SciCat @ DESY

Where are we today and where are we going?

Regina Hinzmann
2024-12-09

HELMHOLTZ

DESY.

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

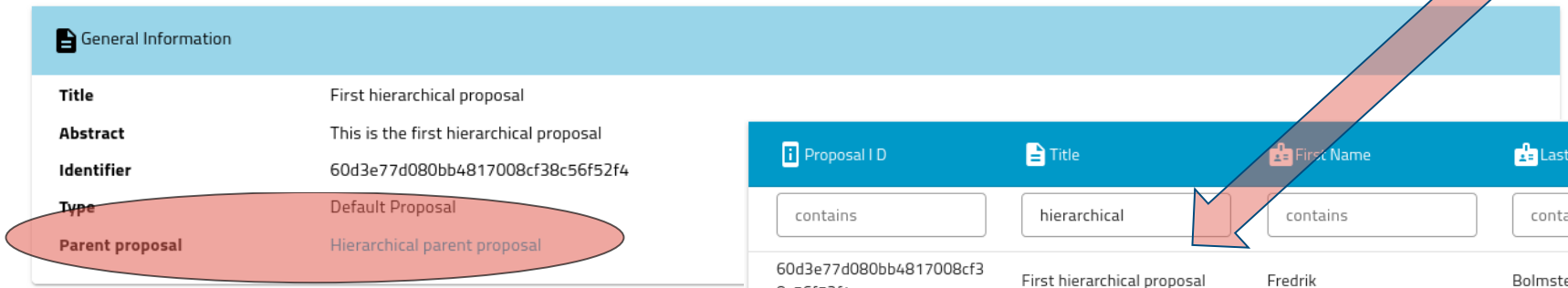
Step by step

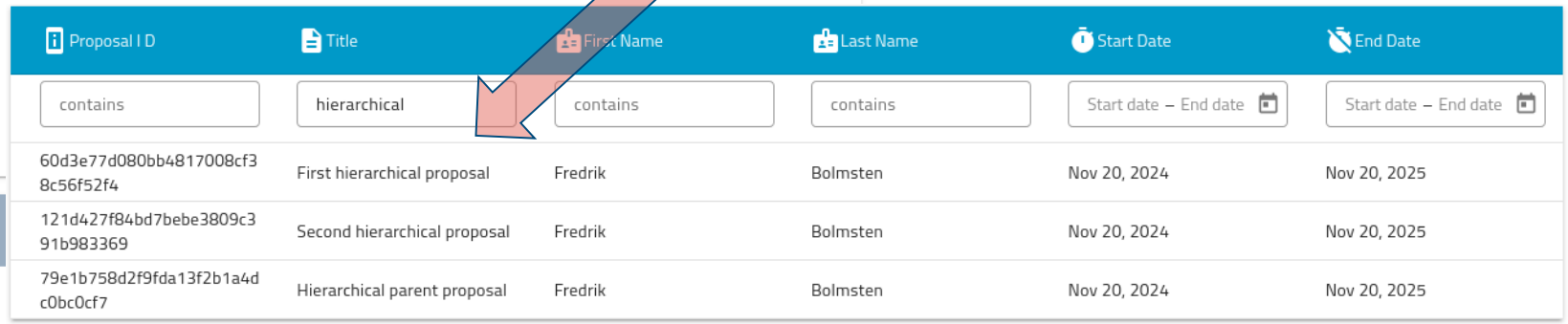
✓ *Proposal hierarchy now being implemented: Can define proposal types just as it was suggested in solution 2 of Max answer*

✅ *Parent proposal, e.g. DOOR proposal ID*

✅ *Child proposal (1st proposal): beamtime ID*

✅ *Grandchildren proposal (2nd proposal): any potential sub beamtime ID*



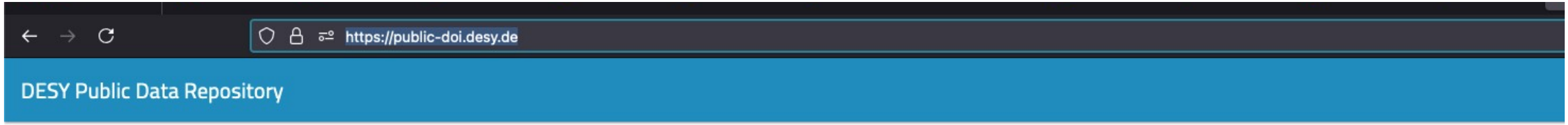


❌ *missing tab that indicates relation of proposals, will be covered soon if not already done.*

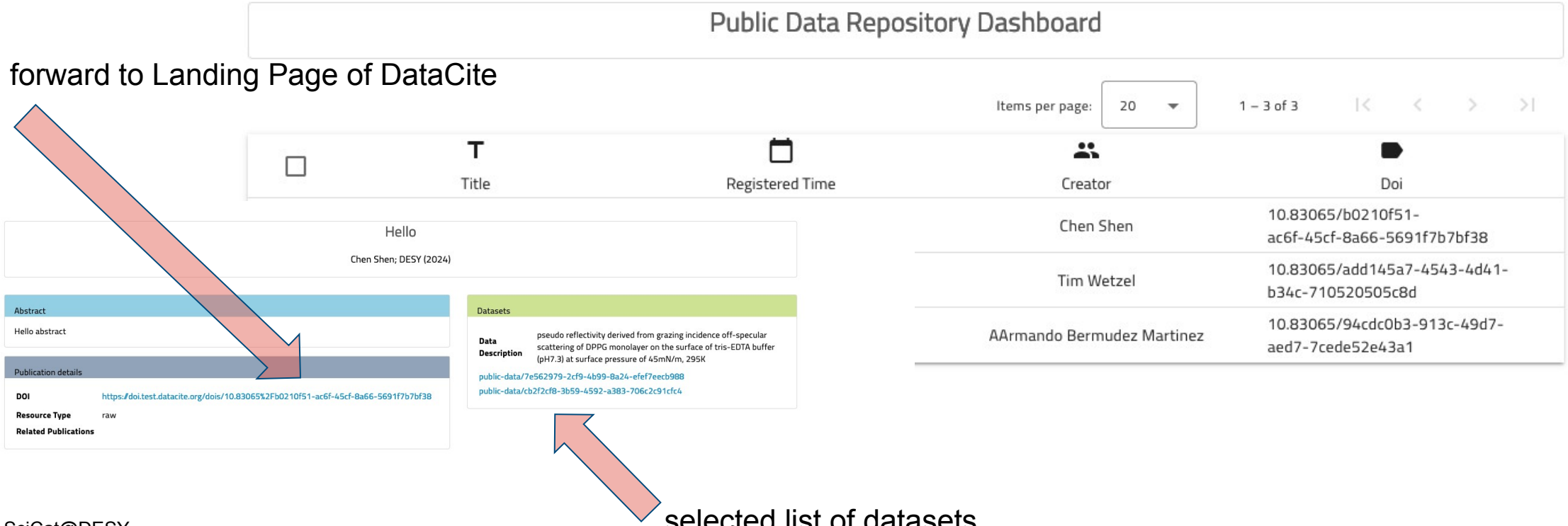
Progress

Step by step

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



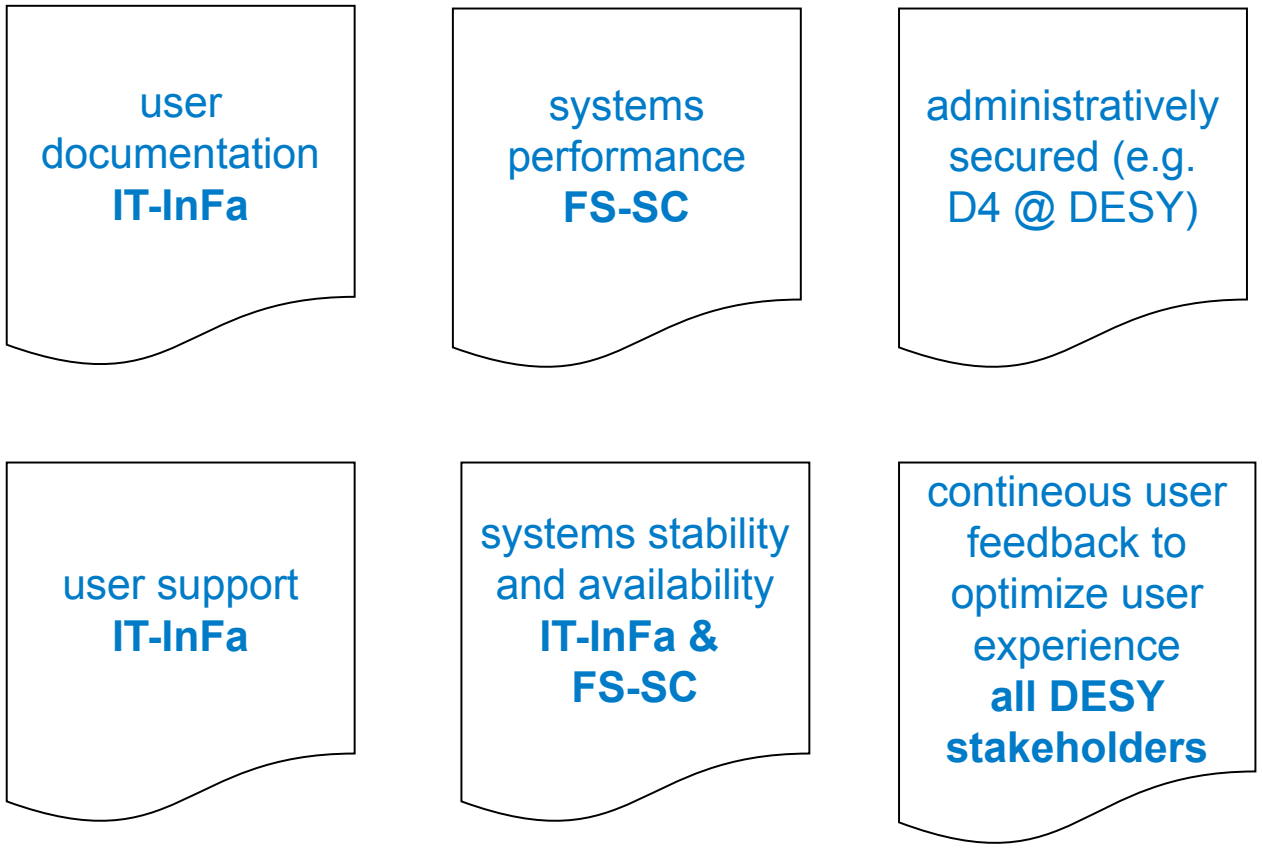
click: forward to Landing Page of DataCite



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

user documentation **IT-InFa**

systems performance **FS-SC**

administratively secured (e.g. D4 @ DESY)

user support **IT-InFa**

systems stability and availability **IT-InFa & FS-SC**

contineous user feedback to optimize user experience **all DESY stakeholders**

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