Metadata at Helmholtz Zentrum Berlin

Workshop PaNSIG ‘Scientific Data Management for Photon and Neutron Facilities’, March 2018

Heike Görzig
Project background - Digital Agenda

• aims to accelerate the digital transformation in science
• expand information infrastructures such as archives and libraries

• increasing requirements of funding agencies with regard to research data management
• application of FAIR data principles
• lack of knowledge in the systematic handling of research data in their life cycle
• lack of knowledge about metadata required for research area overreaching science and to meet FAIR data principles

• systematically record which models exist for systematic research data management and how far the institutions are in the development of such strategies
• research and / or development of solutions to identified challenges for research data management

RDMatDB
BESSY II

- Synchrotron radiation source
- Ca. 50 beamlines
- Undulator, wiggler and dipole source
- Many-faceted mix of beam holes and measuring sites

Total data volume per year: 1-2 PB

BER II

- Neutron beams
- Thermal and cold neutrons
- Several neutron guide tubes
- 10 instruments in user operation
- 8 instruments in house and cooperation use
- Different measuring sites
- Cease operations end 2019
- 2020 starting decommissioning and dismantling

CoreLabs

- multi-user platform
- complex infrastructures with unique and state-of-the-art equipment
- research and development of innovative energy materials
Heterogeneous data sources

**Experimenter – Proposal system:**
GATE

**Sample description:**
Lab book, sample data base, sample creation workflow (LabView)

**Static instrument description:**
Simulation software (RAY-UI), instrument data base

**Analysis software / programming languages:**
(i) Mosflm, XDS(APP), Igor(Pro), Origin, MANTID; Octopus, bean, C/C++, LabView, Python

**Experiment data (intelligibility)**
- Static instrument description
- Instrument calibration data
- Instrument experiment configuration data
- Experiment measurement data
- Sample description
- Pre-analysis process data

**Instrument configuration – experiment configuration / data:**
EPICS, SPEC, TANGO,
CARESS, M2C, MxCuBE,
LISE/M, EMP/2, M, C/C++, LabVIEW, Python
INSTRUMENT SPECIFIC !!!
(depending on producer)

**Metadata catalogue data (find & retrieve)**
- Bibliographic data
- Defined experiment metadata
Workflows for creating and collecting metadata

Data creation workflows

Data selection, gathering and packaging

Ingestion into the repository

The DCC Curation Lifecycle Model

Workflows for creating and collecting metadata I

Simulation configuration, instrument data base

- Relative static description of instruments
- Workflow independent from proposal
- Resulting knowledge available before investigation
- Information required for planning investigation

→ Data sources: instrument data base, simulation software configuration (RAY-UI), device and experiment control
• Investigation specific information ingested
• Workflow before experiments start
• Information required for bibliographic metadata
• Sometimes sample description

→ Data source: GATE
• Experiment specific information and experiment data arise
• Information arises directly before or during the experiment
• Sample description, instrument configuration for experiment and measurement data, pre-analysis data

→ Sources: Device and experiment control, sample description, analysis software
Workflows for creating and collecting metadata IV

- Instrument description workflow
- Proposal workflow
- Experiment workflow
- Curation workflow
- Ingest workflow

**Instrument description workflow**
- Instrument description data
- Instrument status data
- Proposal / user data
- Sample data
- Experiment configuration data
- Experiment execution data

**Curation workflow**

**Ingest workflow**
- RDMatDB
- Automatic and manual selection, gathering and packaging metadata and data from various sources

**Repository**

**Nexus**
Workflows for creating and collecting metadata V

- Instrument description workflow
- Proposal workflow
- Experiment workflow
- Curation workflow
- Ingest workflow

Data collector

- Instrument Control CARESS
- Raw Detector Datasets
- Meta Information

XML-Meta-Info (& Caress)

Repository import file

Archiving and Documentation

TOPCAT

INVENIO

Depositing metadata and data in repository

ICAT

Server

NeXus

ICAT

INVENIO

26.03.2018

Metadata at HZB  Heike Görzig
Existing NeXus File Writer and Packaging Toolkits

Metadata at HZB  Heike Görzig

26.03.2018
Data collector for gathering research data and metadata required

• Categorisation of data sources and workflows
• Existing solutions identified

• Next steps:
  • Revision and testing of existing software
  • Identification of instrument for creating prototype

Thanks for listening!
References

[1] Dank an Jens-Uwe Hoffmann (HZB)

