ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-TUPD111

CONVERTING EXPERIMENT DATA TO NEXUS APPLICATION DEFINITIONS AT BESSY II

S. R. Patel*[1], S. Brockhauser[2,3], M. Kuehbach[2] R. Müller[1], L. Pielsticker[2, 4], L. Porzio[1], W. Smith[1], D. Tomecki[1], S. Vadilonga[1], J. Viefhaus[1], I. Zizak[1],
[1]Helmholtz-Zentrum Berlin für Materialien und Energie (HZB), Berlin, Germany
[2]Humboldt-Universität zu Berlin, Berlin, Germany
[3]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[4]Max Planck Institute for Chemical Energy Conversion, Mülheim an der Ruhr, Germany

Abstract

In our efforts to achieve FAIR data practices at BESSY II, we are leveraging the NeXus standard, a common data exchange format for data obtained in the fields of neutron, muon, and X-ray science. Two core components of this standard are its base classes and application definitions. NeXus base classes serve as building blocks, offering community-agreed names and data structures for all devices required to run an experiment, including those on the beamline. Built upon these base classes, NeXus application definitions specify the minimal required structures and data elements necessary to represent data from a given experimental technique.

In this work, we present preliminary results from the development of an application definition for a multi-modal experiment conducted at the mySpot beamline of BESSY II. This versatile beamline supports measurements with multiple techniques - XRD, SAXS, XRF, EXAFS and XANES performed simultaneously under in-operando conditions. For the data conversion process, we use pynxtools, a tool designed to facilitate FAIR experimental data. Additionally, we discuss the perspective of this development for the Bluesky NeXus package, developed at BESSY II, which enables the automated export of NeXus-compliant HDF5 files for Bluesky-based experiments and beamlines.

INTRODUCTION

The FAIR principles — Findable, Accessible, Interoperable, and Reusable — are now fundamental to effective scientific data management. In experimental science, where data are often highly complex and diverse, adopting community-driven standards alongside a structured research data management system provides a robust foundation for achieving FAIR compliance. In this work, we present a step in this direction through the use of the NeXus standard [1], a widely adopted data format in neutron, X-ray, and muon research, together with NOMAD [2] – an open research data management (RDM) platform. It supports the full lifecycle of experimental data from acquisition to archiving and reuse.

mySpot Beamline

The mySpot Beamline [3] at BESSY II is designed to deliver a stable beam specifically optimized for the mySpot experiment. Figure 1 shows the photon detection methods in a multimodal endstation for the different X-ray techniques and their respective data acquisition system at this beamline. Since the experiment aims to offer multiple methods simultaneously, the beamline properties can be adjusted to provide the ideal beam for a specific combination of experiments. Key parameters, such as total intensity, divergence, energy resolution, suppression of high harmonics, and stability during scans, can be fine-tuned to meet specific requirements. The selected applications at mySpot are as follows:

- Area mapping of the sample using Fluorescence (XRF) and scattering (SAXS, WAXS), eventually Raman with lower spatial resolution
- Same as above, but with absorption spectroscopy (EXAFS, XANES) measurements on the selected points
- 3D fluorescence mapping using confocal capillary setup
- Fluorescence and/or scattering tomography

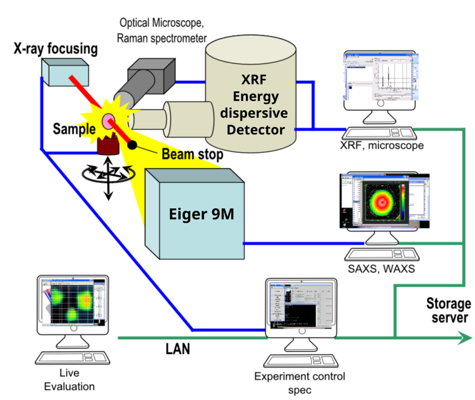

Figure 1: The endstations at mySpot beamline.

In section-Data conversion workflow, we demonstrate the typical workflow that can be smoothly integrated into many of the BESSY II beamlines for exporting the NeXus standard format. The usage of NOMAD facilitates reproducible analysis, is shown in section-Reproducible and Interoperable data analysis. The steps towards automated FAIR data life cycle management at BESSY and conclusions are discussed in section-Towards FAIR research
data at BESSY-II and section-CONCLUSION and OUTLOOK, respectively.

* sonal.patel@helmholtz-berlin.de

20th Int. Conf. Accel. Large Exp. Phys. Control Syst. (ICALEPCS'25), Chicago, IL, USA JACoW Publishing
ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-TUPD111

DATA CONVERSION WORKFLOW

In this work we use the data from the experiment performed at the mySpot beamline. The experiment included simultaneous fluorescence and X-ray diffraction measurements on the liquid sample filled with four cuvettes, as shown in Fig. 2. The experiment studied the correlation between the occurrence of crystallization phases with elemental distribution.

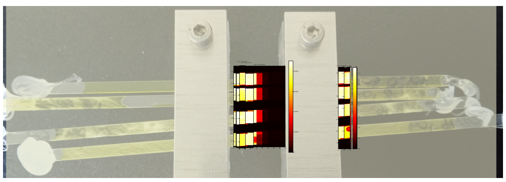

Figure 2: Sample holder with four cuvettes.

Data Reorganization

Currently, for a given experiment, the measurement data are acquired in one directory, with `YYYY_ExperimentName` name, consisting of a different sub-folder for the different experiment technique (ET), as shown in Fig. 3. The data acquisition system is such that it produces one file per scan point per ET and one master file capturing metadata from the beamline control system (`SPEC`[1]). The experiment used in this work had 710 scan points covering the area four cuvettes and at each point on the sample two ET, i.e. Fluorescence and X-ray diffraction, were performed. The Fluorescence measurement data are stored in `mca` files and X-ray diffraction measurement data are stored in the `HDF5` files, in mca and eiger directories, respectively.

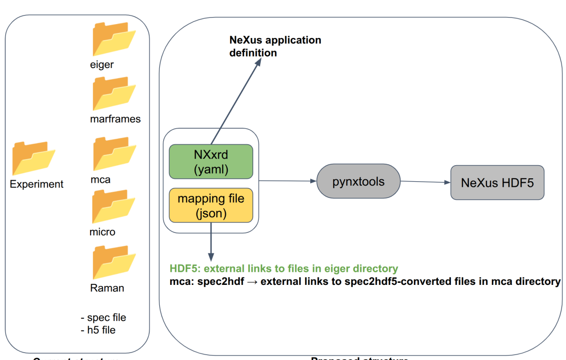

Figure 3: Restructuring mySpot data format.

The proposed data conversion workflow aims to standardize experimental data in the NeXus format to improve interoperability and reuse of data collected at BESSY II beamlines. Depending on the configuration, it produces either one file per ET or one file per experiment, with metadata and data combined in a single `HDF5` file. This approach reduces the complexity and time-consuming effort of handling large numbers of small files.

[1] https://www.certif.com/

NeXus

NeXus [4] is a common data exchange format originally designed for neutron, x-ray, and muon science, which has recently seen an expansion towards a broader spectrum of techniques in materials science. It is being developed as an international standard by scientists and programmers representing major scientific facilities to facilitate greater cooperation in the analysis and visualization of such data. The NeXus standardizes the experiment data, given an ET, in the form of NeXus application definitions, which define the minimum set of community-agreed terms that must be provided in the NeXus data file. Two NeXus application definitions used in this work are NXfluo[2] and NXmonopd[3] corresponding to Fluorescence and X-ray diffraction ET, respectively.

pynxtools

`pynxtools` [5] is an open-source Python package designed to make experimental data FAIR. It parses and integrates various instrument outputs and electronic lab notebook (ELN) formats into an HDF5 file following NeXus application definitions. Besides providing a convenient way to adopt the NeXus format, it addresses the complex challenge of unstructured and non-standardized data in experimental materials science.

Data Conversion

The proposed conversion workflow mainly includes three steps:

Step 1: In the first step large number of small files are combined into one file per ET. This is done by restructuring 1D data from these small files to a 3D array. The first two axes of this 3D array correspond to X, Y positions of the scan on the sample. The third axis is for counts corresponding to either fluorescence measurement or azimuth-integrated diffraction image. In a typical workflow, these two arrays are stored in two separate HDF5 files, one for each ET, along with technique-related metadata.

Step 2: In the second step, configuration files in JSON, one for each ET, are prepared, which store the mapping from HDF5 files produced in the first step to the structure expected by the NeXus application definition. The template for this configuration file for `NXfluo` type of experiment at mySpot is shown in Fig. 4.

Step 3: The third step is to run the conversion for each ET. This can be run from the command line or within a Python environment. For eg. the conversion for a fluorescence type of experiment can be run as shown below from the command line:

```
dataconverter   −−reader json_map
                −−nxdl NXfluo
                −−input-file mySpot_fluo.mapping.json
                −−input-file eln_data_fluo.yaml
```

Here, `eln_data_fluo.yaml` is the input file, which can be used to capture the (meta)data that are available before or

[2] https://manual.nexusformat.org/classes/applications/NXfluo.html
[3] https://manual.nexusformat.org/classes/applications/NXmonopd.html

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-TUPD111


```
{
"/@default": "entry",
"/ENTRY[entry]/@default": "data",
"/ENTRY[entry]/title": "Fluorecence measurement of Cuvvette",
    "/ENTRY[entry]/start_time": "2022-01-22T12:14:12.05018+00:00",
    "/ENTRY[entry]/definition": "NXfluo",
    "/ENTRY[entry]/INSTRUMENT[instrument]/SOURCE[source]/type": "Synchrotron X-ray Source",
    "/ENTRY[entry]/INSTRUMENT[instrument]/SOURCE[source]/name": "BESSY II",
    "/ENTRY[entry]/INSTRUMENT[instrument]/SOURCE[source]/probe": "x-ray",
      "/ENTRY[entry]/INSTRUMENT[instrument]/monochromator/wavelength": {"link":
        "./mySpot_multimodal_fluo_2025-08-14.nxs:/entry_NXfluo/data/energy"},
    "/ENTRY[entry]/INSTRUMENT[instrument]/monochromator/wavelength/@units": "angstrom",
      "/ENTRY[entry]/INSTRUMENT[instrument]/fluorescence/data": {"link":
        "./mySpot_multimodal_fluo_2025-08-14.nxs:/entry_NXfluo/instrument/fluorescence/data"},
      "/ENTRY[entry]/INSTRUMENT[instrument]/fluorescence/energy":{"link":
        "./mySpot_multimodal_fluo_2025-08-14.nxs:/entry_NXfluo/data/energy"},
    "/ENTRY[entry]/SAMPLE[sample]/name": "My wonderful sample",
    "/ENTRY[entry]/MONITOR[monitor]/mode": "monitor",
    "/ENTRY[entry]/MONITOR[monitor]/preset": 0.0,
    "/ENTRY[entry]/MONITOR[monitor]/data": [1, 2, 3],
      "/ENTRY[entry]/DATA[data]/energy": {"link":
        "./mySpot_multimodal_fluo_2025-08-14.nxs:/entry_NXfluo/data/energy"},
    "/ENTRY[entry]/DATA[data]/energy/@units": "eV",
      "/ENTRY[entry]/DATA[data]/data": {"link":
        "./mySpot_multimodal_fluo_2025-08-14.nxs:/entry_NXfluo/instrument/fluorescence/data"},
    "/ENTRY[entry]/DATA[data]/@axes": ["energy"],
    "/ENTRY[entry]/DATA[data]/@signal": "data",
    "/ENTRY[entry]/USER[user]/email": "/user/email",
    "/ENTRY[entry]/USER[user]/name": "/user/name"
}
```

Figure 4: The template of the configuration file for NXfluo experiment at mySpot beamline. Here, the fields under sample and user group are fetched from an ELN file similar to the file shown in Fig. 5.

after the experiment and not captured during the measurement. An example file is shown Fig. 5.

```
{
user:
  email: sonal.patel@helmholtz-berlin.de
  name: Sonal
sample:
  name: "My Wonderful Sample"
process:
  program: bluesky_v0.1.1
  note:
      description: "This program does the
        azimuthal integration of the data
        acquired at mySpot"
}
```

Figure 5: An example input file to inject metadata not acquired during the measurement.

For specialized experiments, such as multi-modal studies where several different methods are performed simultaneously on the sample, `pynxtools` can be used to generate NeXus-compliant files by defining NeXus application definition that bundle each ET in a separate NXsubentry. This approach can then be consistently applied to that experiment type. As a simple demonstration, we use a cuvette scan experiment and create an application definition named `NXmultimodal`. Following the recommendation[4] of the NeXus International Advisory Committee (NIAC) on structuring NeXus files for multi-method data, this application definition combines `NXfluo` and `NXmonopd` as two `NXsubentry` groups within a single `NXentry`, producing one file per experiment. An example NXmultimodal definition file and the corresponding NeXus data file are attached in the Supplementary Material (See APPENDIX).

REPRODUCIBLE AND INTEROPERABLE DATA ANALYSIS

Once the experiment data are obtained in the standard format, it is also crucial to store and provide these in a catalog or archive that provides links to the (meta)data relevant for the experiments reproducibility and complies to the data accessibility policy. Within the BESSY II and HZB user community sample generating laboratory processes are captured in the NOMAD data management for materials science research environment. In addition, existing NOMAD[5] data are leveraged to synthesize promising samples. NOMAD is a web-based platform that fulfills the needs of visualization to reuse with minimal or no effort once data are acquired

[4] https://manual.nexusformat.org/rules.html
[5] https://nomad-lab.eu/

20th Int. Conf. Accel. Large Exp. Phys. Control Syst. (ICALEPCS'25), Chicago, IL, USA JACoW Publishing
ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-TUPD111

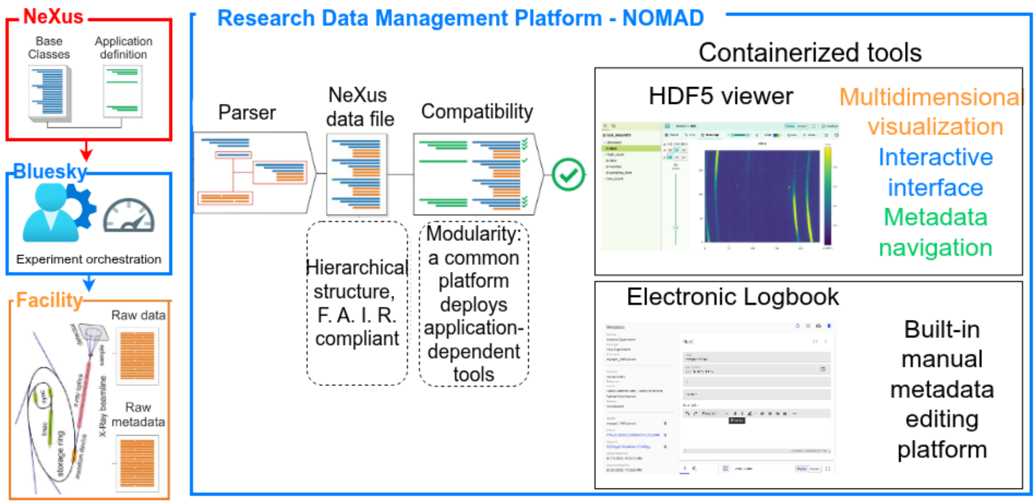

Figure 6: Reproducible and interoperable data analysis through Open RDM platform.

in NeXus format. User-friendly tools to visualize, process, share, and reuse these standard formats reduce the barrier of adopting these standards for various user communities of Synchrotron facilities and leverage the advantage of such an interoperable data format.

TOWARDS FAIR RESEARCH DATA AT BESSY-II

While the data conversion workflow described previously can be deployed across the majority of the BESSY II beamlines, providing an immediate solution also for legacy data, it represents only one component of a larger vision. To achieve a comprehensive, automated FAIR research data lifecycle, BESSY II follows a strategic development plan.

NeXus enshrined into standardized experiment data acquisition and workflow (Bluesky framework) as well as instrument data formats, ensure interoperability in Open Research Data Management Platforms, such as NOMAD as shown in Fig. 6. This approach ensures a holistic transition, from the initial experiments to data analysis and ML tools to the eventual reuse of the data for planning the next experiments and generating samples of interest.

Having NOMAD as a storage of experimental NeXus data nicely fits the integrated approach of the planned materials discovery facility BESSY III. The data conversion workflow described in Section-Data conversion workflow can be directly run on the NOMAD platform. We envision sending the mySpot data from the current structure to the NOMAD server set up as an NOMAD Oasis at the beamline. This workflow is demonstrated by running it in the Jupyter notebook through the NOMAD Remote Tools Hub (See supplementary material in APPENDIX).

Auto-Export to NeXus via Bluesky Framework

Currently, the data aquisition part has been at an advanced stage with a fast-paced roll-out of Bluesky[6]-based [6] device integration and experimental orchestration at nine BESSY-II beamlines. Important novel sub-module of the Bluesky deployment at BESSY-II beamlines is `blueksy_nexus`[7]. This package integrates the NeXus base classes, which provide community-agreed names and standardized data structures for the devices used in the experiments. The package utilizes the Pydantic data validation library to ensure that all device data schemas are consistent with the NeXus standard. A more detailed discussion of this package is available in a separate conference paper [7]. The current Bluesky data acquisition workflow at the BESSY II beamlines is illustrated in Fig 7.

Besides this, NeXus writer of the `bluesky_nexus` ensures the data acquisition in one single file along with metadata. This makes us reduce the first step of converting many small files to one HDF5 file in the data conversion workflow. Moreover, the development of a custom reader in pynxtools, and having it as a sub-module within bluesky deployment at BESSY-II beamlines, will be capable of auto-export of experiment data into to NeXus application definition.

Mapping of all BESSY-II Experiment Techniques to NeXus

A preliminary survey of existing experimental techniques (ET) at BESSY II beamlines is presented in Fig. 8.

The inner distribution of this nested pie chart illustrates the mapping of BESSY II ET to NeXus application definitions. This analysis reveals that over 50% of the ETs have corresponding data standards within NeXus, which can be adopted with minimal modifications. The ongoing and rapid development of the Bluesky project at BESSY II is expected to facilitate a FAIR (Findable, Accessible, Interoperable, Reusable) data lifecycle for a significant number of beamlines. However, the development of standardized data analysis tools remains in its early stages.

[6] https://blueskyproject.io/

[7] https://codebase.helmholtz.cloud/hzb/bluesky/core/source/bluesky_nexus

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-TUPD111

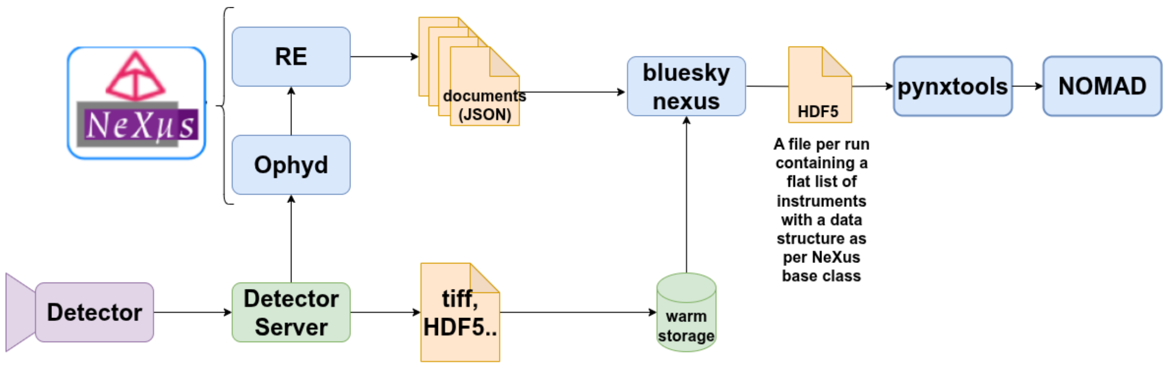

Figure 7: Possible integration of pynxtools to Bluesky data acquisition workflow for auto-exporting experiment data into NeXus application definition.

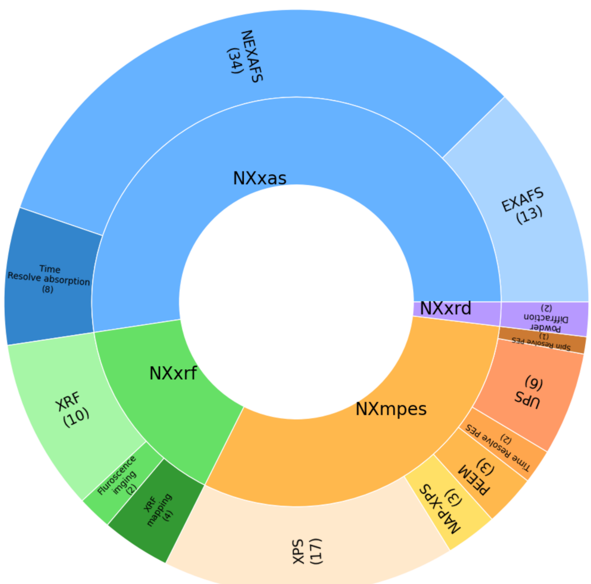

Figure 8: Mapping of current experiment techniques of BESSY-II to NeXus standard. The number in the bracket indicates the number of beamlines performing given ET.

CONCLUSION AND OUTLOOK

The multimodal capabilities and diverse workflow options of the mySpot beamline [3] at BESSY II offer an ideal testbed for developing and validating NeXus application definitions. This framework enables seamless extension to experimental techniques and workflows across other BESSY II beamlines. The implementation of NeXus-formatted experimental data storage within NOMAD oasis represents a significant advancement in FAIR data practices, aligning with national and European Open Science Cloud initiatives (EOSC[8]). Moving forward, HZB remains committed to advancing materials discovery through data-driven scientific approaches that both leverage existing resources and contribute to the broader scientific community.

[8] https://eosc.eu/

ACKNOWLEDGEMENTS

The author thanks the FAIRmat Area B task force and many of the NIAC members for their prompt discussions and support in leveraging NeXus standards at BESSY-II.

APPENDIX

Supplementary material: `https://github.com/hz-b/NeXus_data_examples/tree/main/2025_ICALEPCS`

REFERENCES

[1] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship", *Sci. Data*, vol. 3, p. 160018, Mar. 2016. doi:10.1038/sdata.2016.18

[2] M. Scheidgen *et al.*, "NOMAD: A distributed web-based platform for managing materials science research data", *J. Open Source Softw.*, vol. 8, no. 90, p. 5388, Oct. 2023. doi:10.21105/joss.05388

[3] I. Zizak, "The mySpot beamline at BESSY II", *J. Large Scale Res. Facil.*, vol. 2, p. A102, Dec. 2016. doi:10.17815/jlsrf-2-113

[4] M. Könnecke *et al.*, "The NeXus data format", *J. Appl. Crystallogr.*, vol. 48, no. 1, pp. 301–305, Jan. 2015. doi:10.1107/s1600576714027575

[5] S. Shabih *et al.*, "Pynxtools: A Python Library for NeXus-Compliant Experimental Data Conversion and Integration with NOMAD Platform", Zenodo, May 2025. doi:10.5281/zenodo.15341365

[6] D. Allan *et al.*, "Bluesky's Ahead: A Multi-Facility Collaboration for an a la Carte Software Project for Data Acquisition and Management", *Synchrotron Radiat. News*, vol. 32, no. 3, pp. 19–22, May 2019. doi:10.1080/08940886.2019.1608121

[7] D. Tomecki *et al.*, "Bluesky NeXus: a solution for NeXus-compliant data acquisition in Bluesky", presented at ICALEPCS'25, Chicago, IL, USA, Sep. 2025, paper TUPD110, this conference.