Journal of Physics: Conference Series **3010** (2025) 012143 doi:10.1088/1742-6596/3010/1/012143

Ensuring FAIR data practices by Leveraging NeXus standards at BESSY II

Sonal R. Patel[1], Hector P. Ponce[1], Ivo Zizak[1], Heike Görzig[1], Rolf Krahl[1], Simone Vadilonga[1], William Smith[1], Jens Viefhaus[1], Gerrit Guenther[1]

[1]Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany

E-mail: sonal.patel@helmholtz-berlin.de

Abstract.

In alignment with the FAIR[1] (Findable, Accessible, Interoperable, and Reusable) data principles, BESSY II is embarking on a transformative journey to enhance scientific research capabilities through the integration of standardized device interfaces for metadata capture and the adoption of NeXus[2] as the common data format. NeXus is a common data format specifically designed for neutron, X-ray, and muon science, making it ideally suited for the diverse experimental techniques employed at BESSY II.

A cornerstone of this initiative is the ROCK-IT (Remote, Operando Controlled, Knowledge-driven, and IT-based) project, funded by the Helmholtz Association. This project not only exemplifies the practical application of FAIR principles at BESSY II but also serves as a demonstrator of how advanced IT can significantly enhance control and insights with a focus on the experiments.

We present one of the efforts at BESSY II in curating FAIR data. For this, we are leveraging NeXus data format standards at mySpot beamline of BESSY II within the context of the ROCK-IT project. NeXus standards play a dual role of not only defining all data relevant to the beamline and the experiment but also defining the standards in the form of application definitions which helps in exchanging data between applications. We will discuss the progress of our efforts concerning the mySpot beamline, which is a multi-mode beamline designed to perform different techniques - XRD, SAXS, XRF, EXAFS, and XANES - simultaneously at the sample position.

1 Introduction

GO FAIR[1] is a community-led, stakeholder-driven initiative focused on implementing the FAIR data principles, which promote making data Findable, Accessible, Interoperable, and Reusable (FAIR). This initiative offers an open and inclusive ecosystem for individuals, institutions, and organizations. The acronym and principles [1] were first defined by the Force11[2].

To promote the transition toward an open science system, the Council of the European Union emphasizes[3] that "the opportunities for the optimal reuse of research data can only be realized if data are

[1]https://www.go-fair.org/go-fair-initiative/

[2]https://force11.org/info/guiding-principles-for-findable-accessible-interoperable-and-re-usable-data-publishing-version-b1-0/

[3]https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf

consistent with the FAIR principles (findable, accessible, interoperable, and reusable) within a secure and trustworthy environment". Following that, on February 19, 2020, priorities to implement were stated in the European Data Strategy[4].

At BESSY II, we are working toward making our scientific data as FAIR as possible by leveraging the NeXus[5] standards. NeXus[2] is a common data exchange format for neutron, X-ray, and muon science. Between 1994 and 1996 three parallel developments - Mark Könnecke (Paul Scherrer Institute, Switzerland), Jon Tischler and Mitch Nelson (Oak Ridge National Laboratory, USA), and Przemek Klosowski (National Institute of Standards and Technology, USA), were the foundation stones of the NeXus. The first NeXus API was released in late 1997[6]. Since then scientists and programmers representing major scientific facilities are developing it as an international standard to facilitate greater cooperation in the analysis and visualization of neutron, X-ray, and muon data. The development of NeXus standard is overseen by the NeXus International Advisory Committee (NIAC)[7].

NeXus standards have been attempted to adopt at several facilities[8], mainly X-ray and neutron facilities [5]. As a part of the ESRF upgrade program [6], it was planned to use NeXus/HDF5 for storing all raw data. Currently, all experiments from all beamlines produce HDF5 files following NeXus conventions [7]. As of 2023, 27 out of 37 Diamond Light Source (DLS) beamlines, write files using NeXus standards (Private communication with DLS data acquisition group). The remaining beamlines that do not use NeXus standards are mainly those that have commercial instruments and are unlikely to switch to NeXus in the short term.

As part of the effort toward curating FAIR data, the mySpot beamline at BESSY II was chosen within the context of the ROCK-IT[9] project. ROCK-IT (Remote, Operando Controlled, Knowledge-driven, and IT-based) is a Helmholtz-funded initiative aimed at creating the essential tools for automating and enabling remote access to in-situ and operando catalysis experiments, accessible to users of all skill levels. Key objectives include:

- Remote experimental access for users.
- Experiment automation, including robotics integration.
- Efficient data management.
- Near real-time analysis and optimization through machine learning (ML) and artificial intelligence (AI).

Catalysis operando experiments, which currently lack automation, have been selected as a pilot case for development within ROCK-IT. Catalyst optimization involves exploring a wide range of experimental and material parameters, and automating these processes can greatly improve the efficiency of the development pipeline. Catalysts are integral to the global economy and are crucial in addressing climate change challenges.

2 FAIR structuring of experiment data

The mySpot beamline is designed to deliver a stable beam specifically optimized for the mySpot experiment. The schematic of the mySpot beamline is shown in Figure 1. Various optical devices are employed based on the experiment's needs. 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 the specific requirements.

Figure 2 shows the different X-ray techniques and their respective data acquisition systems at mySpot beamline. 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

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0066from=EN
[5] https://manual.nexusformat.org/index.html
[6] https://manual.nexusformat.org/motivations.htmlmotivationsfornexus
[7] https://www.nexusformat.org/NIAC.html
[8] https://www.nexusformat.org/Facilities.html
[9] https://www.rock-it-project.de/

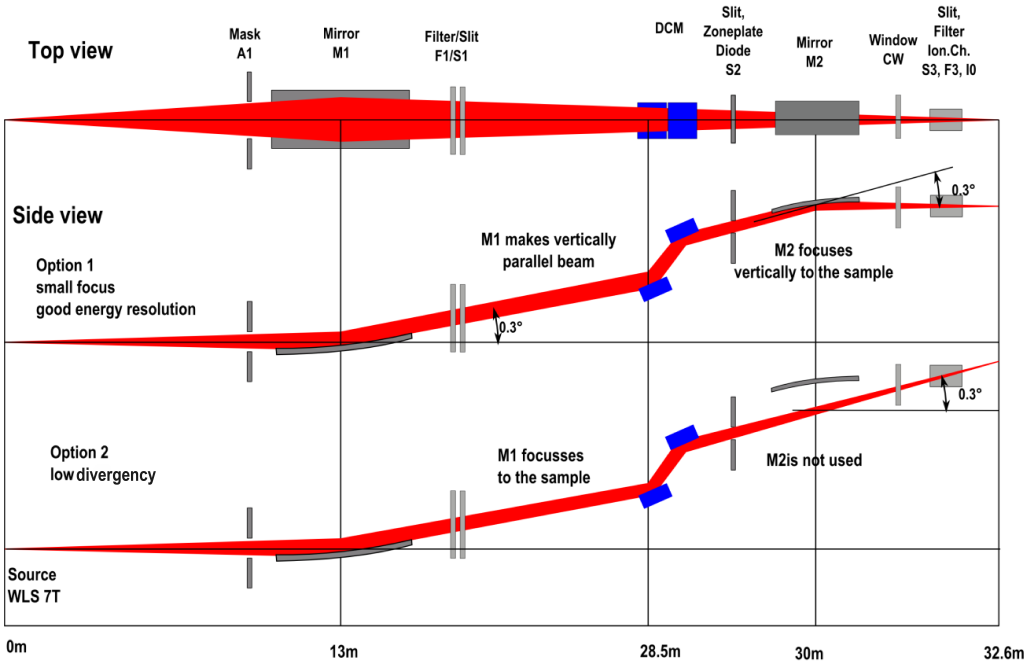

Figure 1: The layout of mySpot beamline at BESSY II

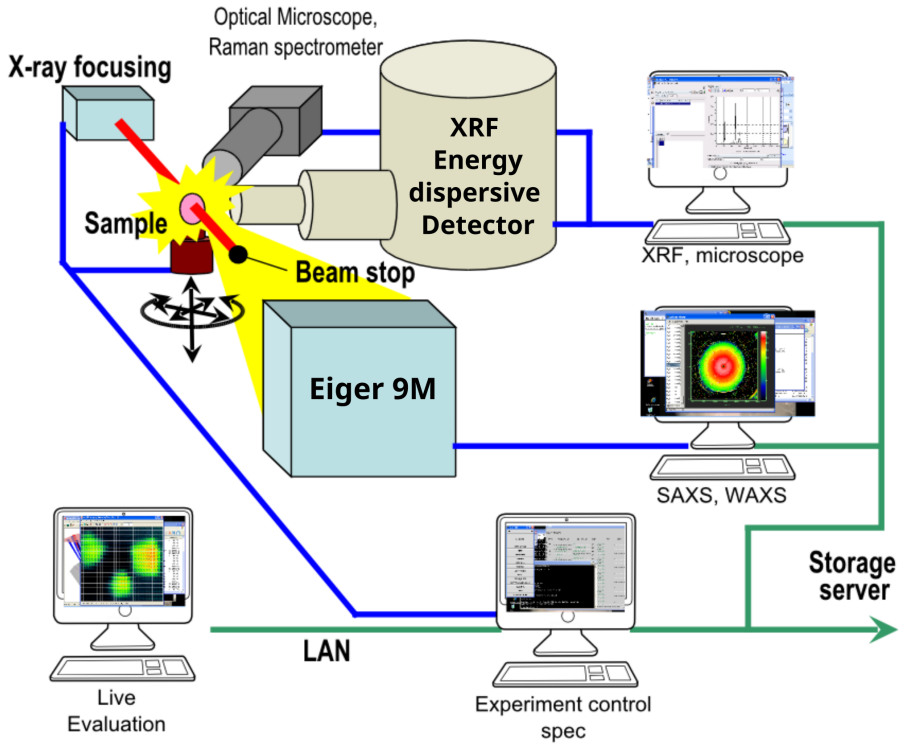

Figure 2: The endstations at mySpot beamline

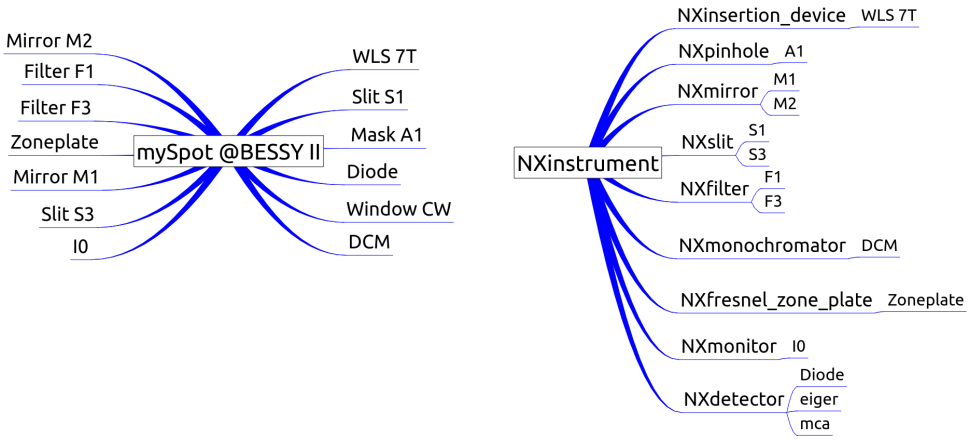

Figure 3: Mapping of instruments at mySpot to NeXus base classes

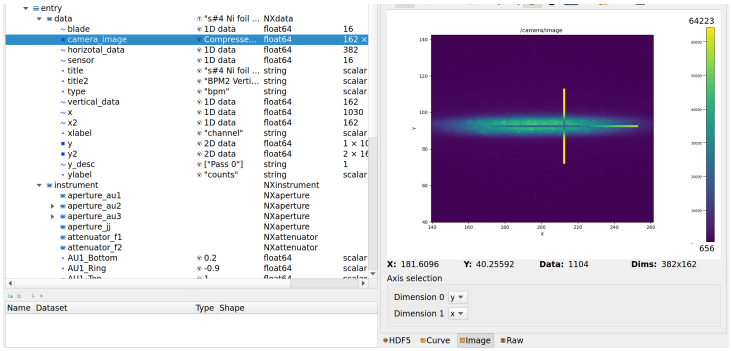

Figure 4: Quick look plot of NXdata

- 3D fluorescence mapping using confocal capillary setup
- Fluorescence and/or scattering tomography

A valuable feature of the mySpot[3] beamline is that it provides users with a choice of several analysis methods simultaneously. However, this poses a challenge for FAIR data capture. The first step toward FAIR data at any given beamline is to populate the instruments with metadata. This is done 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. This structure depends on the technique. Hence, multiple techniques make the NeXus data structure longer compared to a beamline which performs only one technique at a time. 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 any given beamline. An example of FAIR-structured experimental data is shown in Figure 4. Enriching the instrument metadata helps to reproduce ("R" of FAIR) the state of the beamline at a given time and structuring the experiment data following NeXus standards makes the data interoperable ("I" of FAIR).

3 Methods

Two independent efforts are underway to produce an HDF5 file with the same NeXus data structure. One effort makes use of a tool called NexusCreator, while the other uses a Python library called Bluesky.

NexusCreator: NexusCreator[10] is a Javascript/Python-based conversion tool that converts the data to NeXus data format in an HDF5 file given the structure in an ASCII file. As this is an offline data

[10] https://gitlab.com/helmholtz-berlin/nexuscreator

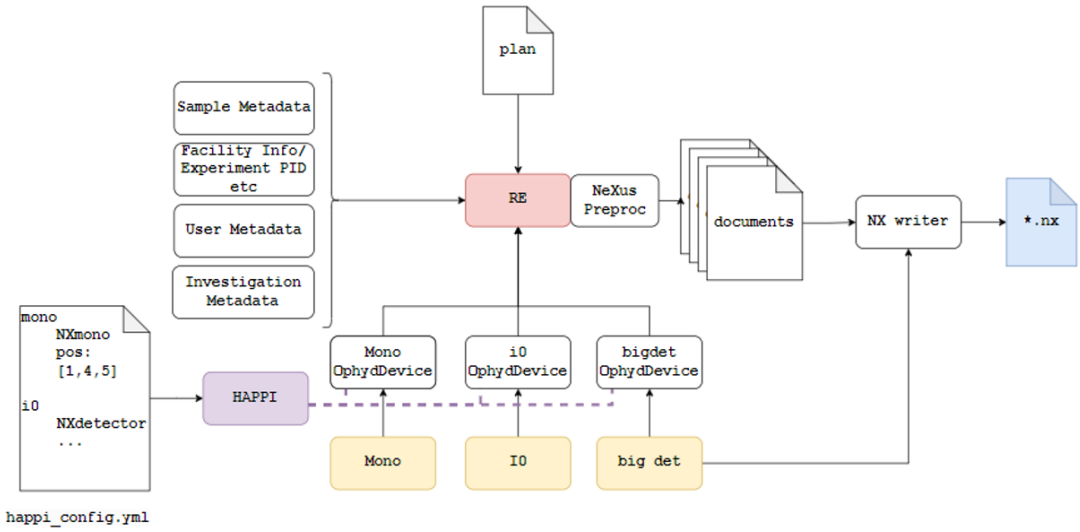

Figure 5: The flowchart of NeXus data export within bluesky environment

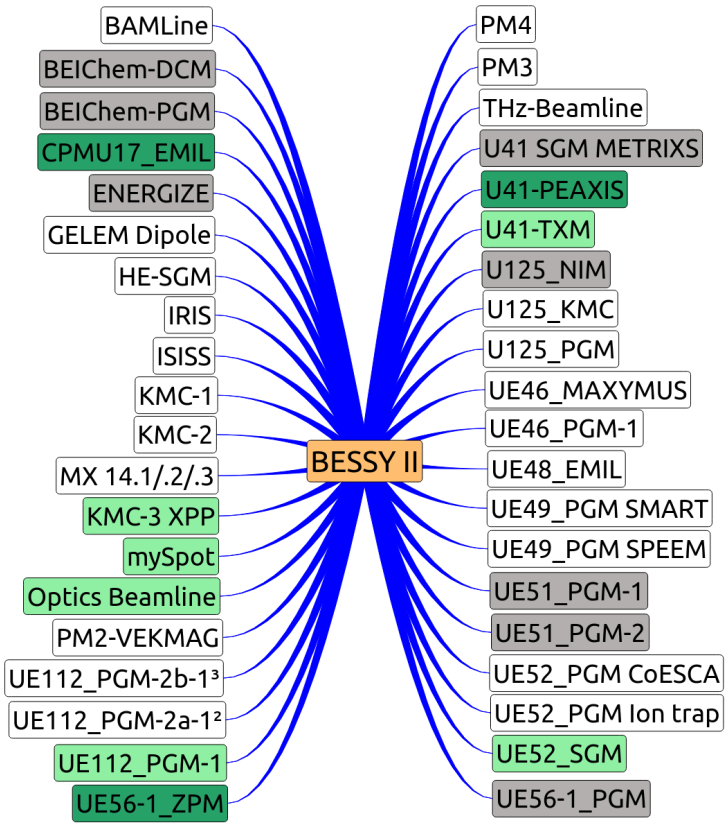

Figure 6: Status of the beamlines capable of producing NeXus data file as of 2024

conversion tool, it can be used as an additional plug-in at any given beamline. Therefore, this effort does not require changes to the existing control and data acquisition infrastructure at the beamline.

bluesky: bluesky[4] is a Python library for experiment control and collection of scientific data and metadata. It can capture and organize the (meta)data to facilitate reproducibility and searchability. Figure 5 shows the flowchart of planned automated (meta)data capture using HAPPI[11] - Heuristic Access to Positions of Photon Instruments. It helps to create and index objects, search through them, and provide relevant metadata based on the object type.

NexusCreator approach will work with any beamline but requires considerable effort for each experiment of each beamline and adds maintenance overhead. While bluesky approach requires changing the control system of a beamline/experiment, but then NeXus files can be created with little additional work and maintenance. Since both efforts are in the development phase, we decided to continue both efforts so that the writing of NeXus HDF5 files at BESSY II beamlines can be maximized by optimizing technical solutions.

4 Conclusions and future outlook

Statistics of the BESSY II beamlines, as of 2024, capable of producing HDF5 files in the NeXus data format, by using one of the two methods described in section 3, are shown in Figure 6. Colors in the figure imply:

- Dark green (3): Conversion to NeXus data format using NexusCreator
- Light green (6): Automated export, thanks to fast-paced bluesky development
- Gray (7): Upcoming beamlines

Additionally, at mySpot beamline, NeXus structure producing NeXus data format using NexusCreator is also maintained to cross-check the files produced using bluesky. The refactored NXxas[12] application definition will be used at mySpot in the near future with blueksy approach. To have an online auto-export of experimental NeXus data into an HDF5 file, we are first prioritizing the structure of the NXinstrument of the beamlines supporting bluesky. Moreover, FAIR data structuring for other beamlines will be started after successful testing of one of these methods at 9 beamlines shown in Dark/Light green in Figure 6.

References

[1] M. Wilkinson et al., Sci Data 2016, 3, 160018 (doi: https://doi.org/10.1038/sdata.2016.18)

[2] M. Konnecke et al., J. Appl. Cryst. 2015, 48, 301-305 (doi: https://doi.org/10.1107/S1600576714027575)

[3] Helmholtz-Zentrum Berlin für Materialien und Energie, Journal of large-scale research facilities, 2016, 2, A102 (doi: https://doi.org/10.17815/jlsrf-2-113)

[4] D. Allen et al., Synchrotron Radiation News, 2019, 32:3, 19-22, (doi: https://doi.org/10.1080/08940886.2019.1608121)

[5] Brief history of NeXus

[6] ESRF upgrade program: ESRF-EBS

[7] R. Dimper et al., Synchrotron Radiation News, 2019, 32:3, 7–12, (doi: https://doi.org/10.1080/08940886.2019.1608119)

[11] https://pcdshub.github.io/happi/v2.5.0/

[12] https://github.com/nexusformat/definitions/pull/1352