Automatic Export of Data from Catalysis Experiments to NeXus in ROCK-IT

Daniel Tomecki[1], Devin Burke[2], Gerrit Guenther[1], Hector Perez Ponce[1], Sonal Ramesh Patel[1], William Smith[1]

[1]Helmholtz-Zentrum Berlin für Materialien und Energie (HZB), Albert-Einstein-Straße 15, 12489 Berlin

[2]Deutsches Elektronen-Synchrotron, Notkestraße 85, 22607 Hamburg

ABSTRACT: Operando Catalysis experiments involve many different devices and processes running simultaneously and on different timescales. Data from sample environment which is changing over time has be correlated with data from measurement techniques acquired using X-Ray beamlines. Users performing these experiments commonly face a daunting task of collating all their data together from different sources. In this poster we describe how as part of the ROCK-IT collaboration this problem is addressed. Metadata from devices is collected using HAPPI and Bluesky. It is then exported automatically to NeXus files which can contain information from the sample environment and multiple measurement techniques.

Motivation

- ROCK-IT (Remote, Operando Controlled, Knowledge-driven, and IT-based) is a **Helmholtz** funded project that aims to develop all necessary tools for the automation and remote-access of in-situ and **operando catalysis experiments**, for users of all experience levels. These will include:
  - Common remote access for experimental users
  - Experimental automation including use of robots
  - **FAIR[1] data management**
  - Near realtime analysis and optimization using AI/ML

- FAIR data management
  - **Complexity** catalysis experiment (Figure 1)
  - Data from multiple sources and on different time scales: Sample environment, beamline devices, elog..
  - **Common standards** for different data sources, different facilities..
  - Existing evolving standards: NeXus[2], SECoP[3], voc4cat[4]..

Bluesky

- Bluesky[5] is a python library for experiment control and collection of scientific data and (meta)data
- It can capture and organize the (meta)data to facilitate reproducibility and searchability
- Figure 2 shows the flowchart of planned automated (meta)data capture using HAPPI (Heuristic Access to Positions of Photon Instruments), which provides a framework for general indexing of devices and things that correspond to python object
- **(Meta)data involved in typical catalysis experiment:**
  - Sample environment: Temperature, pressure, gas analysis, etc.
  - Photon measurements: Beamline instruments (insertion device, monochromator, etc.), techniques such as XAS, XRD, etc.
  - Sample: Name, ID, sample holder ID, preparation, etc.
  - User: Proposal
- **(Meta)data ingestion in bluesky RunEngine:**
  - Start document: Prior information – sample, user, techniques
  - Baseline devices: pre-processor
  - Descriptor
    - NeXusCreator: A Parallel approach at HZB (Poster #94 by H. P. Ponce)
    - ICAT: Metadata catalog to support large scale facility experimental data (Poster #49 by R. Krahl)

Mapping beamline devices to NXinstrument

- The first step towards FAIR data at any given beamline is to populate the instruments with metadata
  - mySpot[6] - a test beamline for ROCK-IT project at HZB
  - A valuable feature of the mySpot beamline is that it provides the users with a choice of several analysis methods simultaneously. However, this poses a challenge for FAIR data capture
  - The beamline instruments are populated with metadata by mapping all the instruments of the beamline to NeXus base classes, as shown in Figure 3

- The second step toward FAIR data is to structure the data based on the experiment to be performed. This structure depends on the technique.
  - The NeXus application definition provides the standards in the form of minimum required fields for a given technique, irrespective of the beamline
  - Following the NeXus recommendation, we have created a NeXus data structure consisting of one NXsubentry for each technique performed at mySpot

NXxas multi-detector

- The experiments taking **measurements in multiple modes**, for eg. total electron yield, partial electron yield, auger electron yield, florescence yield, and transmission

- Structure the data in NeXus data file having as many NXsubentry as number of modes
- Soft-link the NXinstrument within each NXsubentry to one common NXinstrument within NXentry
- Figure 4 shows the example structure of NeXus hdf5 file incorporating multi-mode XAS measurement

Conclusion and future outlook

- The other parallel effort at DESY is serializing the file structure: **structured metadata**[7] (An overview talk by D. Burke)
- Implement metadata validator to ensure mandatory metadata are ingested before running bluesky plan
- Efficient merge of existing standards around ROCK-IT project: SECoP, NeXus, voc4cat, ELNs
- Providing experimental data to users in standardized way

REFERENCES

[1] M. Wilkinson et al., Sci Data **2016**, 3, 160018 (doi: 10.1038/sdata.2016.18)
[2] M. Könnecke et al., J. *Appl. Cryst.* **2015**, 48, 301-305 (doi:10.1107/S1600576714027575)
[3]https://sampleenvironment.org/committees/secop/
[4] https://github.com/nfdi4cat/voc4cat
[5] D. Allen et al., Synchrotron Radiation News, **2019**, 32:3, 19-22 (doi: 10.1080/08940886.2019.1608121)
[6] Helmholtz-Zentrum Berlin für Materialien und Energie, Journal of large-scale research facilities, **2016**, 2, A102 (doi: 10.17815/jlsrf-2-113)
[7] https://codebase.helmholtz.cloud/rock-it/wp2/structured_metadata

ACKNOWLEDGMENT

We gratefully acknowledge funding from the Helmholtz Association HGF for the ROCK-IT project

rock-it-project.de

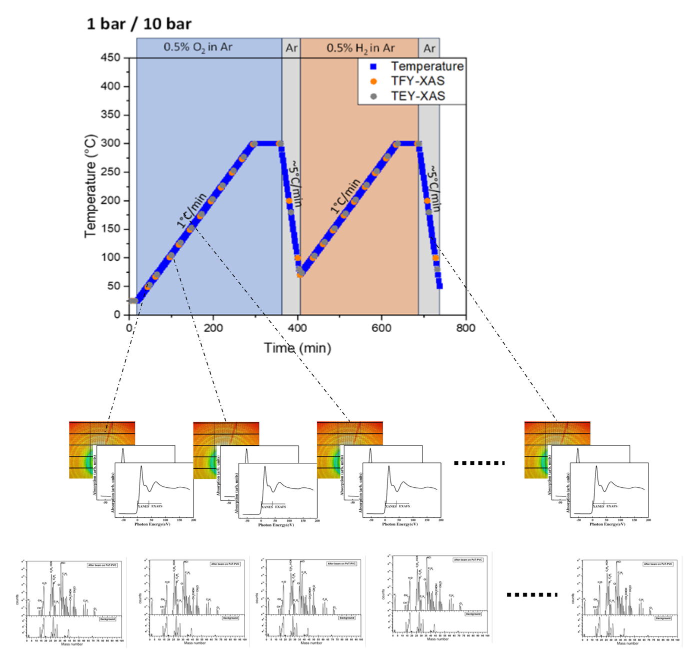

Figure 1: Depiction of source of experimental data from typical operando catalysis experiment

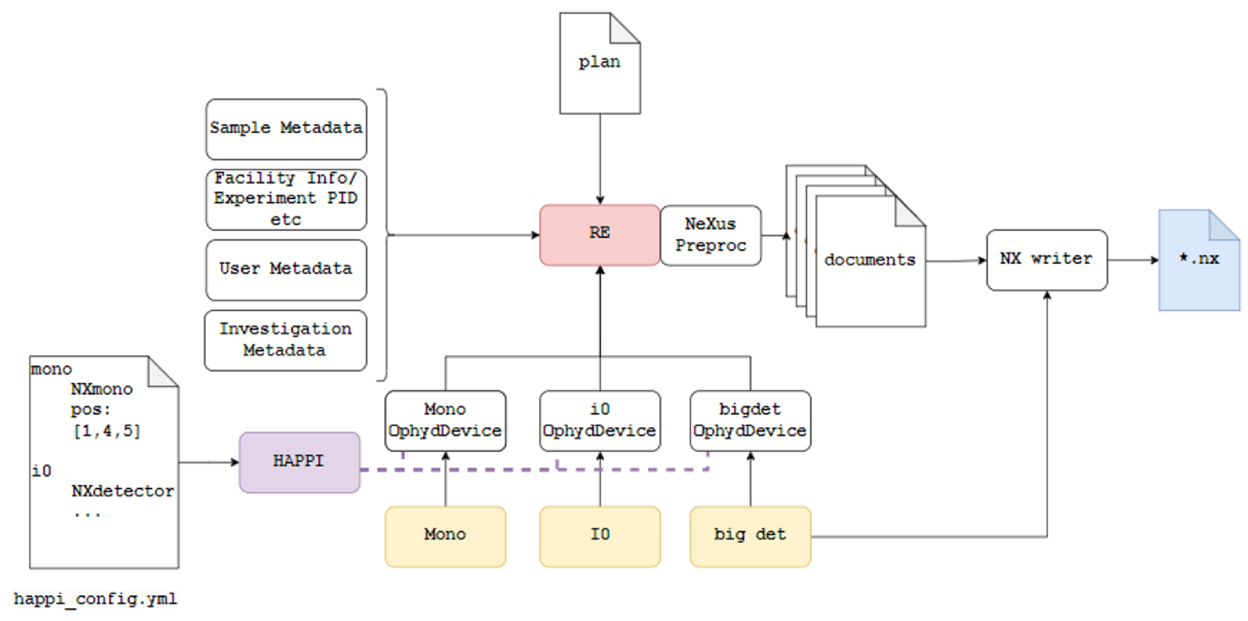

Figure 2: Automated (Meta)data ingestion in bluesky

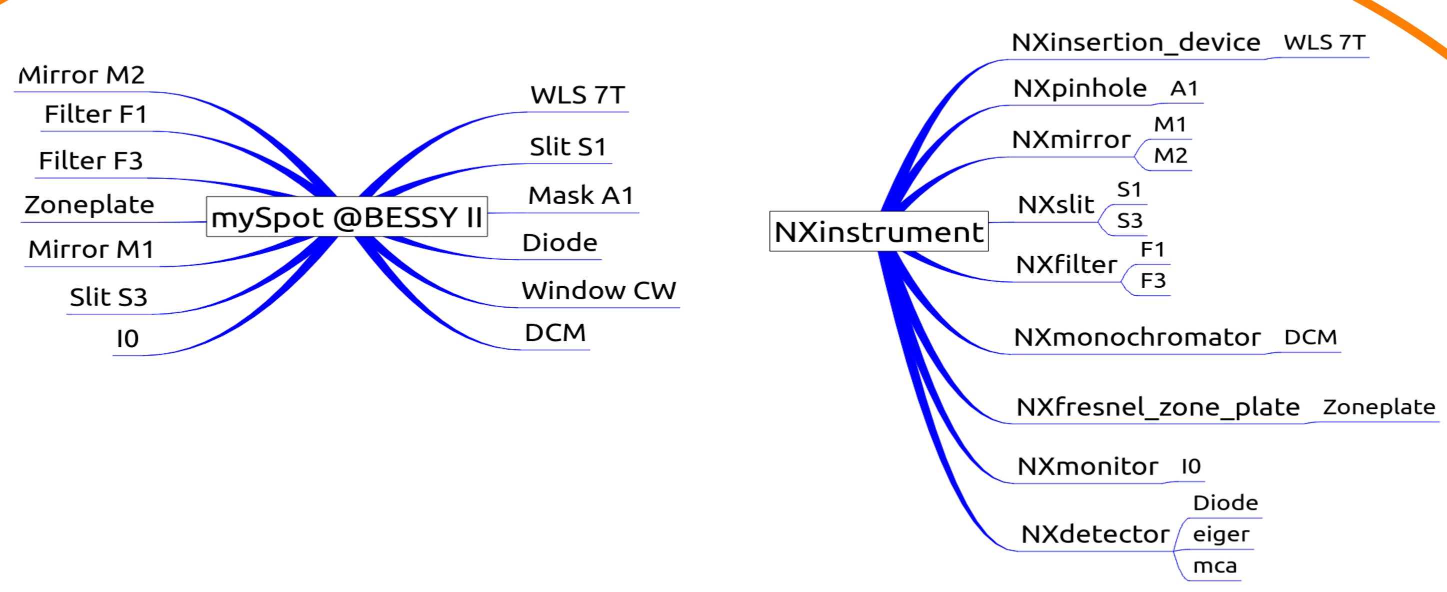

Figure 3: Mapping of mySpot beamline instruments to NeXus base class

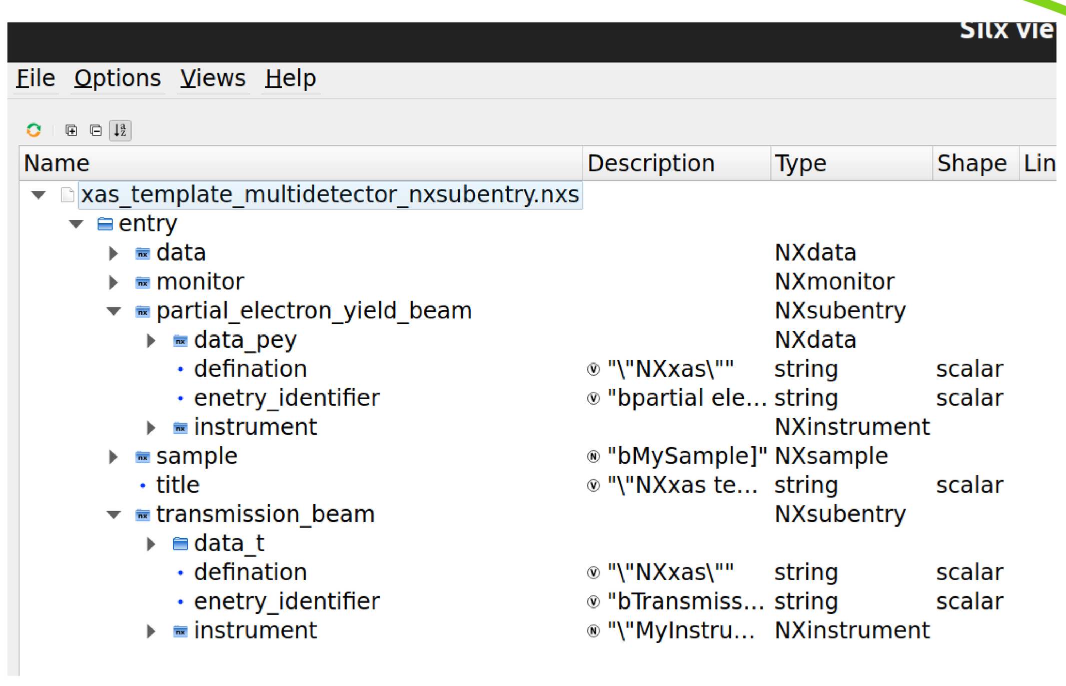

Figure 4: Example structure of NeXus hdf5 file incorporating XAS measurement in multi-mode