ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-WEPD010

SOLEIL II: ENHANCING DATA MANAGEMENT AND COMPUTING FOR TOMORROW'S SCIENCE

Y-M. Abiven†, P. Pereira*, G. Abeillé, A. Buteau, I. Chado, P. Madela,
L.E. Munoz, P. Pierrot, F. Langlois, A. Pain
Synchrotron SOLEIL, Saint Aubin, France

Abstract

Operational since 2008, SOLEIL offers users access to a wide array of experimental techniques through its 29 beamlines, covering a broad energy spectrum from THz to hard X-rays. In response to evolving scientific and societal needs, SOLEIL is undergoing a major upgrade through the SOLEIL II project. This transformative initiative includes the development of a new Diffraction Limited Storage Ring (DLSR) designed to dramatically increase brilliance, coherence, and flux. The upgrade also encompasses the modernization of beamlines to support state-of-the-art experimental techniques, along with a comprehensive digital transformation centered on data and user-oriented workflows.

This poster presents the current status of the digital transformation efforts within the SOLEIL II framework. It outlines the project's overall progress, with a particular focus on advancements in computing and data management. A central element of this transformation is the implementation of a unified Data Platform. Key developments include the deployment of a data catalog, upgrades to the IT infrastructure, user interface (UI) research, and the integration of robotics.

The platform leverages shared infrastructure and software patterns to support both beamline and accelerator teams. Additionally, ongoing evaluations of data streaming technologies—such as ASAPO, LIMA2, and DranSPose—aim to enhance real-time data acquisition and processing capabilities.

SOLEIL II PROJECT ORGANISATION

The SOLEIL facility [1] is currently undergoing a major upgrade of its accelerators and beamlines to achieve unprecedented levels of performance. This upgrade initiative is known as SOLEIL II [2].

SOLEIL II will enable experiments that are up to 10,000 times faster, 1,000 times more sensitive, and capable of achieving nanometer-scale resolution. These enhanced capabilities will make a decisive contribution to addressing key societal challenges in fields such as advanced materials research, energy and sustainable development, health and well-being, and the environment.

As part of the upgrade, the storage ring will be completely rebuilt and its equipment replaced, transforming it into a low-emittance ring capable of delivering an ultra-stable photon beam. Several beamlines will be relocated and upgraded to fully benefit from the enhanced performance of the new photon source.

The project is structured around four main programs:

- Accelerator construction
- Beamline realignment and adaptation
- IT system overhaul
- Infrastructure and logistics

In line with the strategy defined during the TDR (Technical Design Report) phases, the architecture and infrastructure of the Information Systems (IS) are also being modernized to support the facility's evolving needs.

SOLEIL II has now officially entered the construction phase. The shutdown of operations, to allow for the removal of existing equipment and installation of the new accelerator, is scheduled to begin in October 2028 and will last for approximately 18 months. This phase includes the dark period and the commissioning of the storage ring.

Once construction is complete, the next phase, titled *"Towards Full Performance,"* will enable further enhancements to the beamlines to fully exploit the advantages of a coherent, low-emittance electron beam.

STATUS OF IT AND DATA MANAGEMENT PROGRAMME

Introduction

In accordance with the targeted characteristics of the future electron beam and the scientific performance unlocked by Soleil II, the Soleil II project is in charge of delivering an IS capable of providing a state of the art user experience before, during and after the experimental project. This particularly means supporting our users in leveraging the increased capabilities offered to them through Soleil II's beamline instruments, as they will be at the end of the dark period and as the will evolve during Soleil II's lifetime. The IS required to achieve this goal thus needs to be state of the art, to cope with the continuous increase in data generated, and to seamlessly accompany the evolution of Soleil II, while insuring the highest degree of reliability, performance, ease of use and security.

The collaborative approach developed during the Conceptual Design phase has been maintained during the Technical Design Requirements phase, involving many representatives from SOLEIL's IT teams, beamlines and accelerators. The Technical Design Requirements phase was structured into six work packages, each led by a pair of leaders to foster a cross-disciplinary approach that includes accelerators, scientific, and computing teams. These work packages, as summarized in Fig 1, are further divided into 32 sub-tasks. They encompass various aspects, including

† yves-marie.abiven@synchrotron-soleil.fr
* patrick.pereira@synchrotron-soleil.fr

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-WEPD010

organizational improvements, control architecture, future operational tools, data acquisition integration, user experience enhancements, and the implementation of new data processing techniques, including artificial intelligence.

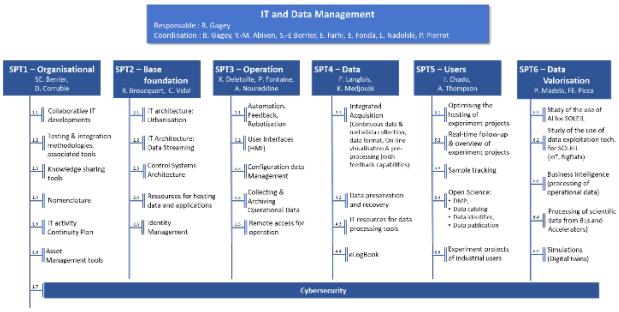

Figure 1: IT and Data management program organisation with the work package description.

Technical Design Report Main Findings

Bridging the CDR [3] phase and the full-fledged SOLEIL II project organization, the Technical Design Requirements (TDR) phase aimed at providing SOLEIL II with a roadmap of the work to be done during the 2 project phases: "Construction" and "Towards Full Performance". During this TDR phase, the IT and Data Management Programme reached essential milestones by putting in production some of the foundational elements required to build SOLEIL II's IT systems.

Among them, the adoption of a DevOps mindset and ecosystem has been identified as a cornerstone for all future software development, including SOLEIL II's specific deliverables. This approach will be offered to every software project originating from SOLEIL in order to streamline software production and ensure quality and security by design. Coupled with a new architectural framework based on PLUSS and on modular software components, this will enable faster and more reliable software delivery for developers with an heterogeneous knowledge of SOLEIL's IT infrastructure.

From an infrastructure standpoint, base services like storage, compute, remote access, operational data archiving of control systems, have been or are undergoing renewal so as to cope with the current and future requirements of the ever-growing generated data and expectances on data analysis of SOLEIL's users and operators. In the same spirit, protection against cyber threats, a priority objective identified as part of the TDR phase, has already been improved with identity federation, multiple factor authentication, logs and metrics analysis initiatives.

The TDR phase also delivered the seeds of services which the SOLEIL II project will accompany to full growth, such as the rollout of FAIR principles at SOLEIL (As A Service data analysis capability with VISA [4], Data Cataloguing with Scicat) or the enhancement of services for SOLEIL's users with the Sample Management Application (SMA).

Additionally, the TDR phase allowed SOLEIL to shape some of the requirements that the SOLEIL II project will need to fulfil, in particular the need for a more integrated control system, able to orchestrate resources (physical and computational) across SOLEIL's whole IS system, establishing a data platform providing end to end analysis abilities (from acquisition to interpretation).

The refinement of some topics outside of the "Construction" phase's critical path, such as the overhaul of SOLEIL's User Office portal, will however need to move forward and be integrated in due time as a deliverable of the Soleil II project.

Target Architecture

Building upon existing expertise and the outcome of the TDR, SOLEIL's IS has been oriented to incarnate a Data oriented Plateform, as illustrated in Fig. 2. The objective of this approach is to improve the automation of experiment, maximize the outcome of the data throughput while carrying out a FAIR (Findable, Accessible, Interoperable, Reusable) data policy by ensuring that SOLEIL can rely on a validated and integrated set of features and services covering all requirements to process its data, be it scientific or operational.

An integrated, cross-functional architecture is being implemented to unite all IT stakeholders, with a primary focus on creating a modular foundation that allows for agile, long-term evolution. This paradigm will be technically realized through service-oriented technologies, such as API management and event driven systems. Cloud-native and big data infrastructure will be implemented and will help further develop existing data-oriented services or create new ones. The envisioned digital transformation will also emphasize automation improvements to integrate smart systems, advanced processing, and the use of artificial intelligence for control, operational maintenance, and data processing.

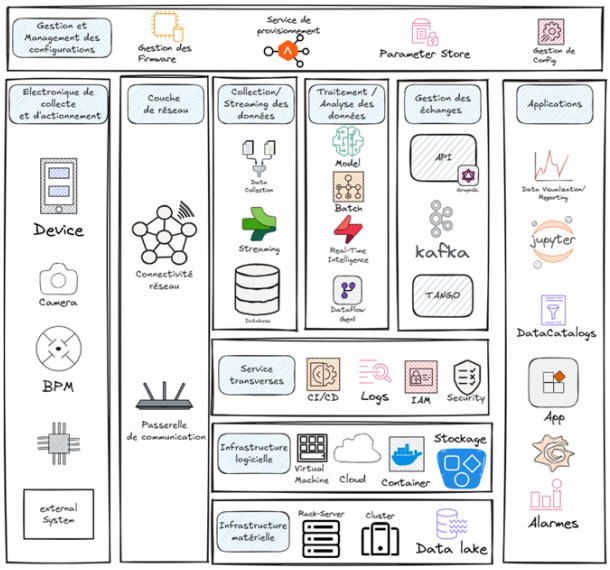

Figure 2: Target architecture for SOLEILII's IT system.

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-WEPD010

MAJOR ACHIEVEMENT

Cybersecurity

A strategy to enhance SOLEIL's cybersecurity has been established. The following sections describe the work that is either in progress or already completed.

Identity Management

To strengthen our cybersecurity, various technical and organizational measures have been gradually implemented:

- Cleaning up our ActiveDirectory, analyzing and gradually optimizing the onboarding/offboarding processes for our employees
- Implementation of a password compliance policy and tools to guarantee and monitor the quality of our passwords
- Gradual deployment of modern authentication methods on our business applications. In particular, the use of MFA (Multi-Factor Authentication) is now the norm for SOLEIL employees.

System Compliance and Monitoring

Protecting our systems also involves using up-to-date workstations and servers. We have therefore defined the required configurations in terms of operating systems, applications and security tools. Computers that cannot be updated are gradually being phased out, while all others must be integrated into automatic update and patch management systems.

Obsolete system and network protocols are now explicitly blocked. A compliance verification and vulnerability detection tool runs periodically, listing all anomalies and ensuring that the overall security level is continuously improved.

Cybersecurity Supervision

In addition to the widespread use of the Zabbix [5] monitoring tool, the projected implementation of an EDR and log centralization tool provides visibility into workstation activity, alerts in the event of anomalies, and quick incident analysis.

Cybersecurity Awareness and Training for Employees

In addition to regular communications to staff on cybersecurity topics, an awareness campaign including phishing test and training tutorials is underway with the help of an external partner.

A monthly report on Cybersecurtity projects is done to the board of directors

Collaborations

Soleil participates to the working group gathering CISOs of most European Photon and Neutron facilities, to discuss best practices with the I.T groups of others scientific facilities and publish recommendations for use by directors.

PLUSS Platform [6]

In order to enhance the way to share information across, Data has been considered in a global approach. As we have to maintain in operation the current information system and build new blocks for SOLEIL II, considering that the amount of data exchange will increase drastically, a project known as PLUSS (PLatform for Urbanization at Synchrotron SOLEIL, see details [6], illustrated in Fig. 3), that proposes a unified approach to interconnect either existing systems or future ones. Technically, we have chosen to setup an Apache Kafka broker for managing asynchronous communications and WSO2 API Manager to handle web APIs. Those two systems are now installed on premise and in operation for the first use cases: Sample Lifecyle management and SciCat as detailed hereafter. As we have gain experience and expertise setting up concrete use cases, we are now confident that this platform is ready to ramp-up for new uses cases.

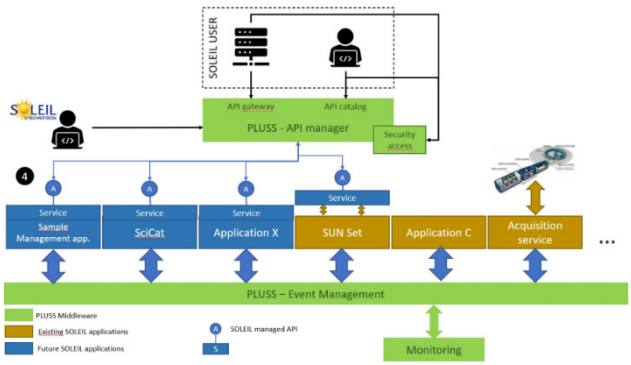

Figure 3: PLUSS architecture.

Sample Lifecycle Management

The TDR study has identified the need for improving the tracking of samples measured on beamlines. A new version of a brand new 'Sample Management Application' (SMA) has been integrated into SOLEIL's IS alongside with the PLUSS initiative which bring some useful architecture methodology and work force. This SMA has many interconnections with other services such as an identity management service, our Digital User Office (SUN Set), a future electronic logbook, the beamline acquisition system as shown in Fig. 4. This first version was deployed recently in a production environment, with its first interconnections to:

- Our Identity Access Management service (KeyCloak [7]) with the OpenID Connect protocol for managing authentication and authorisations,
- Our DUO with some dedicated Kafka producer and consumer to associate samples with the experiment proposal metadata,
- Our beamlines control systems, by exposing on a TANGO device that gather sample's information on a REST API of the SMA, through WSO2 APIM;

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-WEPD010

allowing to associate sample metadata with the experimental data.

The first setup has allowed to successfully create some first samples on one beamline. We are now ready to proceed validations on several beamlines. New functionalities are already identified and will be added in futures developments.

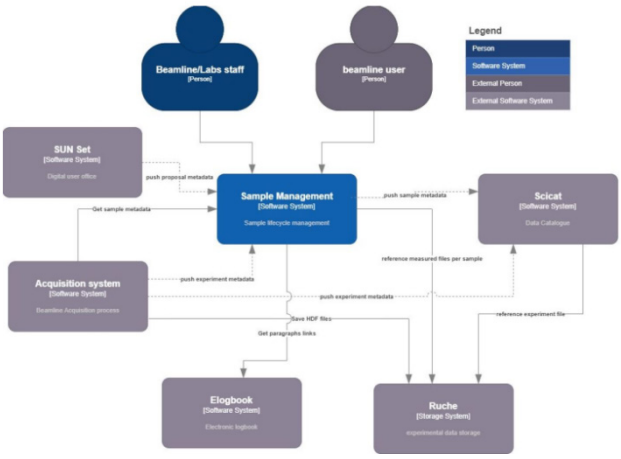

Figure 4: Sample tracking architecture.

Data Catalog (SciCat [8])

The project was launched two years ago in the scope of the digital transformation for SOLEIL II. All integrations have been carried out using PLUSS, TANGO, DUO, and authentication and authorization through IAM (Fig.5).

To integrate the SciCat [8] data catalog into our environment, we extended the scanning system to support dataset creation. The acquisition system, which already collected metadata from Tango for Nexus/HDF5 file generation, was enhanced to handle datasets and capture the necessary metadata. Proposal information from DUO was also enriched in Tango with the metadata required by SciCat.

The system, implemented entirely through extensions of existing Tango devices, is highly flexible, allowing any control system information to be persisted in both Nexus files and SciCat datasets. Proposal and dataset ingestion are integrated with our Kafka infrastructure of PLUSS, while access rights are directly managed through DUO permissions via IAM.

The solution was deployed on the SWING beamline with successful first tests, enabling dataset ingestion into SciCat with minimal changes. Ongoing work focuses on externalizing SciCat access, enabling file access via Globus [9], and implementing DOI creation through DataCite. Building on these results, the components can be instanciated more broadly across SOLEIL, while accommodating the diversity of data management practices. When it comes time to publish the data, SOLEIL is in the process of deploying the SciCat data catalog, aligning with its data policy that adheres to the open science principles defined in the European standard.

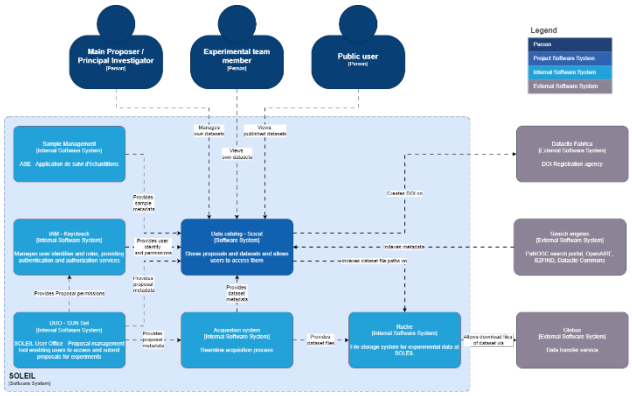

Figure 5: Diagram of the SciCat integration within the SOLEIL environment.

SOLEIL's DUO integration into PLUSS

As part of the urbanization of SOLEIL II's IS data flows, the DUO (Digital User Office) tool, also known as SUN set, plays a central role as it meets the needs of several web applications fed with proposals and samples (i.e.: IAM, SAM, pyISPyB, SciCat, VISA, Eloctronic Logbook, etc.). An unprecedented collaborative effort of different SOLEIL groups has been made to extend its capabilities, through the implementation of new interfaces based on standard protocols:

- One interface is based on Kafka for real-time data flow processing.
- Another is based on REST APIs for synchronous application exchanges.

Regarding data flows, a Kafka producer was designed, developed, tested, and interconnected with the SMA application, then integrated into the new PLUSS platform. This Kafka producer is embedded in SUN set in the form of an internal module. For greater resilience, it has been reinforced by a cache mechanism: messages are saved in a database and transmitted to the Kafka broker as soon as it becomes available again, including in the event of maintenance, unexpected interruption, or failover. The same work has been carried out for the REST interface, with particular attention paid to the generic nature of the APIs to maximize reusability, limit specific developments, reduce maintenance complexity, and ensure better scalability. The next step will be to integrate this interface into the WSO2 platform. Fully operational these two interfaces, Kafka and REST, now reinforce the interoperability of the IT system, facilitate maintenance and ensure better scalability. They are already used in the development of IAM, SAM and pyISPyB applications.

Logs Centralization

The Elasticsearch infrastructure [10] for log analysis has been stabilized and improved, initially focusing on TANGO logs by resolving issues like duplicates and dashboard instability. Data collection is now reliable, structured, and efficiently managed with a rotation policy. A comparison with Loki/Grafana confirmed Elastic as the preferred solution. The system was extended to include web cluster logs (Kafka, APIM, Scicat, Webcontrol), with

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-WEPD010

a modernized architecture using Elastic Agent and Ingest Pipeline for better efficiency and flexibility. A POC was launched to migrate TANGO logs to this setup, and a prototype was developed to track TANGO database changes for improved traceability. The use of Elasticsearch for logs centralization is also one of the projected architectures for Cybersecurity logs centralization and analysis.

Control System Enhancement

Significant enhancements have been made to the SOLEIL control system, addressing both obsolescence issues and laying the groundwork for SOLEIL II. In the software factory [11] domain, the C++ build system has been successfully migrated from Maven to Conan and CMake, providing a modern, cross-platform foundation for dependency and build management. This transition enables the use of up-to-date compilers and standards across all platforms.

Since late 2023, several automation applications with robotic arms have been implemented on beamlines: SWING uses a robot for liquid pipetting in BioSAXS [12]; NANOSCOPIUM completed commissioning of its X-ray detector positioning system for Bragg CDI and began expert-user experiments; and a measurement bench for supermodule [13] characterization was deployed for the magnetism and insertions group in 2024. Additionally, promising developments include monochromator-undulator synchronization for optimal performance, and a magnetic levitation-based translation–rotation demonstrator developed with an industrial partner under the LEAPS-INNOV project.

In order to improve automation several new tools and applications have also been developed, including PyImgProcessor, a Python Tango device for modular image processing now in production on multiple beamlines. It is use at Cristal beamline, for an AI-based application for automatic capillary centering; and at Swing beamline, it provides tools for sample centering and image file integration into the Tango ecosystem.

Additionally, the unification of the historical ScanServer and Flyscan [14] is nearing completion, with successful tests on two beamlines and operational interfacing with SPYC. The next step involves deployment on a new beamline, followed by the migration and cleanup of legacy configurations. In parallel, a strategic effort is underway to define the future of graphical user interfaces (GUIs) at SOLEIL, involving the evaluation and prototyping of new solutions tailored to various control system applications.

CONCLUSION

The TDR phase has structured and validated several key transformations that will be deployed at scale during the SOLEIL II construction phase. To support the ambitious overhaul of the information system, a scaled Agile approach is currently being organized, aiming to build SOLEIL II through a collaborative and iterative development process.

ACKNOWLEDGEMENTS

The authors extend their sincere gratitude to all the teams involved in the achievements described in this paper for their valuable contributions.

Special thanks are due to former IT Director Brigitte Gagey, who led the TDR phase and defined the main drivers of this digital transformation for SOLEIL.

REFERENCES

[1] Synchrotron SOLEIL facility, http://www.soleil.fr

[2] SOLEIL II project, https://www.synchrotron-soleil.fr/en/news/conceptual-design-report-soleil-upgrade

[3] SOLEIL CDR report, https://www.synchrotron-soleil.fr/fr/file/13803/download?token=OUzsp46P

[4] Panosc, Data Analysis, https://www.panosc.eu/services/data-analysis/

[5] Zabbix, https://www.zabbix.com/.

[6] G. Abeillé *et al.*, "Rolling Out a New Platform for Information System Architecture at SOLEIL," in *Proc. ICALEPCS'23*, Cape Town, South Africa, Oct. 2023, pp.~1301-1305. doi:10.18429/JACoW-ICALEPCS2023-THPDP007

[7] Keycloak, https://www.keycloak.org/.

[8] SciCat, https://scicatproject.github.io/.

[9] Globus, https://www.globus.org/.

[10] Elastisearch, https://www.elastic.co/elasticsearch

[11] P. Madela, G. Abeillé, Y.-M. Abiven, X. Elattaoui, J. Pham, and F. Potier, "Conan for building C++ Tango devices at SOLEIL", in *Proc. ICALEPCS'23*, Cape Town, South Africa, Oct. 2023, pp.~1227-1230. doi:10.18429/JACoW-ICALEPCS2023-THMBCM015

[12] Y. M. Abiven *et al.*, "SOLEIL'S process automation improvement using industrial robots", *Synchrotron Radiat. News*, vol. 34, no. 4, pp. 10–17, Jul. 2021. doi:10.1080/08940886.2021.1968268

[13] L.E. Munoz *et al.*, "Robotic process automation: on the continuity of applications development at SOLEIL", in *Proc. ICALEPCS'23*, Cape Town, South Africa, Oct. 2023, pp. 1275-1279. doi:10.18429/JACoW-ICALEPCS2023-THMBCM032

[14] N. Leclercq *et al.*, "Flyscan: A fast and multi-technique data acquisition platform for the SOLEIL beamlines", in *Proc. ICALEPCS'15*, Melbourne, Australia, Oct. 2015, paper WEPGF056, pp. 826-829. doi:10.18429/JACoW-ICALEPCS2015-WEPGF056