ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

myMdC: A CENTRALISED METADATA CATALOGUE FOR SUSTAINABLE DATA MANAGEMENT AT EUROPEAN XFEL

L. Maia*, S. Aplin, D. Boukhelef, F. Dall'Antonia, I. Derevianko, L. Gelisio, D. Goeries, S. Hauf, N. Liss, J. Malka, M. Manetti, G. Previtali, P. Schmidt, V. Singh, J. Szuba, W.-T. Tsai, B. Vanganuru, K. Wrona, European X-Ray Free-Electron Laser†, Schenefeld, Germany

Abstract

Large-scale research facilities such as the European XFEL rely on multiple distributed services to manage complex experimental workflows. Consistent, reliable and timely metadata exchange between these services is essential to ensure automation and data integrity.

This paper presents **myMdC**, the centralised metadata backbone that supports scientific data services in European XFEL. *myMdC* unifies and orchestrates metadata flows across core facility services, enabling interoperability and ensuring alignment with the facility's long-term data management strategy.

In operation since the first day of user experiments in 2017, myMdC ensures real-time injection, documentation, and cataloguing of scientific datasets, forming the basis of data and metadata operations at European XFEL. Its integration with facility services enables immediate dataset tracking from acquisition through publication. It supports automation across the data inventory, control access during embargo periods, metadata preservation, Digital Object Identifier (DOI minting and integration of Findable, Accessible, Interoperable, and Reusable (FAIR) data workflows.

By consolidating metadata management into a unified platform, myMdC facilitates scalable, transparent, and future-proof data practices, ensuring that scientific output remains robust and interoperable within the global research community.

INTRODUCTION

Modern accelerator-based facilities such as the European XFEL operate a diverse ecosystem of services responsible for experiment scheduling, configuration, control, as well as data acquisition and processing. Ensuring that these services have consistent and validated metadata at the right time is essential for automation, interoperability, reproducibility, and overall efficient operation of the facility.

Historically, metadata management has often relied on domain-specific tools or custom scripts, which can lead to duplication, inconsistencies, and operational complexity [1]. At the scale of the European XFEL, such limitations become critical: experiments involve many interdependent services, and produce large volumes of data and metadata such as configuration parameters and provenance information.

To overcome these challenges, *myMdC* [2] was developed at European XFEL as a centralised metadata backbone, designed to deliver a unified and reliable infrastructure for consistent metadata preservation, seamless integration, and orchestration across facility services. *myMdC* serves as the authoritative source of metadata, ensuring consistent and reliable information for all dependent systems. Since its deployment in 2017, it has become a cornerstone of the European XFEL infrastructure, providing high availability across all experiment stations and demonstrating flexibility through seamless integration with newly introduced services.

This paper describes the most important design and operational features of myMdC. We first present its layered architecture, followed by details of the services integration approach and operational lessons from seven years of production use. Finally, we outline future developments.

European XFEL Overview

The European X-ray Free-Electron Laser (European XFEL) is a facility featuring a 3.4 km long linear accelerator producing pulsed, ultra-intense, and coherent X-rays in hard and soft regimes. Spanning from DESY[1] (Hamburg) to Schenefeld (Schleswig-Holstein), it hosts seven scientific instruments powered by three self-amplified spontaneous emission (SASE) sources [3,4]. At any time, one instrument operates per SASE, in either 24 h continuous or 16 h user beamtime / 8 h commissioning cycles. This scheduling maximises beamline efficiency, ensuring full-capacity delivery for each experiment.

In this context, myMdC is responsible for gathering and documenting experiment metadata and for ensuring that proposal-related resources and team assignments are prepared to support a successful experimental campaign.

Data Management

European XFEL generates several petabytes of experiment data per week, amounting to tens of petabytes annually, driven by MHz-repetition-rate X-ray large-area detectors [5]. Data systems perform real-time capture, annotation, and quality control, feeding high-throughput pipelines for calibration, reduction, and feedback during experiments [6–8]. Raw and processed data are stored on European XFEL data systems integrated into the DESY data centre and are managed jointly with DESY [9]. The tiered storage architecture balances rapid access for active experiments and cost-efficient long-term archival, but it requires coordinated data migration between layers. Common services (such as data analysis clusters) are available for users who can access them on-site and remotely.

* Corresponding author: luis.maia@xfel.eu
† https://www.xfel.eu/

[1] Deutsches Elektronen-Synchrotron

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

The metadata describing these workflows and artefacts (critical to ensuring data integrity, reproducibility, and compliance with open-data policies) are captured or tracked in myMdC.

Metadata Capture and Extraction

Metadata capture is integral to preserving the scientific value of the datasets generated at European XFEL. Relevant metadata include beamline configurations, SASE pulse parameters, detector geometries, calibration constants [7], environmental conditions, and experiment control logs, as well as user-provided parameters.

Experiment data and associated metadata acquired during experiments at European XFEL are stored as a facility-specific data format based on HDF5 using Karabo [10]. Data processing pipelines, such as DAMNIT[2] [11], play a key role in enriching datasets with additional metadata. Beyond extracting facility- and user-relevant parameters from control and acquisition systems, DAMNIT also integrates user-provided annotations, adding contextual information.

MyMdC serves as the central point for correlating data repositories and files with their computed metadata, associated identifiers, and high-level metadata (experiment identifiers, proposal numbers, team information, and related publications). This approach ensures interoperability across instruments, supports reproducibility by preserving the complete experimental context, and facilitates the centralised reuse of information and datasets (e.g. enabling recalibration, comparative studies, or re-analysis with improved algorithms) even years after acquisition [7].

Metadata Catalogue

Metadata catalogues are central infrastructure components for organising and preserving the diverse metadata generated throughout the experimental lifecycle. By enforcing facility policies such as embargo periods and access controls, metadata catalogues balance openness with compliance requirements. Modern catalogues increasingly adopt community standards, persistent identifiers, and FAIR principles to enable interoperability across facilities and to link raw data to derived results, publications, and archives.

myMdC PROJECT

The myMdC project was initiated at the European XFEL to develop a versatile metadata catalogue that supports both facility operators and researchers throughout the entire data lifecycle. Its primary objectives include preserving and enabling traceability of information, providing centralised lifecycle management of data across layered infrastructure, ensuring continuous user access to data, and enforcing facility policies such as embargo periods and controlled access. In addition to preservation and reuse, myMdC acts as a metadata backbone that connects key facility services, ensuring each service receives the required information and infrastructure support at the right time – for example, during experiment preparation or data migration. The project has evolved alongside the facility and its user community to support workflows ranging from proposal confirmation to data acquisition, analysis, logbook integration, publication, and long-term archival. It adheres to scientific data management best practices, including the FAIR principles, the use of DOI persistent identifiers, and interoperability with external systems in alignment with initiatives like PaNOSC [12, 13].

Service Overview

MyMdC plays a central role in the European XFEL service landscape by unifying metadata and ensuring consistent traceability across data operations. It provides a policy-compliant framework that integrates diverse services into a cohesive, efficient, and reliable environment for research.

At its foundation, myMdC is a comprehensive metadata catalogue that organises scientific data and enables structured storage, retrieval, and querying of run[3] related datasets (raw, processed, etc.) while ensuring alignment with the facility data management policies. Beyond catalogue features, it supports data migration and location tracking, DOI publication, data workflows, data export, and policy-driven authorisation to restrict access to experiment data during the embargo period to the experiment team.

A critical function of *myMdC* is to manage globally unique entities that underpin experiment consistency and reproducibility: proposal identifiers, sample identifiers, and run types. Enforcing these unique identifiers reduces possible conflicts and preserves integrity across experiments. In addition, it serves as a central integration hub for numerous facility services and subsystems, as illustrated in Fig. 1.

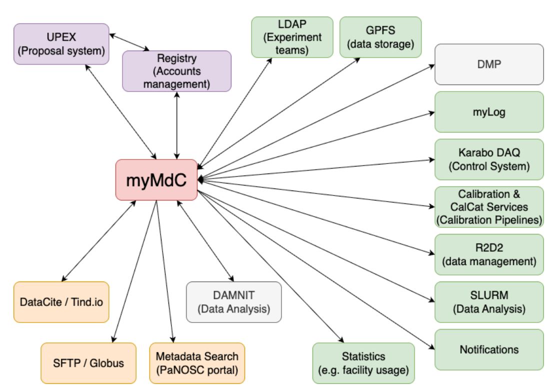

Figure 1: myMdC services orchestration and integration.

MyMdC interfaces with several key systems that support proposal and user management, authentication, storage, and computation. It connects to the proposal and user management platform UPEX [14] and the DESY Registry [15], and supports authentication through LDAP [16] and Keycloak [17]. The system integrates with storage solutions using GPFS [18] and dCache [19], the data-migration service *R2D2*, and the compute scheduler SLURM [20].

[2] Data And Metadata iNspection Interactive Thing

[3] In the context of European XFEL, a *run* refers to a contiguous data acquisition period during an experiment, typically associated with a specific sample or experimental configuration

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

In addition, *myMdC* interfaces with systems supporting the full experiment lifecycle, including the Karabo control system [10], DAQ [21], and calibration services [7], as well as external services such as calCat [22] (calibration constants database), myLog [23,24] (experiment logbooks using Zulip as backend), and data export platforms (SFTP [25,26], Globus [27]). It also supports data publication via DataCite [28,29] and Tind.io [30,31], and facilitates open-data discovery through the PaNOSC data portal [12,13].

Moreover, myMdC exposes RESTful APIs [32,33] for programmatic access, provides notification mechanisms, and supplies statistical reporting for operational insight. The development team is currently working on additional interfaces with the facility's forthcoming Data Management Plan (DMP) solution [34] and with DAMNIT [11].

Service Design

MyMdC is a Ruby on Rails application following the Model–View–Controller (MVC) architecture, separating data, business logic, and presentation layers [35] (Fig. 2).

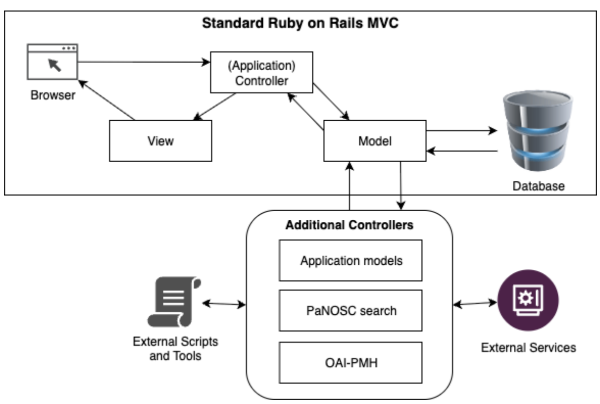

Figure 2: Ruby on Rails standard MVC architecture in myMdC, alongside its additional custom controllers for API integration and service interoperability.

The backend defines the data model and uses schema migrations to maintain consistency and adaptability during development. Application functionality is exposed through controllers as RESTful APIs and web endpoints, enabling interaction with external services and facility infrastructure via HTTPS and TCP. Views are server-side rendered; additionally, myMdC provides controllers that expose further RESTful endpoints (e.g., OAI-PMH, PaNOSC search endpoints, and application models) as shown in Fig. 2.

This layered design enables integration with front-end frameworks, external scripts, and standalone applications, with the Python client *metadata-client* [36] supporting integration with facility environments and user workflows.

The service runs on five dedicated production servers and uses a three-node Percona MySQL [37] cluster to ensure myMdC's high availability and redundancy, especially during facility operation.

Overall, the layered approach ensures modularity, scalability, and maintainability.

FEATURES AND FUNCTIONALITIES

MyMdC provides a broad set of features to support the full data lifecycle at European XFEL, from proposal preparation to data publication. The functionalities ensure consistent metadata management, secure access control, and tight integration with other facility systems, aligned with established Data Management policies. Key features are organised below by thematic categories.

Proposal and User Management

Proposal and user management in myMdC enforces secure, policy-driven access to experimental data and metadata. The system implements hierarchical roles for users (members of experiment teams), facility experts (European XFEL personnel involved in experiments), and systems, connected to proposal-specific responsibilities that manage team activities throughout the proposal lifecycle.

myMdC Roles are organised in three main categories:

- **Administrative:** Includes *Admin*, which grants full control to myMdC administrators, IT, and Data Management experts, and *Trusted*, which provides broad read access for operational monitoring and auditing.
- **Users & Staff:** The *read_only* role is the default for all users, allowing basic access to the system.
- **System Integration:** Roles assigned to service accounts that enable automated interactions via APIs and backend processes using OAuth2.

myMdC Proposal Responsibilities Responsibilities are defined in the context of individual proposals and determine the actions that users and facility experts can perform throughout the experiment lifecycle (see Fig. 3).

- **PI, MP**: Principal Investigator / Main Proposer – full control of team management and proposal settings; can publish metadata and data.
- **EDC**: Experiment Data Contact – team-designated contact for data topics throughout the lifecycle.
- **LC**: Local Contact – facility-designated Instrument Expert; during beamtime holds permissions comparable to PI/MP; can add/update samples and finalize proposals after beamtime.
- **LDC**: Local Data Contact – facility-designated specialist supporting data-related tasks.
- **Experiment Team**: members may have up to three responsibilities: *Data Manager*, who can access online cluster data, install software under the experiment folder, assess run quality, update run metadata, and request calibrations; *Participant*, who can access data in the experiment folder, use experiment resources, and define run types; *Lab User*, who indicates physical access to the instrument laboratory.
- **Support Team**: members who can access metadata and the logbook, but have no access to proposal data.

Instrument Roles Responsibilities are contextual to scientific instruments and define what facility experts can do during the experiment lifecycle:

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

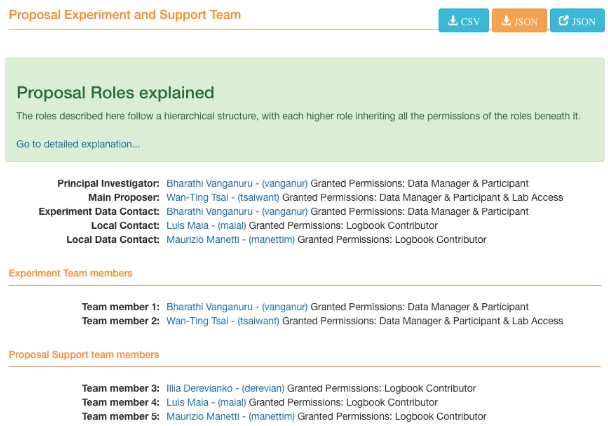

Figure 3: *myMdC* interface displaying proposal team members along with their responsibilities and associated access rights.

- **Leading Scientist** and **Deputy Leading Scientist** are instrument experts with permissions comparable to PI/MP and LC; they can set facility-related proposal parameters such as the number of allocated shifts.
- **Instrument Experts** are all the remaining instrument specialists who, during the beamtime (usually 3–6 days), hold permissions comparable to the LC.
- **Instrument Support Group** members are defined per instrument and have read permissions to all instruments' proposals data and metadata.

Hierarchical Permissions Roles and responsibilities follow an inheritance model: higher roles include the permissions of lower roles. The mapping between *myMdC roles* and *proposal responsibilities* defines the overall security model of myMdC.

Integration with UPEX Proposal System

myMdC is tightly integrated with *UPEX*, the European XFEL proposal submission and evaluation system. UPEX manages the early proposal lifecycle (submission, review, and beamtime allocation). Once a proposal is approved and beamtime is awarded, myMdC leverages this planning information to create the operational environment for the experiment (see Fig. 4).

All users accounts and associated roles originate from UPEX, ensuring that only users with valid credentials and who have accepted the European XFEL Scientific Data Policy can access myMdC services and proposal data. Continuous synchronization between UPEX and myMdC guarantees policy compliance, accurate team membership, and consistent identity management across both systems.

When a Main Proposer (MP) confirms assigned beamtime in *UPEX*, myMdC is notified and translates proposal planning into operational reality by:

- Creating the proposal record in myMdC, and associated metadata such as samples, instrument, experiment cycle, and team members;
- Pre-assigning storage locations and generating LDAP groups for access control, which will be later applied to proposal directories through ACLs[4];
- Initialising integration hooks for downstream systems, including logbooks, data repositories, DAQ, calibration workflows, and DOI assignment.

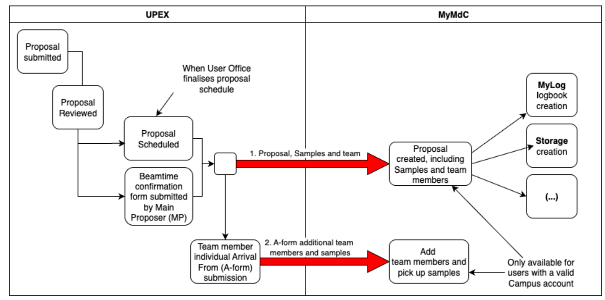

Figure 4: Integration workflow between *UPEX* and *myMdC* for team and sample synchronization after beamtime confirmation and arrival form submission.

This automation reduces administrative overhead, lowers error rates, and aligns experimental data management with scientific planning. It also documents actual execution, as real conditions sometimes deviate from initial plans.

Facility, Instruments and Data Repositories

MyMdC maintains structured metadata about the facility and its instruments, instrument cycles, and data repositories, which is essential for enabling consistent data management and a soft link between proposals and their produced data. This information allows myMdC to determine proposal-specific storage locations. A typical storage path might look like `XFEL/raw/FXE/201701/p000010/r0002`. This structure encodes several hierarchical components: the facility (**XFEL**), the data type (**raw**), the instrument (**FXE**), the instrument cycle (**201701**), the proposal identifier (**p000010**), and the run number (**r0002**).

Proposal storage directories are created, automatically by myMdC, four weeks before the experiment starts. This gives users enough time to familiarise with the facility infrastructure, organise their workflows, and prepare necessary software.

Access control for these storage locations that is implemented on the filesystems level uses LDAP-based group management. For each proposal, myMdC creates three LDAP groups (*Data Manager*, *Participant*, and *Lab User*) that define permission levels, and applies them to proposal directories via ACLs, ensuring secure and policy-compliant data access during the experiment lifecycle.

Workflow and Notification Services

The platform streamlines proposal handling with well-defined workflows and automated notifications. Users receive alerts for key events — proposal-folder readiness, cali-

[4] Filesystem Access Control Lists

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

bration requests, or quality assessments — ensuring transparency and timely operational responses.

A key workflow is the proposal lifecycle, which starts with infrastructure preparation, proceeds through proposal execution, and ends with data assessment and cleanup of data as illustrated in Fig. 5.

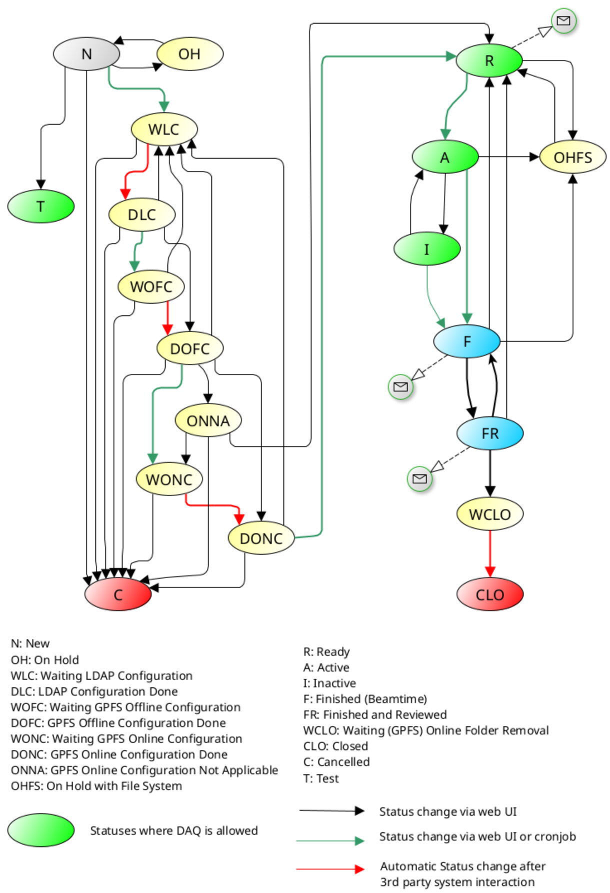

Figure 5: Overview of proposal lifecycle in *myMdC*, illustrating all possible status transitions and associated notifications.

Integration with myLog Experiment Logbook

The experiment logbook, provided by *myLog* [23, 24], documents operational and scientific activities before, during and after beamtime. myMdC orchestrates the creation, configuration, and maintenance of these logbooks to ensure consistency and synchronization across systems.

Creation of the Experiment Logbook After beamtime acceptance, myMdC allows the creation of the corresponding myLog, which enables Instrument Experts to engage early with the experiment team. The logbook provides a structured platform to document preparatory decisions in collaboration with facility experts such as the Local Data Contact. Each logbook includes a default set of topics configurable per instrument to support instrument-specific workflows and data capture.

Dynamic Team Synchronization Team membership in myLog is automatically synchronised with myMdC: adding a person to the experiment team grants logbook access, while removing them revokes it. Instrument Experts retain access to all logbooks of their instrument to provide oversight and support as long as they serve this role.

Access and Information Sharing MyMdC efficiently propagates and share logbook information, supporting transparency and real-time troubleshooting. Logbooks will be exported for archiving or publication, aligning with open-data initiatives. This integration ensures that experimental activities are documented, traceable, and persistently accessible, reinforcing reliability and long-term data stewardship.

Integration with DAQ System

The Data Acquisition (DAQ) system records raw data during beamtime and integrates with myMdC to permanently link proposals, runs, and acquired data [21]. This bidirectional exchange eliminates manual post-run imports, reduces inconsistencies, and ensures integrity, and traceability across the experiments, enhancing reproducibility.

Run Start and Metadata Exchange To initiate a run in DAQ, a *run type* (e.g. flat-field, Diffraction) must be selected. Optional fields such as sample and technique enable richer metadata capture from the start.

The DAQ interacts with myMdC via RESTful APIs to send run metadata and request the next available run identifier for a proposal. In response, myMdC registers a globally unique run ID, records its metadata, and returns the run ID and storage path.

Run Stop At run completion, DAQ uses myMdC APIs to finalize metadata, ensuring complete and consistent records for subsequent data assessment and calibration.

Run Management and Experiment Context

myMdC provides a unified framework for run management, integrating quality assessment, calibration workflows (primarily addressing detector corrections), and descriptive annotations. Each run is assigned a globally unique identifier and linked to contextual metadata, including proposal number, run type, sample description, techniques, associated datasets, files, and relevant parameters.

When a run ends, myMdC allows **Data Managers** to assess run quality (see Fig. 6) and trigger calibration workflows [7]. Each run consolidates metadata (type, sample, techniques, data groups, files, configuration), calibration references for reproducible data correction/processing, and a quality status: *Good*, *Unclear*, or *Not Interesting*.

This status drives retention workflows: runs marked *Good* or *Unclear* are migrated to long-term storage via the data-migration service (*R2D2*) service, while non-essential data marked as *Not interesting* may be skipped. *MyMdC* also tracks data locations, repository synchronisation, and retention policies.

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

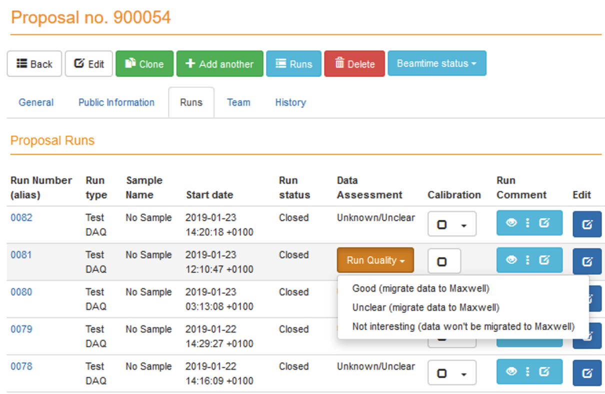

Figure 6: Interface for run management in *myMdC*, including tools for data quality assessment and calibration requests.

PaNET Techniques Integration

myMdC imports the PaNET ontology of experimental techniques [38], which is curated at the facility level and for each instrument. Users are encouraged to select the appropriate technique in the DAQ for each run. This metadata will enable future workflows where suitable calibration pipelines are automatically applied based on the selected technique.

Data Source Management

MyMdC provides mechanisms to manage which data sources (detectors, sensors, Karabo devices) are captured during an experiment, offering both flexibility and consistency. Central to this are *Data Source Groups* — collections of related sources deemed relevant for recording. These groups provide fine-grained control data relevant to the run information and reproducibility.

Integration with Karabo via RESTful APIs provides real-time synchronization, automatically reflecting planned acquisitions. Before acquisition, users define the required Data Source Groups. If necessary, during experiments, authorised team members can update these groups, with changes instantly propagated to maintain alignment between planning and live operations.

Data Publishing and DOI Management

To support FAIR principles, myMdC integrates with DataCite [28, 29] for DOI assignment. This enables publication of persistent identifiers, making them discoverable and citable.

Currently, DOI assignment is automatically triggered for instruments and proposals. Future extensions aim to assign DOIs at the dataset level so data sets within *RAW*, *RED*, and *PROC data* [39] can be individually referenced and preserved.

Technical Integrations with External Services

myMdC exposes comprehensive RESTful APIs [32, 33] for programmatic access to all its models, which include proposal data, run metadata, and data transfer operations. These APIs enable automated integration with custom workflows and third-party applications.

A service-oriented architecture combined with open APIs allows myMdC to adapt rapidly to evolving scientific and operational requirements, ensuring flexibility, scalability, and long-term sustainability.

Open Data and Long-Term Access

Aligned with European XFEL data policies, myMdC supports open-data workflows, including automated release of embargoed datasets and archiving of related publications. A pilot project enables published datasets to be discoverable via the PaNOSC portal[5] through a public OAI-PMH endpoint, allowing programmatic metadata harvesting.

FUTURE DEVELOPMENTS

myMdC will continue evolving as a central platform for metadata management and open science at European XFEL. The next development priorities target both institutional requirements and the broader community's expectations:

- **Dataset inventory and DOI assignment:** Enable DOIs for individual datasets to enhance citation and traceability. To preserve the consistency between filesystem and myMdC records, periodic checks will help detect and fix any discrepancies (e.g. emergency data movement).
- **DMP integration:** Enhancing metadata annotation and documenting key decisions.
- **DAMNIT integration:** Enriching dataset metadata to enhance searchability and support advanced metadata preservation, search and data analysis workflows.
- **New Scientific Data Policy [34,39]:** Implementing features like the *RED data* concept to identify and preserve critical data for long-term public access.
- **Open-data workflows:** Extending programmatic interfaces (e.g., OAI-PMH) for better metadata harvesting, interoperability, and AI/ML-driven discovery through portals like PaNOSC.

CONCLUSION

MyMdC is the central backbone for metadata management and system integration at European XFEL, supporting all stages of the experimental lifecycle—from proposal submission to data acquisition, data processing, and data publication. By providing automated workflows, role-based access control, and seamless integration with systems such as UPEX, Karabo, myLog, and PaNET, myMdC ensures traceable, reproducible, and FAIR-compliant data management. Its DOI workflows and planned DMP/DAMNIT integration will support the facility's efforts to open science and long-term data stewardship.

Overall, *myMdC* streamlines operations, enforces consistent policies across all instruments, reduces administrative overhead, serves as the single user entry point, and maximizes the scientific impact of European XFEL experiments.

[5] `https://www.panosc.eu/`

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

Looking ahead, upcoming developments will strengthen its role as a cornerstone of metadata governance and open science, enhancing transparency and empowering the community with sustainable access to open data.

ACKNOWLEDGEMENTS

The authors would like to thank all colleagues and collaborators at European XFEL who contributed to the development, deployment, and maintenance of *myMdC*. Their expertise, feedback, and continuous support have been invaluable in shaping the system and ensuring its success across the facility. We also acknowledge the broader user community for their engagement and constructive input, which has guided the evolution of *myMdC* and its integration with complementary services. Furthermore, we extend our gratitude to international initiatives such as PaNOSC and DAPHNE and related collaborative projects for promoting interoperability and best practices in data management, which have significantly influenced the design and development of *myMdC*.

REFERENCES

[1] R. Eichler*et al.*, "Enterprise-Wide Metadata Management: An Industry Case on the Current State and Challenges,"Business Information Systems, pp. 269–279, Jul. 2021. doi:10.52825/bis.v1i.47

[2] myMdC – European XFEL Metadata Catalog, https://in.xfel.eu/metadata

[3] T. Tschentscher *et al.*, "Photon Beam Transport and Scientific Instruments at the European XFEL", *Appl. Sci.*, vol. 7, no. 6, p. 592, Jun. 2017. doi:10.3390/app7060592

[4] W. Decking *et al.*, "A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator", *Nat. Photonics*, vol. 14, no. 6, pp. 391–397, May 2020. doi:10.1038/s41566-020-0607-z

[5] The European X-ray Free Electron Laser, https://www.xfel.eu

[6] M. Kuster *et al.*, "Detectors and Calibration Concept for the European XFEL". *Synchrotron Radiat. News*, vol. 27, no. 4, pp. 35–38, Jul. 2014. doi:10.1080/08940886.2014.930809

[7] P. Schmidt *et al.*, "Turning European XFEL raw data into user data", *Front. Phys.*, vol. 11, Jan. 2024. doi:10.3389/fphy.2023.1321524

[8] E. Sobolev *et al.*, "Data reduction activities at European XFEL: early results", *Front. Phys.*, vol. 12, Feb. 2024. doi:10.3389/fphy.2024.1331329

[9] J. Malka *et al.*, "Data Management Infrastructure for European XFEL", in *Proc. ICALEPCS2023*, Cape Town, South Africa, Sep. 2023, p. 952-957. doi:10.18429/JACoW-ICALEPCS2023-WE1BCO02

[10] D. Göries *et al.*, "The Karabo SCADA System at the European XFEL", *Synchrotron Radiat. News*, vol. 36, no. 6, pp. 40–46, Nov. 2023. doi:10.1080/08940886.2023.2277650

[11] DAMNIT – Data And Metadata iNspection Interactive Thing, https://damnit.readthedocs.io/

[12] PaNOSC: Photon and Neutron Open Science Cluster, https://www.panosc.eu

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

[14] UPEX – User Portal for Experiment Scheduling, https://in.xfel.eu/upex

[15] D. Jahnke-Zumbusch and D. Knittel, "DESY's service for the management of central systems with regard to accounts, groups and mailboxes – the DESY Registry (ver. R2)", presented at HEPiX 2023, Victoria, Canada, Oct. 2023, unpublished. https://bib-pubdb1.desy.de/record/597025

[16] LDAP – Lightweight Directory Access Protocol, https://ldap.com/

[17] Keycloak: Open-Source Identity and Access Management, https://www.keycloak.org

[18] GPFS – IBM Spectrum Scale (General Parallel File System), https://www.ibm.com/products/spectrum-scale

[19] M. Ernst, P. Fuhrmann, M. Gasthuber, T. Mkrtchyan, C. Waldman, "dCache, a distributed data storage caching system", presented at CHEP2001, Beijing, China, Sep. 2001, unpublished.

[20] SLURM – Simple Linux Utility for Resource Management, https://slurm.schedmd.com/

[21] S. Hauf *et al.*, "Lessons learned refactoring the EuXFEL's central data acquisition system", presented at ICALEPCS'25, Chicago, IL, United States, Sep. 2025, paper WECG003, this conference.

[22] calCat – Calibration Constants Database, https://in.xfel.eu/calibration

[23] myLog – European XFEL Experiment Logbook, https://mylog.connect.xfel.eu

[24] L. Maia *et al.*, "myLog: A Modern, Integrated Logbook for Scientific Collaboration at European XFEL", presented at ICALEPCS'25, Chicago, IL, United States, Sep. 2025, paper THMG013, this conference.

[25] SFTP – Secure File Transfer Protocol, https://en.wikipedia.org/wiki/SSH_File_Transfer_Protocol

[26] European XFEL SFTP Service, sftp.xfel.eu

[27] Globus – Research Data Management Service, https://www.globus.org/

[28] DataCite – DOI Registration for Research Data, https://datacite.org/

[29] European XFEL DataCite Repository, https://commons.datacite.org/repositories/xfel.eu

[30] TIND – Digital Repository Solutions, https://tind.io/

[31] European XFEL TIND Repository, https://xfel.tind.io/

[32] RESTful APIs and myMdC API Documentation, https://restfulapi.net/

[33] myMdC API, https://in.xfel.eu/metadata/api-docs/index.html

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-THCR005

[34] J. Malka *et al.*, “Towards the implementation of a new European XFEL scientific data policy – challenges and chances,” presented at ICALEPCS’25, Chicago, IL, United States, Sep. 2025, paper TUMG015, this conference.

[35] Ruby on Rails Guides, https://guides.rubyonrails.org

[36] Python Client for the European XFEL Metadata Catalogue (related to myMdC), https://pypi.org/project/metadata-client/

[37] Percona Server for MySQL, https://www.percona.com/software/mysql-database/percona-server

[38] S. P. Collins *et al.*, “ExPaNDS ontologies v1.0”, *Zenodo*, Jun. 2021. doi:10.5281/zenodo.4806026

[39] Scientific Data Policy (2023), https://www.xfel.eu/users/policies/index_eng.html