Infrastructure for Research Data Management at HZB

Rolf Krahl

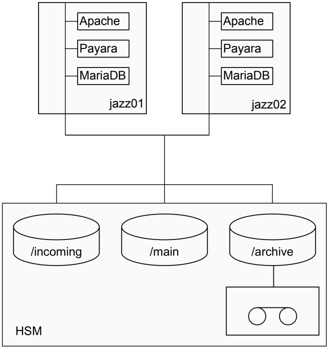
Workshop on Research Data Management, HZB, 11th June 2019

HZB Data Policy

- HZB Data Policy regulates management of scientific data from public research at HZB's large-scale facilities.
- Distinguish raw data, results, and metadata.
- Raw data and associated metadata will be curated and stored by HZB for at least ten years.
- Raw data and associated metadata are placed in the public domain (Creative Commons CC0 Dedication).
- Access to raw data and associated metadata is restricted to their creators for an embargo period of five years. After that, they become openly accessible.
- Results and associated metadata may be stored with the raw data. They will not be curated by HZB. They may be made openly accessible upon request of their creators.

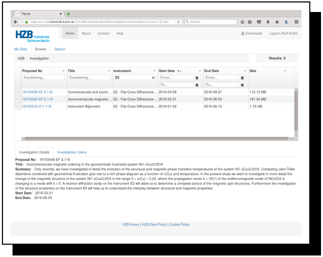
Server Systems

- Storage systems: hierarchical storage management (HSM), the `/archive` area consists of disks and tape libraries.
- ⇒ Most data will reside on tapes most of the time.
- Two dedicated servers in active/passive configuration.
- Data volume: we calculate 2 PB/Jahr.



ICAT Metadata Catalogue

- Access to the data is provided by the ICAT metadata catalogue.
- Search for the user's own and for public data.
- A request to download data automatically triggers the staging of that data from tape to disk. The data may be downloaded after the staging is complete.
- ICAT is developed as free software in cooperation with other Photon and Neutron sources (STFC, DLS, ISIS, ESRF, HZB).

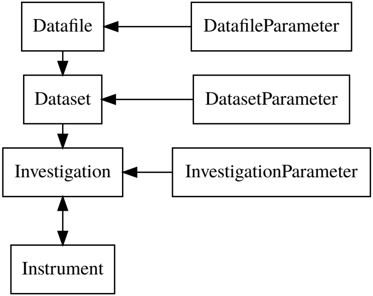

Topcat
https://ro-icat.helmholtz-berlin.de:8181/#/browse/facility/HZB/investigation?uiGridState={"columns":[{"nam-
Suchen
HZB Helmholtz Zentrum Berlin
Home
About
Contact
Help
Downloads
Logout (Rolf Krahl)
My Data
Browse
Search
HZB / Investigation
Results: 3
Investigation Details
Investigation Users
HZB Home | HZB Data Policy | Cookie Policy

Proposal No	Title	Instrument	Start Date	End Date	Size
Containing...	Containing...	E2	From... To...	From... To...	
16103436-ST-2.1-N	Commensurate and incom...	E2 - Flat-Cone Diffractome...	2016-04-06	2016-09-27	112.73 MB
16103546-EF-2.1-N	Incommensurate magnetic...	E2 - Flat-Cone Diffractome...	2016-03-21	2016-08-03	161.55 MB
16103319-IT-1.1-N	Instrument Alignment	E2 - Flat-Cone Diffractome...	2016-01-05	2016-06-10	1.75 GB

Proposal No: 16103546-EF-2.1-N
Title: Incommensurate magnetic ordering in the geometrically frustrated system Ni1-xCuxCr2O4
Summary: Only recently, we have investigated in detail the evolution of the structural and magnetic phase transition temperatures of the system Ni1-xCuxCr2O4. Competing Jahn-Teller distortions combined with geometrical frustration give rise to a rich phase diagram as a function of x(Cu) and temperature. In the present study we want to investigate in more detail the change of the magnetic structure of the system Ni1-xCuxCr2O4 in the range 0 < x(Cu) < 0.20, where the propagation vector k = (001) of the antiferromagnetic mode of NiCr2O4 is changing to a mode with k = 0. A neutron diffraction study on the instrument E2 will allow us to determine a complete picture of the magnetic spin structures. Furthermore the investigation of the structure properties on the instrument E9 will help us to understand the interplay between structural and magnetic properties.
Start Date: 2016-03-21
End Date: 2016-08-03

ICAT Schema

- Central elements for organizing the data in ICAT are: `Investigation` ← `Dataset` ← `Datafile`.
- Correspondence:
  - `Investigation` ≙ Proposal,
  - `Dataset` ≙ Measurement,
  - `Datafile` ≙ File.
- `DatasetParameter` allows storing physical metadata of the measurement in a simple keyword/value schema.



Metadata

We distinguish two classes of metadata:

Administrative metadata

- Proposal, title, abstract, user, access rights.
- Get imported from the user office portal GATE beforehand. Analogous arrangements will be made for measurements not related to a GATE proposal.
- Relevant for the control of internal ICAT workflows.

Physical metadata

- Parameter of the measurement, sample etc. See other talk.
- Will be collected (preferably) automatically and stored in the datafiles before ingestion into ICAT.
- A selection of the metadata may be additionally stored in ICAT as `DatasetParameter`, if they are relevant for the search of data.

Data Life Cycle

Steps in the life cycle for scientific data at HZB:

1. User submits a proposal in GATE.
2. Proposal gets accepted.
3. Administrative metadata get imported from GATE into ICAT. (For data not related to a proposal, create an `Investigation` in ICAT instead.)
4. User comes to HZB and performs experiments. Data is collected and curated at the instrument.
5. Data are ingested from the instrument into ICAT.
6. User have exclusive access to their data.
7. User may optionally upload results from data anaylsis into ICAT.
8. After expiring of the embargo period, the raw data and associated metadata become openly accessible. The user may decide to make also the results openly accessible.

Current Status & Implementation

Current status:

- The storage systems are available.
- ICAT operates in test production. Mostly ready for production.
- Two BER II instruments (E2 and E9) register routinely their data.
- Implementation at the first BESSY II station (NanoclusterTrap) is in preparation and will begin shortly.

Implementation:

- For each instrument, we need to hook up (the data curation and) the data ingestion to ICAT into the measurement workflow.
- This requires consideration of the prerequisites and the workflows at the instrument individually.
- We will proceed intrument by instrument.

Issues:

- We lack a proper identity management at HZB.