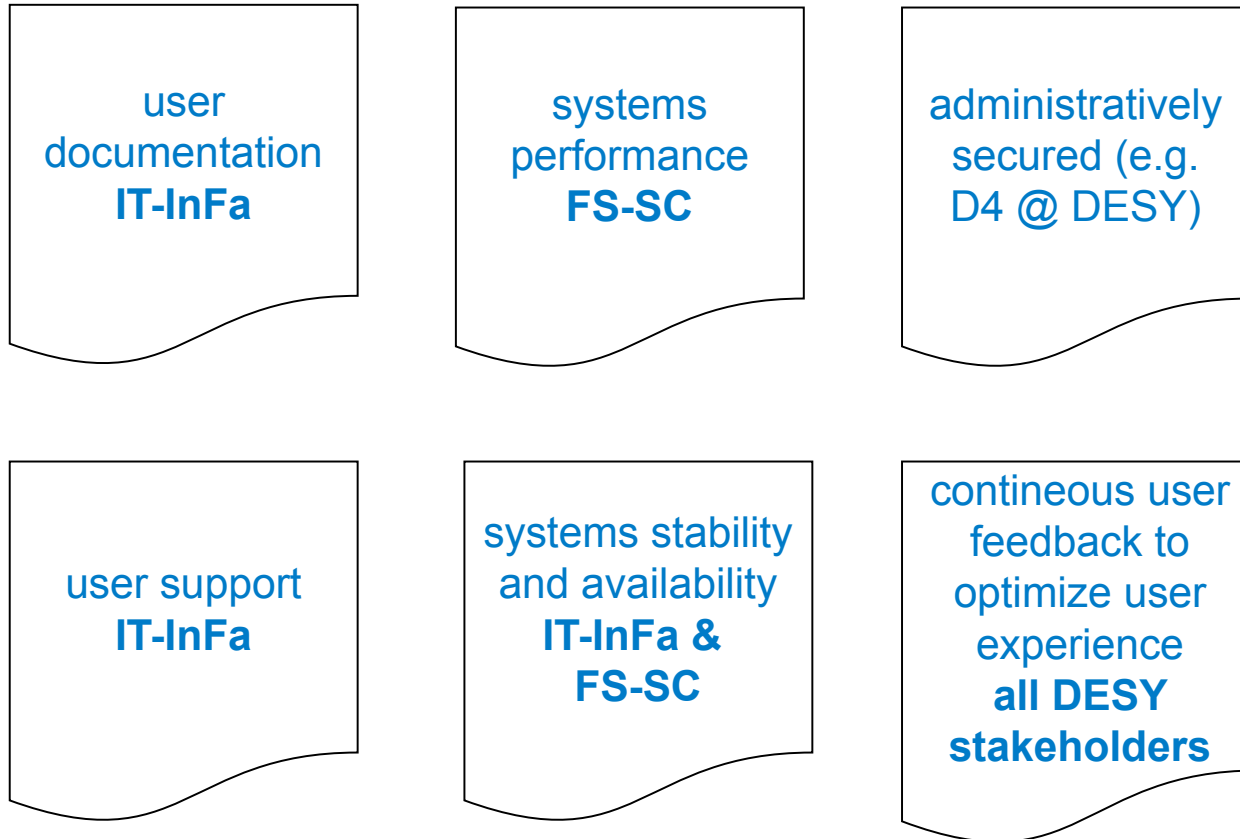
[public-data/cb2f2cf8-3b59-4592-a383-706c2c91cfc4](#)



selected list of datasets

System Requirements

What do I expect from SciCat, what do we want to use SciCat for?



Two main parts

1. Core part: meta data ingestion

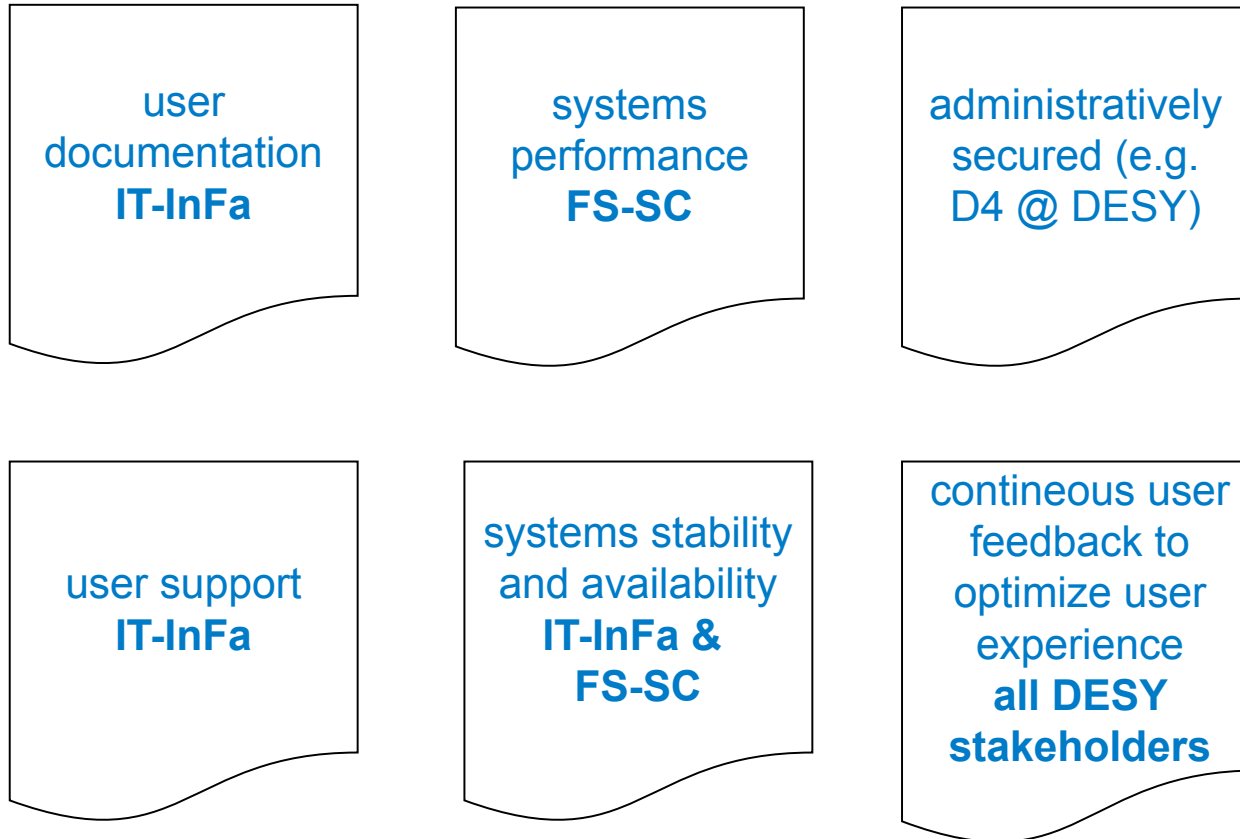
1. *Which meta data is available?*
2. *Can we store them in a way that actually is descriptive?*

2. Representing meta data to be findable and (re-)usable.

1. *At which point does the user update the meta data?*
2. *Which info is integrated from DOOR? (e.g. sample, instrument information)*

Features of SciCat Service to be implemented

Open questions to SciCat provider



2024 highlights and prospects for 2025

End of Year goals

- DOI workflow for beamlines
- 1 database for all PETRA III beamlines!
- 1 database for all DESY beamlines?

Moving away from test instances:

- get missing features in (e.g. data access: integration with other IT-systems/services)
- promote users to use SciCat e.g. through documentation, user support, use cases, regular exchange, etc.

Outlook and Conclusion

SciCat @ DESY by end of project (September 2026)

Data curative and *SciCat systems* parts. We do not wait and cannot allow us to wait for one of them to finish first, but to move forward in parallel in a continuous feedback loop. This is what we are doing (demonstrator concept) and need to be doing in future even more/better (for stable and useful service).

FS contributions:

- SciCat **Ingestion** needs to be ready in a general way: what is the roadmap of integration of several beamlines? Is the meta data descriptive?
- What meta data is taken from DOOR/which from real-time? Estimate of load (requests per time unit) to define system requirements. Stress tests. Obvious: Good communication to/with IT vital.

IT contributions:

- 2025: Design and integrate SciCat data access.
- Design performant SciCat infrastructure and get it operational: 1 db for all DESY data (several PETRA and FLASH beamlines) plus 1 db only public datasets DAPHNE wide.

With DAPHNE 2: (amongst others) **user support**.

Thanks