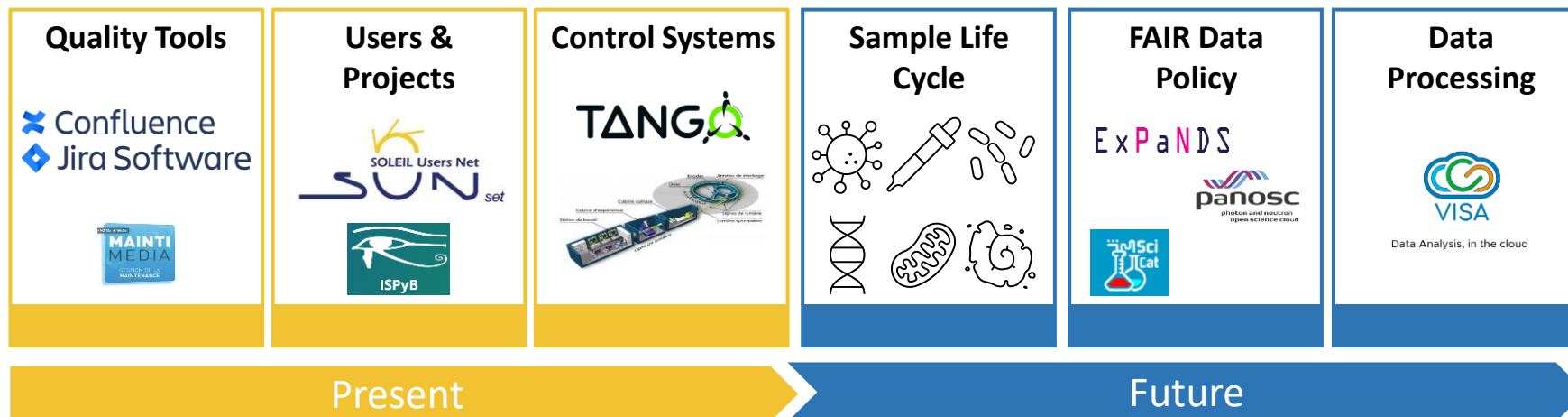


Scicat Soleil status update

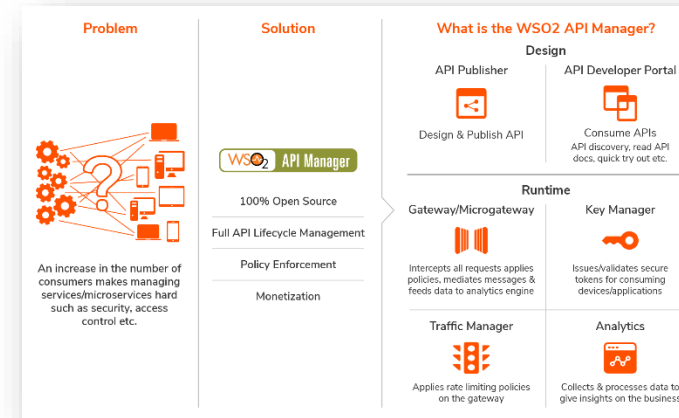
SciCatCon Meeting 2024 – P. Madela – F. Potier



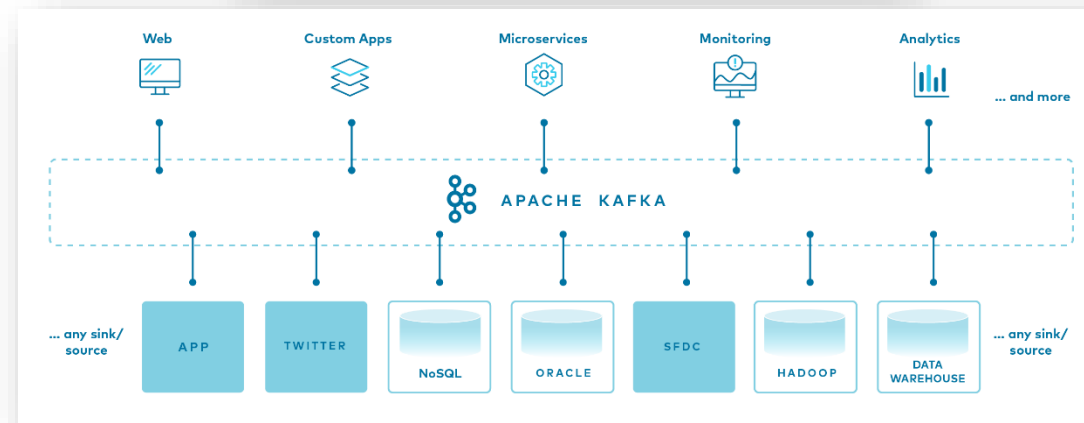
- **Historical Development:** Solutions developed based on evolving needs, leading to siloed systems.
- **Deployment Challenges:** SciCat, Sample Tracking Application, VISA, ...
- **Goal:** Implement transversal processes with common tools.
- **SciCat:** Part of this transformation, not just deploying application components.

- Objective
 - A platform to exchange data between heterogeneous applications, existing or future.

- WSO2 API Manager (Open Source)
 - Synchronous communications
 - Web API catalog (REST, SOAP, GraphQL, etc.)
 - Controls usage (security, quotas, etc.)
 - Will be exposed on the external web



- Apache Kafka (Open Source)
 - Asynchronous communications
 - Supports high-throughput messaging
 - Decouples producers and consumers
 - Many existing connectors
 - Data transformation
 - Secure

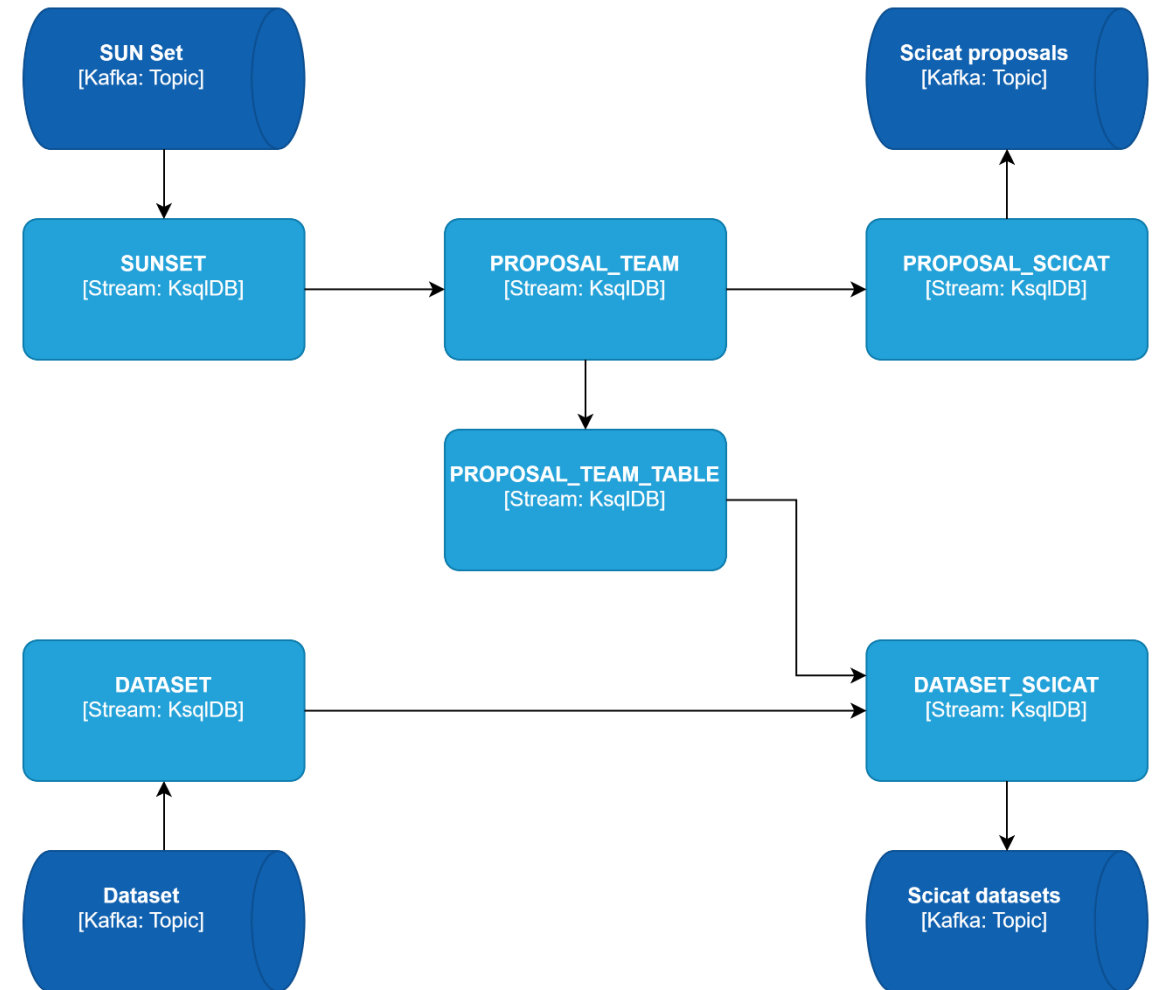


Contact: *Gwenaëlle Abeillé*

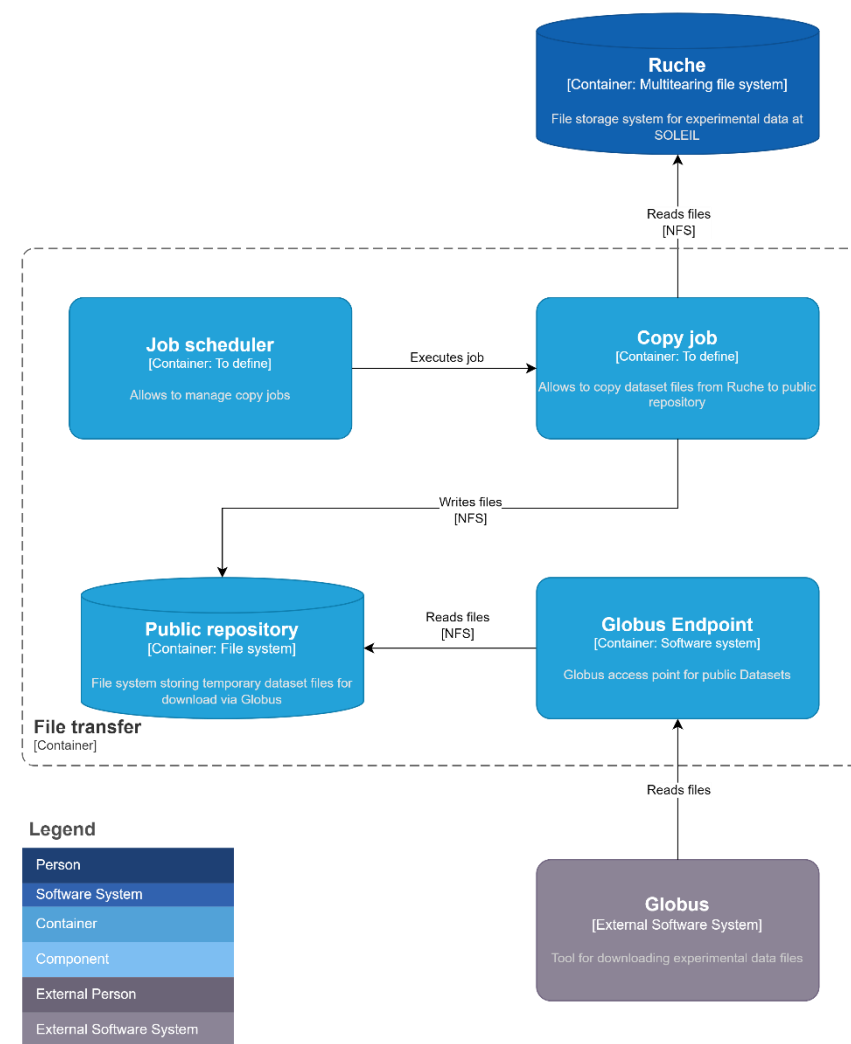
- **SUN Set API:**
 - **SUN Set:** Our Digital User Office
 - **REST API** to directly query information.
 - **Common producer** to publish DUO events on a single Kafka topic with multiple schemas.
 - **Topic used by various applications:** SciCat and sample tracking application.
- **SciCat consumer:**
 - Ingests proposals from DUO Kafka topic.
 - Handles both simple proposals and BAG (Block Allocation Group) proposals.
 - BAG proposals:
 - Multiple experiments over multiple beam time cycles.
 - Each BAG includes a coordinator and multiple teams, each led by a Principal Investigator (PI).
 - Target for BAG projects : create a proposal in SciCat for each team.

- IAM with Keycloak
 - Implementing Keycloak-based IAM for access management based on DUO information
- Evaluation of authorization retrieval solutions
 - **DUO's REST API**: Utilizing API endpoints
 - **Database Integration**: Direct access to a DUO's database view
- SciCat authorization strategy
 - **Access Control**:
 - Users will have read-only access to SciCat objects through "access groups".
 - Write access restricted to ingestion processes; no direct write access through "owner groups".
 - **Access groups** assigned to SciCat objects based on roles from DUO
 - **Proposals**: Identified by proposalId[-teamId]
 - **Beamlines**: Defined by beamline-branch

- **Device evaluation**
 - Ongoing work to create and modify control system devices on beamlines to gather necessary dataset information.
- **Kafka producer and consumer**
 - Exploring the use of KsqlDB for integrating data from DUO topics for dataset ingestion.
- **Goal**
 - Retrieve proposal-related information directly from DUO for dataset creation, bypassing the need to fetch from the control system.



- **Current Work:**
 - Implementing access to datasets
 - Awaiting job release
- **Two Strategies Under Consideration:**
 - **Mirroring system:**
 - Mirror the experimental data storage system for requested dataset files.
 - **Archive creation:**
 - Create an archive of requested dataset files
- **Access Method:**
 - Public endpoint access via Globus



- Progress
 - Significant time and effort spent on proposal ingestion and developing an authorization strategy.
 - Work is ongoing, but substantial progress has been made.
- Current Work:
 - Focus on deploying infrastructure components and dataset ingestion.
- Acknowledgments:
 - Thanks for the existing application, which has been a great help.
 - The main challenge is in collecting data from existing systems, not in deploying SciCat.





- Thank you to all the contributors of Scicat
- Main Soleil Scicat project contributors: J. Perez, M. Ounsy, Y.M. Abiven, G. Abeillé, S. Poirier, I. Chado, Y. Taheb, A. Buteau, A. Vandenbossch, C. Vidal