Transition from ISPyB to ICAT at ESRF (II)

Alex DE MARIA
On behalf of
Software Group and Structural Biology group
ELETTRA 20/11/2024

Context

- March 2023 at Soleil
  - **“A proposal for a Next Generation of ISPyB Software”**
    - How a generic metadata catalog could replace ISPyB
    - Requested further evaluation and testing of a solution based on ICAT

- November 2023 at ALBA
  - “**MX on ICAT**”
    - Evaluation was satisfactory
    - Further development and deployment in real conditions

- May 2024 at MAXIV
  - “**Transition from ISPyB to ICAT at ESRF**“
    - Developments on the UI and Backend
      - Shipment and sample tracking
      - Multi-sweep
    - Integration with MXCuBE
    - Plan for deployment in a few beamlines

Missing bits and future plans (MAXIV meeting)

- Data catalog/LIMS
  - Add Phasing and SAD (ongoing)
  - Add missing metadata
  - Test with friendly users

- MXCuBE
  - ICAT sample synchronization (ongoing)
  - Do some abstraction that allows to send data to both ISPyB and ICAT (ongoing)

- Plans
  - Roll out with full functionality on the beamlines by September
  - Use both ISPyB/ICAT in parallel for the time being
  - Help other partners willing to adopt this software solution

Highlights

Software renamed

DRAC: Data Repository for Advancing open sCience

Logo thanks to **Marie SPITONI, Ludovic BROCHE**

Implements the FAIR principles

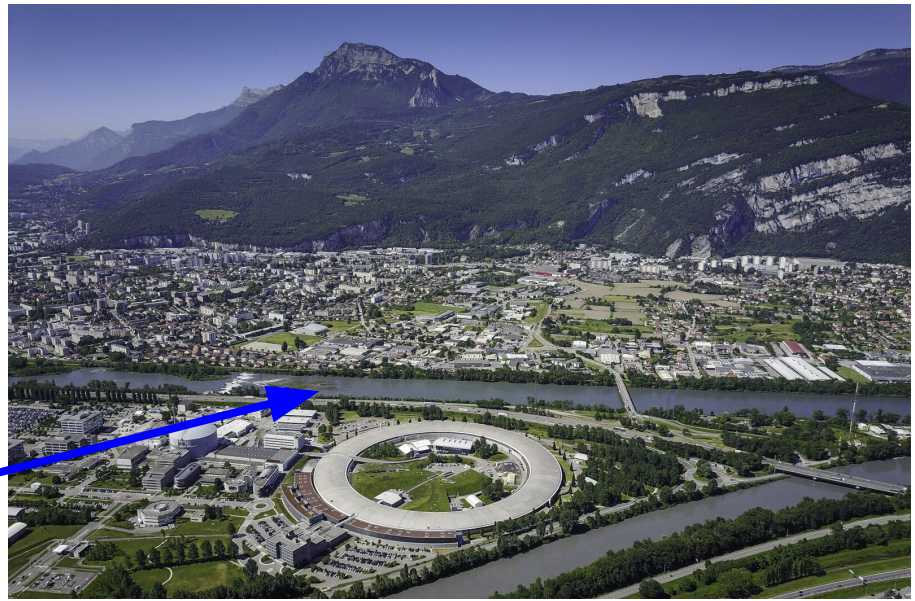
- Findable
- Persistent identifiers
- Data is preserved (interface to tape)
- etc…

Composed by:

- Catalog (ICAT)
- Sample Tracking
- E-Logbook
- Reprocessing
- Viewers. E.g: H5Viewer
- Technique Specific Viewers
  - Mx
  - BioSAXS
  - CryoEM/ET
  - Tomography
  - Etc..

Software renamed

DRAC: Data Repository for Advancing open sCience



Credits:

Marie SPITONI
Ludovic BROCHE

Data Portal v2 launched

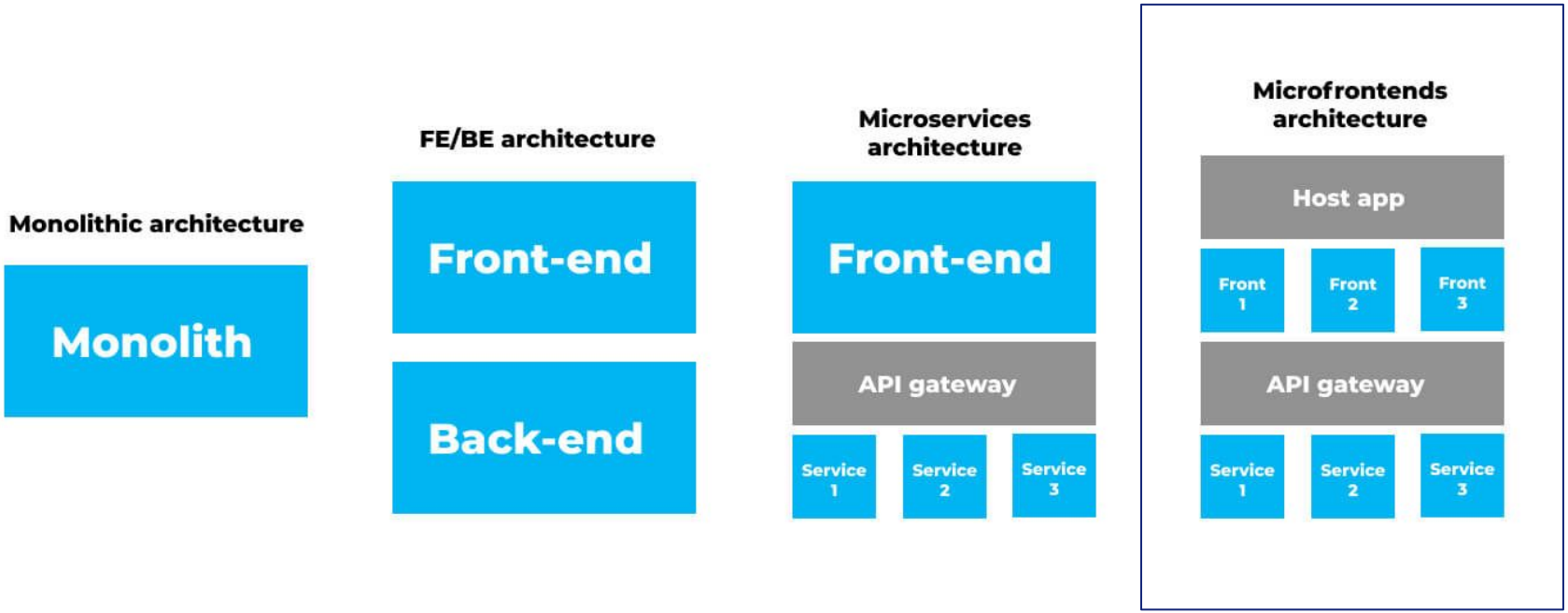
New version:

https://data.esrf.fr

- Upgraded version of the data portal
  - Launched October 2024

- All-in-one solution
  - Single entry point for users (MX and non-MX)
    - Data visualization
    - FAIR (public + private data)
      - Search
      - DOI
      - Logbook
    - Sample Tracking
    - Interface with TAPE (restoration + Globus)
    - ….

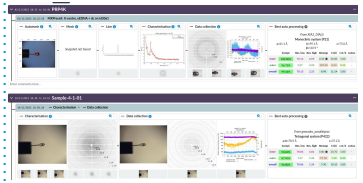
Microfrontend Architecture



Microfrontend Architecture

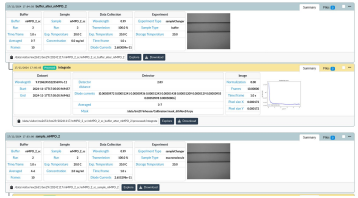
Frontend

https://data.esrf.fr



MX

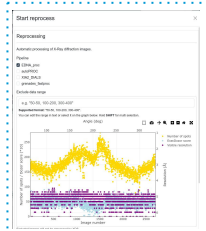
CRYO-EM/ET



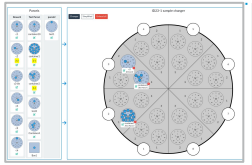
BIOSAXS

TOMOGRAPHY

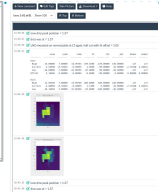
CUSTOM VIEWS/TECHNIQUE



REPROCESSING

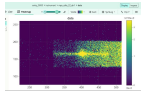


SAMPLE TRACKING



LOGBOOK

TOOLS



H5Viewer

VIEWERS

ICAT + API (https://icatplus.esrf.fr/api-docs)

Backend

SAMPLE TRACKING

ICAT CATALOG

H5GROVE

LOGBOOK

...

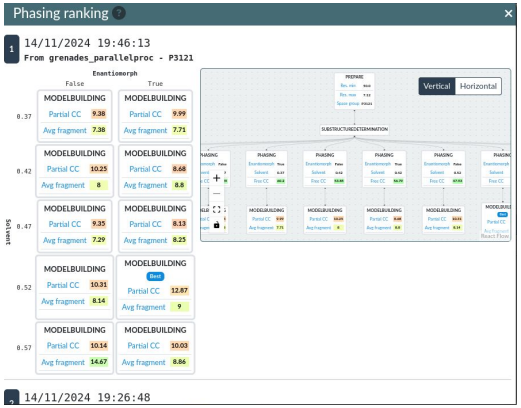
PANOSC API

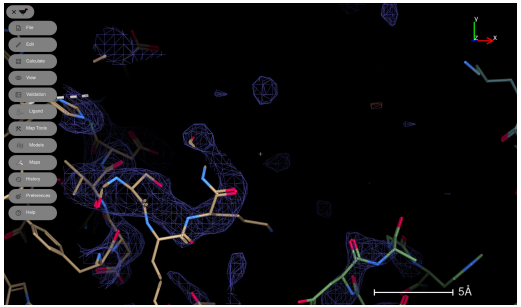
EWOKS

OAI-PMH

SAD/Phasing

- Deployed on all beamlines
  - Triggered when anomalous signal detected
    - Runs SHELXC+D+E
- Use case for extending features with 0-backend development
- Moorhen as visualization tool
  - Almost full coot functionality
  - Use of MTZ instead of maps

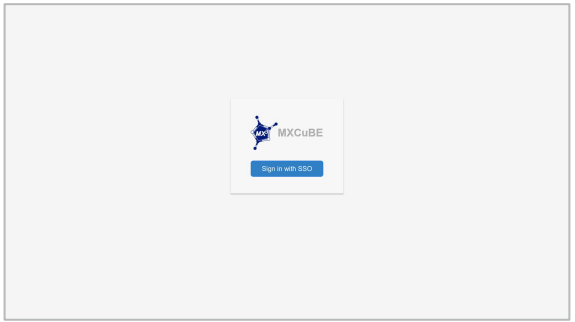




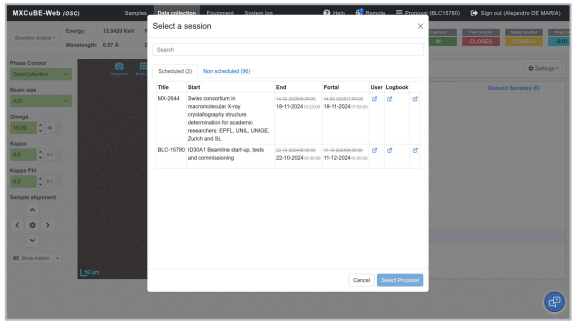
Phasing

SSO, user authentication and sessions selected from DRAC

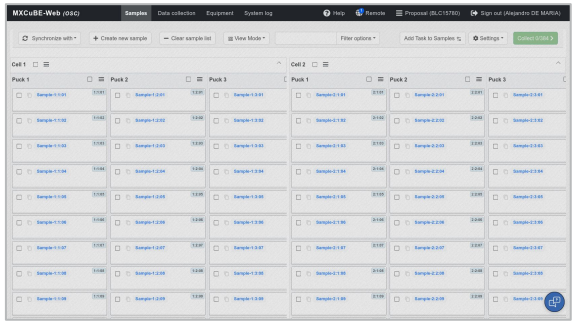
- Deployed and tested on **ID30A-1** since **September 2024**



User authentication with **SSO**



Session selection from **DRAC**



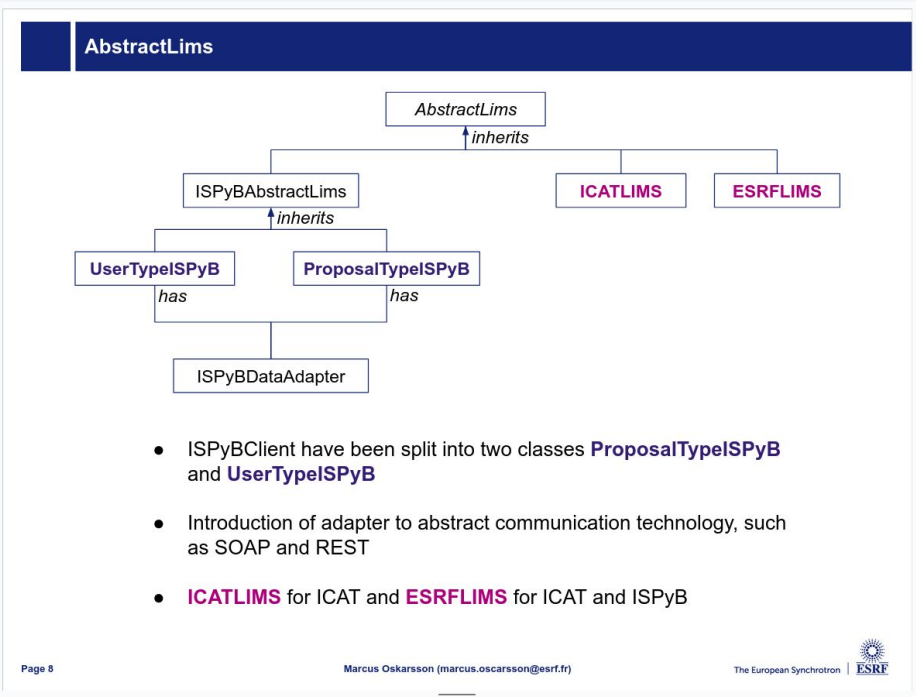
Sample tracking compatible with both **ISPyB and DRAC**

- MXCuBE is backwards compatible
  - Proposal authentication
  - Session selection via ISPyB

- Results are pushed to I**SPyB and DRAC**

Lowlights

MXCuBE integration with DRAC



Credits: **Marcus Oscarsoon**

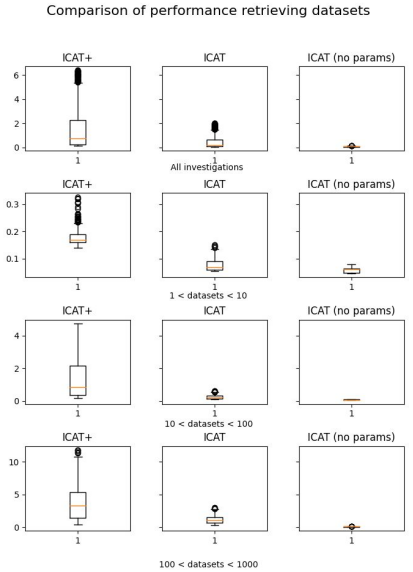
- Extract interface
- Implement non-abstract classes
  - UserTypeISPyB
  - ProposalTypeISPyB
  - ICAT
  - ESRF
- Better (hopefully) management of:
  - Session
  - Proposals
  - Users

Performance

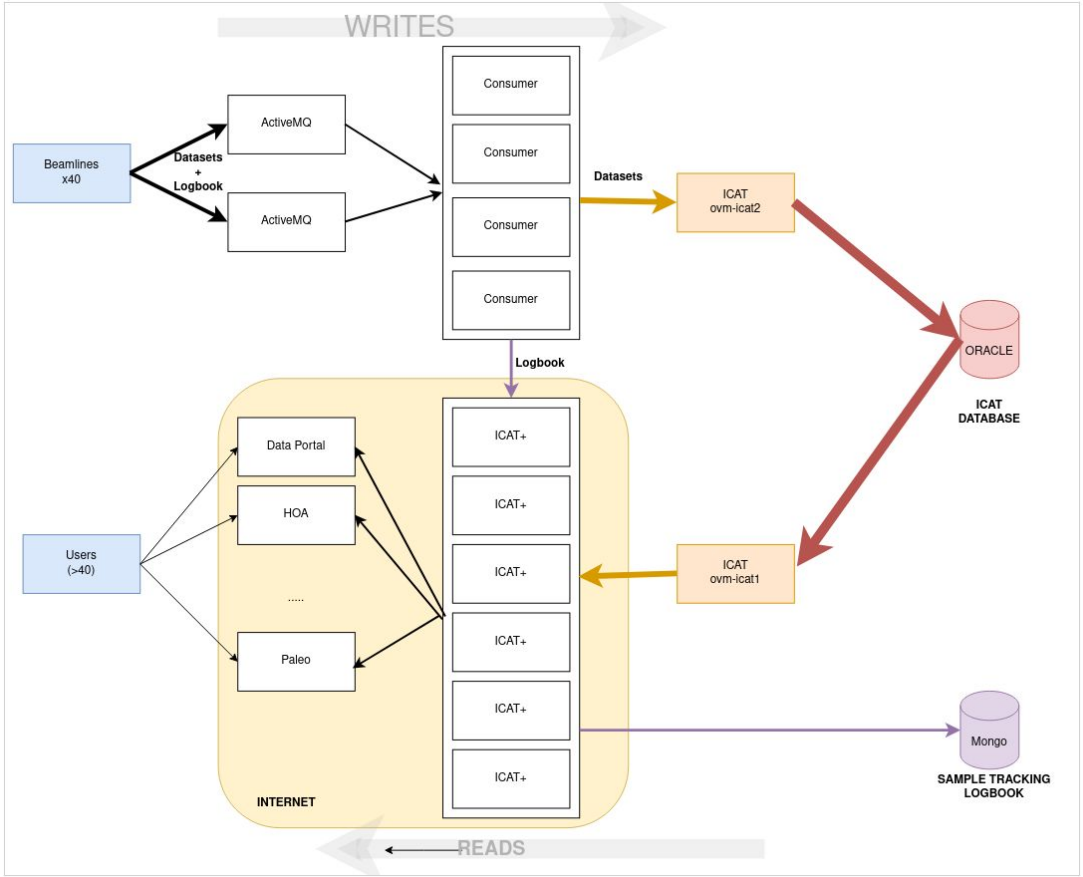
- Lot of time invested in getting better performance
  - **Queries**
  - User interface
    - Smart Lazy loading
  - MXCuBE -> DRAC communication
    - Removed strong dependency
    - Speed up data acquisition
  - **Infrastructure**
  - ….

Performance

- Queries
  - Huge impact in speed for the whole synchrotron
  - Public/private data makes user's role (user, staff, manager) to have a considerable impact



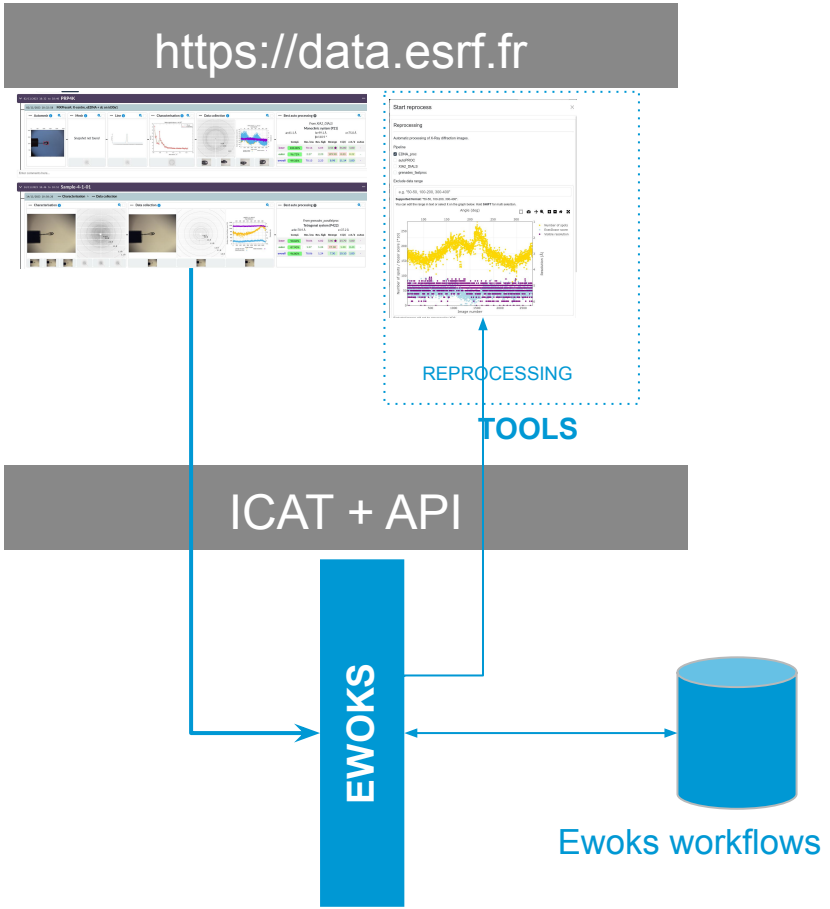
Performance (III)



Ewoks

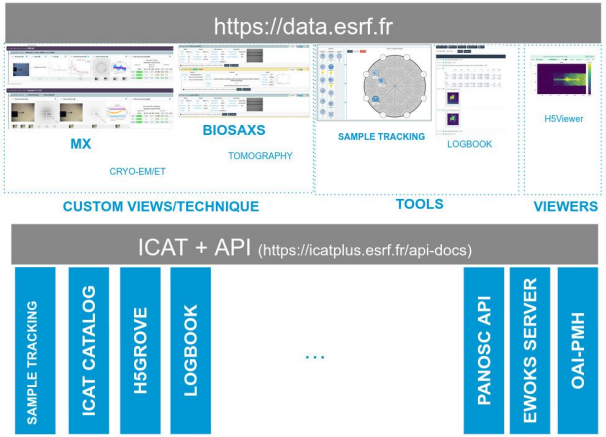
- Automatic rendering based on workflow definition
- Filtering workflows by:
  - Technique
  - Dataset type
  - Beamline

- Still work in progress
  - Example MX
  - Example BioSAXS



Packaging

- DRAC converted into docker compose stack
  - https://gitlab.esrf.fr/icat/drac
- Configuration done with .env
- Used for integration tests
- Help adoption of the software



Run your own instance easily:

```
Docker compose –profile database sample_tracking mx biosaxs h5viewer up
```

Future plans

- Update the MXCuBE configuration to fully enable ICAT on all beamlines by the January start-up
    - User authentication with SSO
- Bug fixing and user support will occupy much of our time in the coming months

- Improving dewar tracking
    - Labels/barcode scanning
    - Simplified workflow

- Making restored open data work seamlessly with Globus

- Linking samples to AlphaFold models

- Linking DOIs to PDB entries

Future plans

- Reprocessing
  - Single datasets
    - Prepopulate parameters based on
      - initial parameters used by the pipeline OR
      - final parameters
        - X beam/ybeam
        - Detector distance
        - Mosaicity
  - Multiple datasets
    - Merge datasets across samples and maybe sessions
      - Simple merge
      - HCA
      - GA
- Continue providing support or assistance in any matter related to ISPyB, py-ISPyB, or DRAC

Conclusions

- Development is progressing smoothly, and the feedback has been positive
- So far, the developments have only been made available to a small group of users. DRAC is expected to be publicized for all MX users soon, reaching an important milestone.
- In terms of functionality, DRAC has surpassed ISPyB for MX, but there is still much work to be done. These developments should be understood in the context of a project start rather than a final product.
- New lines of work are opening up, such as extending information related to the samples, reprocessing, or the automated use of the e-logbook.

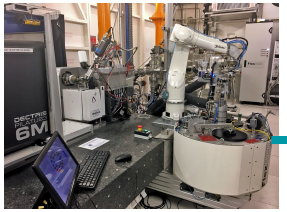
Acknowledges

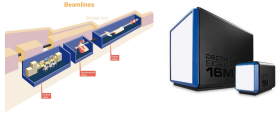
- DRAC
  - Mael Gaonach
  - Marjolaine Bodin
- Workflows and automation
  - Olof Svensson
- Scientists
  - Didier Nurizzo
  - Estelle Mossou
  - Matthew Bowler
  - Romain Talon
- MXCuBE
  - Marcus Oscarsson
  - Antonia Beteva
- Steering
  - Max Nanao
  - Andy Gotz
  - Vicente Rey

Backup slides

Architecture and services

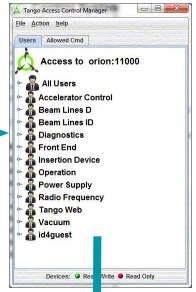
INSTRUMENTS





Beamline trigges:
SPEC, detectors,
Sample
environment

TANGO

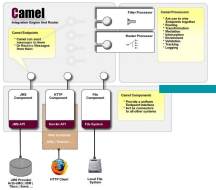


HDF

Nexus / HDF5

ACQUISITION

Apache ACTIVEMQ



TAPE SYSTEM

doi

INGESTION

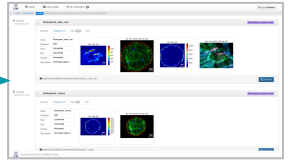
CATALOGUE

ICAT

ORACLE

DATABASE

DATA PORTAL



https://data.esrf.fr

CATALOGUE

SAD/Phasing visualization with Moorhen

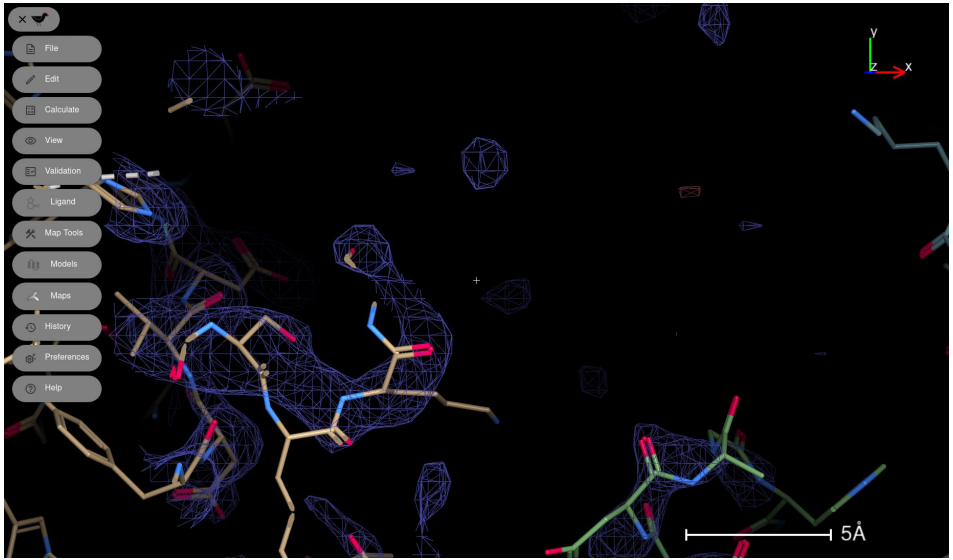
Moorhen is a web browser molecular graphics program based on the Coot desktop program.

CCP4 and Coot into Webassembly + REACT programs, Coot and their dependencies to **Web Assembly** and then combining with a **React user interface.**

Repository:
https://github.com/moorhen-coot/Moorhen

Two major benefits:

- Almost full coot functionality
- Use of MTZ instead of maps



Moorhen