Journal of Physics:
Conference Series

PURPOSE-LED
PUBLISHING™

PAPER • **OPEN ACCESS**

Status and implementation of the MXCuBE experiment control and ISPyB information system at the P11 macromolecular crystallography beamline at PETRA III (DESY, Hamburg)

To cite this article: A. Gruzinov *et al* 2025 *J. Phys.: Conf. Ser.* **3010** 012139

Status and implementation of the MXCuBE experiment control and ISPyB information system at the P11 macromolecular crystallography beamline at PETRA III (DESY, Hamburg)

A. Gruzinov[1*], O. Merkulova[1], V. Rey-Bakaikoa[2], G. Pompidor[1], H. Taberman[1], S. Chatziefthymiou[1], A. Grebentsov[1], J. Meyer[1], T. Melson[1,3], J. L. Song[1], V. Delsoglio[1], S. N. Potturu[1], C. Borges[1], M. Savko[4], T. A. White[1,5], S. Flewett[1,5], F. Schluenzen[1], A. Rothkirch[1], G. Bourenkov[6], M. Nikolova[6], M. Agthe[6], S. Storm[6], I. Karpics[7], T. R. Schneider[6] and J. Hakanpää[1]

[1] Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany
[2]European Synchrotron Radiation Facility - ESRF, 71 Av. des Martyrs, 38000 Grenoble, France
[3] Hamburg University of Applied Sciences, Stiftstraße 69, 20999 Hamburg, Germany
[4] Synchrotron SOLEIL, L'Orme des Merisiers Départementale 128, 91190 Saint-Aubin, France
[5] Center for Data and Computing in Natural Science CDCS, Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany.
[6] European Molecular Biology Laboratory, Hamburg Unit c/o Deutsches Electronen-Synchrotron DESY, Notkestraße 85/Geb. 25A, 22607 Hamburg, Germany
[7] European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany

*E-mail: andrey.gruzinov@desy.de

Abstract. The P11 beamline at PETRA III (DESY, Hamburg) is a versatile instrument dedicated to High Throughput Macromolecular Crystallography. The combination of the Eiger detector and sample changer robot allows data collection in about two minutes per sample. MXCuBE is used as a standard data acquisition software at several synchrotron beamlines, including the beamlines P13 and P14 operated by the European Molecular Biology Laboratory (EMBL) at PETRA III. A unified interface at various MX beamlines improves the cross-facility user experience and the convenience of conducting experiments. This work describes the current status and challenges of implementing MXCuBE and ISPyB as standard data collection and representation interfaces at the P11 beamline. In particular, we address the unique, in-house designed hardware of the beamline. This work has been jointly carried out by DESY and EMBL staff to harmonize the data acquisition interfaces at the Macromolecular Crystallography beamlines at the PETRA III storage ring in Hamburg.

1. Introduction

Macromolecular crystallography (MX) is a well-established technique for obtaining the high-resolution structures of proteins from crystals. The wide use of cryo-cooling and standardized sample mounts have enabled advanced automation of beamline instruments in a similar fashion as combined software tools have done for data handling and representation. The whole process, from the sample shipping to the data processing, can be standardized for many use cases. Naturally, complex experiments still require manual intervention by the experimenter, but the complete automatization achieved at beamlines with the ability to collect unattended data matches the experimental necessities for a large fraction of samples [1–3].

Sample and shipment tracking [4], in combination with the robotic sample mounting [5–7] and data processing pipelines [8–15], have already prepared the synchrotron users for the standardized interfaces. As a natural continuum, it is advantageous to share the development and use of intuitive data acquisition software with several MX synchrotron beamlines, harmonizing the data collection process across sites.

Macromolecular Xtallography Customized Beamline Environment (MXCuBE) is a synchrotron beamline control environment for macromolecular crystallography experiments [16,17]. MXCuBE originated as a common user interface designed for crystallographers working on various beamlines. The same researchers often work across different synchrotrons, so having a common interface across these facilities offers significant advantages. MXCuBE standardizes the data collection methods and simplifies the user experience, providing an intuitive, easy-to-use interface.

MXCuBE has evolved into a collaborative project involving multiple stakeholders. Originally designed with a user interface based on the PyQt framework, it has expanded to offer both PyQt and web-based interfaces [16–19] for the graphical user environment (GUI). The software consists of three main components: the core MXCuBE library and two distinct user interfaces - one based on web technologies and the other on Qt. The core library is shared between these two interfaces and is written entirely in Python.

The MXCuBE core provides a software abstraction layer that simplifies interactions with various hardware components. This abstraction covers hardware and protocols, making it easier for newcomers to adapt the system to their specific needs. The library defines an API for the PyQt and web interfaces, allowing it to build on top of the core framework as needed.

The core architecture consists of abstract classes that provide flexibility in integrating new control systems, sequences, or protocols. MXCuBE architecture supports full control system integration, such as with BLISS [20,21], while offering channels and commands for other cases like EPICS [22,23], TANGO [24,25], and TINE [26,27]. This structure enables developers to quickly extend the platform with minimal custom code, focusing only on the unique aspects of their system.

Apart from providing a software-architectural template for beamline GUI, MXCuBE comes with a complete design of an MX experiment itself. The design is rather complex in its entirety. It has been validated and tested over many years by thousands of users at multiple MX facilities. It would be difficult to recreate the design at a single beamline from scratch.

ISPyB is an Information System for Protein Crystallography Beamlines that integrates with beamline control systems and databases and captures information related to the experimental setup, data collection, and results [4]. MXCuBE has a strong integration with the data representation database ISPyB that allows the users an overview and immediate feedback on the experiment's status and processed results also in the remote or unattended mode of operation.

Today, MXCuBE is actively supported by the following partners: ESRF, SOLEIL MAXIV, HZB, EMBL, Global Phasing Ltd, DESY, ALBA, ELETTRA, LNLS, NSRRC, SARI, ARINAX, and ANSTO. Overall, MXCuBE offers a robust, flexible, and user-friendly environment for macromolecular crystallography, supporting the needs of both researchers and beamline operators.

2. Beamline data collection and software system

2.1 MX endstation

The P11 beamline [28,29] at PETRA III (DESY, Hamburg) is designed for high-throughput macromolecular crystallography. It shares a canted undulator sector of the PETRA III storage ring with the P12 BioSAXS beamline operated by the European Molecular Biology Laboratory (EMBL).

The beamline uses a U32 undulator as a source, with downstream optics including a double-crystal monochromator (DCM) and two mirror systems that deflect the primary beam and focus the beam toward the MX endstation 73 meters downstream from the X-ray source. The Kirkpatrick-Baez (KB) mirror system close to the sample position supports adjustable beam sizes (200 x 200 μm^2 to 5 x 10 μm^2). The beam energy offered at the beamline ranges from 6 to 28 keV and flux of up to 10^{13} ph/s.

A high-precision single-axis goniometer with an air-bearing spindle positions samples with sub-micrometer accuracy. The system also has an on-axis microscope for sample alignment and various apertures for beam shaping.

Data is collected using an Eiger2 X 16M detector (Dectris, Switzerland) capable of shutterless collection. Datasets can typically be collected in 18-36 seconds, depending on the number of frames and angular range. The beamline uses an Oxford Cryostream 1000 (Oxford Cryosystems, UK) to keep the samples at 100 K during data collection. Additionally, fluorescence spectra can be acquired using a Vortex-EM silicon drift detector (Hitachi, Japan) to support experimental phasing experiments. The sample changer accepts samples loaded in uni-puck standard [30], with a robotic arm and gripper, and can hold up to 368 samples in total (23 pucks). The entire hardware is controlled through the TANGO [24,25,31,32] control system.

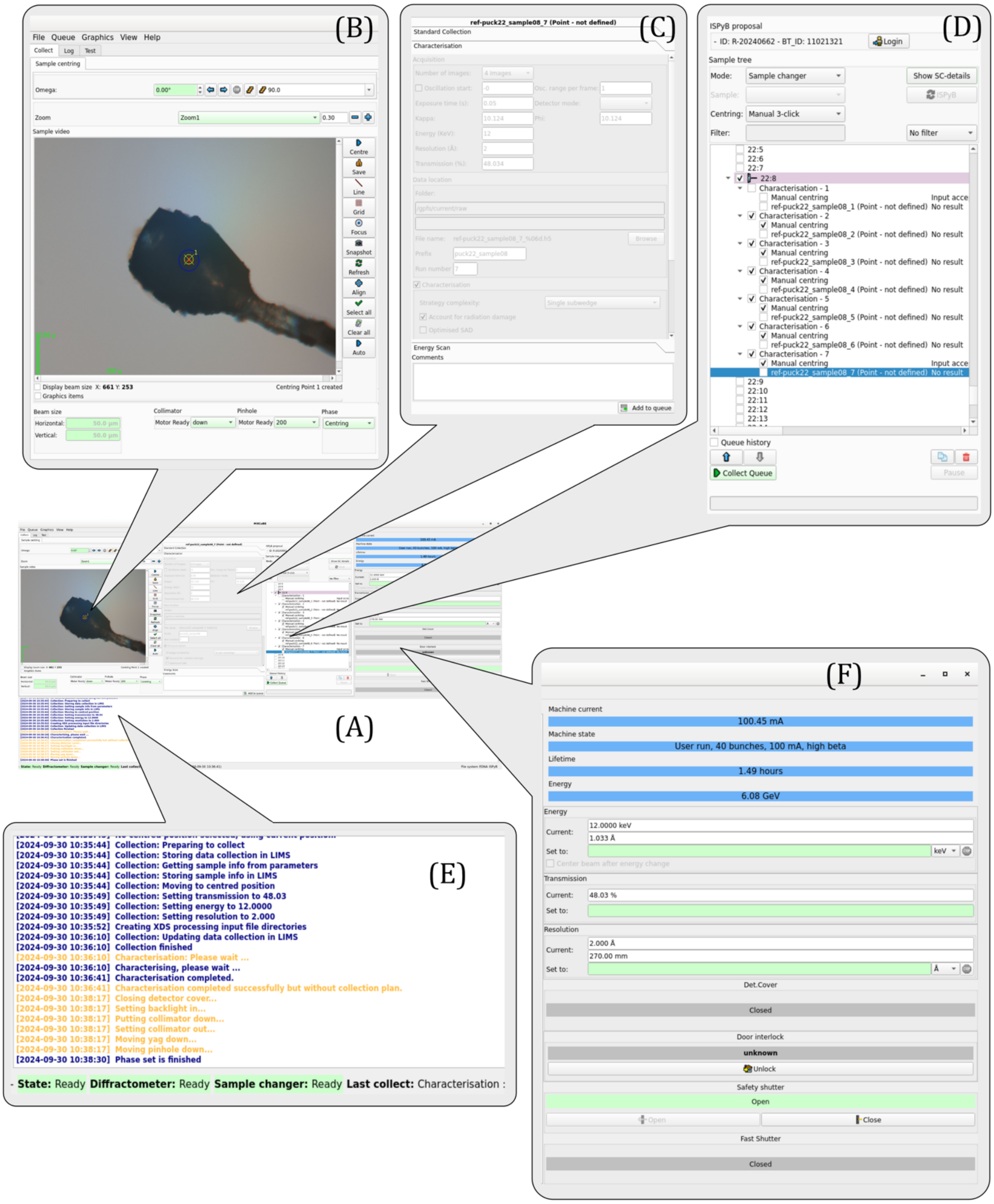
2.2 MXCuBE as graphical user interface

The MXCuBE version at the P11 beamline is a part of the official repository [33,34] containing a general backend used for the web and desktop interfaces. The MXCuBE core repository provides abstract classes and localized implementations for hardware components across various beamlines.

At P11, the hardware stack is primarily developed in-house to support unique instruments, presenting specific challenges in implementation and testing. While standard components enable some level of code reuse, custom-made components offer greater flexibility and are highly adaptable to specific needs.

Figure 1A shows the graphical user interface of MXCuBE of the P11 beamline. The graphical widgets used are combined from the existing library of typical beamline components [34], such as the video image of the sample (Figure 1B), data collection parameters in different modes (Figure 1C), execution queue (Figure 1D), status messages window (Figure 1E), controls of the beam shutter, detector position, resolution, and accelerator status (Figure 1F).

At the early stages of MXCuBE deployment at P11, it was decided to base it on the latest development version of the project, allowing it to keep up with the fast-paced, collaborative development and easier incorporation of P11-specific software components to the main development branch of the project. The Qt variant was chosen as a graphical user interface rather than the web version to harmonize the user experience between DESY and EMBL MX beamlines.



Omega: 0.00° 90.0
Zoom: Zoom1 0.30
Sample video
Display beam size X: **661** Y: **253**
Centring Point 1 created
Graphics items
Beam size
Horizontal: 50.0 µm
Vertical: 50.0 µm
Collimator: Motor Ready down
Pinhole: Motor Ready 200
Phase: Centring

ref-puck22_sample08_7 (Point - not defined)

ISPyB proposal
- ID: R-20240662 - BT_ID: 11021321
Sample tree
Mode: Sample changer
Centring: Manual 3-click
Filter: No filter

```
☐ 22:5
☐ 22:6
☐ 22:7
☑ 22:8
  ☐ Characterisation - 1
    ☐ Manual centring                                     Input acce
    ☐ ref-puck22_sample08_1 (Point - not defined)  No result
  ☑ Characterisation - 2
    ☑ Manual centring
    ☐ ref-puck22_sample08_2 (Point - not defined)  No result
  ☑ Characterisation - 3
    ☑ Manual centring
    ☐ ref-puck22_sample08_3 (Point - not defined)  No result
  ☑ Characterisation - 4
    ☑ Manual centring
    ☐ ref-puck22_sample08_4 (Point - not defined)  No result
  ☑ Characterisation - 5
    ☑ Manual centring
    ☐ ref-puck22_sample08_5 (Point - not defined)  No result
  ☑ Characterisation - 6
    ☑ Manual centring
    ☐ ref-puck22_sample08_6 (Point - not defined)  No result
  ☑ Characterisation - 7
    ☑ Manual centring                                     Input acce
    ☐ ref-puck22_sample08_7 (Point - not defined)  No result
☐ 22:9
☐ 22:10
☐ 22:11
☐ 22:12
☐ 22:13
```

```
[2024-09-30 10:35:44] Collection: Preparing to collect
[2024-09-30 10:35:44] Collection: Storing data collection in LIMS
[2024-09-30 10:35:44] Collection: Getting sample info from parameters
[2024-09-30 10:35:44] Collection: Storing sample info in LIMS
[2024-09-30 10:35:44] Collection: Moving to centred position
[2024-09-30 10:35:49] Collection: Setting transmission to 48.03
[2024-09-30 10:35:49] Collection: Setting energy to 12.0000
[2024-09-30 10:35:49] Collection: Setting resolution to 2.000
[2024-09-30 10:35:52] Creating XDS processing input file directories
[2024-09-30 10:36:10] Collection: Updating data collection in LIMS
[2024-09-30 10:36:10] Collection finished
[2024-09-30 10:36:10] Characterisation: Please wait ...
[2024-09-30 10:36:10] Characterising, please wait ...
[2024-09-30 10:36:41] Characterisation completed.
[2024-09-30 10:36:41] Characterisation completed successfully but without collection plan.
[2024-09-30 10:38:17] Closing detector cover...
[2024-09-30 10:38:17] Setting backlight in...
[2024-09-30 10:38:17] Putting collimator down...
[2024-09-30 10:38:17] Setting collimator out...
[2024-09-30 10:38:17] Moving yag down...
[2024-09-30 10:38:17] Moving pinhole down...
[2024-09-30 10:38:30] Phase set is finished
```

- **State:** Ready **Diffractometer:** Ready **Sample changer:** Ready **Last collect:** Characterisation :

Machine current
100.45 mA
Machine state
User run, 40 bunches, 100 mA, high beta
Lifetime
1.49 hours
Energy
6.08 GeV

Energy
Current: 12.0000 keV
1.033 Å
Set to: keV
☐ Center beam after energy change

Transmission
Current: 48.03 %
Set to:

Resolution
Current: 2.000 Å
270.00 mm
Set to: Å

Det.Cover
Closed

Door interlock
unknown

Safety shutter
Open

Fast Shutter
Closed

Figure 1. Display of the user view of the experiment control at P11 using MXCuBE. (A) MXCuBE interface is being used for controlling and monitoring experiments. (B) MXCuBE panel for live display of a crystal sample under the microscope. (C) MXCuBE sample information panel with experiment parameters setup. (D) MXCuBE sample tree panel with information about ISPyB proposal, sample centering selection, and sample changer details button. (E) MXCuBE user logging information panel displaying beamline operations

executed so far and their status. (F) MXCuBE panel for the current beamline setup, including accelerator status, current energy, transmission, resolution, and various beam shutter statuses.

2.3 ISPyB as information management system

To enable sample information import and metadata archiving at P11, MXCuBE will interact with the EXI version of the ISPyB to provide users with browser-based access to their experiments' results.

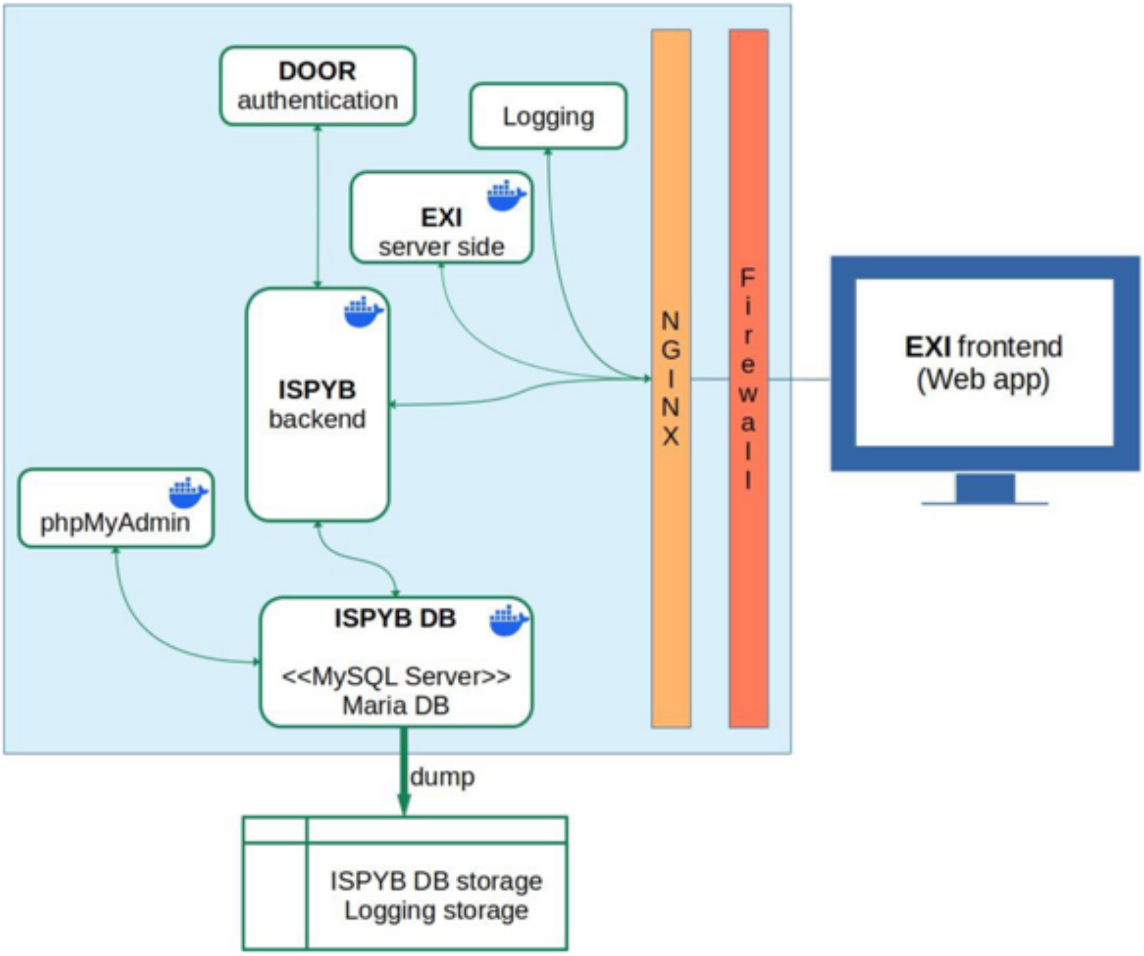
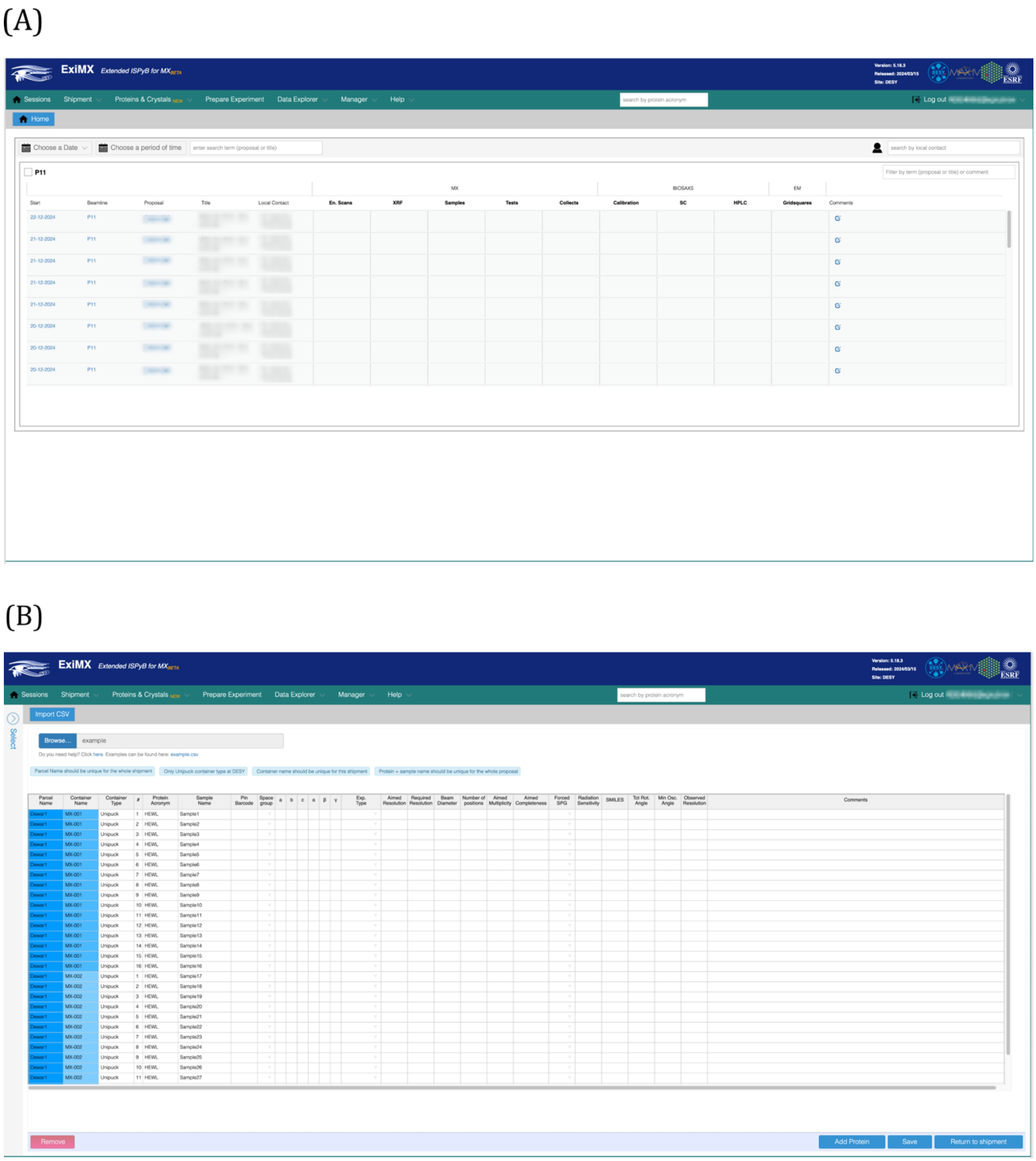


Figure 2. The ISPyB infrastructure outline is mapped on the DESY IT infrastructure, showing data flows between different components.

At DESY, a Python component was developed to synchronize data between ISPyB and the DESY Online Office for Research with Photons (DOOR) portal. ISPyB is a complex, multi-component application consisting of SQL databases and components built with Java, Python, and JavaScript, which interact with each other directly or through a load balancer (Figure 2). Following best practices, we have deployed and integrated an Elastic-based observability stack [35].

At P11, a custom authentication model was developed to integrate ISPyB with DOOR, verifying user credentials before accessing experimental data and sample management. A custom sample changer widget was also created for ISPyB-MXCuBE interaction and is being integrated into the EXI interface.



Log out [redacted]

☐ **P11**

					MX					BIOSAXS			EM	
Start	Beamline	Proposal	Title	Local Contact	**En. Scans**	**XRF**	**Samples**	**Tests**	**Collects**	**Calibration**	**SC**	**HPLC**	**Gridsquares**	Comments
22-12-2024	P11	[redacted]	[redacted]	[redacted]										
21-12-2024	P11	[redacted]	[redacted]	[redacted]										
21-12-2024	P11	[redacted]	[redacted]	[redacted]										
21-12-2024	P11	[redacted]	[redacted]	[redacted]										
21-12-2024	P11	[redacted]	[redacted]	[redacted]										
20-12-2024	P11	[redacted]	[redacted]	[redacted]										
20-12-2024	P11	[redacted]	[redacted]	[redacted]										
20-12-2024	P11	[redacted]	[redacted]	[redacted]										

Log out [redacted]

example

Do you need help? Click here. Examples can be found here: example.csv

Parcel Name should be unique for the whole shipment

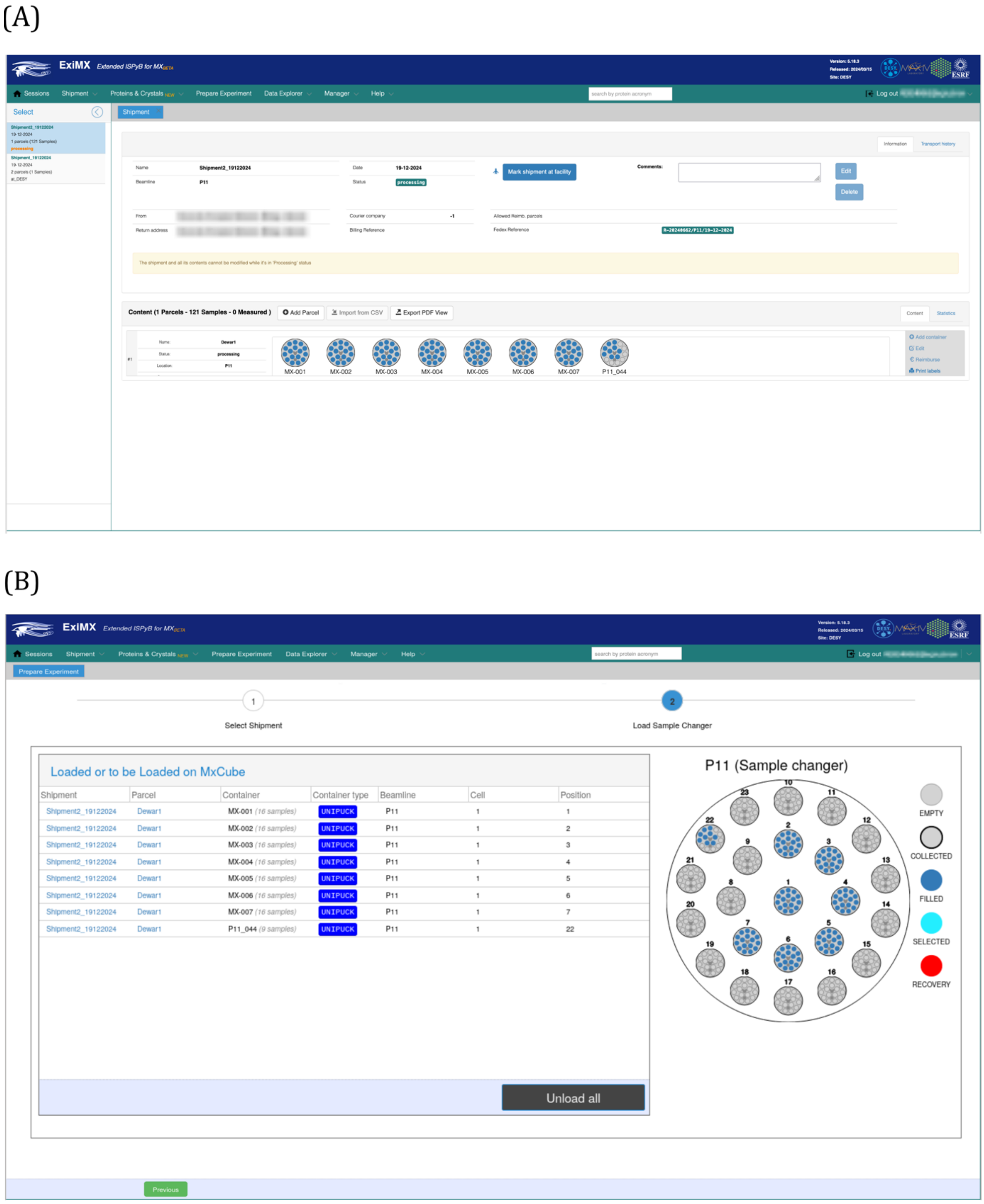
Only Unipuck container type at DESY

Container name should be unique for this shipment

Protein + sample name should be unique for the whole proposal

Parcel Name	Container Name	Container Type	#	Protein Acronym	Sample Name	Pin Barcode	Space group	a	b	c	α	β	γ	Exp. Type	Aimed Resolution	Required Resolution	Beam Diameter	Number of positions	Aimed Multiplicity	Aimed Completeness	Forced SPG	Radiation Sensitivity	SMILES	Tot Rot. Angle	Min Osc. Angle	Observed Resolution	Comments
Dewar1	MX-001	Unipuck	1	HEWL	Sample1																						
Dewar1	MX-001	Unipuck	2	HEWL	Sample2																						
Dewar1	MX-001	Unipuck	3	HEWL	Sample3																						
Dewar1	MX-001	Unipuck	4	HEWL	Sample4																						
Dewar1	MX-001	Unipuck	5	HEWL	Sample5																						
Dewar1	MX-001	Unipuck	6	HEWL	Sample6																						
Dewar1	MX-001	Unipuck	7	HEWL	Sample7																						
Dewar1	MX-001	Unipuck	8	HEWL	Sample8																						
Dewar1	MX-001	Unipuck	9	HEWL	Sample9																						
Dewar1	MX-001	Unipuck	10	HEWL	Sample10																						
Dewar1	MX-001	Unipuck	11	HEWL	Sample11																						
Dewar1	MX-001	Unipuck	12	HEWL	Sample12																						
Dewar1	MX-001	Unipuck	13	HEWL	Sample13																						
Dewar1	MX-001	Unipuck	14	HEWL	Sample14																						
Dewar1	MX-001	Unipuck	15	HEWL	Sample15																						
Dewar1	MX-001	Unipuck	16	HEWL	Sample16																						
Dewar1	MX-002	Unipuck	1	HEWL	Sample17																						
Dewar1	MX-002	Unipuck	2	HEWL	Sample18																						
Dewar1	MX-002	Unipuck	3	HEWL	Sample19																						
Dewar1	MX-002	Unipuck	4	HEWL	Sample20																						
Dewar1	MX-002	Unipuck	5	HEWL	Sample21																						
Dewar1	MX-002	Unipuck	6	HEWL	Sample22																						
Dewar1	MX-002	Unipuck	7	HEWL	Sample23																						
Dewar1	MX-002	Unipuck	8	HEWL	Sample24																						
Dewar1	MX-002	Unipuck	9	HEWL	Sample25																						
Dewar1	MX-002	Unipuck	10	HEWL	Sample26																						
Dewar1	MX-002	Unipuck	11	HEWL	Sample27																						

Figure 3. Screenshots of the ISPyB EXI interface (sensitive user information is blurred). (A) The session overview page displays user session details; (B) The sample management page for input and editing of sample-specific data. This example shows the option to load a CSV file containing information about samples.



Shipment2_19122024
19-12-2024
1 parcels (121 Samples)
processing

Shipment_19122024
19-12-2024
2 parcels (1 Samples)
at_DESY

Name: Shipment2_19122024
Beamline: P11
Date: 19-12-2024
Status: processing
Comments:
From: [redacted]
Return address: [redacted]
Courier company: -1
Billing Reference
Allowed Reimb. parcels
Fedex Reference: R-20240662/P11/19-12-2024

The shipment and all its contents cannot be modified while it's in 'Processing' status

Content (1 Parcels - 121 Samples - 0 Measured)

#1
Name: Dewar1
Status: processing
Location: P11

Loaded or to be Loaded on MxCube

Shipment	Parcel	Container	Container type	Beamline	Cell	Position
Shipment2_19122024	Dewar1	MX-001 *(16 samples)*	UNIPUCK	P11	1	1
Shipment2_19122024	Dewar1	MX-002 *(16 samples)*	UNIPUCK	P11	1	2
Shipment2_19122024	Dewar1	MX-003 *(16 samples)*	UNIPUCK	P11	1	3
Shipment2_19122024	Dewar1	MX-004 *(16 samples)*	UNIPUCK	P11	1	4
Shipment2_19122024	Dewar1	MX-005 *(16 samples)*	UNIPUCK	P11	1	5
Shipment2_19122024	Dewar1	MX-006 *(16 samples)*	UNIPUCK	P11	1	6
Shipment2_19122024	Dewar1	MX-007 *(16 samples)*	UNIPUCK	P11	1	7
Shipment2_19122024	Dewar1	P11_044 *(9 samples)*	UNIPUCK	P11	1	22

Figure 4. Screenshots of the ISPyB EXI interface (sensitive user information is blurred). (A) The shipment management page shows shipment details, sample containers, and associated metadata; (B) Sample changer loading page displaying container positions and statuses on the P11 beamline sample changer.

Figure 3A shows the session overview page, where users can browse experiment sessions by criteria such as date range. This page provides a structured summary of session details, including experiment metadata and associated activities.

Figure 3B displays the sample management page, where users can input, edit, and organize sample data. This interface allows for the import of sample lists via CSV files and shows a detailed table for managing sample-related metadata, such as protein type, sample containers, and experiment-specific attributes.

Figure 4 highlights the shipment and sample loading functionalities. Figure 4A depicts the shipment management page, which displays details of active shipments, including shipment status, associated parcels, and individual sample containers. Figure 4B shows the sample changer loading page for the P11 beamline. Here, users can view and manage the loading status of sample containers, including their positions within the sample changer. A graphical representation of the sample changer visually indicates the current state of each position.

2.4 Usage

The combination of MXCuBE/ISPyB has already been in user operation at P11 by users requesting it; however, only standard data collections and characterizations are currently implemented in MXCuBE. ISPyB has been used for the sample name import only by staff in the local environment. Implementing raster scans in MXCuBE and enabling user access to ISPyB containing the metadata and auto-processing results is planned in the near future.

2.5 Automatic optical sample centering

Sample centering is an important part of the data collection procedure. It is usually done manually using the so-called n-click centering procedure, where the sample position is manually chosen on the camera display with the mouse pointer. By repeating this step at several different crystal rotations, the three-dimensional position of the crystal relative to the beam is determined. This allows the crystal to stay in the beam throughout a full 360-degree rotation. Errors in the sample centering can cause the crystal to move out of the beam at certain rotation angles, compromising the dataset quality. Automating these processes is important for developing the unattended data collection mode at P11.

This optical centering method usually becomes problematic when the view of the crystal is obscured by ice or the intrinsic properties of the solution where the crystal is growing, for example, in the case of membrane proteins in the lipid cubic phase. In these cases, X-ray centering can be used instead [36–38] as, for example, the next step after determining the approximate crystal position or loop area with a fast optical centering. Implementing X-ray centering is the ultimate goal for complete data collection automation at the beamline.

We have chosen the recently developed library for the optical sample centering for the crystal coordinates recognition developed at synchrotron SOLEIL - *murko* [39]. It is based on machine learning (ML) and uses a pre-trained model whose input is the image of the sample from the camera. It returns a most likely click position in relative coordinates. If the crystal is not detected, it returns the detected crystal's relative coordinates or the pin's furthest outstanding point. This information is used in the MXCuBE in real time to calculate the positions of the motors for the centering.

We tested the *murko* software on both CPU and GPU nodes using a V100 GPU. On the multi-threaded CPU, the prediction step of the auto-centering process took approximately 0.73 seconds per frame, while on the GPU, it took only 0.11 seconds per frame. However, since rotating the crystal also takes time, the overall time savings from using the GPU are minimal using the current beamline setup. Recent beamline tests showed that the manual centering process takes approximately 18 seconds. The complete ML-based centering using the GPU takes a similar amount of time, while the ML-based centering on a CPU takes about 22 seconds. The crystal rotation and *murko* calculations are performed sequentially, but running them parallel could improve efficiency.

In our implementation, users must click 'Auto' to initiate the centering process, whereas, at SOLEIL, the process begins automatically when the sample is loaded. *Murko* is running as a separate service on the dedicated P11 node.

2.6 Live view interface

The *yamone* (Yet Another MONitoring interface for the Eiger) is a C++ program developed to act as an intermediary between the Eiger 16M X-ray detector and ADXV software [40] for real-time data visualization. The available code of the Eiger receiver [41] was used as a basis for implementation. The program connects to the detector monitoring interface to retrieve image files in TIFF format and handles the display with explicit headers. It stores files in a local directory, typically /tmp/eiger_monitor, for immediate display during data collection. *Yamone* launches in the background ADXV as a stand-alone executable in socket mode, enabling real-time visualization of diffraction images while the experiment progresses.

The program also interacts with ADXV, continuously updating crucial experimental parameters such as the beam center, energy, and detector distance that are read in real-time from the detector TANGO server. These parameters are used to append the proper header to the saved monitoring images to update and display resolution rings on the screen. As soon as the data collection is finished, the MXCuBE sends the information about the last collected dataset through the socket interface to ADXV to display it after the data collection is finished.

Combined with ADXV's familiar user interface, it provides an efficient live view tool for online data collection monitoring.

2.7 Sample Shipment and Result Representation

ISPyB simplifies the management of synchrotron experiments, connecting user portals like DOOR at DESY and supporting data collection, tracking, and analysis. It integrates with MXCuBE, which uses ISPyB APIs for data input and provides a detailed interface for reviewing data and analysis. The Java-based version of ISPyB is widely adopted, and efforts are focused on aligning it with EMBL beamlines (P13 and P14) using the EXI interface.

It is important to ensure easy access to data, and ISPyB makes sure processed data is stored separately from raw data for long-term access. Sample information can be uploaded via CSV and synchronized with MXCuBE to minimize errors, speed up data collection, and guide auto-processing. ISPyB also enables P11 staff to review data collection through a dedicated viewer interface.

A Lean-Lift (Hänel, Germany) was acquired for housing shipments to improve sample logistics. In collaboration with Hamburg University of Applied Sciences (HAW) and EMBL, a barcode-based logistics chain will be set up using modified software from Diamond Light Source to track dewars. The chain will soon connect to ISPyB for shipment management

2.8 Data processing pipelines

At P11, the conventional crystallographic data is currently autoprocessed with XDSAPP [9,42], and recently, a parallel processing pipeline with autoPROC [43] (Global Phasing) has been established. The hardware for autoprocessing is two dedicated CPU nodes with AMD EPYC processors on Maxwell, the DESY computational cluster. During the standard user operation, processing with optimized parameters for the XDSAPP or autoPROC is triggered on the dedicated cluster nodes using SLURM when the data collection is finished. In the near future, we aim to substitute custom-written SLURM submission scripts with the EDNA-based processing pipelines with the online ISPyB results reporting.

EDNA-characterization [11] (Mosflm [44,45], Raddose [46,47], BEST [48]) is used in MXCuBE. Grid scans will be implemented in MXCuBE shortly. The scoring of diffraction images (with data from each line of the grid stored as an HDF5 container) runs very fast with DOZOR scores [49–51] and produces a heatmap.

We also optimized the cluster usage for different processing pipelines, including parallel runs of the XDSAPP and autoPROC pipelines simultaneously. It allowed us to reduce the data processing time for the typical data sets by approximately a factor of two, making the data processing for the users more comfortable.

3. Conclusions and outlook

We have established the MXCuBE/ISPyB data collection software at the P11 beamline as part of the standard user operation procedures. Implementing the machine-learning-driven centering procedures will make the operation more robust in the unattended mode. A large part of this work was done as part of the project to harmonize the data collection and processing user experience.

Characterizations and standard data collections are implemented in MXCuBE, and ADXV supports the data collections as a live viewer of the diffraction images. To enable sample information import and metadata archiving at P11, MXCuBE will standardly interact with ISPyB to provide the users with EXI GUI—browser-based access to the results of their experiments. In the current test phase, a combination of MXCuBE/ISPyB has been used by users specifically requesting this. Implementing raster scans in MXCuBE and enabling user access to ISPyB containing the metadata and auto-processing results is in progress.

While finalizing the MXCuBE PyQt and EXI implementations, we will continue developing the automated centering, auto-refinement, and structure representation processes, as well as the unattended data collection and workflows. This shows the potential to implement further collaborative developments into the MXCuBE, including the option for time-resolved data collection and custom database connections.

All recent code changes are currently located in the official repositories of the MXCuBE core and desktop versions.

4. Acknowledgments

The authors are grateful to the MXCuBE/ISPyB collaborators for their constant interest in development, useful comments, and code reviews, which allowed the fast integration of the P11 beamline into the MXCuBE community. This work was supported in part through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. This work was supported in part by UHH, DESY, TUHH and HamburgX grant LFF-HHX-03 to the Center for Data and Computing in Natural Sciences (CDCS) from the Hamburg Ministry of Science, Research, Equalities and Districts.

5. Software availability

Software	**Link**
MXCUBE Core	https://github.com/mxcube/mxcubecore
MXCUBE Qt	https://github.com/mxcube/mxcubeqt/
ISPyB (P11 mod)	https://github.com/scientific-software-hub/mxhub-ispyb
EXI (P11 mod)	https://github.com/scientific-software-hub/mxhub-exi
murko	https://github.com/MartinSavko/murko
yamone	https://github.com/agruzinov/yamone
ISPyB Logistics	https://github.com/DiamondLightSource/flask-ispyb-logistics

References

[1] Bowler M W, Nurizzo D, Barrett R, Beteva A, Bodin M, Caserotto H, Delagenière S, Dobias F, Flot D, Giraud T, Guichard N, Guijarro M, Lentini M, Leonard G A, McSweeney S, Oskarsson M, Schmidt W, Snigirev A, Von Stetten D, Surr J, Svensson O, Theveneau P and Mueller-Dieckmann C 2015 MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules *J Synchrotron Rad* **22** 1540–7

[2] Svensson O, Gilski M, Nurizzo D and Bowler M W 2019 A comparative anatomy of protein crystals: lessons from the automatic processing of 56 000 samples *IUCrJ* **6** 822–31

[3] Bowler M W, Svensson O and Nurizzo D 2016 Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data *Crystallography Reviews* **22** 233–49

[4] Delagenière S, Brenchereau P, Launer L, Ashton A W, Leal R, Veyrier S, Gabadinho J, Gordon E J, Jones S D, Levik K E, McSweeney S M, Monaco S, Nanao M, Spruce D, Svensson O, Walsh M A and Leonard G A 2011 ISPyB: an information management system for synchrotron macromolecular crystallography *Bioinformatics* **27** 3186–92

[5] Cipriani F, Felisaz F, Launer L, Aksoy J-S, Caserotto H, Cusack S, Dallery M, di-Chiaro F, Guijarro M, Huet J, Larsen S, Lentini M, McCarthy J, McSweeney S, Ravelli R, Renier M, Taffut C, Thompson A, Leonard G A and Walsh M A 2006 Automation of sample mounting for macromolecular crystallography *Acta Crystallogr D Biol Crystallogr* **62** 1251–9

[6] Lazo E O, Antonelli S, Aishima J, Bernstein H J, Bhogadi D, Fuchs M R, Guichard N, McSweeney S, Myers S, Qian K, Schneider D, Shea-McCarthy G, Skinner J, Sweet R, Yang L and Jakoncic J 2021 Robotic sample changers for macromolecular X-ray crystallography and biological small-angle X-ray scattering at the National Synchrotron Light Source II *J Synchrotron Rad* **28** 1649–61

[7] Abiven Y-M, Bouvel S, Bucaille T, Chavas L, Elkaim E, Gourhant P, Liatimi Y, Medjoubi K, Pierre-Joseph Zéphir S, Pilliaud B, Pinty V, Somogyi A and Thiam F 2020 Robotizing SOLEIL Beamlines to Improve Experiments Automation *Proceedings of the 17th International Conference on Accelerator and Large Experimental Physics Control Systems* **ICALEPCS2019** 4 pages

[8] Kabsch W 2010 *XDS Acta Crystallogr D Biol Crystallogr* **66** 125–32

[9] Krug M, Weiss M S, Heinemann U and Mueller U 2012 *XDSAPP* : a graphical user interface for the convenient processing of diffraction data using *XDS J Appl Crystallogr* **45** 568–72

[10] Agirre J, Atanasova M, Bagdonas H, Ballard C B, Baslé A, Beilsten-Edmands J, Borges R J, Brown D G, Burgos-Mármol J J, Berrisford J M, Bond P S, Caballero I, Catapano L, Chojnowski G, Cook A G, Cowtan K D, Croll T I, Debreczeni J É, Devenish N E, Dodson E J, Drevon T R, Emsley P, Evans G, Evans P R, Fando M, Foadi J, Fuentes-Montero L, Garman E F, Gerstel M, Gildea R J, Hatti K, Hekkelman M L, Heuser P, Hoh S W, Hough M A, Jenkins H T, Jiménez E, Joosten R P, Keegan R M, Keep N, Krissinel E B, Kolenko P, Kovalevskiy O, Lamzin V S, Lawson D M, Lebedev A A, Leslie A G W, Lohkamp B, Long F, Malý M, McCoy A J, McNicholas S J, Medina A, Millán C, Murray J W, Murshudov G N, Nicholls R A, Noble M E M, Oeffner R, Pannu N S, Parkhurst J M, Pearce N, Pereira J, Perrakis A, Powell H R, Read R J, Rigden D J, Rochira W, Sammito M, Sánchez Rodríguez F, Sheldrick G M, Shelley K L, Simkovic F, Simpkin A J, Skubak P, Sobolev E, Steiner R A, Stevenson K, Tews I, Thomas J M H, Thorn A, Valls J T, Uski V, Usón I, Vagin A, Velankar S, Vollmar M, Walden H, Waterman D, Wilson K S, Winn M D, Winter G, Wojdyr M and Yamashita K 2023 The *CCP* 4 suite: integrative software for macromolecular crystallography *Acta Crystallogr D Struct Biol* **79** 449–61

[11] Incardona M-F, Bourenkov G P, Levik K, Pieritz R A, Popov A N and Svensson O 2009 *EDNA* : a framework for plugin-based applications applied to X-ray experiment online data analysis *J Synchrotron Rad* **16** 872–9

[12] Winter G, Beilsten-Edmands J, Devenish N, Gerstel M, Gildea R J, McDonagh D, Pascal E, Waterman D G, Williams B H and Evans G 2022 DIALS as a toolkit *Protein Science* **31** 232–50

[13] White T A, Kirian R A, Martin A V, Aquila A, Nass K, Barty A and Chapman H N 2012 *CrystFEL* : a software suite for snapshot serial crystallography *J Appl Crystallogr* **45** 335–41

[14] Keegan R, Wojdyr M, Winter G and Ashton A 2015 *DIMPLE* : a difference map pipeline for the rapid screening of crystals on the beamline *Acta Crystallogr A Found Adv* **71** s18–s18

[15] White T A, Mariani V, Brehm W, Yefanov O, Barty A, Beyerlein K R, Chervinskii F, Galli L, Gati C, Nakane T, Tolstikova A, Yamashita K, Yoon C H, Diederichs K and Chapman H N 2016 Recent developments in *CrystFEL J Appl Crystallogr* **49** 680–9

[16] Gabadinho J, Beteva A, Guijarro M, Rey-Bakaikoa V, Spruce D, Bowler M W, Brockhauser S, Flot D, Gordon E J, Hall D R, Lavault B, McCarthy A A, McCarthy J, Mitchell E, Monaco S, Mueller-Dieckmann C, Nurizzo D, Ravelli R B G, Thibault X, Walsh M A, Leonard G A and McSweeney S M 2010 *MxCuBE* : a synchrotron beamline control environment customized for macromolecular crystallography experiments *J Synchrotron Rad* **17** 700–7

[17] Oscarsson M, Beteva A, Flot D, Gordon E, Guijarro M, Leonard G, McSweeney S, Monaco S, Mueller-Dieckmann C, Nanao M, Nurizzo D, Popov A N, Von Stetten D, Svensson O, Rey-Bakaikoa V, Chado I, Chavas L M G, Gadea L, Gourhant P, Isabet T, Legrand P, Savko M, Sirigu S, Shepard W, Thompson A, Mueller U, Nan J, Eguiraun M, Bolmsten F, Nardella A, Milàn-Otero A, Thunnissen M, Hellmig M, Kastner A, Schmuckermaier L, Gerlach M, Feiler C, Weiss M S, Bowler M W, Gobbo A, Papp G, Sinoir J, McCarthy A A, Karpics I, Nikolova M,

Bourenkov G, Schneider T, Andreu J, Cuní G, Juanhuix J, Boer R, Fogh R, Keller P, Flensburg C, Paciorek W, Vonrhein C, Bricogne G and De Sanctis D 2019 *MXCuBE2* : the dawn of *MXCuBE* Collaboration *J Synchrotron Rad* **26** 393–405

[18] Oskarsson M, Beteva A, Bolmsten F, De Sanctis D, Eguiraun M, Guijarro M, Leonard G, Milan-Otero A, Nan J and Thunnissen M 2018 MXCuBE3 Bringing MX Experiments to the WEB *Proceedings of the 16th Int. Conf. on Accelerator and Large Experimental Control Systems* **ICALEPCS2017** 6 pages

[19] Mueller U, Thunnissen M, Nan J, Eguiraun M, Bolmsten F, Milàn-Otero A, Guijarro M, Oscarsson M, De Sanctis D and Leonard G 2017 MXCuBE3: A New Era of MX-Beamline Control Begins *Synchrotron Radiation News* **30** 22–7

[20] Guijarro M, Felix L, De Nolf W, Meyer J and Götz A 2023 A Marriage Made in BLISS—An Integrated Beamline Control System *Synchrotron Radiation News* **36** 12–9

[21] Guijarro M, Beteva A, Coutinho T, Dominguez M-C, Guilloud C, Homs A, Meyer J, Michel V, Papillon E, Perez M, and others 2017 Bliss-experiments control for esrf ebs beamlines *Proc. 16th Int. Conf. on Accelerator and Large Experimental Control Systems (ICALEPCS'17)* pp 1060–6

[22] Dalesio L R, Kozubal A and Kraimer M 1991 *EPICS architecture* (Los Alamos National Lab., NM (United States))

[23] White G, Shankar M, Johnson A N, Rivers M L, Shen G, Veseli S, Shroff K, Sekoranja M, Cobb T, Korhonen T, and others 2019 The EPICS software framework moves from controls to physics *10th International Particle Accelerator Conference (IPAC2019)* (JACoW)

[24] Chaize J, Meyer J, Pérez M, Götz A, Taurel E and Klotz W 1999 TANGO: An Object Oriented Control System Based on CORBA

[25] Götz A, Taurel E, Pons J, Verdier P, Chaize J, Meyer J, Poncet F, Heunen G, Götz E, Buteau A, and others 2003 TANGO a CORBA based Control System *ICALEPCS2003, Gyeongju, October* 94–8

[26] Duval P 1999 TINE: An Integrated Control System for HERA

[27] Bartkiewicz P and Duval P 2007 TINE as an accelerator control system at DESY *Measurement Science and Technology* **18** 2379

[28] Meents A, Reime B, Stuebe N, Fischer P, Warmer M, Goeries D, Roever J, Meyer J, Fischer J, Burkhardt A, Vartiainen I, Karvinen P and David C 2013 Development of an in-vacuum x-ray microscope with cryogenic sample cooling for beamline P11 at PETRA III SPIE Optical Engineering + Applications ed B Lai (San Diego, California, United States) p 88510K

[29] Burkhardt A, Pakendorf T, Reime B, Meyer J, Fischer P, Stübe N, Panneerselvam S, Lorbeer O, Stachnik K, Warmer M, Rödig P, Göries D and Meents A 2016 Status of the crystallography beamlines at PETRA III *Eur. Phys. J. Plus* **131** 56

[30] The Universal Container Project: https://smb.slac.stanford.edu/robosync/Universal_Puck/

[31] Verdier P, Pons J, Poncet F and Leclercq N 2011 Tango control system management tool *Proceedings of ICALEPCS*

[32] Meyer J, Claustre L, Petitdemange S, Svensson O, Götz A, Coutinho T, Klora J, Picca F, Ounsy M and Buteau A 2012 Tango for experiment control *Proc. PCaPAC* pp 1–3

[33] mxcubecore https://github.com/mxcube/mxcubecore

[34] mxcubeqt https://github.com/mxcube/mxcubeqt

[35] Khokhriakov I, Mazalova V and Merkulova O 2023 Improving Observability of the SCADA Systems Using Elastic APM, Reactive Streams and Asynchronous Communication 6 pages

[36] Cherezov V, Hanson M A, Griffith M T, Hilgart M C, Sanishvili R, Nagarajan V, Stepanov S, Fischetti R F, Kuhn P and Stevens R C 2009 Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 μm size X-ray synchrotron beam *J. R. Soc. Interface.* **6**

[37] Aishima J, Owen R L, Axford D, Shepherd E, Winter G, Levik K, Gibbons P, Ashton A and Evans G 2010 High-speed crystal detection and characterization using a fast-readout detector *Acta Crystallogr D Biol Crystallogr* **66** 1032–5

[38] Miller M S, Maheshwari S, Shi W, Gao Y, Chu N, Soares A S, Cole P A, Amzel L M, Fuchs M R, Jakoncic J and Gabelli S B 2019 Getting the Most Out of Your Crystals: Data Collection at the New High-Flux, Microfocus MX Beamlines at NSLS-II *Molecules* **24** 496

[39] Savko M 2024 https://github.com/MartinSavko/murko

[40] Arvai A https://www.scripps.edu/tainer/arvai/adxv.html

[41] Grimm S 2023 https://github.com/SaschaAndresGrimm/EIGERStreamReceiver

[42] Sparta K M, Krug M, Heinemann U, Mueller U and Weiss M S 2016 *XDSAPP2.0 J Appl Crystallogr* **49** 1085–92

[43] Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T and Bricogne G 2011 Data processing and analysis with the *autoPROC* toolbox *Acta Crystallogr D Biol Crystallogr* **67** 293–302

[44] Battye T G G, Kontogiannis L, Johnson O, Powell H R and Leslie A G W 2011 *iMOSFLM* : a new graphical interface for diffraction-image processing with *MOSFLM Acta Crystallogr D Biol Crystallogr* **67** 271–81

[45] Powell H R, Johnson O and Leslie A G W 2013 Autoindexing diffraction images with *iMosflm Acta Crystallogr D Biol Crystallogr* **69** 1195–203

[46] Paithankar K S and Garman E F 2010 Know your dose: *RADDOSE Acta Crystallogr D Biol Crystallogr* **66** 381–8

[47] Zeldin O B, Gerstel M and Garman E F 2013 *RADDOSE-3D* : time- and space-resolved modelling of dose in macromolecular crystallography *J Appl Crystallogr* **46** 1225–30

[48] Bourenkov G P and Popov A N 2010 Optimization of data collection taking radiation damage into account *Acta Crystallogr D Biol Crystallogr* **66** 409–19

[49] Svensson O, Malbet-Monaco S, Popov A, Nurizzo D and Bowler M W 2015 Fully automatic characterization and data collection from crystals of biological macromolecules *Acta Crystallogr D Biol Crystallogr* **71** 1757–67

[50] Melnikov I, Svensson O, Bourenkov G, Leonard G and Popov A 2018 The complex analysis of X-ray mesh scans for macromolecular crystallography *Acta Crystallogr D Struct Biol* **74** 355–65

[51] Zander U, Bourenkov G, Popov A N, De Sanctis D, Svensson O, McCarthy A A, Round E, Gordeliy V, Mueller-Dieckmann C and Leonard G A 2015 *MeshAndCollect* : an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines *Acta Crystallogr D Biol Crystallogr* **71** 2328–43