ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2AO04

EXPERIMENTAL DATA TAKING AND MANAGEMENT: THE UPGRADE PROCESS AT BESSY II AND HZB*

R. Müller†, Heike Görzig, Gregor Hartmann, Klaus Kiefer, Ruslan Ovsyannikov, William Smith‡, Simone Vadilonga, Jens Viefhaus, HZB, Berlin, Germany
Daniel Allan, BNL, Upton, NY, USA

Abstract

The endeavor of modernizing science data acquisition at BESSY II started 2019. Significant achievements have been made: the Bluesky software ecosystem is now the accepted framework for data acquisition, flow control and automation. It is operational at an increasing number of HZB beamlines, endstations and instruments. Participation in the global Bluesky collaboration is an extremely empowering experience. Promoting FAIR data principles at all levels developed a unifying momentum, providing guidance at less obvious design considerations. Now a joint demonstrator project of DESY, HZB, HZDR and KIT, named ROCK-IT (Remote Operando Controlled Knowledge-driven, IT-based), aims at portable solutions for fully automated measurements in the catalysis area of material science and is spearheading common developments. Foundation there is laid by Bluesky data acquisition, AI/ML support and analysis, modular sample environment, robotics and FAIR data handling. This paper puts present HZB controls projects as well as detailed HZB contributions to this conference into context. It outlines strategies providing appropriate digital tools at a successor 4th generation light source BESSY III.

INTRODUCTION

From the origins of BESSY II there was no holistic control system environment, spanning accelerator, beamlines and instruments. Fragmentation of data acquisition (DAQ) tools led to specific, intertwined solutions, that blocked sustainable development plans. A concept for modernizing science data acquisition at BESSY II became an obvious need. A process has been started and plans have been documented in 2019 [1]. Due to the existing diversity of device access and software tools, EPICS channel access could be identified as necessary common denominator for integrating simple, complex and legacy systems (Fig. 1). A clear picture for flow control and the data handling software stack was far less obvious then.

Today significant progress has been made in software core areas, leveraging Bluesky [2] and ophyd enabled device integration and abstraction. Achieved status, and HZB relations to wider Bluesky community developments [3], are at the core of this update paper. Specific HZB contributions to Bluesky are presented in other submissions to this conference [4–8].

* Work funded by BMBF and Land Berlin
† roland.mueller@helmholtz-berlin.de
‡ william.smith@helmholtz-berlin.de

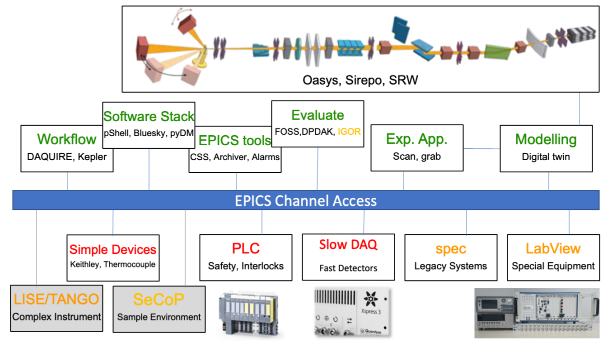

Figure 1: General device, legacy sub-system and digital twin integration guidelines of the 2019 modernisation proposal.

Today, other topics addressed in Ref. [1], like fly scan and machine learning are emphasized within a broader BESSY II facility upgrade plan BESSY II+, dubbed BII+. That plan aims at an elevated level of automation, lowering access barriers for non-expert users, higher instrument science data output and seeks to allow for more complex operando studies while being more sustainable (Fig. 2).

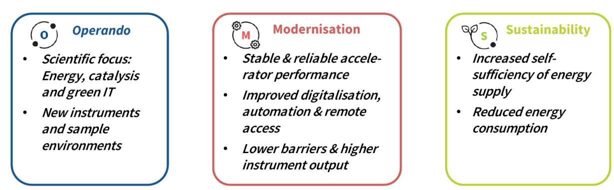

Figure 2: Intended goals of the BII+ facility re-adjustment plan, in terms of scientific focus, general resource saving, modernisation and climate protection.

In an unexpected turn of events, HZB was the victim of a ransomware attack in June 2023. While this caused a challenging set-back, the pause in operation and rebuilding process presents an unprecedented opportunity to speed up the implementation of the more unified controls infrastructure at BESSY II.

STATUS OF BLUESKY AT BESSY II

Unusual compared to other facilities, progress in the experimental usability of the facility relies very much on grass root type initiatives from engineers and scientists, due to a lack of central steering and under-staffing of experimental control support.

Deployment and Roll Out

Following initial success at the accelerator, EMIL beamlines and SISSY end stations in 2021, a team formed, en-

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2AO04

abling the rollout of Bluesky to another 7 beamlines and 2 more end stations (see Fig. 3). Of these, 3 were already running automated sequences using spec or other tools. In all cases the roll-out was mainly supported by interested beamline scientists and engineers. This had the advantage that staff involved were highly motivated, but presented some challenges in standardisation. These were overcome with regular meetings and the use of common repositories for ophyd devices and Bluesky plans. Work continues to develop a development and deployment workflow, based on containers.

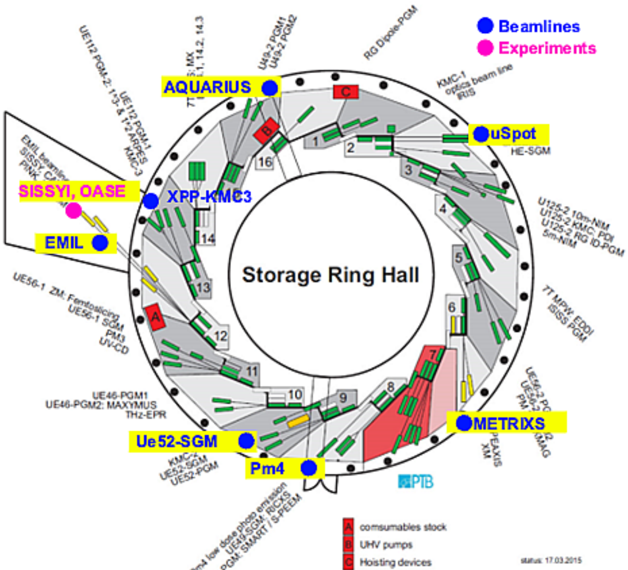

Figure 3: Bluesky rollout overview.

Deploying to new beamlines was much easier than to end stations, because they share common components. Endstations required more work (Fig. 4), and also often necessitated an initial deployment of EPICS or integration of e.g. detectors using AreaDetector. The support of the Bluesky community and the availability of skilled external developers present on Slack work spaces, was invaluable in helping to solve encountered problems quickly.

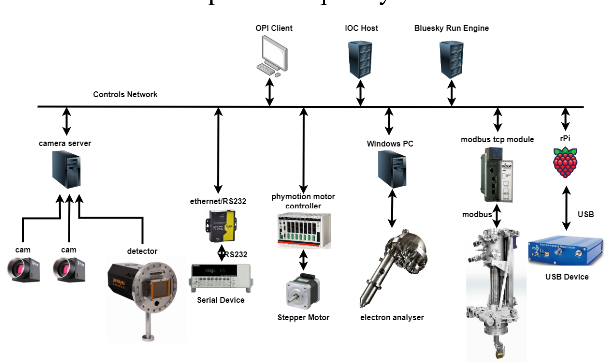

Figure 4: Complex device variety at instruments.

Regular updates from this team to the rest of BESSY II helped drive adoption in other areas. Undulator commissioning has evolved from using dedicated Perl scripts, through a home-grown Python interface using pyEPICS, and have now been integrated with ophyd. Hard to maintain scripts, which had a single developer, were refactored to use Bluesky plans. EPICS and Bluesky solutions are beginning to be developed also for the undulator magnet measurement lab. A team of motion control engineers adopted Bluesky in their lab to perform commissioning tasks of motors and motor controllers. This was used in the development of an update to the Phytron EPICS Motor record.

A team in the accelerator group continued work on the development of a digital twin for the machine [7]. Commissioning plans were developed using Bluesky that could interact with either real or simulated accelerator devices using a consistent data model.

The Bluesky Collaboration

The global Bluesky collaboration has rapidly increased scope and extended to many research centers and universities world wide. The need of orchestration of major developments and securing of resources and commitment at the various labs became obvious to the core development team. Thus a governance structure [9–11] has been created in order to guarantee the long term success of the project. It comprises a project advisory board (PAB), focused on strategic and managerial issues, and a technical steering committee (TSC), dealing with solutions best suited to changing requirements of contributing and beneficiary projects. HZB already sits on the TSC and strives to become part of the PAB soon.

BESSY II is the first European lightsource to adopt Bluesky extensively in production, i.e. user beam mode, but it supports and learns from a rich community in Europe and elsewhere. Diamond Light Source contributes heavily to the project and will adopt Bluesky for Diamond II, replacing GDA and Malcom e.g. for flyscanning. Bluesky is used extensively by our FHI/MPG[1] stakeholders, controlling reactors creating samples for catalysis research. Within the FAIRmat project FHI leads the work package D5: Configurable Experiment Control System [12]. There a tool matching small lab needs, called CAMEL (Control Application for Measurements, Experiments and Laboratory Systems) [13], has been developed with Bluesky at it's core. CAMEL aims to ease the process of setting up experiments typical for this smaller environment. PSI has recently made impressive progress exploring using Bluesky for experiment control at beamlines to be operated at the upgraded SLS II.

IHEP in Beijing developed an experiment control system called Mamba [14], tailored to the use cases expected at the 4th generation light source HEPS. Despite somewhat isolated from other projects, IHEP have made great strides in developing their experiment control and data acquisition system Mamba [14] around the Bluesky core. Data management [15] is handled somewhat differently than at other labs, nevertheless the ideas behind are worth a second thought. There is also growing interest from other labs based on controls tools other than EPICS, notably the TANGO sites Soleil,

[1] Fritz-Haber Institute, Max Planck Gesellschaft, Berlin

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2A004

MAX IV and DESY. In all cases, the responsive community and modular nature of Bluesky prove invaluable. In the attempt to offer a wider and persistent communication platform, HZB and HZDR have worked together to provide an instance of Mattermost that it's hoped in the near future will become the home platform to the Bluesky community, replacing the free instance of Slack which now looses messages after 90 days.

HZB Contributions

Within the large Bluesky collaboration there is a wide span of opportunities to contribute, ranging from raising questions, that help sharpen the mind, to maintenance of whole packages. At BESSY the amount of basic integration work on site is huge and we are kind of short in man power, thus visibility within the community is limited yet. Nevertheless it is worthwhile to name what we are working at

- Bluesky Web Client and GUI to the queue server [4]
- SECoP integration for the ophyd hardware abstraction layer [5]
- FAIR management of beamline configuration data [8]
- Advancements in digital twins at BESSYII [6, 7]

For details see the papers contributed to this conference.

UNIFYING MOMENTUM OF FAIR DATA PRINCIPLES

In an unregulated environment, like at BESSY II, agreement on standards is not easy and time consuming. Even if **FAIR** data principles are imposed by funding agencies for good reasons, for HZB it is less a nasty duty then a great motivating metric to standardisation opportunities.

Findable **F**: One of the main objectives of the European Open Science Cloud (EOSC) is to make Open Data from photon and neutron facilities **F**AIR. Some portal, like the European Photon and Neutron Open Data Search Portal [16] should implement the **F**(indable) part of FAIR via a federated search engine. The search results should provide a link to the landing page of the data DOIs through which the other data services for data downloading, analysis, notebooks and simulation can be accessed. Leveraging as much self-documentation e.g. via control system historical databases, electronic lab books and aggregation and analysis of system logs are certainly a good basis for a comprehensive meta data description. On top, proper linking of all data sources involved, from sample tracking, experiment conditions and execution to data acquisition, analysis and interpretation are required to complete metadata sets. Bluesky has the documentation of the data acquisition process built in.

Accessible **A**: to work with the scientific data aquired requires a secure WEB portal to all beamtime data. Typically, one has to authenticate against a single sign-on mechanism (e.g. keycloack) and authorize with your digital user office beamtime priviliges. At BESSY this is a missing link that hampers data move from the *hot area* at the instrument to a *warm intermediate* for analysis and process tuning and finally to the *long term storage* (e.g. ICAT, SciCat) where it should be publicly retrievable when the embargo time is over.

Interoperable **I**: any application or tool can only work with scientific data if the data sets are self describing in a standard way. Accepted standards for acquisition and storage are still emerging (NeXus, HDF5). Since for e.g. NeXus there is already a standardisation committee (NIAC) in place, funding agencies start to have an eye on instruments capabilities to generate NeXus compliant data. Within the Bluesky collaboration work towards standards for various synchrotron experiments was succesful (NXxas) or is on a good path (NXxpcs, NXptycho, NXtomo...). NOMAD of the FAIRmat consortium offered another starting point.

Reusable **R**: experiment comparability and repeatability require calibrated data sources, well characterized samples and in the case of transients (operando techniques) the capturing of intermediate conditions. This is certainly the most challenging goal, but indispensable for efficient and sustainable usage of valuable beamtime. Even process tuning relies on safe discrimination of fluctuations originating from the experimental set-up and the signals to be studied.

At BESSY we need to concentrate on the **I** and **R**. We made good progress in converting instrument data to NeXus format transferred to accessible, long term storage ICAT e.g. analyzing underlying spec scripts. For the **R** changes in the beamline configuration are captured [8]. For sample and detector characterization HZB leverages the partnership with the German National Bureau of Standards (PTB). Having their expertise on absolute calibration in house is an invaluable asset.

For **F** we insist to roll out EPICS and tools to get as much devices online as reasonable. EPICS tools, like the archiver appliance historic database, support self documentation if device status allowing correlations and providing basic meta data. For **A**, i.e. easy and secure access via a WEB portal, we lack commitment and support of out IT department yet. Remote experimentation is limited to secure access to the experiment consoles in data acquisition mode only. But at least the demand and relevance of **A** for a good, reliable data policy, HZB has promised his users, is understood.

DEMONSTRATOR PROJECT ROCK-IT

Another unifying momentum well beyond the modernizing plans at BESSY II [1] has been triggered by post-pandemic demands on user operation at large scale facilities in Germany, which manage together about 10.000 user visits per year. Travel restrictions and hygiene protocols required mail-in and remote-access experiments at levels, that created a non-sustainable load on existing beamlines staff. The ROCK-IT demonstrator project is meant to evaluate and develop common necessary tools for the automation and remote-access, even for more complicated in-situ and operando experiments.

Remote-access protocols should be established to allow for a holistic experiment and beamline control software,

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2A004

including also aspects of artificial intelligence for automatically conducting the experiments, evaluation of the data in real-time for adjusting experimental parameters, and suitable robotic and automation for changing samples.

Four Helmholtz centers, DESY, HZB, HZDR, and KIT have identified catalysis operando experiments as a pilot development case. So far, no complete automation exists for such experiments and since the optimization of catalysts requires to evaluate a large parameter space of experimental and material conditions, a suitable automation of such experiments will allow for a more effective development workflow. High attractiveness to non-expert users and users from industry, resulting in lowered access barriers and accelerated innovation cycles is the aim of this project. Most of the software tools developed should be easily transferable to other synchrotron and neutron radiation experiments.

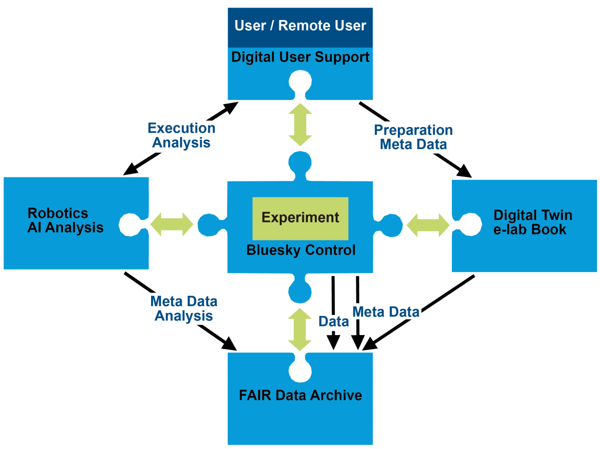

Figure 5: Interplay of the ROCK-IT application interfaces and data streams. Like the other centers HZB contributes to all tasks. Nevertheless, this paper concentrates on the aspects of experiment flow control, center of this sketch, and interfaces to the other packages.

Intended solutions should improve the efficiency of catalysis research, tailor-made, yet composed of generic building blocks portable enough to be applicable to a wide range of measurements. The building blocks – partly interfacing to other ongoing projects in the field of data handling and science - include standardized and interchangeable data formats, standardized metadata collection, interfaces to electronic lab books, sample tracking and handling, AI-based experiment control and data evaluation, and automation of experiments under remote control (Fig. 5). ROCK-IT lives up to its role as a demonstrator with first implementations for user operation at PETRA III (DESY) and BESSY II (HZB) as well with a test setup at KARA (KIT) and DRACO (HZDR).

Agreement on Bluesky

The four institutes of the collaboration use different control system protocols. EPICS, TANGO and SECoP are all used. In order to collaborate and create consistent approaches to performing measurements an intermediate control layer is required which can communicate with all of these control systems. An agreement has been reached to use the Bluesky package ophyd to do this. The development of an interface to SECoP by HZB positively influenced the developers of ophyd.v2 and showed that it was sufficiently agnostic of control system architecture to interface to both EPICS and SECoP [5]. Work progresses at DESY to integrate TANGO into ophyd.v2 in collaboration with Diamond Light Source and NSLS II.

Having agreed to use ophyd as a common interface to different control systems the project is well placed to use the other elements of Bluesky to achieve its goals. The Queueserver and associated REST API along with the Tiled Server and various existing work integrating AI into Bluesky offer a compelling solution to the challenges faced by the collaboration. Reaching agreement is already a big achievement because each facility already has it's own solutions and constraints.

PATH TOWARDS 4TH GENERATION BESSY III

The case for a 4th generation successor facility BESSY III has already been made [17–20]. The preCDR [17] has been reviewed and was well perceived by the steering and funding bodies. The three pillars of the new facility are (1) the state-of-the-art 4th generation light source, (2) embedment in an integrated research campus at Berlin-Adlershof and (3) quantitative and metrological materials science capabilities. HZB is charged to present a full CDR and start work on the TDR before the end of 2026. Since BESSY III will not be available for users before 2030 the BII+ upgrade of BESSY II has been started to adequately serve our soft x-ray science community in between.

BII+, Upgrade of BESSY II

The upgrade proposal for BESSY II, named BII+, builds in large parts on the HZB involvement in ROCK-IT. It has been outstanding positively reviewed and was forwarded to the funding agencies. It is meant to bridge the time until BESSY III gets into user operation. It is addressing adjustment of scientific focus, enlarging the operando footprint from catalysis to other green energy and materials studies, also laying grounds to the automation and autonomous experimentation needs of a modern SR light source, modernizing by emboldening fly scan capabilities (see Fig. 6), remote

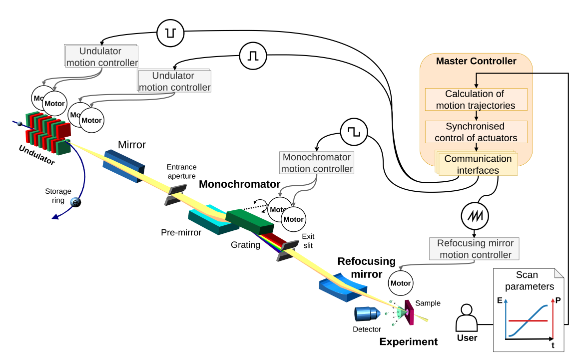

Figure 6: Orchestration of motion controllers are essential means to enhance fly scan performances.

access modes and efficient usage of CO_2 intense beamtimes (Fig. 2).

Demands on robust, well tested, remotely accessible automation and data handling modules will be very high, while testing and commissioning time at the facility running user mode will be scarce. Reliable and responsive digital twins, also for the beamlines, become increasingly mandatory for performance maintenance of the x-ray characteristics shining focused on the sample area of interest (see Fig. 7).

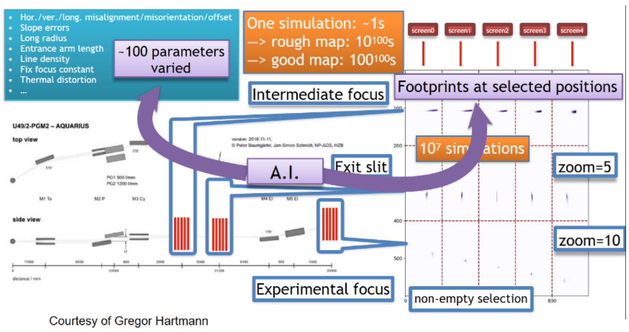

Figure 7: Surrogate models of the beamlines generated with machine learning methods (AI) are more efficient in process tuning than e.g. ray tracing predictions.

Plans for BESSY III Digitalization

The digitization plans for the integrated research facility BESSY III endorses three aspects of experimentation:

(1) *automated* and *remote*: covers IT environment for basic device control, data acquisition and handling, scientific computing. High standards of availability, reliability, ease of use, fundamental to these operation modes, are mandatory for all components to be integrated or adapted. Experimenting remotely will hardly mean any disadvantage.

(2) *smart* and *autonomous*: manual, simple raster scans are no more adequate to vast parameter spaces, faint signals or rare events. Intelligent data acquisition will utilize various knowledge sources to detect regions of interest rapidly and safely (see Fig. 8). Robust data analysis and data evaluation will justify trust in complex automated sequences and decisions no more supervised by humans.

(3) *data quality* and *quantity*: Sample complexity and tiny structures of novel functional materials push theory, simulation results and machine learning predictions beyond validity limits. For decisive hypotheses verification, probing with the 4^{th} generation light source beam, capable to resolve finest details, is required. Detailed, precise and high quality data have to be acquired to provide the urgently needed unambiguous ground truth. Thus veracity of data, BESSY III research adds to the FAIR archives, is ensured.

Implications of (1)-(3) on data management and automation is already sketched in Ref. [17]. On the general level of *digital HZB*, material research at BESSY III can leverage the achievements of the ubiquitous digital transformation process of the society. For material design and characterization, relevant data pools are already well orchestrated, available to research institutes and industry. To analysis and planning

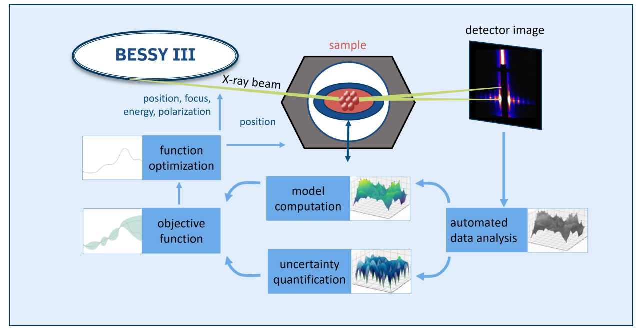

Figure 8: A schematic of autonomous experimentation balancing exploration, i.e. learning the landscape of the parameter space, approaching the objective function, and exploitation, i.e. attempting to rapidly achieve relevant results, trying functions as optimized. Sample model development and uncertainty removal are best guided by simulations and checked against FAIR data bases.

they add data-driven models and guidance by physical understanding. For BESSY III it is relevant, that HZB is well embedded into this rapidly growing global digital infrastructure. Distinction of the domain specific IT BESSY III infrastructure from the hosting general IT helps to define project scope and phases.

SUMMARY

HZB started a lengthy upgrade process. The goal that needs to be achieved with BESSY III is clear and promising. Digital means are key to optimal use of available beam time, saving operational costs. Cutting down unproductive dead times maximizes output during given access to the overbooked facility. Intelligent autonomous experimentation reduces the turnaround time from idea to scientific verification and demonstration of innovative applications. AI supported sample synthesis and handling reduces valuable material spending. A big share on digital tools within the overall project plan elevates sustainability of the facility significantly: most valuable in terms of relevance and longevity of scientific data stored in FAIR databases. A high degree of automation has a democratizing effect on science made available by this new light source. Costly expert training on instrument specifics is no more prerequisite for access to state-of-the-art x-ray experiments. Common remote operation tools and a familiar user access experience brings facilities to the scientists, saving travel time and energy.

ACKNOWLEDGEMENTS

The authors thank all HZB staff members involved for many fruitful and clarifying discussions. This includes beamline and instrument scientists, accelerator controls, beamline optics as well as central IT networking and storage. Special thanks go to many members of the Bluesky collaboration. Their comments and hints are highly appreciated.

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-MO2A004

REFERENCES

[1] R. Müller *et al.*, "Modernization of Experimental Data Taking at BESSY II", in *Proc. ICALEPCS'19*, New York, NY, USA, Oct. 2019, pp. 65–69. doi:10.18429/JACoW-ICALEPCS2019-MOCPL02

[2] Bluesky Data Collection Framework https://blueskyproject.io/bluesky/

[3] Bluesky is a collection of Python libraries that are co-developed but independently useful and may be adopted a la carte. https://blueskyproject.io

[4] H. L. He *et al.*, "Bluesky Web Client at Bessy II", presented at ICALEPCS'23, Oct. 2023, Cape Town, South Africa, paper TUPDP014, this conference.

[5] P. Wegmann *et al.*, "SECoP Integration for the ophyd Hardware Abstraction Layer", presented at ICALEPCS'23, Oct. 2023, Cape Town, South Africa, paper THMBCMO10, this conference.

[6] S. Vadilonga *et al.*, "Advancements in Beamline Digital Twin at BESSY II", presented at ICALEPCS'23, Oct. 2023, Cape Town, South Africa, paper THMBCMO18, this conference.

[7] W. Sulaiman Khail *et al.*, "Developing a Digital Twin for BESSY II Synchrotron Light Source Based on EPICS and Microservice Design", presented at ICALEPCS'23, Oct. 2023, Cape Town, South Africa, paper THSDSC06, this conference.

[8] G. Günther *et al.*, "FAIR Data of Physical and Digital Beamlines", presented at ICALEPCS'23, Oct. 2023, Cape Town, South Africa, paper THMBCMO17, this conference.

[9] Bluesky Governance, https://github.com/bluesky/governance

[10] Bluesky Project Advisory Board (PAB), https://github.com/bluesky/governance/blob/main/PROJECT-ADVISORY-BOARD.md

[11] Bluesky Technical Steering Committee (TSC), https://github.com/bluesky/governance/blob/main/TECHNICAL-STEERING-COMMITTEE.md

[12] FAIRmat, building a FAIR data infrastructure for condensed-matter physics and the chemical physics of solids, Area D: Digital Infrastructure, D5: Configurable Experiment Control System, https://www.fair-di.eu/fairmat/areas-fairmat/structure-fairmat

[13] NOMAD-CAMELS, Configurable Application for Measurements, Experiments and Laboratory-Systems, https://github.com/FAU-LAP/NOMAD-CAMELS

[14] Y. Liu *et al.*, "*Mamba*: a systematic software solution for beamline experiments at HEPS", *J. Synchrotron Rad.*, vol. 29, no. 3, pp. 664–669, 2022. doi:10.1107/S1600577522002697

[15] P.-C. Li *et al.*, "PandA(Box) flies on Bluesky: maintainable and user-friendly fly scans with Mamba at HEPS", *arXiv*, 2023. doi:10.48550/arXiv.2302.08304

[16] European Photon and Neutron Open Data Search Portal, https://data.panosc.eu/

[17] P. Goslawski *et al.*, "BESSY III The Materials Discovery Facility", HZB, Berlin, Germany. doi:10.5442/r0004

[18] O. Schwarzkopf *et al.*, "Materials discovery at BESSY", *Europ. Phys. J. Plus*, vol. 138, no. 4, p. 348, 2023. doi:10.1140/epjp/s13360-023-03957-8

[19] P. Goslawski *et al.*, "BESSY III & MLS II - Status of the Development of the New Photon Science Facility in Berlin", in *Proc. IPAC'21*, Campinas, SP, Brazil, May 2021, pp. 451–454. doi:10.18429/JACoW-IPAC2021-MOPAB126

[20] P. Goslawski *et al.*, "BESSY III - status and overview", in *Proc. IPAC'23*, Venice, Italy, May 2023, pp. 457-460. doi:10.18429/JACoW-IPAC2023-MOPA174