ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-TUPD009

INTEGRATED DATA ACQUISITION AND PROCESSING PIPELINES FOR USERS AT ELETTRA 2.0: A CASE STUDY AT SYRMEP, THE µCT BEAMLINE

A. Contillo*, M. Belletti, R. Borghes, V. Chenda, A. Hafner, G. Kourousias, E. Longo, A. Olivo, M. Prašek, M. Turcinovich, Elettra-Sincrotrone Trieste S.C.p.A., Trieste, Italy

Abstract

Elettra, the Italian Synchrotron in Trieste, is about to undergo a major upgrade of the facility. To effectively exploit such improvement, data acquisition and processing are being integrated into single automated pipelines, subject to a facility-wide standard and yet flexible enough to accommodate the specific usage at each beamline. The entire procedure of data acquisition and processing spans a vast ecosystem of different structures and frameworks, which are being standardized across the whole facility: the GeCo control system, handling the safety and the operation of the beamlines; the TANGO Controls framework, which allows distributed control of the data acquisition process; the architecture of data storage and processing, whose accessibility is mediated by VUO, the Elettra unified portal; a modular adaptive processing infrastructure (MAPI) for analysis workflows; and an overview of the data lake data@Elettra. The intertwining of all these components results in the integrated pipeline experienced by the users. The acquisition/processing sequence presently in place at SYRMEP, the microtomography beamline, is presented as a case study of the standard structure that is being designed. Beamline and acquisition control, on-the-fly, and post-acquisition processing are described in the light of the general landscape proposed for Elettra 2.0, the upgraded facility.

INTRODUCTION

The third generation Italian Synchrotron facility Elettra has been running since the first light in 1993, until the last shutdown in July 2025. During these decades, it has served the international scientific community by providing X-ray light to almost thirty beamlines, offering a variety of thechniques such as imaging, spectroscopy, diffraction, scattering, etc, for 160 thousand hours of operation. More than 11 thousand beamtimes were conducted during this period, resulting in almost 10 thousand peer-reviewed publications. Despite the constant upgrades and improvements, the facility has finally turned old, requiring a deep renovation in order to maintain its appeal to an international audience. Thus the upgrade to a fourth generation synchrotron light source, with highly enhanced brilliance and coherence. The first light is expected to be produced by the fall of 2026 and the first external users are expected by early 2027.

SYRMEP (SYnchrotron Radiation for MEdical Physics) [1] is the beamline of Elettra dedicated to microtomography, mainly devoted to life sciences research, but also to material science, geology and other neighboring fields. Tomography is one of the setups with the highest throughput, producing data in the order of ten megabytes per projection (the exact figure depending on the surface area of the detector). With the currently available fluxes, SYRMEP was able to acquire up to 100 projections per second, which brought the data acquisition rate to about 800 MB/s. Depending on the specific details of the beamtime (sample size, quickness of sample switching, etc.), this has allowed at times the acquisition of several terabytes per day. Following the upgrade, the photon flux available to the beamlines is expected to increase by three orders of magnitude. Therefore, even in the most conservative estimates (in which the limiting factor, rather than the signal integration time, is the dead time due to motor movements and/or data readout), the rate of data acquisition is expected to increase by at least a factor 10.

SYRMEP-LS (Life Sciences) is the upcoming embodiment of the imaging beamline in the upgraded facility. The sample-to-detector distance will be increased drastically, so as to make the best use of the increased beam coherence in a free space propagation-based phase contrast setup [2, 3], thus reducing the dose of X-rays needed to produce meaningful images. The surface area of the deployed detectors will also be enlarged, possibly even doubled. All things considered, the expected data throughput is expected to increase to 10-20 GB/s. Such rates pose challenges at several different levels: I) the data transmission infrastructure, II) the data storage capacity, III) the data access speed, IV) the data processing power. Each of these aspects must be tackled in such a way to avoid bottlenecks that would slow the whole pipeline of data acquisition and processing.

The Information Technology department of Elettra is enforcing a facility-wide standard for all components of the above-mentioned pipeline. The massive requirements of the beamline SYRMEP in terms of data throughput, storage, and processing make it an exemplary case study to describe these components in detail. In particular, the vacuum and beam transport elements are handled by the GeCo control system [4]. The control of the instrumentation and their coordination into data acquisition sequences are governed through the TANGO Controls [5] infrastructure of the beamline. The processing of raw data is handled by STP3 Web, a distributed software based on a Modular Adaptive Processing Infrastructure (MAPI) [6]. The access to stored data is mediated by the Virtual Unified Office (VUO) of Elettra, a multipurpose portal which (among many other things) handles the data access permissions via user authentication.

---

* adriano.contillo@elettra.eu

Finally, a Data Lake [7] structure is in place to act as a repository for simplified storage and access. Each of the following sections is devoted to one of said components.

GECO CONTROL SYSTEM

The name GeCo is an acronym for the Italian words Gestione (management) and Controllo (control). It designates a beamline interlock system, that monitors the beamline components and the vacuum elements and operates them according to a set of predefined safety conditions.

This safety logic is enforced by a Programmable Logic Controller (PLC) architecture, consisting of a master PLC (Siemens S7 1500 – CPU 1513 - 1PN) and one or more peripheral slave units (Siemens ET200MP - IM155-5 PN ST3-1). A specialized 6U crate has been engineered in-house to host the PLC instrumentation. The master and the slave units are connected *via* PROFINET fieldbus, while a network communication module on the master unit (Siemens CM 1543–1) allows software clients to interact with the system using a Modbus protocol. The master unit controls the vacuum and beam transport elements of the initial part of the beamline, including the front-end. The slave units are dedicated to the control of the final parts of the beamline, that are often split into several branches.

The GeCo PLC firmware has been developed with the Siemens programming tool Tia Portal [8], using the Structured Control Language (SCL). A template project, suitable for every beamline, is customized on the basis of the interlock rules of each beamline. A set of Python scripts is used to translate the interlock logic (which is provided in the form of a spreadsheet) into the datablocks and I/O tag tables used by the PLC program.

The aforementioned network communication module allows the binding of the PLC logic to the TANGO infrastructure, which governs the data acquisition - and is described in detail in a dedicated section of this paper. A Python-based TANGO DS has been developed specifically in integration with the GeCo system: at startup, the DS parses the content of the *data exchange* datablock and creates a dynamic TANGO attribute for each of its elements. In addition, it instantiates and updates attributes for logs, alarms, and interlocks. This automatic configuration eliminates the need to intervene on the DS source code when a new beamline is installed or an existing beamline setup is modified.

The status of the beamline can be monitored and controlled through a synoptic web platform, DonkiWeb, which integrates a web server written in Python and a GUI Javascript library that provides widgets for the graphical interface. The multithreaded web server offers a REST interface to the TANGO infrastructure. Through a defined URL format it is possible to read and write attributes, send commands, or subscribe to a data stream web-socket server.

TANGO INFRASTRUCTURE

TANGO Controls is a well-known framework for the development of distributed, object-based control systems. The building block of a TANGO infrastructure is the Device Server (DS), a software implementing custom code and a network communication with the rest of the infrastructure. The structure of a DS allows the execution of TANGO commands, as well as the storage of local variables into TANGO attributes, both accessible from anywhere within the infrastructure.

Ideally, the beamlines at Elettra have a TANGO DS for each instrument deployed on the beamline, connected to the physical device and acting as its digital interface. Variables such as detector readouts or motor positions are translated into TANGO attributes of the corresponding DS. Instruments are operated (started and stopped, moved, opened and closed, etc.) *via* TANGO commands.

On top of this layer of *physical* TANGO DS, each beamline has a number of *abstract* DS whose purpose is to coordinate the interaction between the physical ones. Such DS is called Executer [9] as it executes custom scripts that handle complex actions, like opening and closing the lightpath or even an entire data acquisition sequence, from the preliminary checks to the triggering of the post-acquisition data processing. A beamline like SYRMEP has a given number of "standard" acquisition modalities, each encoded in one instance of the Executer TANGO DS. The most common modality (basic tomography scan) includes 1) a check of storage space; 2) a dark field acquisition; 3) a flat field acquisition; a synchronized sample rotation and tomographic acquisition ; 4) the writing of the acquired data into an HDF5 file [10]. All these tasks are accomplished by TANGO DS and coordinated by an instance of the Executer.

Apart from the servers, a TANGO infrastructure obviously allows clients. These can interact with the servers in several ways, from dedicated software to scripting (mainly through the Python module PyTango [11]). Beamline users are provided with a series of graphic interfaces, based on the custom-made utility DonkiGui (relying on PyTango and Taurus [12]), which are easily constructed from ready-made widgets and allow intuitive interactions with both the physical DS and the abstract ones. As an example, SYRMEP users have access to panels that display the positions of the sample stage motors, panels that govern the alignment of the sample stage and the rocking procedure of the monochromator, panels that trigger and and monitor the unfolding of the tomographic acquisition.

STP3 WEB ANALYSIS SYSTEM

STP3 Web represents a transformative implementation of the MAPI framework, specifically designed to address the computational tomography challenges of SYRMEP. Built upon proven ASTRA Toolbox [13] reconstruction algorithms, this web-based system modernizes traditional desktop workflows while maintaining algorithmic integrity and performance standards.

The architecture leverages NiceGUI [14] with Material UI components for intuitive user interfaces, enabling users to access reconstruction capabilities from any web-capable

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-TUPD009

device. This platform-agnostic approach eliminates the installation of any libraries or dependencies and facilitates remote collaboration during experimental sessions. Interactive workflows allow users to optimize reconstruction parameters using preview sinograms before submitting full-volume processing tasks.

Central to STP3 Web effectiveness is its distributed computing capability through Celery [15] task queues and MariaDB database integration. Job parameters are stored persistently, enabling asynchronous execution where computational resources are allocated dynamically based on availability. The system seamlessly integrates with VUO (see dedicated section) for user authentication and data repository access, ensuring secure operations within the infrastructure of the facility.

Performance improvements have been substantial, with interactive workflow speeds increasing by over 600% through responsive interfaces, intelligent caching, and optimized data handling. Batch reconstruction performance improved by 200% in sequential mode, with greater gains achievable through parallel execution. These enhancements stem from eliminating file conversion bottlenecks, implementing intelligent GPU scheduling across available Tesla T4 and L40S cards, and streamlining data paths between storage and processing nodes.

Currently serving multiple daily users in full production, STP3 Web processes up to 3 TB of tomographic data per day while maintaining responsive multiuser operations. The system operates on INTEGRA compute nodes featuring 64-96 CPU cores, multiple GPU accelerators, and high-speed 100 Gbit/s network connectivity to the storage system. Continuous monitoring through Zabbix ensures reliable operation during critical experimental periods, while automated deployment from GitLab repositories maintains version control and streamlined updates.

VIRTUAL UNIFIED OFFICE

The Virtual Unified Office, usually abbreviated as VUO, was deployed in the late 1990s to digitalize the (physical) Users Office. It integrates several Elettra management systems into a single portal, based on an Oracle database and running on Linux OS with Apache [16].

Among the tools available through VUO, a few are worth mentioning as the most useful from a user perspective. The menu "*My investigations*" allows access to the raw data acquired during a given beamtime, arranged in a tree-like directory structure of the form Investigation/Experiment/Dataset. Access is granted either through the web interface of VUO, or by means of remote file browsing services. Permissions are handled by VUO itself, based on an Access Control List (ACL) that can be updated dynamically by the administrators of the relevant *tag* (a sort of label that can be applied to any structure within VUO to link it to a specific beamline or laboratory). Currently, VUO allows the upload of raw data in any format, but the goal for Elettra 2.0 is to force the HDF5 format as a facility-wide standard.

VUO also provides access to computational resources. Through the menu "*My applications*", users can connect remotely to the beamline control workstations, and access a variety of web interfaces for scientific applications like Fiji [17], Jupyter Notebook [18], or custom Python Scientific toolboxes. Access to these services is also handled *via* the ACL and tag system.

Rocket.chat [19] is an open source communication platform that Elettra has adopted as the standard for all activities involving direct messaging. Originally introduced during the recent COVID-19 pandemic to facilitate the interactions among colleagues suddenly working mostly from remote, it has survived the general return to the workplace and spread its application to other contexts. VUO allows the association of dedicated chat channels to the users of each beamtime and the dispatch of automatic messages through the TANGO control system of the beamline, allowing the users to receive timely warnings about the progress of the experiments.

SCIENTIFIC DATA LAKE

The Elettra Scientific Data Lake (EDL) provides the foundational data management infrastructure supporting the high-throughput experimental workflows of SYRMEP. This specialized implementation of modern data lakehouse architecture addresses the unique requirements of synchrotron facilities, combining data lake flexibility with data warehouse governance to manage heterogeneous scientific data streams effectively.

EDL operates entirely on local infrastructure, ensuring data sovereignty and microsecond-level latencies critical for real-time experimental feedback. The heterogeneous ecosystem spans from edge computing devices at beamlines handling multi-gigabyte-per-second data streams to centralized HPC resources optimized for intensive reconstruction calculations. Storage employs a tiered architecture with NVMe for active datasets, disk arrays for frequently accessed data, and automated migration to tape libraries for long-term preservation.

Scientific data management within EDL embraces FAIR (Findable, Accessible, Interoperable, Reusable) [20] principles through sophisticated metadata layers and persistent Digital Object Identifiers (DOI). HDF5 serves as the primary container format for complex experimental datasets, preserving crucial metadata relationships between acquisition parameters and processed results. The architecture supports streaming ingestion for real-time monitoring, automated quality assessment, and comprehensive provenance tracking linking raw projections to reconstructed volumes.

For offline archiving, EDL utilizes an IBM Spectrum Archive 4500 tape library with 8 LTO-8 drives, providing 14 petabytes of uncompressed storage across 1200 tapes. Custom software built on RESTful APIs with Python and Celery workers ensures data integrity through double-copy storage and SHA512 checksums. Raw experimental data are automatically archived in duplicate immediately after acquisition, while Principal Investigators retain autonomous control over data migration and restoration processes.

The effectiveness of the system manifests through specialized web applications including STP3 Web for tomographic reconstruction, enabling researchers to transform raw experimental data into scientific insights through intuitive interfaces. Integration with VUO provides unified authentication and role-based access control, ensuring compliance with ISO27001 security standards while facilitating seamless data access across the computational infrastructure of Elettra.

CONCLUSION

In order to qualify as a state-of-the-art research facility and a thrustworthy scientific institution, the upgraded Elettra 2.0 will need to comply to the modern-day requirements that are expected by the international scientific community. From a control system perspective, these include the simplification and optimization of beamline usage, the accessibility of the acquired data, and the availability of computing resources for (almost) real-time processing and analysis of such data. As a consequence, the Information Technology department of Elettra put an effort in defining optimized, facility-wide standards for instrument control, data storage and handling.

With these systems and structures in place, Elettra 2.0 will be as appealing to the present day audience as the original installation has been since the 1990s until today. The whole crew of Elettra is looking forward to that moment, and is working hard to build the best possible version of it.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the entire staff of the SYRMEP beamline and of the Information Technology department (in particular the ICT group, the Software for Experiments group, the Business Intelligence group, and the Scientific Computing group) for their support and contributions to the development and testing of the solutions presented in this publication.

REFERENCES

[1] E. Longo *et al.*, “SYRMEP beamline: state of the art, upgrades and future prospects”, *Eur. Phys. J. Plus*, vol. 139, no. 10, p. 880, Oct. 2024. `doi:10.1140/epjp/s13360-024-05489-1`

[2] T. E. Gureyev, Y. I. Nesterets, A. Kozlov, D. M. Paganin, and H. M. Quiney, “On the ‘unreasonable’ effectiveness of transport of intensity imaging and optical deconvolution”, *J. Opt. Soc. Am. A:*, vol. 34, no. 12, p. 2251, Nov. 2017. `doi:10.1364/josaa.34.002251`

[3] L. Brombal *et al.*, “Phase-contrast breast CT: the effect of propagation distance”, *Phys. Med. Biol.*, vol. 63, no. 24, p. 24NT03, Dec. 2018. `doi:10.1088/1361-6560/aaf2e1`

[4] V. Chenda *et al.*, “GeCo: The Elettra 2.0 Beamline Control System”, in *Proc. 19th Int. Conf. Accel. Large Exp. Phys. Control Syst. (ICALEPCS’23)*, Cape Town, South Africa, Oct. 2023, pp. 583–587. `doi:10.18429/JACoW-ICALEPCS2023-TUPDP034`

[5] TANGO Controls system, `https://www.tango-controls.org`

[6] A. Hafner *et al.*, “Modular Adaptive Processing Infrastructure (MAPI): a blueprint for interconnecting generic workflows with modern interfaces”, submitted for publication.

[7] R. Pugliese *et al.*, “Introducing the Elettra Scientific Data Lake: Concepts, Architecture and Select Applications”, presented at the GARR Conference 2025, Bari, Italy, May 2025, unpublished.

[8] Tia Portal, `https://www.siemens.com/it/it/prodotti/automazione/industrysoftware/automationsoftware/tiaportal.html`

[9] M. Scarcia, R. Borghes, “Executer – A TANGO based tool for experiment control”, presented at the 39th Tango Community meeting at INAF, Giulianova, Italy, May 2025, unpublished.

[10] The HDF5 Library, `https://www.hdfgroup.org/solutions/hdf5`

[11] Python module PyTango, `https://tango-controls.readthedocs.io/projects/pytango`

[12] Taurus framework, `https://taurus-scada.org`

[13] W. van Aarle *et al.*, “The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography”, *Ultramicrosc.*, vol. 157, pp. 35–47, Oct. 2015. `doi:10.1016/j.ultramic.2015.05.002`

[14] NiceGUI, `https://github.com/zauberzeug/nicegui`

[15] Celery - Distributed Task Queue, `https://docs.celeryq.dev`

[16] R. Pugliese, F. Billè, D. Favretto, N. Guidi, and M. Turcinovich, “Managing by Objectives a Research Infrastructure”, in *Proc. 14th Int. Conf. on Accelerator and Large Experimental Physics Control Systems (ICALEPCS’13)*, San Francisco, CA, USA, Oct. 2013, paper MOPPC083, pp. 262-265.

[17] Fiji image processing package, `https://imagej.net/software/fiji`

[18] Project Jupyter, `https://jupyter.org`

[19] Rocket.chat, `https://www.rocket.chat`

[20] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship”, *Sci. Data*, vol. 3, no. 1, p. 160018, Mar. 2016. `doi:10.1038/sdata.2016.18`