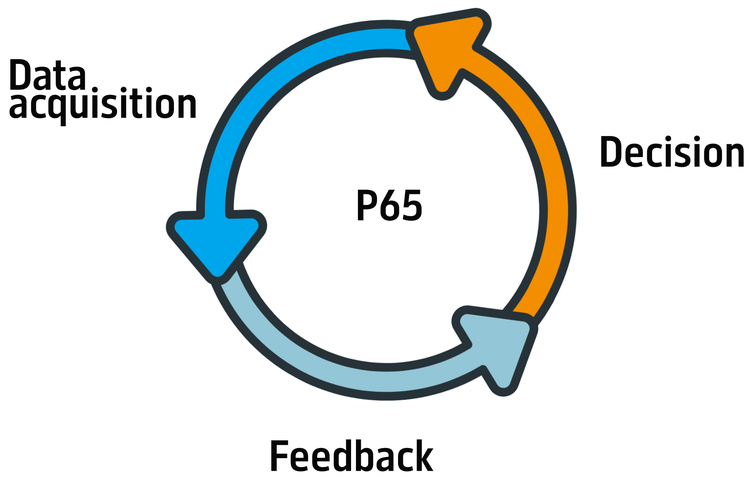
Near Real-Time Analysis Pipeline for X-ray Absorption Spectroscopy

Diana Rueda, Marc-Olivier Andrez, Tim Schoof, Mikhail Karnevskiy, Sergiu Levcenco,
Martin Gasthuber, Edmund Welter, Anton Barty
Deutsches Elektronen-Synchrotron DESY,
Hamburg, Germany

Near Real-Time Data Analysis in Experimental Science

The demand for **rapid feedback and automated decision-making** has led to the development of our near real-time analysis pipeline, initially implemented for X-ray Absorption Spectroscopy (XAS) and **potentially extendable to other experimental setups**.

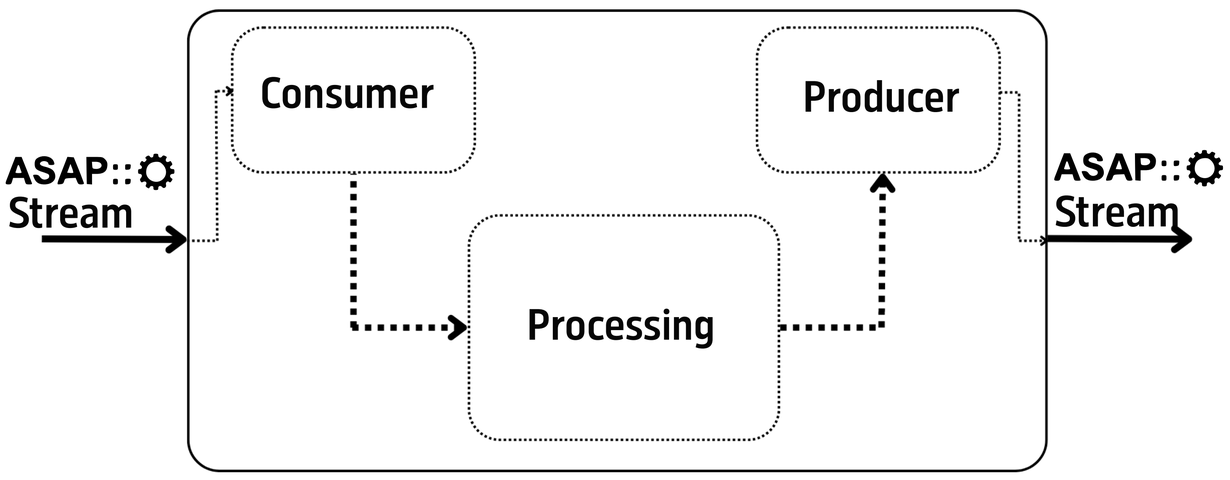


The ROCK-IT Project

As part of the broader ROCK-IT Project, this pipeline is a key component of a **fully automated experimental workflow.** The project is aligned with specific work packages focused on data processing and automation, aiming to standardize the processing steps across various experiments. This standardization is expected to **reduce the complexity and cost** associated with **development and maintenance**, making it easier to replicate and adapt the workflow for different experimental setups, and other Helmholtz research institutions.

Workers

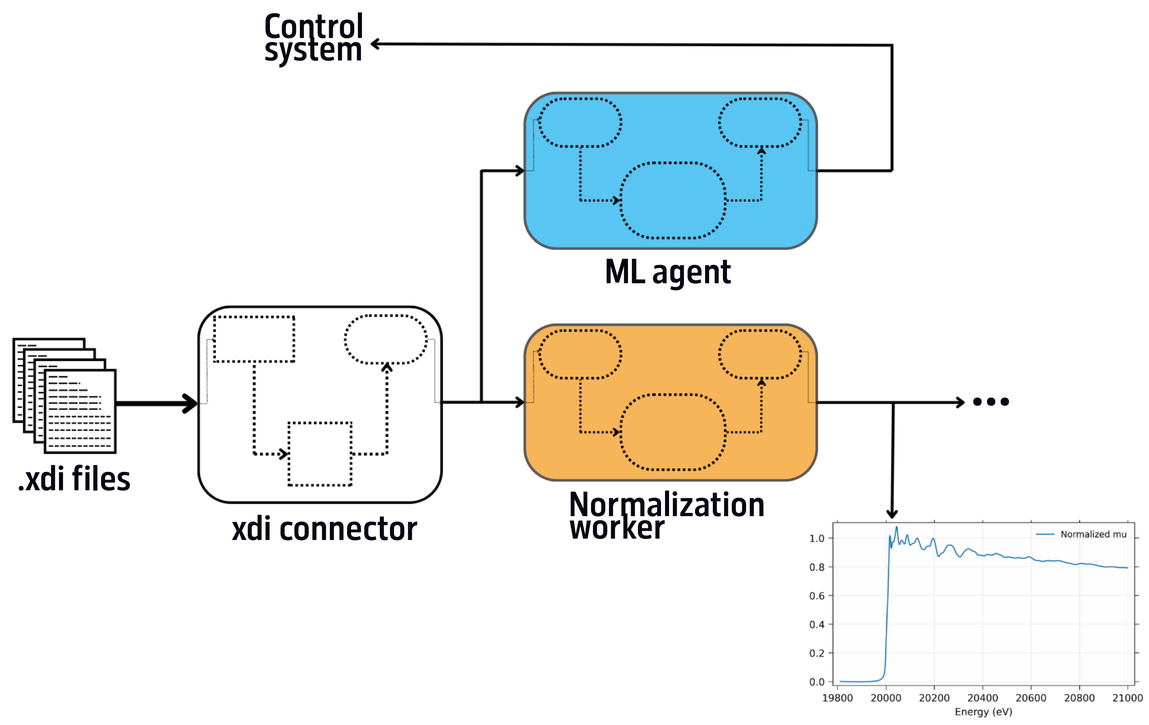
Central to this pipeline is the concept of **workers - modular processing units that handle specific tasks**. Each worker has inputs, outputs, and configurable parameters, and can represent either a **conventional analysis step or a pre-trained machine learning agent**. The pipeline API ensures seamless data conversion processes, such as **serialization and deserialization**, allowing scientists to wrap their analysis functions in Python for ease of integration.



Conceptual diagram of a worker in our pipeline

The pipeline is built on top of **ASAP::O,** a **robust distributed streaming platform** designed for high-throughput data environments. ASAP::O's unique architecture supports **scalable and efficient data processing**, which is particularly beneficial for experiments that generate large volumes of data at high rates, such as serial crystallography.

Processing Pipeline



Conceptual diagram of the analysis pipeline

To aid in the analyses, we have chosen to use the existing **Python-based Larch library**. The workers could include Normalization, Linear Combination Analysis, or custom Machine Learning agents that perform denoising or provide feedback to the control system.

ASAP::

Future Developments

To further enhance the pipeline, several advancements are planned:

1. **Live Visualization**: Integrating live data visualization tools to provide immediate feedback during experiments.
2. **Dynamic Parameter Updates**: Real-time adjustment of control system parameters (*e.g.*, energy offset (E0) calibration)
3. **Worker Development Library**: Specialized library to simplify the creation and integration of new workers, further streamlining the process for scientists and engineers.
4. **Pipeline Manager:** For orchestrating the starting, stoping, monitoring, and restarting of workers, ensuring smooth operation even in multi-node environments.

ROCK-IT Project https://www.rock-it-project.de/
P65 Beamline https://photon-science.desy.de/facilities/petra_iii/beamlines/p65_applied_xafs/
ASAP::O https://asapo.pages.desy.de/asapo/
Larch library https://github.com/xraypy/xraylarch
Questions/comments diana.rueda@desy.de

ROCK-IT

HELMHOLTZ