Elettra Sincrotrone Trieste

DATA LIFECYCLE IN LARGE EXPERIMENTAL PHYSICS FACILITIES

the approach of the **Elettra** synchrotron and the **Fermi** free electron laser

Synchrotron and free electron laser facilities like Elettra and FERMI have **special requirements** regarding their data flow. In order to face problems related to data deluge, Elettra implements a complete **data lifecycle**. This covers all the experimental stages; from proposal submission to beamline access and data analysis. In order to do so, key elements like **proposal management, acquisition software, and data analysis** need to be seamlessly **interconnected**. This brings to the introduction to the concept of data as «***a set of connected information nodes from heterogeneous sources***» in contrast to the traditional file-centric view. Future work will extend and build on this idea while the planned upgrade to **Elettra 2.0** may be the ground for implementing the potential advances.

Proposal
submission, evaluation, accounts

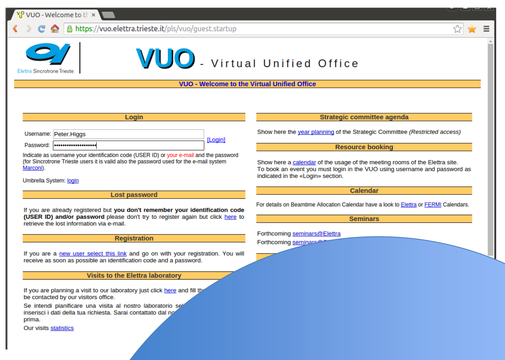
Experiment
control, DAQ, acquisition

Data
reduction, analysis, cataloguing

Access
data, operations, SW & HW

An experiment in a synchrotron facility requires a very specific procedure. While the core remains the actual experiment, the complete process consists of multiple subprocesses like proposal submission and evaluation, access request, data access and processing, remote operations and general accounting. In **1997** Elettra has been among the first institutes of its kind, to develop a complete system named Virtual Unified Office (**VUO**) which serves as the backbone for all such processes. The VUO is involved in:

- Proposal evaluation
- Publications record
- Access request
- User accounts
- Beamtime scheduling
- Resource access (WiFi, HPC)
- Remote operations
- Single sign-on & Umbrella
- Dataset storage organization
- ERP integration



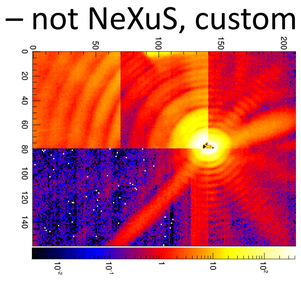
The actual **acquisition** of data is a process of paramount importance. The beamline end-stations of synchrotrons and free electron lasers are complex instruments of interconnected components like motors, pumps, and detectors. The data acquisition software is an advanced system capable of commanding all of these components while at the same time provides the user with a graphical interface. The Scientific Computing team of Elettra has designed a modern, flexible and extensible endstation control based on technologies compatible with the **TANGO** distributed control system. In order to avoid multiple developments, the system has a common core that aims at easy adaption to each beamline. Special attention has been paid for allowing easy integration with data analysis pipelines. See poster: "*A flexible system for end-user data visualisation, analysis prototyping and experiment logbook*" Borghes et al.

- TANGO 8
- Python & C++
- HDF5
- QT
- DonkiTools (in-house)
- High performance (due to FEL frequency)
- Modular and Extensible
- Electronic logbook

In most experiments, the data produced require various stages of processing prior reaching a stage where they can be interpreted. In their initial form they are often referred to as RAW data but this is an oversimplification. During an experiment there are multiple detectors that acquire data (e.g. images, spectra, beam characteristics, machine parameters etc). As a unique set they get assembled and processed. This processing may include multiple steps and later be separated in different operation pipelines. It soon becomes clear that the **workflow is neither standard nor static** thus it requires dynamic approaches that often make optimized solutions really difficult to be achieved.

- Python, C, Fortran, CUDA and OpenCL in Linux - not much MPI
- Matlab and IDL in Windows - even if we try to avoid them

- HDF5 formats for CT, CDI, XRF and STXM – not NeXuS, custom to our needs



- Analysis in Cluster and Desktop
- GPGPUs (but require lots of memory)

- Good Network and Storage/Filesystem

types of access:
1. **supervised**, where a user access them on will
2. **unsupervised**, when it is part of an automated process like a data analysis workflow.

A special class of access requirements, that of **remote** access.

Any form of off-site data access or operation is considered remote. It usually takes place after a beamtime experiment. The simplest form regards **data download**. This is already included as a service in the VUO.

- More complex forms of remote access may be necessary.
- Beyond remote desktop (NX)
- Complete beamline automation and remotisation (Xpress)
- Data analysis Software & Hardware as a Service (Pore3D)

Fulvio Billè, Roberto Borghes, Francesco Brun, Valentina Chenda, Alessio Curri, Venicio Duic, Daniele Favretto, George Kourousias, Marco Lonza, Milan Prica, Roberto Pugliese, Martin Scarcia, Michele Turcinovich
Scientific Computing team, IT Group, Elettra Sincrotrone Trieste, Basovizza, Trieste, Italy
sci.comp@elettra.eu