ROLLING OUT A NEW PLATFORM FOR INFORMATION SYSTEM ARCHITECTURE AT SOLEIL

G. Abeillé, B. Gagey, Y. M. Abiven, Synchrotron SOLEIL, Gif-sur-Yvette, France
P. Grojean, C. Rognon, F. Quillien, V. Szyndler, EMOXA*, Boulogne-Billancourt, France

Abstract

SOLEIL Information System is a 20-year legacy with multiple software and IT solutions following constantly evolving business requirements. Lots of non-uniform and siloed information systems have been experienced increasing the IT complexity. The future of SOLEIL (SOLEIL II) will be based on a new architecture embracing native support for continuous digital transformation and will enhance user experience. Redesigning an information system given synchrotron-based science challenges requires a homogeneous and flexible approach.

A new organizational setup is starting with the implementation of a transversal architectural committee. Its missions will be to set the foundation of architecture design principles and to foster all projects' teams to apply them. The committee will support the building of architectural specifications and will drive all architecture gate reviews.

Interoperability is a key pillar for SOLEIL II. Therefore, a synchronous and asynchronous inter-processes communications is being built as a platform to connect existing systems and future ones; it is based both on an event broker and an API manager. An implementation has been developed to interconnect our existing operational tools (CMMS, Computerized Maintenance Management System and our ITSM portal, Information Technology Service Management). Our current use case is a brand new application dedicated to samples' lifecycle interconnected with various existing business applications.

This paper will detail our holistic approach for addressing the future evolution of our information system, made mandatory given the new requirements from SOLEIL II.

PLUSS JOURNEY

SOLEIL has launched a Proof Of Concept (POC) since early 2021 to test new technologies helping the design of the future SOLEIL Information System (IS). The main driver was to anticipate new needs for SOLEIL II [1] where we foresee a global strong requirement to answer the increasing demands from the business. Consequently, Soleil's current information system may require a deep revamping to provide a fluid integration between all components.

To achieve it, SOLEIL seeked consulting support from external experts from a French company, EMOXA, specialized to foster organisations for transforming their IS. EMOXA has primarily analysed our current IS which has confirmed our assumptions: SOLEIL has been working mostly in a silo-ed organisation that will inevitably end up with the 'spaghetti stack syndrome', i.e. resulting of point to point,

---

* https://emoxa.fr

synchronous, ad'hoc communication links between all the interconnected systems. So a transformation becomes mandatory for SOLEIL II stakes.

Consequently, EMOXA and SOLEIL have jointly deployed a first simple use case to challenge this hypothesis and identify how to adapt and not revamp the whole IS. A small platform, named PLUSS (PLaform d'Urbanisation du Synchrotron SOLEIL, or SOLEIL's PLatform for Information System Design), has been set up with Apache Kafka [2] for managing asynchronous communications and WSO2 API Manager [3] for synchronous communications (cf Fig. 1). The result is an interconnection between existing applications by aggregating data coming for our DUO tool (Digital User Office, SUN Set [4]) and a beamline acquisition system (FlyScan [5]) with a limited amount of work. The outcome was to continue the process of setting up such a platform and to extend it to real use cases.

PLUSS is now in MVP (Minimum Viable Product [6]) stage (, still with the support from EMOXA company. Before going to production, we have done a gate review of the existing use cases and the expected ones. The spotted use cases will be addressed thereafter.
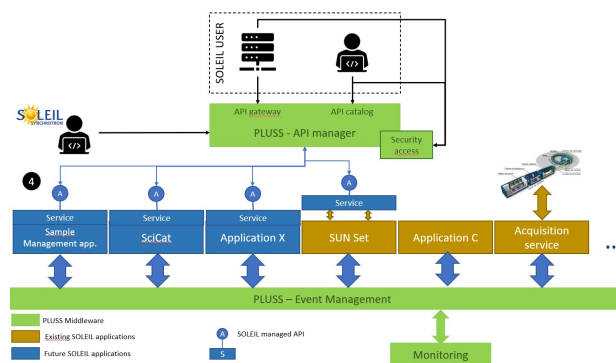

Figure 1: PLUSS global overview.

During our audit, we have discovered over already heterogeneous systems, numerous people or teams with diverse development methodologies or technical backgrounds. So PLUSS challenges are coming both from organisational and technical angles.

TOWARDS SOLEIL INFORMATION SYSTEM ARCHITECTURE'S GOVERNANCE

A proficient Information System (IS) architecture does not only rely on new technologies but also leans on a human organisation. The PLUSS initiative have proposed to set up a committee called CAI (Cellule d'Architecture Informatique

or Informatics Architecture Committee) with people from each computing team (infrastructure, databases, software...) to ensure coherence of the IS from the architectural and technical standpoints.

IS architecture activities can not only be based on experience nor competences; a common foundation embodied by a set of architectural principles is mandatory; they act as a guideline for every IS component's designer who apply them proactively. They are not set in stone, they are meant to be adapted with the evolving SOLEIL's stakes. CAI is in charge of defining, maintaining and rolling out those principles into daily developers activities. Here is an extract of currently proposed ones:

- *Theme*: Interoperability. *Statement*: Use Open API definition for REST API *Outcome*: APIs seamless sharing across the organisation.
- *Theme*: Performance. *Statement*: Plan for scalability of applications. *Outcome*: Guarantee availability of the service.
- *Theme*: Data management. *Statement*: Ensure data is shared properly. *Outcome*: Guarantee that accredited persons have seamless access to data.

Another key mission of the CAI is to ensure that every new piece of software, IT component or major refactoring introduced in the IS has either to follow the architectural practices as proposed by the CAI, or to explicitly seek derogation. Thus, we have proposed an architectural record template for helping project teams to make the right architectural decisions. The committee will guide the project teams whilst building the record and when the project is sufficiently mature the CAI will drive a formal architecture review. This review will validate the application of architectural principles or will accept temporary derogation that may be necessary given the constraints of the project (deadlines, not yet ready technical component...). The main purpose of this approach is to ensure smooth communications between technical experts, make decisions knowingly, and finally document and share knowledge. The ultimate justification being that the architecture of a system is not coded and so it must be described somewhere to avoid maintenance or evolution issues.

A UNIFIED PLATFORM FOR SOLEIL INFORMATION SYSTEM COMMUNICATIONS

To address our interoperability challenge, SOLEIL is deploying a central interconnection platform, called PLUSS, with two components to cover our major needs.

API Management Platform

Our first component is an API Manager to handle all deployed API and their consumers. It aims to control their life-cycle with policy enforcement. The selected product is WSO2 API Manager [3] which is open source. It supports many API standards (SOAP, REST, GraphQL..) and provides a portal that lists all available APIs and their doc-

umentation. It can be deployed as a single instance or in a clustered architecture to support high-availability and load-balancing.

Streaming Platform

Our second component is an event streaming bus, Apache Kafka [2]. It is based on a highly scalable broker for high throughput that supports high availability and load-balancing. A strong added value of Kafka for SOLEIL is the large ecosystem of off-the-self connectors [7], and its built-in functionalities for on-line data aggregations (with KSQLDB or Stream API [8]). Another added value is its message schema management allowing to control the data format and message schema evolution with a central repository called "schema registry". The SOLEIL broker will mainly transport metadata but not directly experimental data as Kafka is not designed for large messages. Kafka is a on-the-fly service and not a hard real-time message broker for low-latency communications that may be required in the future for data streaming on SOLEIL experimental stations.

Operating the PLUSS Platform

To operate WSO2 API Manager and Apache Kafka, a strong expertise is required but also a SRE (Site Reliability Engineering) [9] approach. So a training plan is in progress, where all persons that will operate the PLUSS platform learn how to deploy, maintain and operate WSO2 API Manager and the Kafka broker. As SOLEIL has separated teams for software and system administration, the PLUSS initiative embarks several teams in the production launch, but it is still a challenge as such a platform demands new skills which are not yet spread across all computing teams. Moreover, given the wide spectrum of necessary expertise, a trained and dedicated team is compulsory for at least the following techniques and technologies:

- Automation with Continuous Integration/Continuous Delivery (CI/CD) techniques and infrastructure for deploying the solution and perform upgrades.
- Containers (Docker, Podman...) architecture and their orchestration (Kubernetes, MicroK8S ...).
- Monitoring and alerting tools (Prometheus/Grafana, ELK...) to guarantee the quality of service.
- Cybersecurity to configure all the security of the stack.
- Web architecture to have a global knowledge about how this kind of solutions operates.
- Deep knowledge of Apache Kafka and WSO2 architectures and configuration.

Today, WSO2 API Manager is in operation with a single on-premise node on a virtual machine and started with Docker Compose [10] for a basic use-case detailed further on. We are also in the process of deploying a Kafka broker for operation (also with Docker Compose), starting with a single node then with three nodes. For the long term perspective, we may deploy the PLUSS platform upon Kubernetes as the opportunity to benefit on containers orchestration have been proposed by SOLEIL CI/CD colleagues. Nevertheless, the journey to a fully operational PLUSS platform with a

24/7 support is still at its early stage. Setting up such an initiative demonstrates tangibly what we need to address how to fulfills our stakes, especially the limits of our current organisation where developers and system administration people work in separated ways.

ROLL OUT OF USE CASES

Service Desk Management

Since 2017, we have developed an interconnection to automatically create issues from our service desk product (Jira) directly to our CMMS one (Maintimedia) [11]. For this purpose, a Jira dedicated plugin has been developed to call the SOAP Web service of Maintimedia.

But in 2022, due to the mandatory upgrade of Jira that was no more supporting the connection to a SOAP service, this interconnection was broken. WSO2 API Manager came to the rescue by providing a tool called "micro-integrator" [12] that is able, by configuration, to perform transformations. We have used it to create a REST endpoint that facades the SOAP end-point. This web service is called from an off-the-shelf Jira extension, Script Runner [13]. This solution is now in operation since 2022, and will be revamped soon as our CMMS now provide a new REST API.

Sample Life-Cycle

During our initial evaluation of potential use cases, a particular use case stood out: following the whole life-cycle of samples measured at SOLEIL. A dedicated ISpyB [14] application already exists for 3 beamlines of SOLEIL but it does not comply with other beamlines needs. In SOLEIL II TDR (Technical Design Report) [1] starting beginning of 2022, a dedicated work-package began to collect all needs for sample management.
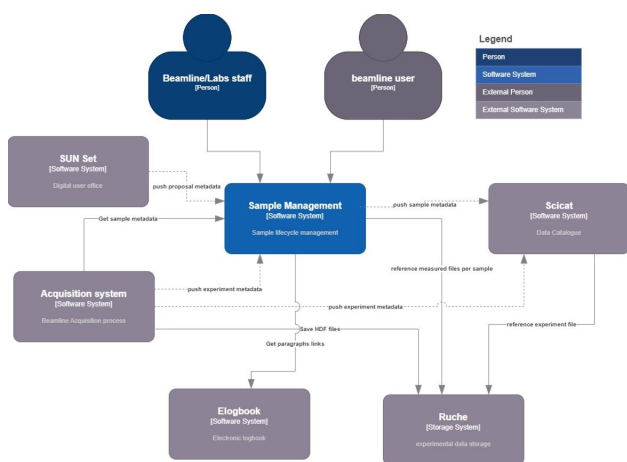

Figure 2: Sample management global architecture.

Rapidly, it appeared that the PLUSS initiative could bring some useful architecture methodology and manpower. PLUSS offered to guide to whole process by writing the requirements, defining the architecture and developing a

MVP [6] of a brand new 'Sample Management Application' (SMA). Actually, this SMA has many interconnections with other services such as an identity management service, our DUO (SUN Set [4]), a future electronic logbook, the beamline acquisition system as shown in Fig. 2.

EMOXA has taken in charge the writing of specifications and architectural documents, the development of a wireframe mock-up (cf Fig. 3), a full stack MVP application for a delimited number of functions. The wireframe mock-up was shared during demo sessions with many beamlines scientists and lab staff to get feedback on the specified functionalities. In parallel EMOXA has started the development of the application composed by: a front-end in React.JS/Refine, several back-end micro services in Java/SpringBoot and a PostGreSQL database [15]. The specifications and the wireframe are now in a stable version, after many iterations with the TDR work-package. The MVP application will be tested in a staged environment by the end of October 2023.
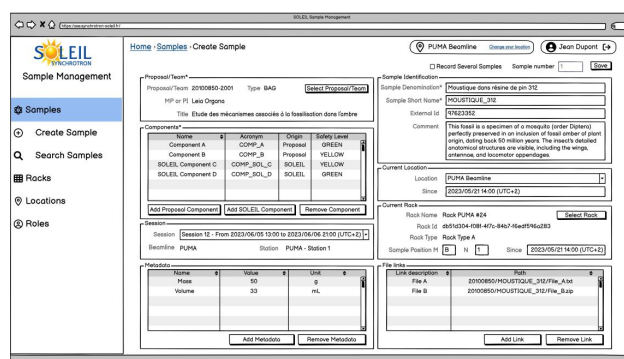

Figure 3: 'Create sample' screen wireframe.

In parallel, we are working of the MVP interconnections: identity server, SUN Set [4], and our acquisition systems.

Identity server connection   To manage authorisations and authentications, the SMA must be connected to a service that managed all beamlines users, SOLEIL staff and their memberships to SUN Set proposals, beamlines or laboratories. Unfortunately, SOLEIL currently operates at least 3 different directories:

- A local one in the SUN Set that contains beamlines users and scientists. It contains also important information about complex proposals called 'BAG proposals' [4] that will be necessary for the SMA.
- A LDAP server, which is a basic extract of the SUN Set for beamline users organised by proposal which does not contains BAG proposal data.
- And an Active Directory server for internal staff.

The unaligned directories challenge is already known at SOLEIL, a dedicated TDR work-package is currently analysing it but no federated solution has been decided yet. For the SMA, we are currently working for a temporary solution that could be based on KeyCloak [16].

SUN Set connection   The SMA needs to be informed of all validated proposals and their declared 'components'

for safety assessment; as components are the constituents of sample, actual measured samples will be registered in the SMA from their components. The earlier mentioned POC from 2021 was done with an off-the-shelf Kafka source connectors showing that the possibility to get basic proposal data directly from the SUN Set database, simply by configuring SQL queries [17]. But, the SUN Set developer team has stated that this solution is not possible as the database schema is way too complex and not adapted for the prerequisites of this Kafka connector. So, they have proposed to develop their own producer inside the SUN Set application. A first message specification has been realised using the Apache AVRO [18] message format. The development will be hopefully starting soon. It has be noted that another solution based on Change Data Capture [19] was quickly dropped as it does not support the currently SUN Set database version (Oracle 10g).

Acquisition systems connection   Tango [20] is used at SOLEIL for the orchestration of the acquisition processes, so the connection with the SMA will be available through Tango devices:

- One device that facades the SMA REST API to select from the acquisition system environment the current measured sample and its metatada.
- One device to send Kafka events to inform the SMA each time a new acquisition has been realised.

The first versions are already available and will be tested soon on the staging platform.

SMA Outcomes   The sample management application is a showcase for SOLEIL of how to design a new product from scratch based on modern technologies and methodologies. It has shown the importance of a "frond-end first" approach where we have focused on users feedback directly on a mock-up ahead of the development phase. It has also demonstrated the importance of architectural activities that must be driven before and all along the product development. A challenge that is still to be addressed is to organise the computing teams to efficiently build transversal interconnections and integration of all components in a seamless process. Today we still separate teams with a lot of legacy to daily maintain, missing some dedicated people to encompass this complex technical stack. With the endorsement of the architectural committee, we hope to smoothen this process in a near future.

Data Catalog

SciCat [21] is open source scientific data catalog that has been selected by SOLEIL to comply with the European Open Data policy. This product is deployed in a development environment at SOLEIL and will be deployed in a staging environment and validated on 2 beamlines in 2024. As first sight, SciCat must be at least interconnected with an identity server, our SUN Set, our acquisition system, our service to download experimental data files, and a DOI (Digital Object Identifier) generator service. So the PLUSS initiative

has proposed in 2022 to realise a first POC to ingest data coming from mock data sources of a beamline and SUN Set proposals through Apache Kafka and its on-the-fly data transformation capabilities with KSQDB (cf Fig. 4) and Kafka Streams [8]. This POC realised by EMOXA has demonstrated that Kafka is a good candidate for data ingestion with two complementary solutions: KSQLDB for basic data transformation (realised with 'à-la' SQL requests) and Kafka Streams for advanced use cases (Java code required).
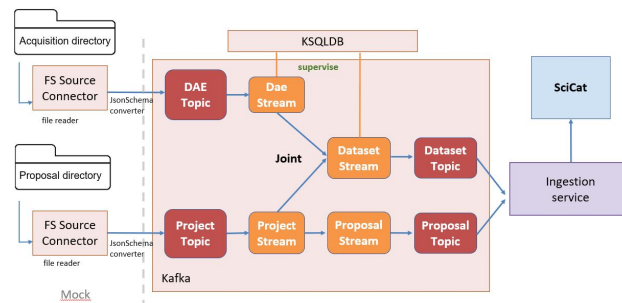

Figure 4: Scicat metadata ingestion POC with KSQLDB.

For the next steps, the SciCat team will have to mostly follow the ready-made path of the sample management application. It will reuse the same identity server, the same templates for interconnections specifications; and most importantly get the support for the architecture committee to guarantee its coherency regarding the whole information system.

Next Use Cases

The PLUSS initiative is a large investment for SOLEIL but we sense the first use cases for sample management followed by SciCat will draw a path for all the already identified use cases:

- *Operational data analysis*: We have created a GraphQL API on top of our existing Tango archiving system to provide access to modern Web tools. The API will be integrated in our API manager to handle its security and quotas.
- *On-the-fly proposal follow-up*: The first POC of PLUSS from 2021 has shown the added value of such a reporting called 'Vision 360'; a TDR work-package is now currently gathering specifications.
- *Electronic logbook interconnections*: Another TDR work-package is currently working on selecting our future electronic logbook. It must be interconnected with many systems like our future identity server, the sample management application, SciCat, the acquisition and control systems.
- *Experimental Data treatment*: SOLEIL has already on line data analysis on a few beamlines but with heterogeneous solutions. A study will be conducted to measure the added value of PLUSS for this use case.

OUTCOMES

With all major projects going through the architectural committee, a continuous improvement cycle has been engaged. We have created a framework where we can address the challenges of digital transformation from the central architectural perspective.

By driving some concrete implementations on real products that are part our current stakes, we gain experience and feedback that are now mandatory in a modern and complex computing technologies ecosystem.

For instance, the first sample management implementation has shown the limits of our current way of doing, and was able to clearly identified some missing pieces of our digital transformation journey. It has, for instance, highlighted to the management board the importance of an aligned service for 'Identity and Access Management' and thus to pull up its priority. On the other hand, concerning our current gap between developers and system administration people, an initiative is currently starting to identify first solutions to drive concrete joint activities such as a Kubernetes training.

As PLUSS is directly reporting to SOLEIL board of directors, it has also made visible some concrete impact of a digital transformation that leads to future wise decisions.

PERSPECTIVES

The main lesson learned is that we must always address the information system problematic with an ubiquitous sociotechnical point of view. Introducing a new technology without considering the stakes and implied organisation transformations is a risk that is not worth taken. The PLUSS initiative has emphasised the central role of architectural activities. It has also highlighted the importance of an aligned strategic organisation that must closely define, prioritise business stakes and sponsor the roll-out of a modern scientific organisation information system.

ACKNOWLEDGEMENTS

REFERENCES

[1] Y. M. Abiven *et al.*, "SOLEIL II: Towards A Major Transformation of the Facility", presented at the ICALEPCS23, Cape Town, South Africa, Oct. 2023, paper TUMBCMO21, this conference.

[2] Apache Kafka, `https://kafka.apache.org/`

[3] WSO2 API manager, `https://wso2.com/api-manager/`

[4] SUN Set, `https://www.synchrotron-soleil.fr/en/users/application-beamtime/how-prepare-proposal`

[5] 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, pp. 826–829. `doi:10.18429/JACoW-ICALEPCS2015-WEPGF056`

[6] Minimum Viable Product, `https://en.wikipedia.org/wiki/Minimum_viable_product`

[7] Kafka connect, `https://docs.confluent.io/platform/current/connect`

[8] KSQLDB, Kafka Streams, `https://ksqldb.io`, `https://kafka.apache.org/documentation/streams`

[9] Site Reliability Engineering, `https://sre.google/sre-book/table-of-contents`

[10] Docker compose, `https://docs.docker.com/compose`

[11] A. Buteau, G. Abeille, *et al.*, "Inception of a Learning Organization to Improve SOLEIL's Operation", in *Proc. ICALEPCS'19*, New York, NY, USA, Oct. 2019, p. 1001. `doi:10.18429/JACoW-ICALEPCS2019-WEMPL004`

[12] WSO2 micro-integrator, `https://wso2.com/micro-integrator`

[13] Jira Script Runner, `https://marketplace.atlassian.com/apps/6820/scriptrunner-for-jira`

[14] ISpyB, `https://exi.esrf.fr`

[15] Refine, Spring Boot, PostGreSQL, `https://refine.dev`, `https://spring.io/projects/spring-boot`, `https://www.postgresql.org`

[16] KeyCloak, `https://www.keycloak.org`

[17] Kafka JDBC source connector, `https://docs.confluent.io/kafka-connectors/jdbc/current/source-connector`

[18] Apache AVRO, `https://avro.apache.org`

[19] Debezium, `https://debezium.io`

[20] Tango, `https://tango-controls.org`

[21] SciCat, `https://scicatproject.github.io`