

Assessment Results for HZB

Rolf Krahl

ExPaNDS Meeting for WP3 KPI Assessment, 7th December 2021



On the following slides, we mark the assessment by KPI. Explanation of symbols:

- ✓: We have accomplished this one,
- ⚠: We plan to accomplish this by the end of the project,
- ✗: Not considering this for the moment.

Prerequisites

Data Infrastructure ✓

Server and storage systems are in production.

Facility Proposal info ✓

Proposal data is managed by user office systems.

Standard Data !

Rather heterogeneous data formats by now.

But: we are working on that, there is significant progress.

PID issuer !

We don't issue PIDs for raw data yet.

Need to setup issuing of ePIC handles. Need to work on the PID metadata for raw data.

Data Policy ✓

HZB Data Policy adopted in June 2016.

Resources !

Still lack personal resources for the production deployment. There is light at the end of the tunnel.

Choosing a Catalogue

Clear and agreed requirements ✓

Solutions researched ✗

Solutions Tested ✓

Decision ✓

We did not consider alternatives. Back then when we decided for a data catalogue, ICAT was the only metadata catalogue used in the community. We tested it and found it fit for our needs.

File ingest ✓

We have a workflow for ingesting files.

Authenticated data access ⚠

Authenticators and access rules are in place. We lack a centralized ID management though. Work in progress.

Domain agnostic metadata ✓

Proposal metadata are automatically imported into ICAT.

PaN API ⚠

ICAT does not implement the PaN API yet.

Stability ✓

Two dedicated servers to run the metadata catalogue.

10% < 25% instrument coverage 

25% < 50% 

50% < 75% 

75% < 100% 

We are clearly lacking behind our own schedule.

Federated access to data 

No PaN API yet.

OAI-PMH Implemented 

The OAI-PMH component for ICAT is working, but we need to work on the metadata to disseminate first.

Harvesting by EOSC services 

Need OAI-PMH first.

EOSC Services access 

Not considered yet.

EOSC statistics 

Not considered yet.

The prerequisites to work on EOSC integration are not yet in place.

PaN Specific Implementations

Domain metadata ingest 

Workflow to ingest domain specific metadata into ICAT is in place.

RAW and derived data catalogued 

Only considering raw data for the moment.
Yet, we have curated data publications in place.

Data provenance 

Only considering raw data for the moment.

Calibration data 

Only considering raw data for the moment.

Link to instrument reference information 

Need to to work on the PID metadata for raw data first.

Link to publications 

Need to to work on the PID metadata for raw data first.

eLogBook data 

Work in progress: testing ESRF E-logbook.

What missing items are hampering advancement the most:

- PIDs for raw data. This one is blocking advancement for many of the other KPIs.
- ID management. This is urgently needed by our users.

Next most important steps:

- PIDs for raw data: define a schema for the automatically collected metadata for the raw data to be registered with the PIDs. Implement minting of ePIC handles for raw data.
- Setup ID management with Keycloak.
- Go in production with E-logbook. Gather feedback from instrument scientists.

Things where help from the project would be the most useful:

- Data standards: definition of standards and formats for PaN specific metadata.
- ID management: integrate Keycloak with UmbrellaID. See upcoming workshop in spring.