Status report on Pilot for DOIs service at DESY

What are the features you can expect?

What are the prospects?

What is left to do to get there?

Regina Hinzmann with input from many
esp. IT collegues Johannes Reppin, Armando B Martinez, Tim Wetzel and Tobias Klann.

Pilot: DOI minting service and catalogue functionality at DESY

benefiting from DESY involvements in European project that triggered SciCat activities (DESY-IT-RIC)
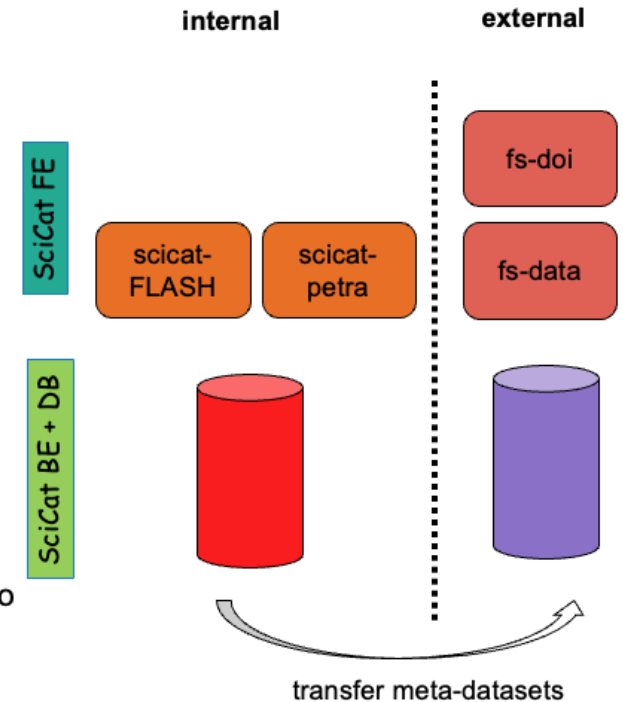
Beginning Oct - Mar
- IT and FS agreed on a set up: to handle data **under embargo and those for which a DOI will be minted**.
- we worked on such a setup with this concept:
  - to have an **internal catalogue** in which (**all) beamline data** is ingested into, just as the demonstrator instance at P08.
  - From that internal catalogue the user shall **select those datasets** (of metadata) which are to be published and DOI minted.
  - The user would need to **transfer the metadatasets to an external catalogue**, being exposed to the www without authentication and configured to point to a Landing page.

Sketch of DOI service for FS

- Coordinate efforts to set up a DOI service for FS: benefit from IT-RIC efforts **who is involved in several European projects (e.g. EOSC beyond, NFDI, Helmholtz, etc).**

- **Pilot DOI service fo FS** will be able to
  - transfer user selected meta-datasets from internal, existing SciCat to externally accessible SciCat,
  - publish meta-datasets; mint meta-datasets with DOI.

- **Productive SciCat DOI service for FS** will be able to
  - transfer selected datasets in a more user friendly way,
  - enable a more refined DOI minting process, introducing reservation of DOIs for DESY data, automated after embargo period, for better data curation.

- **Timeline:** pilot expected to be ready late Q1/2025
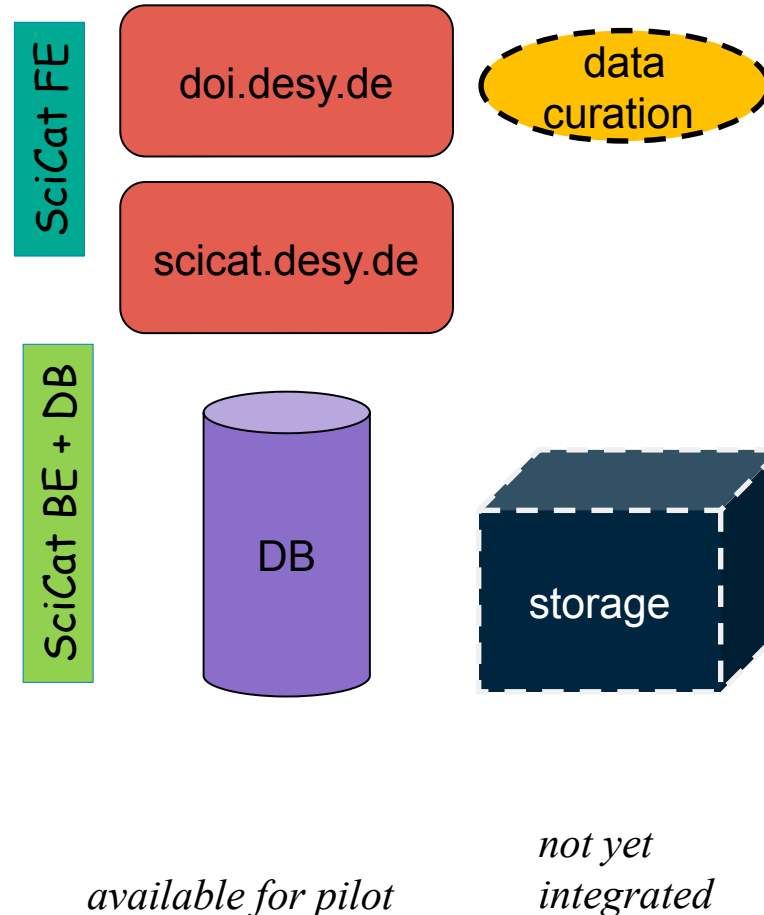
R Hinzmann | 20.01.2025

DESY.   Seite 2

Pilot: DOI minting service and catalogue functionality at DESY

Status update since start in Okt 2024

Status
- We moved away from OpenStack and use **already the productive environment** (systems Kubernetes cluster).
- We will deploy a **single instance** capable of dealing with internal and published meta data.
- Proprietary/industry metadata will not be ingested at all.
- We moved from fs-data.desy.de to **scicat.desy.de** to cover also cases beyond FS.
- a Landing Page has been configured and set up: **doi.desy.de***
- For the next 1 ½ - 2 months: **test** workflow, collect missing features and benefit from the second round of DAPHNE contribution to SciCat.

SciCat FE
- doi.desy.de
- scicat.desy.de

data curation

SciCat BE + DB
- DB
- storage

available for pilot   *not yet integrated*

Outlook
- **pilot anticipated (early) Q2**.
  - catalogue function for meta datasets for all beamlines where ingestion to SciCat is setup.
  - DOIs of selected metadata
  - **no direct data access**
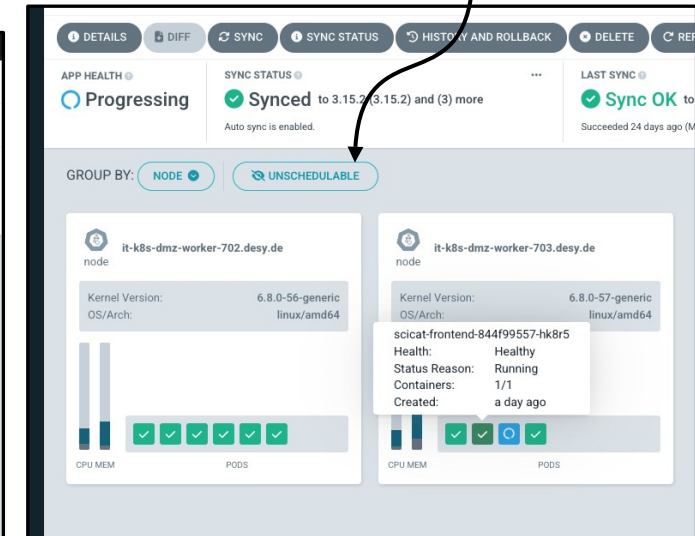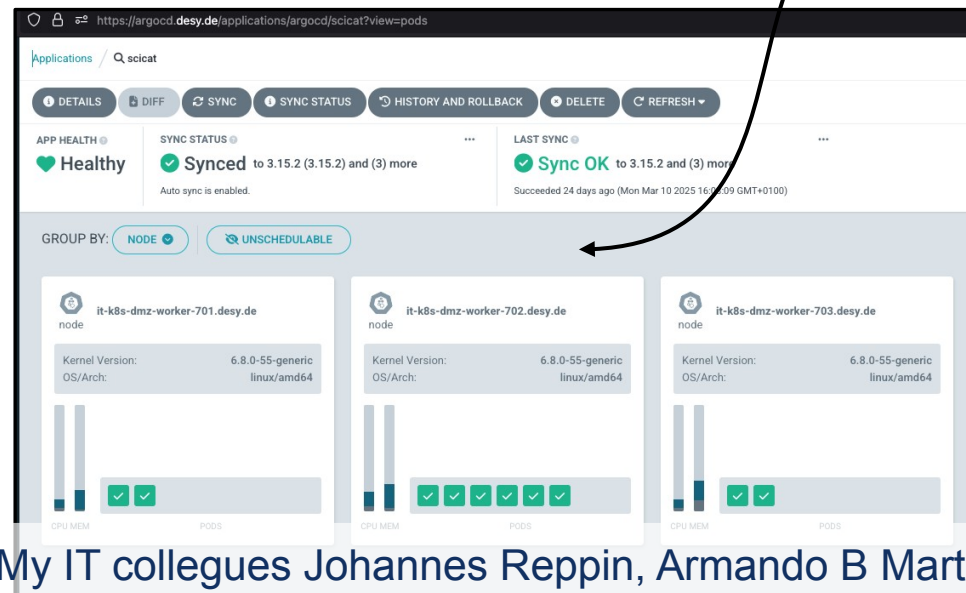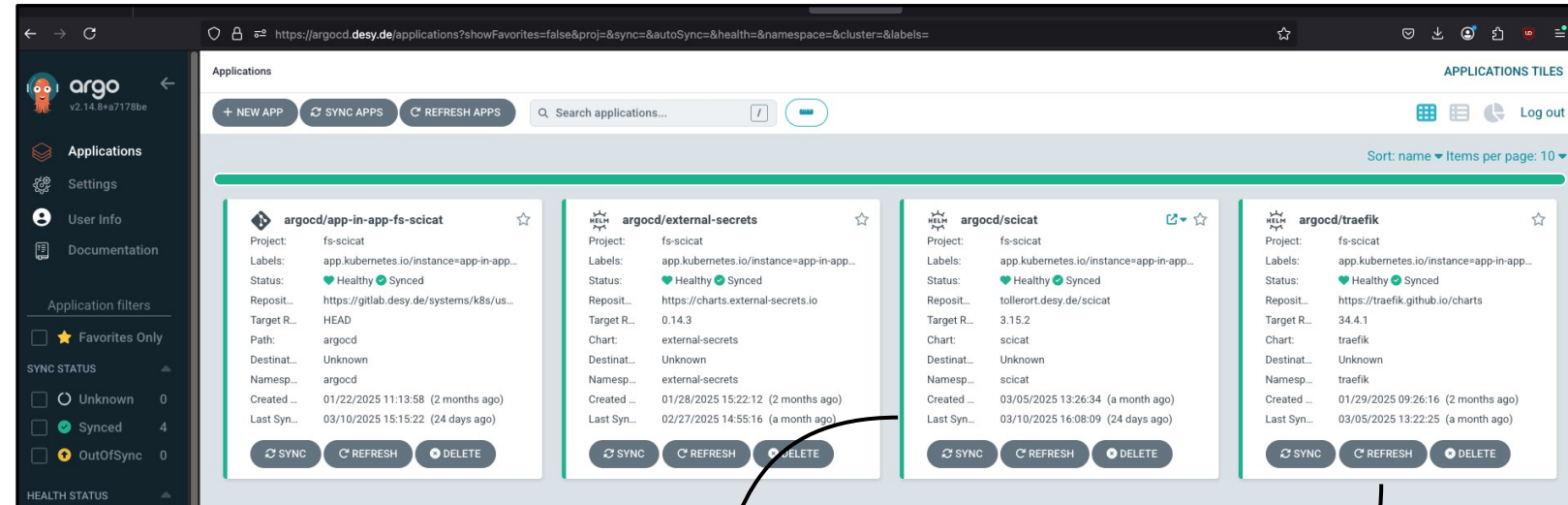 - **no data curation**

Prospects/to be decided
- To provide **core functions** of a metadata catalogue DESY must agree on
1. workflow of DOI minting at DESY. The pilot offers a crude way to mint DOIs. But: **combine it with curation service of library**?
2. Needs agreement: What are automated and what manual parts in the workflow?
3. Define **concept of data access** or "download functionality".

TEST instance scicat.desy.de on systems cluster

Under development and continous improvement. Don't expect stable behavior for now!

- Configured to **enable each user to ingest** a dataset:

  - no functional account for ingestor is needed which makes it more secure.

- **Tweak performance**: requested new features for SciCat for better administration, e.g. move the credentials of BE to DB so one can have several BE API servers for 1 DB.

- **Next in April**: Test the DOI minting workflow with larger test data.

- Plan is to **dump TEST data** of current SciCat instances into our pilot cleaned from any proprietary data.

credits for set up & running: My IT collegues Johannes Reppin, Armando B Martinez, Tim Wetzel and Tobias Klann.

DESY-wide metadata catalogue

for all FS data under embargo

scicat.desy.de

scicat can handle public metadatasets

scicat handles externally defined access rights (me as part of this group have ownership for this dataset too)

Pilot: DOI minting service and catalogue functionality at DESY

First impressions.

DESY Public Data Repository

A New Aera

Max Novelli,Regina Hinzmann; DESY (2025)

Abstract	Datasets
of abstracts	**Data Description** a new way
	scicat/893f2965-6f39-429a-9408-50dd8b1dfe0e
	scicat/9721096a-e570-4d91-be30-cf528e19dc23

Publication details

DOI	https://doi.test.datacite.org/dois/10.83028%2Fb0059540-ed1f-47ab-a0af-7bfb56ce2ddb
Resource Type	raw
Related Publications	

Selection of several datasets

SciCat covers only a minimum of DataCite fields.

cannot edit all fields
e.g. authors, for now Creator is author

https://doi.test.datacite.org/dois/10.83028%2Fb0059540-ed1f-47ab-a0af-7bfb56ce2ddb

DataCite Fabrica

About   Support   Sign in

10.83028/b0059540-ed1f-47ab-a0af-7bfb56ce2ddb

resolved to DataCite catalogue

URL

https://doi.desy.de/detail/10.83028%2Fb0059540-ed1f-47ab-a0af-7bfb56ce2ddb

Metadata   Summary View

A New Aera Dataset
Max Novelli, Regina Hinzmann,
Raw published 2025 via DESY
of abstracts

https://handle.test.datacite.org/10.83028/b0059540-ed1f-47ab-a0af-7bfb56ce2ddb

Metadata Export
DataCite XML
DataCite JSON
Schema.org JSON-LD
BibTeX

Pilot: DOI minting service and catalogue functionality at DESY

Next tasks for pilot early O3

1. **I**teration with developers, e.g. no search on global Landing Page is currently possible.
   - Be able to edit more fields than currently possible, e.g. authors.
   - Remove selection column (left) or better: get rid of LP but integrate the more general approach published Data as home page
   - Fix DOI minting workflow (e.g. user can still edit record after registration).
   Collect all missing features for production. With Nicolas H we worked on an improved workflow.
2. Test findability of datasets.
3. Disclaimer that no direct data access (download of data) is granted until IT/FS/L agrees and realises a concept.
4. Define and perform tests to test stabilty of DOI workflow. Involvement of test user and L (and FS).

Prospects: within the next (4) month, ie. early O3, significant improvement will be made. With some grace time, the pilot is considered ready (w/ agreed plan for data access).

https://doi.desy.de

doi.desy.de

Public Data Repository Dashboard

Items per page: 20   1 – 10 of 10

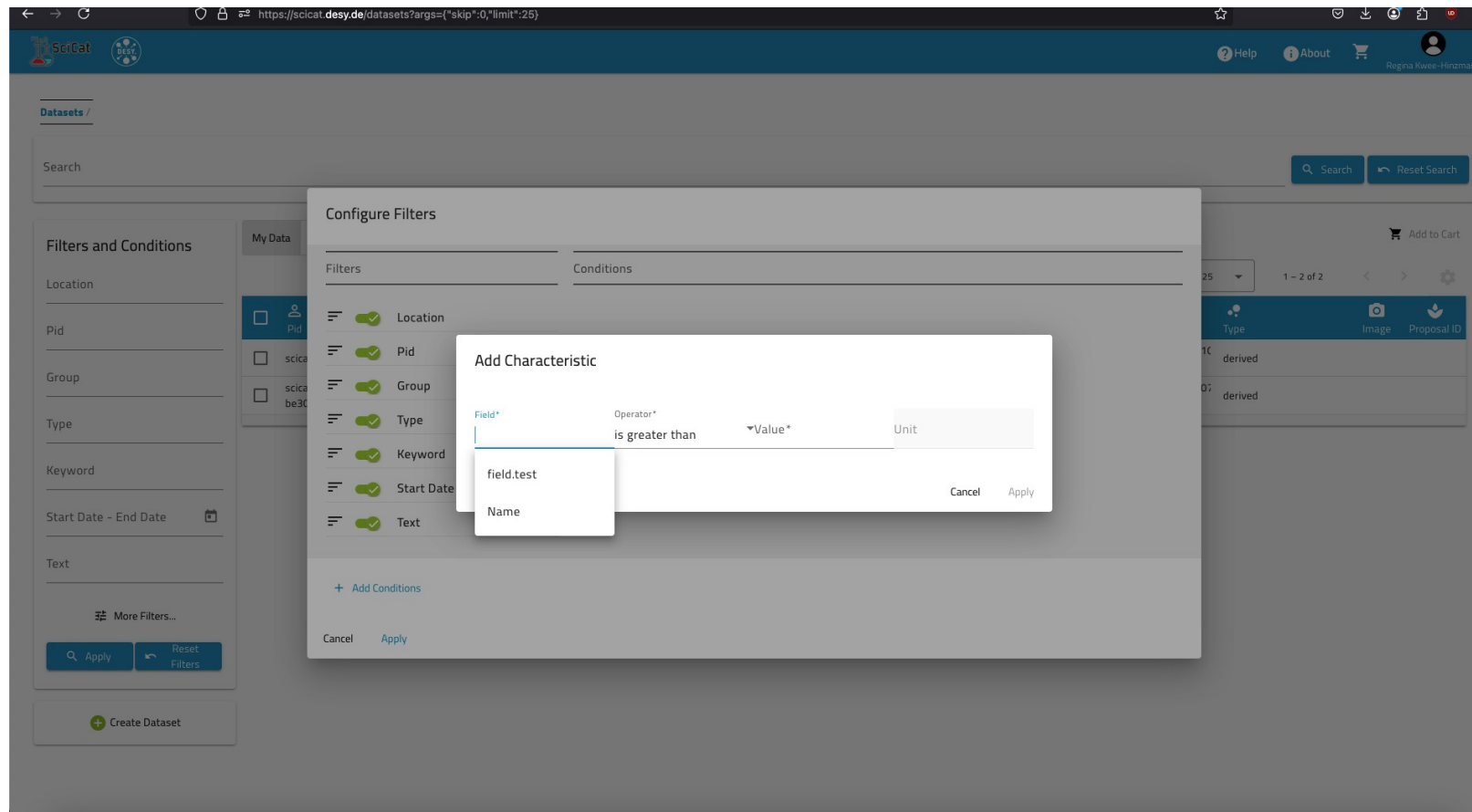| Title	Registered Time	Creator	Doi
Test Dataset for Scicat	2025-03-07 12:01	https://keycloak.desy.de/auth/realms/production_63832465-c0f3-4465-b4df-aa635e86d282	028/81c398fd-0906-47aa-907a-39e51eb3f
ein Titel	2025-03-07 12:32	rkwee	10.83028/ca543b64-fc85-4eb1-939c-ab691f7bab8e
A new Test begins.	2025-03-07 13:00	https://keycloak.desy.de/auth/realms/production_63832465-c0f3-4465-b4df-aa635e86d282	10.83028/53689143-49ee-4426-b2ce-881161f7e8ad
Let's try to publish this	2025-03-07 13:35	reppinjo	10.83028/6be2bb80-2eb8-449c-a34d-7f792a4c5958
new DOI URLs	2025-03-10 15:17	reppinjo	10.83028/24036135-8800-489f-a459-256f33be0ef6
This is the last test.	2025-03-10 15:40	reppinjo	10.83028/235d0c2e-83b9-4164-8f8e-bc13212d8628
pko'	2025-03-10 17:48	reppinjo	10.83028/f8eb6aaf-f92e-4016-9348-d8757f74a287
a title	2025-03-21 13:22		0.83028/957d835-9de4-40d5-9328-9e9edd141b51
A New Aera	2025-03-21 13:35	Max Novelli,Regina Hinzmann	10.83028/b0059540-ed1f-47ab-a0af-7bfb56ce2ddb
some nice title	2025-03-25 15:16	Regina	10.83028/4c56eb94-a118-4fe4-b238-b3fa3420b9cc

additional info (reminder)

Outlook Pilot

When SciCat becomes useful…

Next items on the table:
- Further improvements on UI/UX especially the **search**.
- Refactor DOI minting procedure.

shown at <u>SCT 2025-03-31</u> <u>https://indico.desy.de/event/48566 /</u>

Improvements for the second round of DAPHNE contributions to SciCat (DCS).

Outlook Pilot

Usefulness of SciCat: DOIs

Improvements for the second round of DAPHNE contributions to SciCat (DCS).

Next items on the table:
- Further improvements on UI/UX especially the search.
- Refactor **DOI minting procedure**.

From annual meeting last week we worked on a list which Max is now preparing for the next contract.

A big THANK YOU to all who contributed!