Metadata at Helmholtz Zentrum Berlin

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

Heike Görzig

- 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

Project motivation

- 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

Helmholtz-Zentrum Berlin für Materialien und Energie (HZB)

BESSY II

- Synchrotron radiation source
- Ca. 50 beamlines
- Undulator, wiggler and dipole source
- Many-faceted mix of beam holes and measuring sites
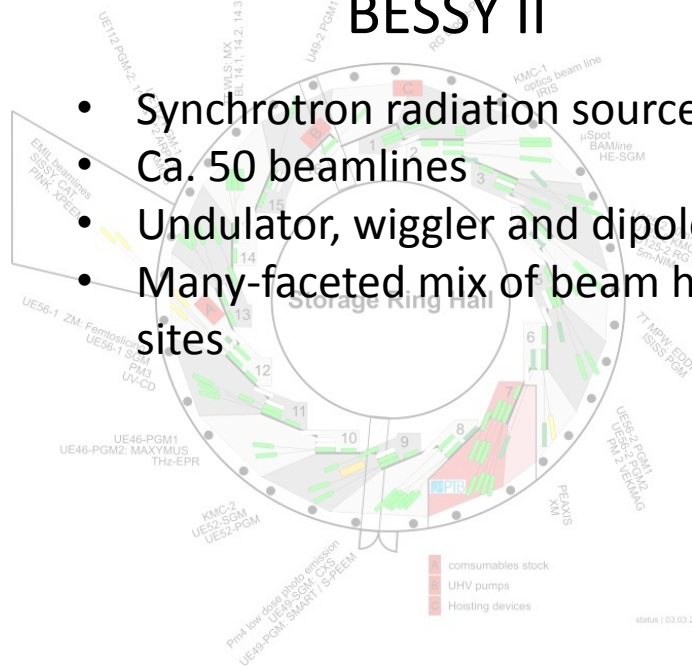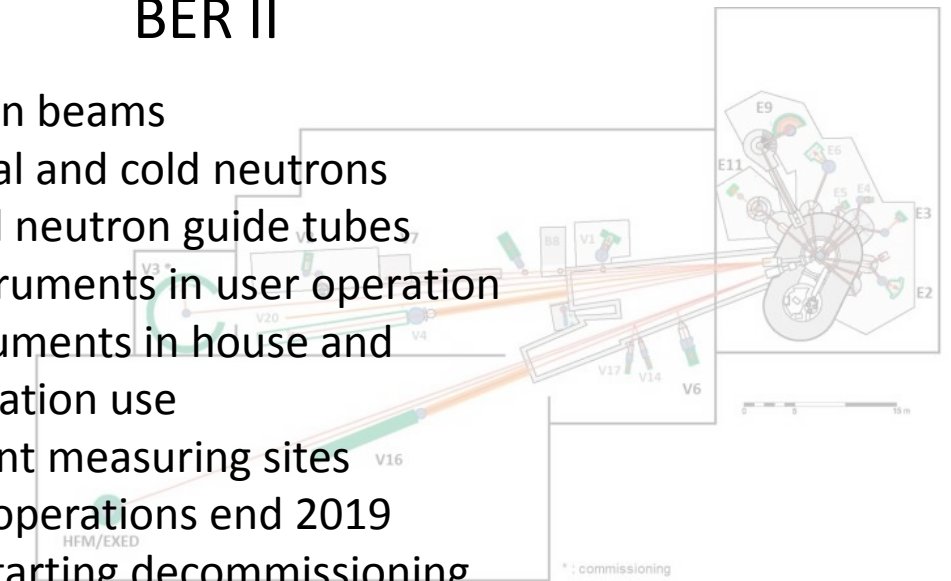
Total data volume per year:
1-2 PB

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

Berlin Energy Recovery Linac Prototype
- electron beam having high-brightness
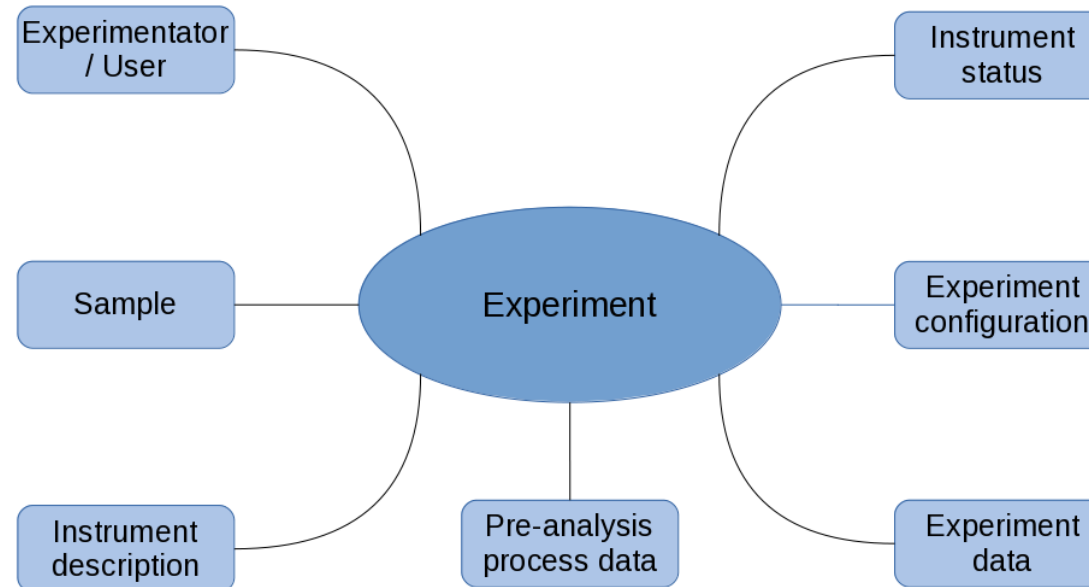- high-power capabilities
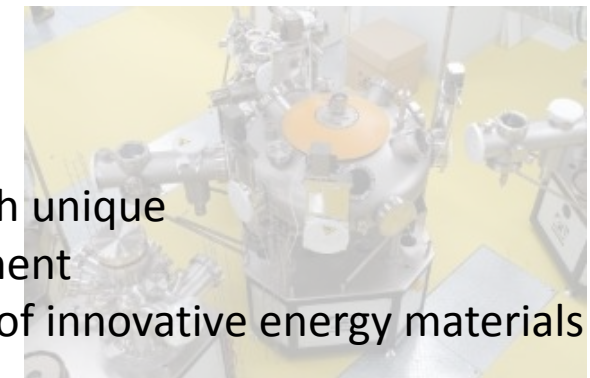
Heterogeneous data sources

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

Experimentator / User

Instrument status

Sample

Experiment

Experiment configuration

Instrument description

Pre-analysis process data

Experiment 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)

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

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

Metadata catalogue data (find & retrieve)
Bibliographic data
Defined experiment metadata

Workflows for creating and collecting metadata

Data creation workflows

- Instrument description workflow
- Proposal workflow
- Experiment workflow

Data selection, gathering and packaging

- Curation workflow

Ingestion into the repository

- Ingest workflow

The DCC Curation Lifecycle Model

http://www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf

Workflows for creating and collecting metadata I

Instrument description workflow

Proposal workflow

Experiment workflow

Curation workflow

Ingest workflow

Simulation configuration, instrument data base

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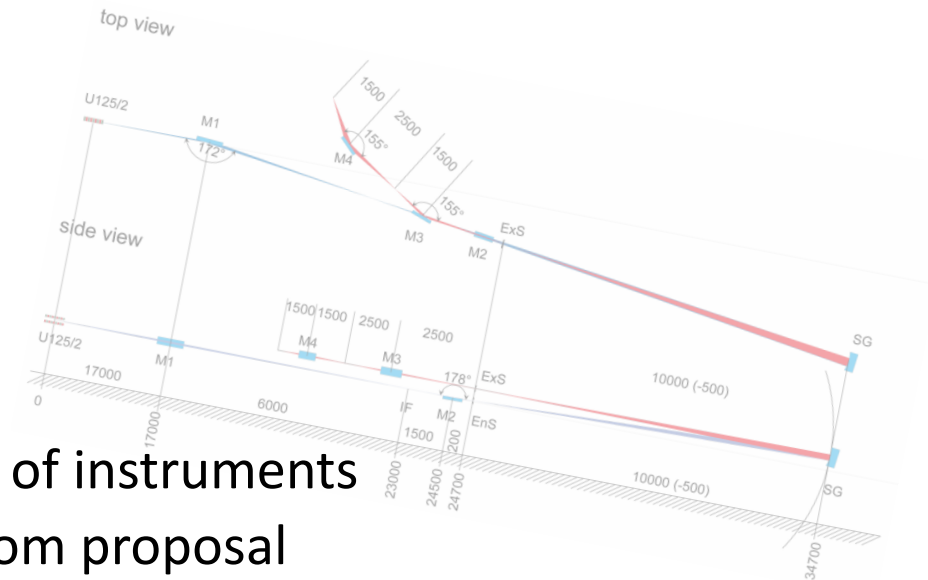
- Relative static description of instruments
- Workflow independent from proposal
- Resulting knowledge available before investigation
- Information required for planning investigation

→ 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

GATE

Creation of bibliographic metadata

The proposal deadline for round 2014/1 is 1 September 2013.

You have ...
- Any new publications?

Prepare beamtime

After the beamtime

My proposal
- Submit a new proposal
- Proposal list
- Publication list / submit publications

Personal section
- Personal data
- Change password
- Change username
- Online trainings

Instrument description workflow

Proposal workflow

Experiment workflow

Curation workflow

Ingest workflow

Instrument description workflow

Proposal workflow

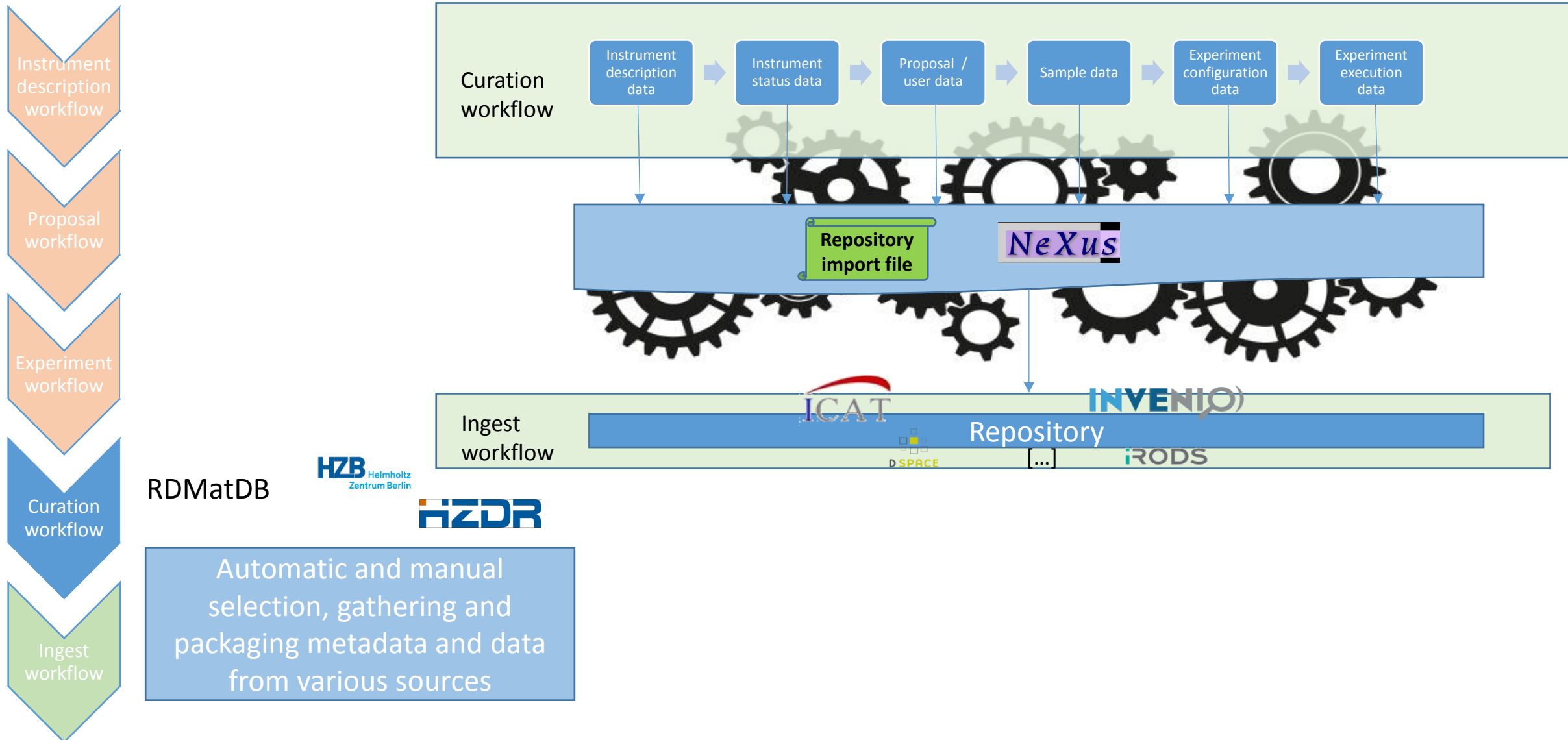Experiment workflow

Curation workflow

Ingest workflow

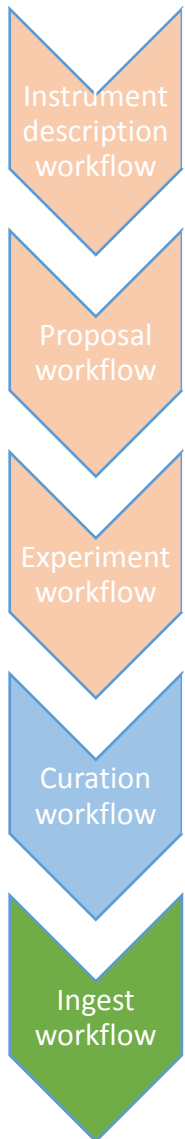Device and experiment control, sample description, analysis software

Creation of experimental metadata and data

- 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
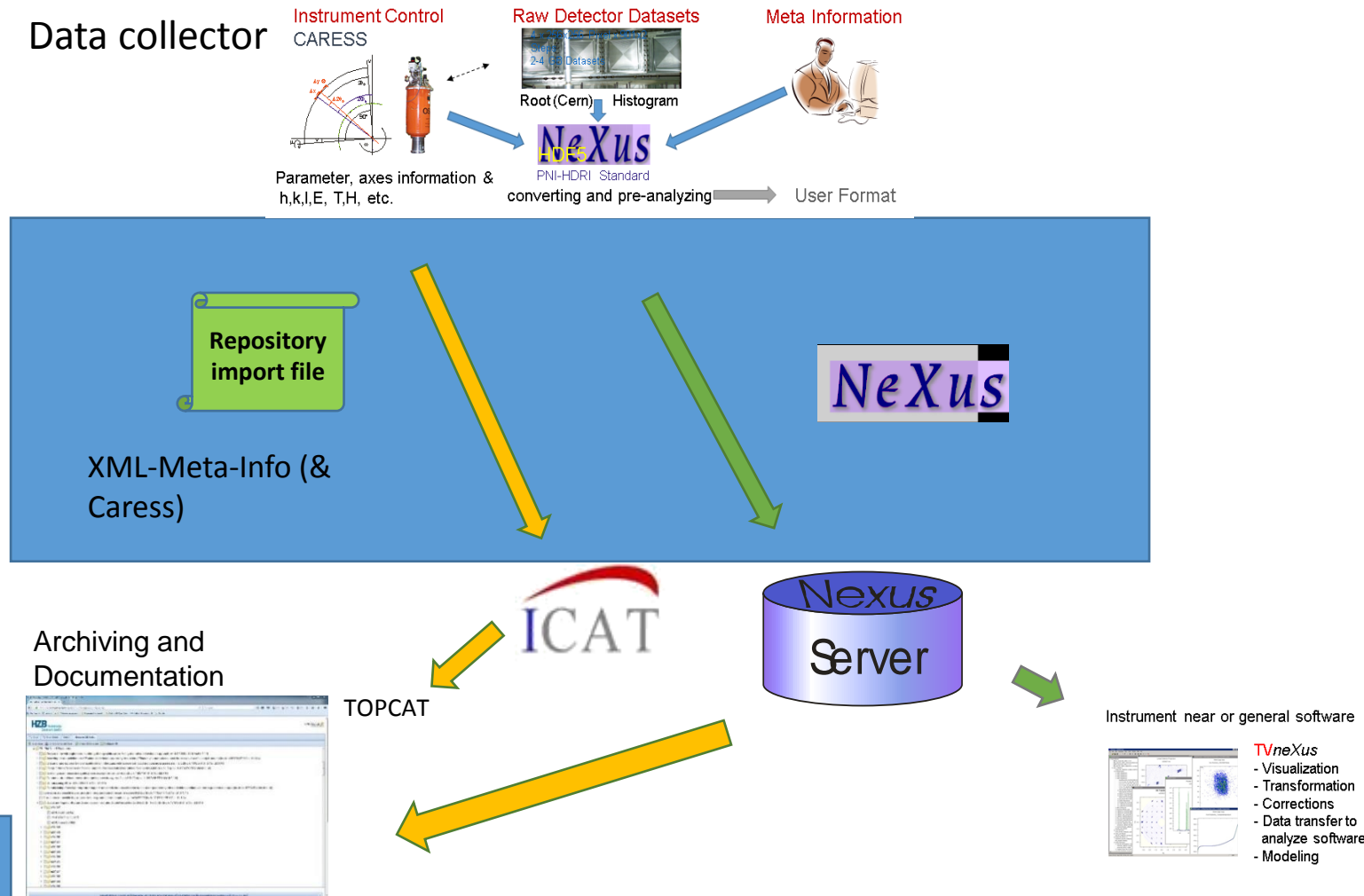
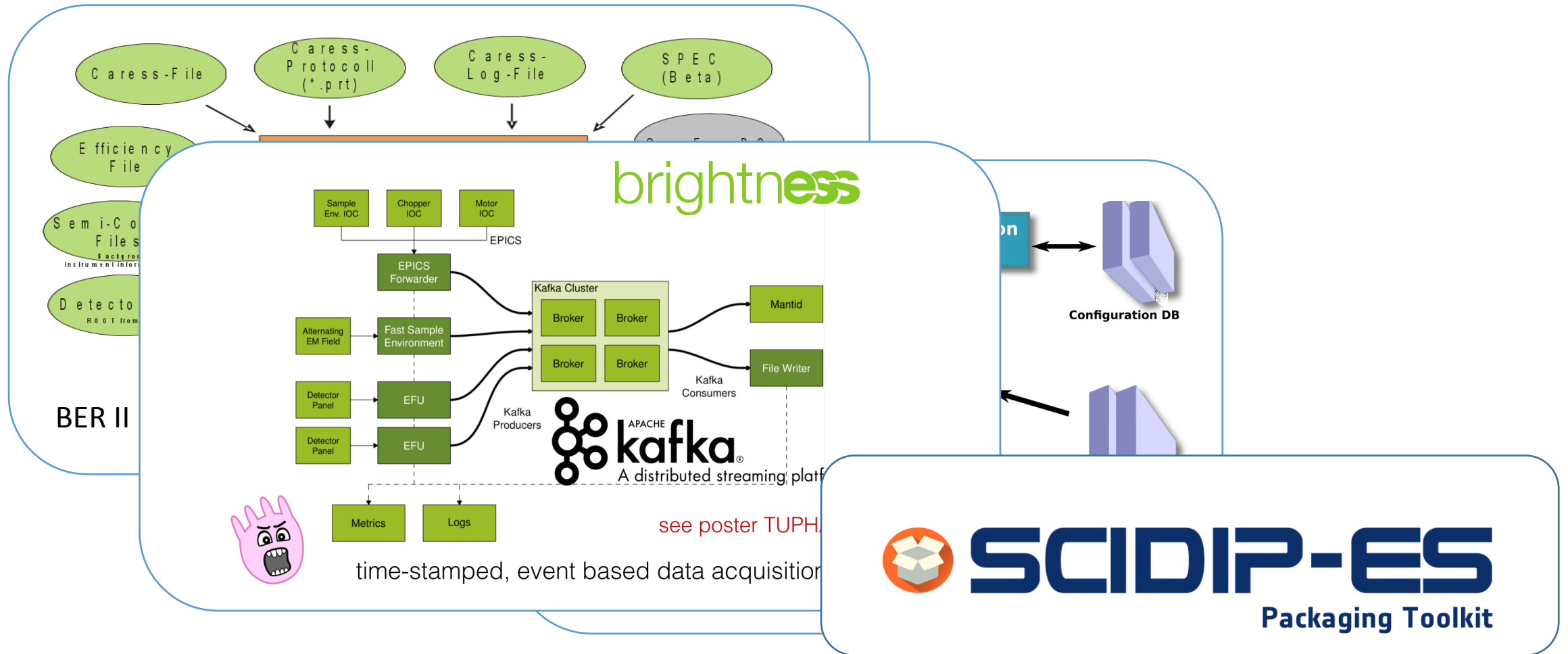# Workflows for creating and collecting metadata IV

Instrument description workflow

Proposal workflow

Experiment workflow

Curation workflow

Ingest workflow

RDMatDB

HZB Helmholtz Zentrum Berlin

HZDR

Curation workflow

Instrument description data → Instrument status data → Proposal / user data → Sample data → Experiment configuration data → Experiment execution data

Repository import file

NeXus

Ingest workflow

ICAT   INVENIO

Repository

DSPACE   [...]   iRODS

Automatic and manual selection, gathering and packaging metadata and data from various sources

Workflows for creating and collecting metadata V

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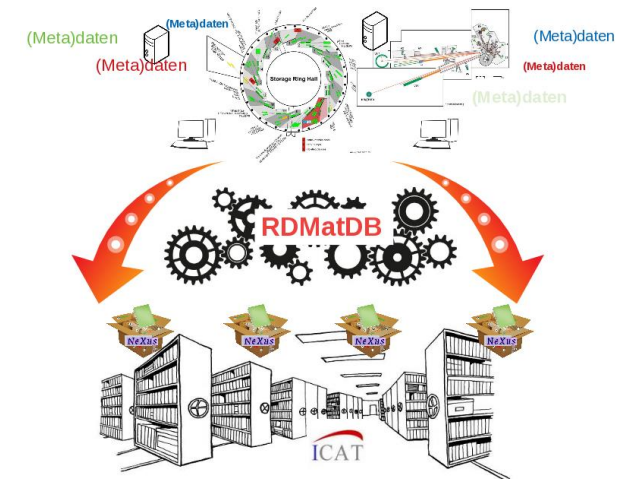
[1] Dank an Jens-Uwe Hoffmann (HZB)

[2] Richter, Tobias (ESS): NeXus/HDF5 @ ESS, HDF5 Workshop @ ICALEPCS 2017. Last accessed on 4.1.2018: https://indico.esrf.fr/indico/event/12/session/2/contribution/17/material/slides/0.pdf

[3] Kotański, Jan (DESY): Metadata composer for NeXus, HDF5 Workshop @ ICALEPCS 2017. Last accessed on 4.1.2018:
https://indico.esrf.fr/indico/event/12/session/2/contribution/15/material/slides/0.pdf

[4] http://int-platform.digitalpreserve.info/dashboard/packaging-toolkit/