ALBA SYNCHROTRON

Towards FAIR data management principles

Oriol Vallcorba *on behalf of*

Emilio Centeno, Fernan Saiz, Fulvio Becheri, Gemma Rosas, Marc Armenter, Nicolas Soler, Rodrigo Cabezas, Zbigniew Reszela

NOBUGS 2024, 23-27 September, ESRF and ILL, Grenoble

Outline

- FAIR at ALBA
- Challenges
- Design and implementation
- Remarks & future considerations

Outline

- **FAIR at ALBA**
- Challenges
- Design and implementation
- Remarks & future considerations

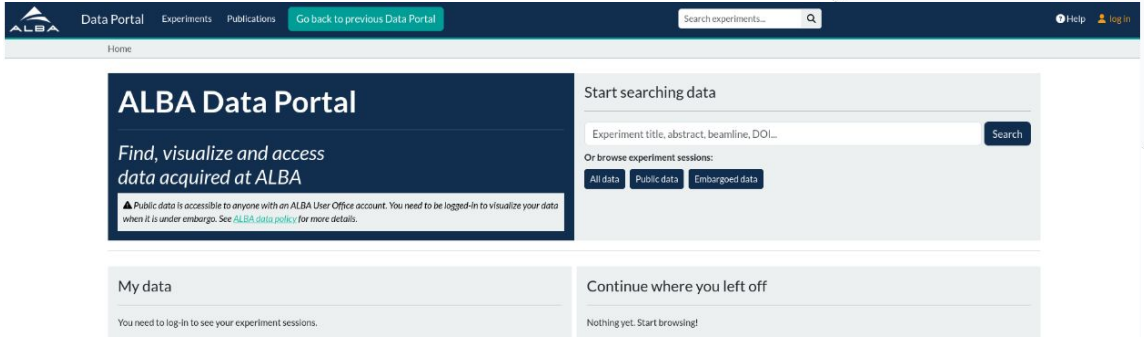
Current Data Policy for public experiments (raw data)

A few key points

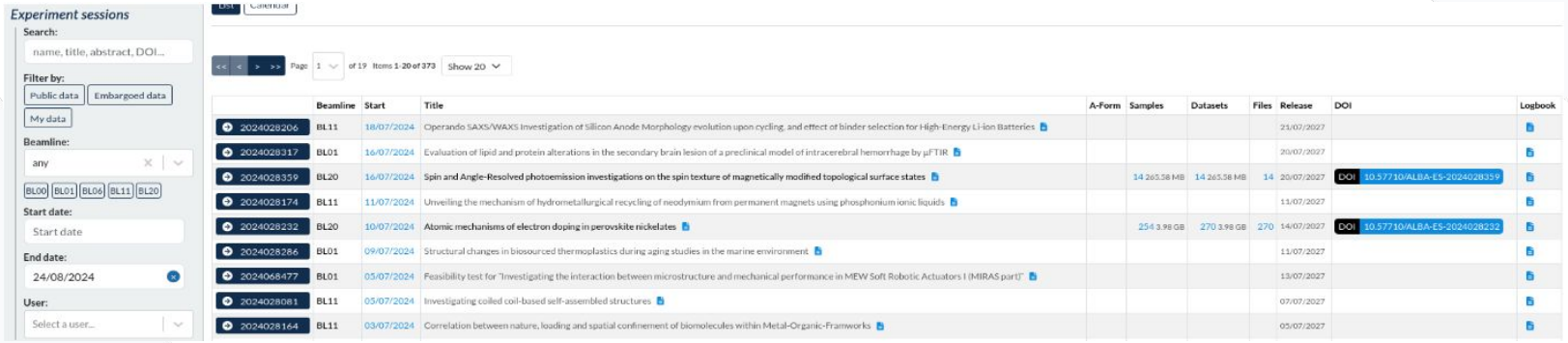
- **No need to copy data home**, ALBA custodies it
- Raw data is **retained** for a minimum period of **5 years** (archived after 1 year)
- A **3-year embargo period** is applied to raw data, afterwards it becomes **publicly available**
- High-level metadata, such as proposal title, authors and abstracts, are visible immediately after the experiment is completed.

ICAT: ALBA's data catalogue

https://data.cells.es



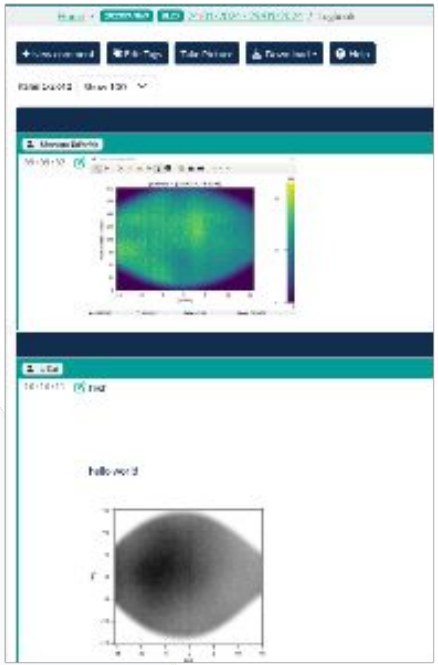
Previously, users accessed their data via SFTP or portable disks. Now, with the data catalogue, users can **browse** and **access** their data directly from anywhere.



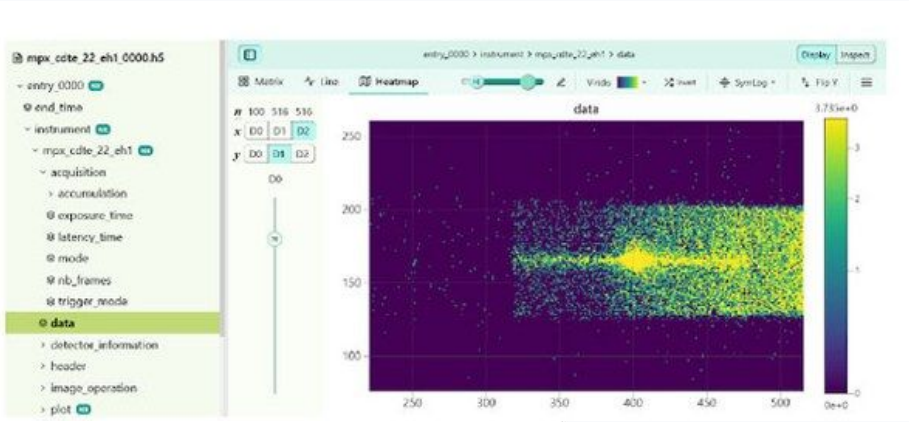
ICAT: ALBA's data catalogue

integrated tools

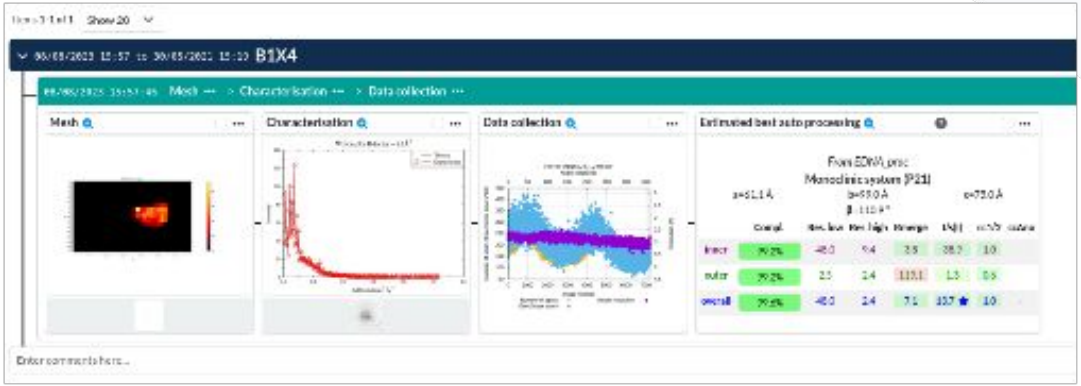
e-logbook to take notes during experiments



NeXus data **visualizer** (h5web)

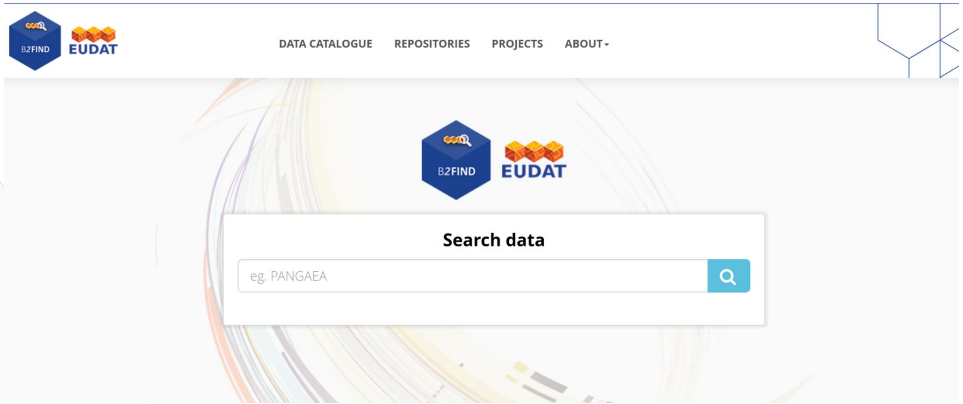


LIMS for MX

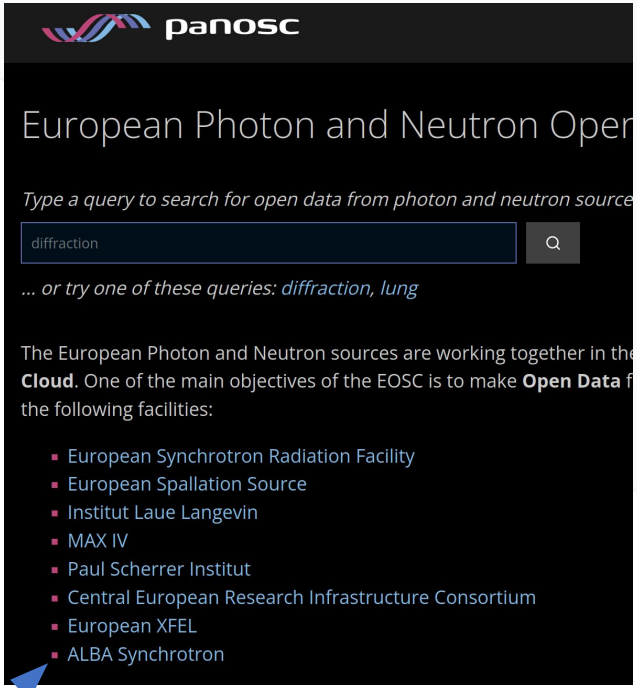


ICAT: ALBA's data catalogue

- Experiments are referenced by a **Digital Object Identifier (DOI)** and a **landing page**. DOIs can be also assigned to a **subset of datasets**.
- Photon and Neutron (PaN) **data sharing in Europe**



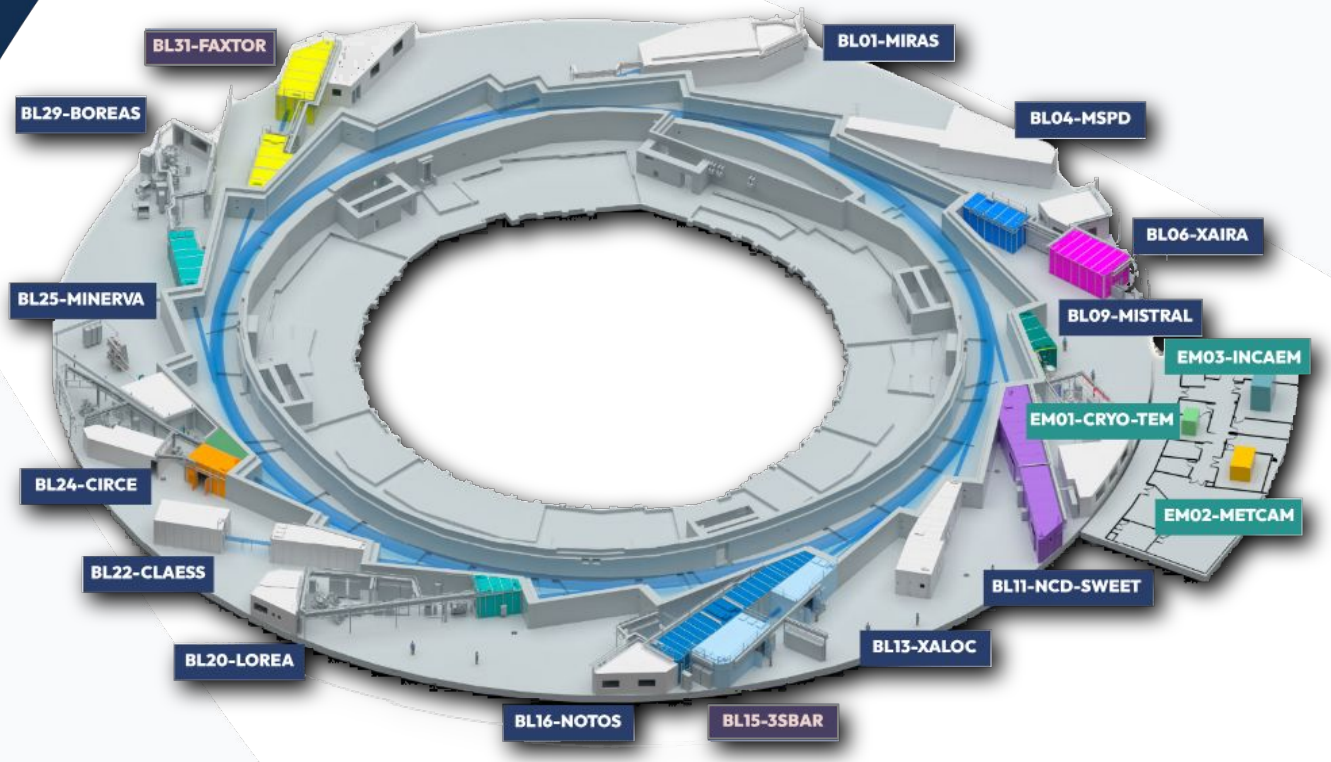
https://b2find.eudat.eu



https://data.panosc.eu/

ICAT: ALBA's data catalogue

Status and Roadmap

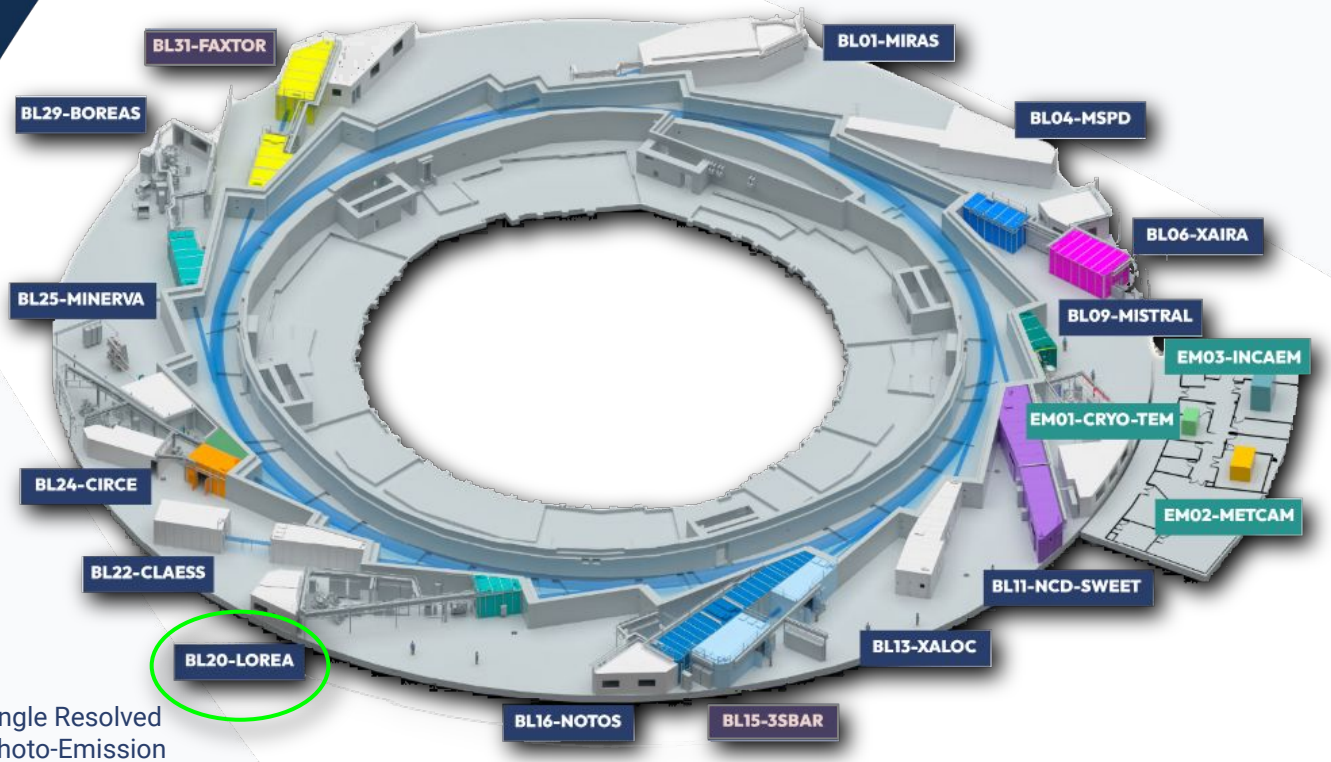


12 BLs in operation

2 BLs in construction

ICAT: ALBA's data catalogue

Status and Roadmap



Angle Resolved Photo-Emission Spectroscopy (ARPES)

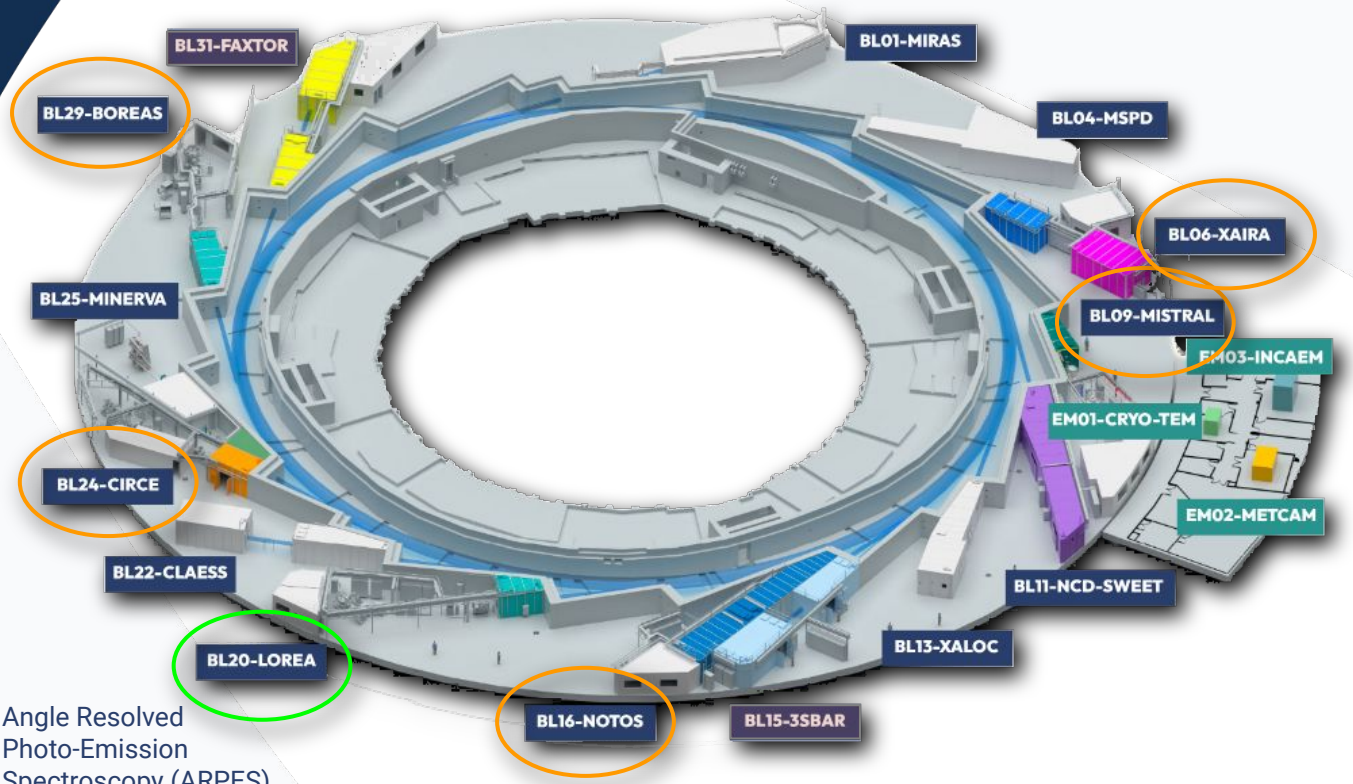
12 BLs in operation

2 BLs in construction

1 BL integrated in ICAT

ICAT: ALBA's data catalogue

Status and Roadmap



Angle Resolved Photo-Emission Spectroscopy (ARPES)

12 BLs in operation

2 BLs in construction

1 BL integrated in ICAT

5 in process (end 2024)

All other shortly after (2025)

Outline

- FAIR at ALBA
- **Challenges**
- Design and implementation
- Future considerations and Conclusions

Challenges
Organizational & Technical

Project spanning multiple **ALBA Computing Division sections**

- Systems, Controls, Scientific Data Management (SDM) & Management of Information Systems (MIS)
- Information transfer and debugging across several software systems

Challenges
Organizational & Technical

Project spanning multiple **ALBA Computing Division sections**

- Systems, Controls, Scientific Data Management (SDM) & Management of Information Systems (MIS)
- Information transfer and debugging across several software systems

International collaborations & inputs

Project tracking tools

Regular follow-ups

Small working groups

Avoid overlapping Identify boundaries

Clear API definitions

Documentation

Challenges
Organizational & Technical

Coordination with **Scientists**

- Balance FAIR requirements with existing workflows, aiming to **minimize interference** and additional workload. Point out the benefits to the users.
- Instrumentation context. Software to exploit the data.
- Identify relevant metadata to publish in ICAT for searches (findables).
- **Standardize** solutions as much as possible. **Maintenability**.

Challenges
Organizational & Technical

Coordination with **Scientists**

- Balance FAIR requirements with existing workflows, aiming to **minimize interference** and additional workload. Point out the benefits to the users.
- Instrumentation context. Software to exploit the data.
- Identify relevant metadata to publish in ICAT for searches (findables).
- **Standardize** solutions as much as possible. **Maintenability**.

Coordination with **Users**

- Sample description & Metadata
- Facilitate tools for this (before/during/after experiment)

Outline

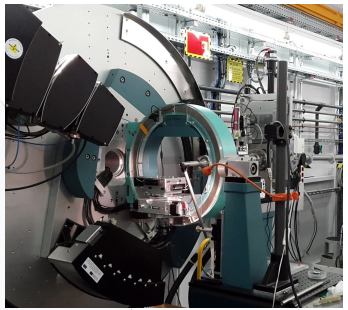
- FAIR at ALBA
- Challenges
- **Design and implementation**
- Remarks & future considerations

Design and implementation
the full picture

ALBA Data Portal

Find, visualize and access data acquired at ALBA

Beamline
Instrument



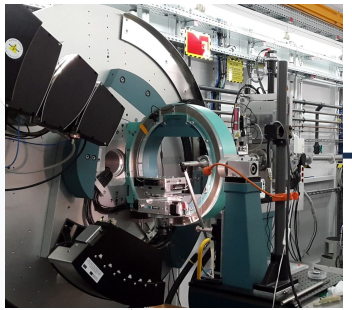
Design and implementation
the (meta)data flow

ALBA

ALBA Data Portal

Find, visualize and access data acquired at ALBA

Beamline
Instrument



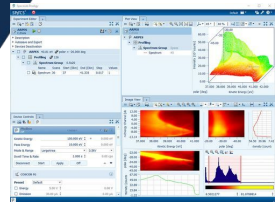
data/metadata

Control System

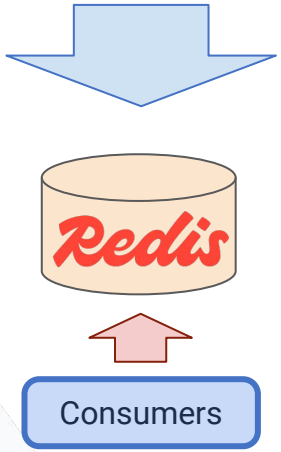
TANGO CONTROLS

Design and implementation

Controls



Proprietary endstation software

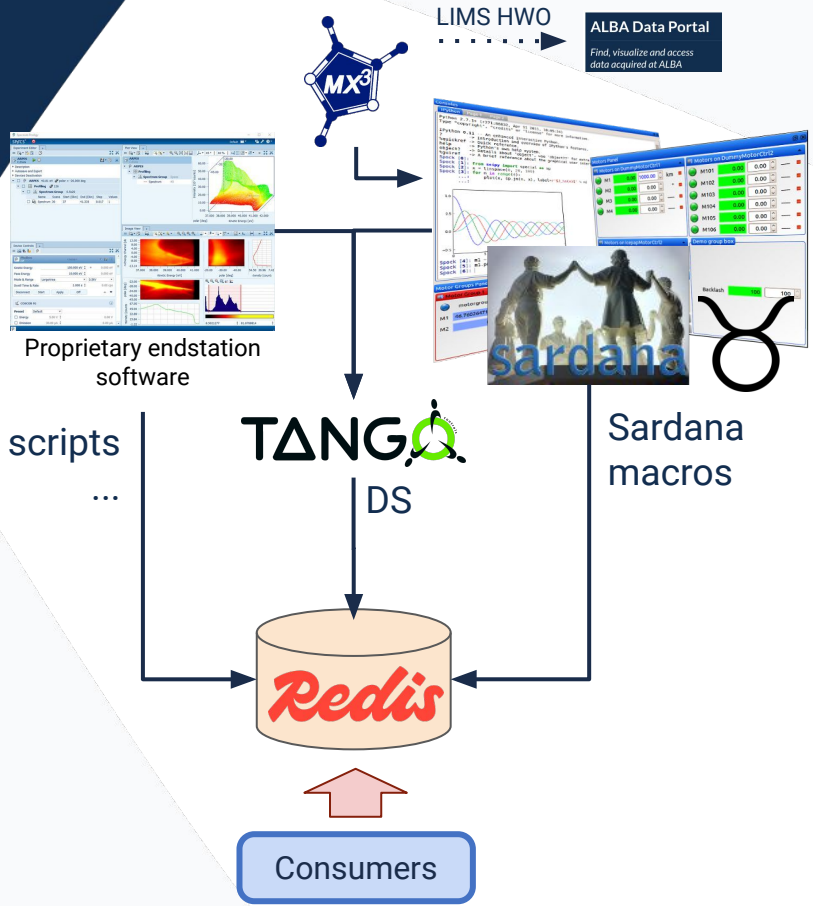


- Store metadata in **Redis** Streams.
  - Append-only data structure
  - Consumers act as processing queue.
  - Stream keys (e.g. *elogbook* or *meta*)
  - No data is lost. Consumers can retake processing from specified timestamps.
- Stream may contain:
  - Minimum info and point to NeXus fle.
  - Extended data definition with all metadata.

```
meta = {
    'beamline': 'BL24',
    'endstation': 'PEEM',
    'proposal_number': 2024000001,
    'metadata_file': '/path/to/nexus/file.h5',
    'metadata_dir': '/path/to/other/data/',
    'fetch_info_from_uos': True
}
```

Design and implementation

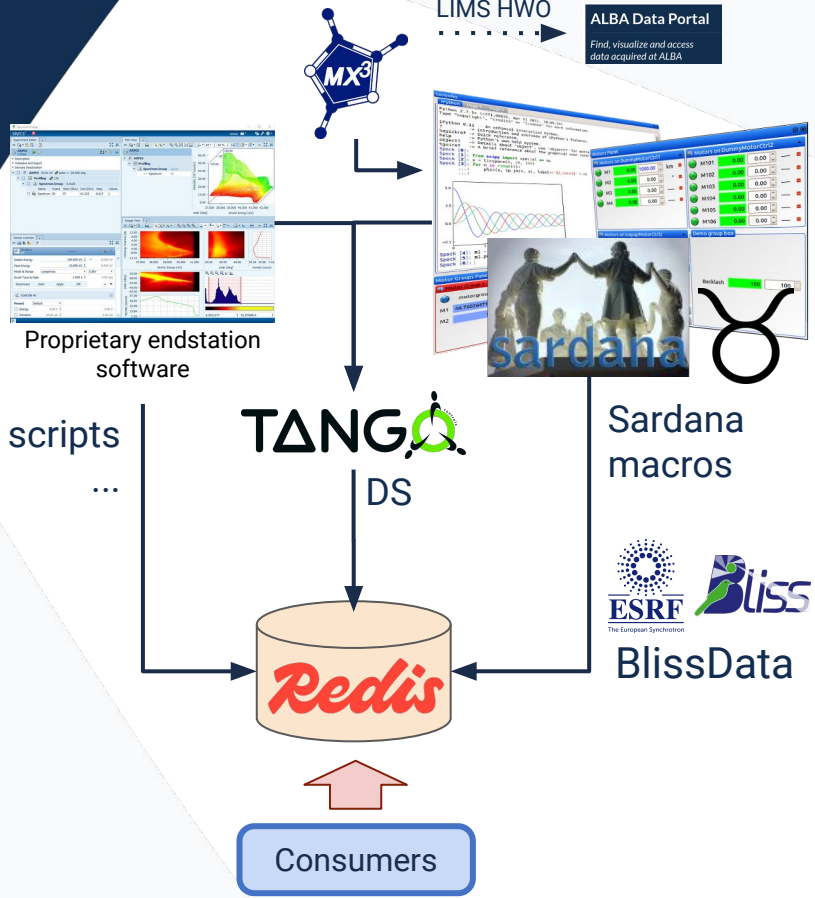
Controls



- Store metadata in **Redis** Streams.
  - Append-only data structure
  - Consumers act as processing queue.
  - Stream keys (e.g. *elogbook* or *meta*)
  - No data is lost. Consumers can retake processing from specified timestamps.
- Stream may contain:
  - Minimum info and point to NeXus fle.
  - Extended data definition with all metadata.
- Store from Sardana or from Tango device server.

Design and implementation

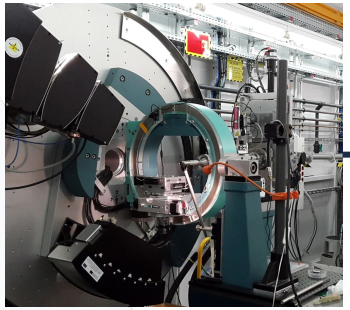
Controls

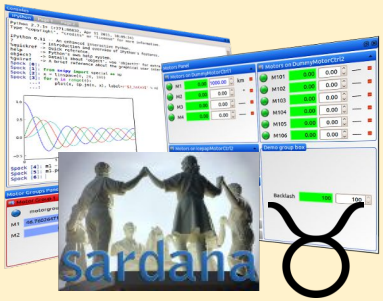


- Store metadata in **Redis** Streams.
  - Append-only data structure
  - Consumers act as processing queue.
  - Stream keys (e.g. *elogbook* or *meta*)
  - No data is lost. Consumers can retake processing from specified timestamps.
- Stream may contain:
  - Minimum info and point to NeXus fle.
  - Extended data definition with all metadata.
- Store from Sardana or from Tango device server
- Improving the generation of **NeXus files from Sardana**
- Redis integration in Sardana with **BlissData**

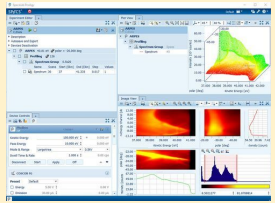
Design and implementation
the (meta)data flow

Beamline
Instrument





data/metadata



Proprietary endstation software

EDF
SPEC

...

Control System

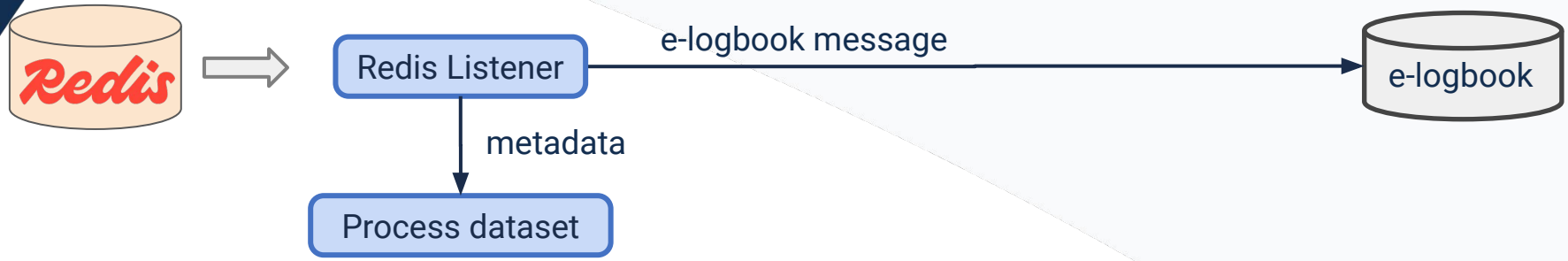
Scientific Data Management

ALBA Data Portal

Find, visualize and access data acquired at ALBA

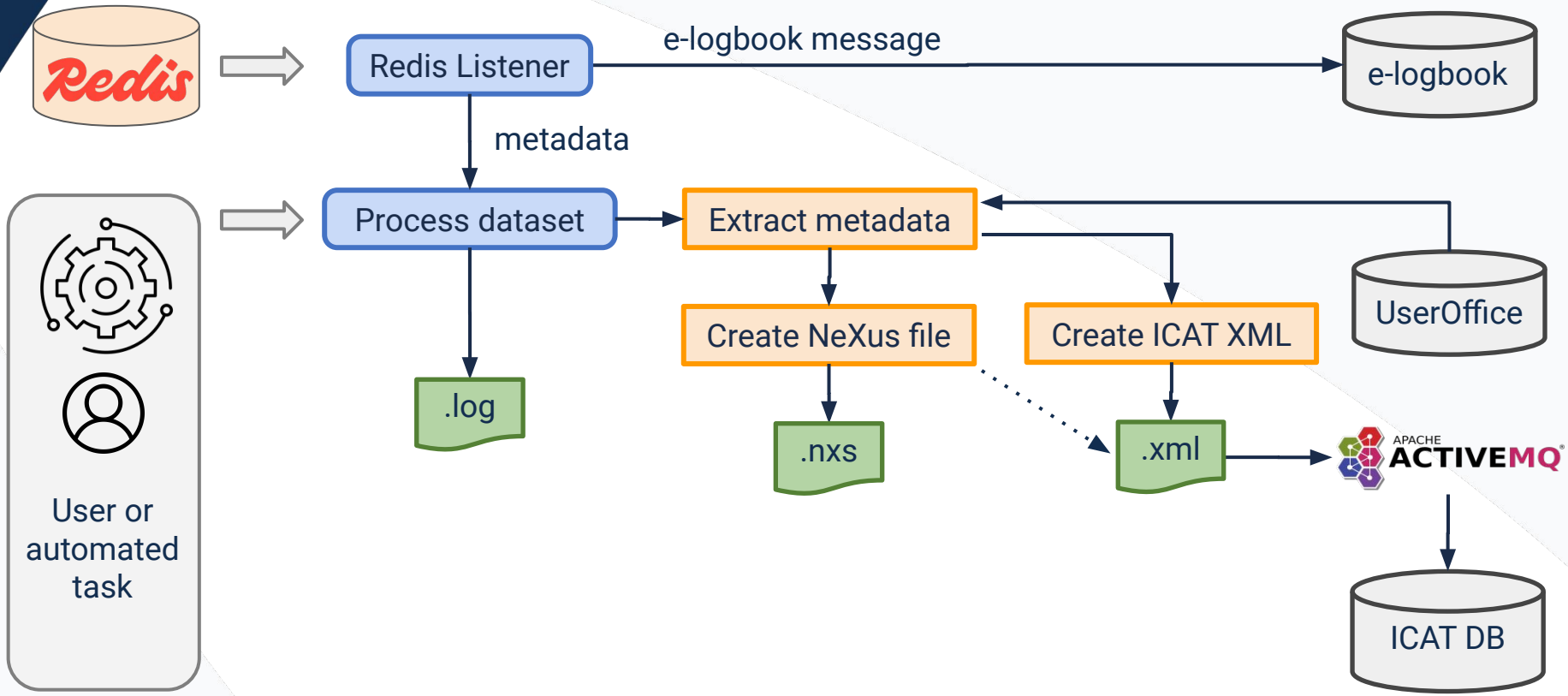
Design and implementation

Scientific Data Management



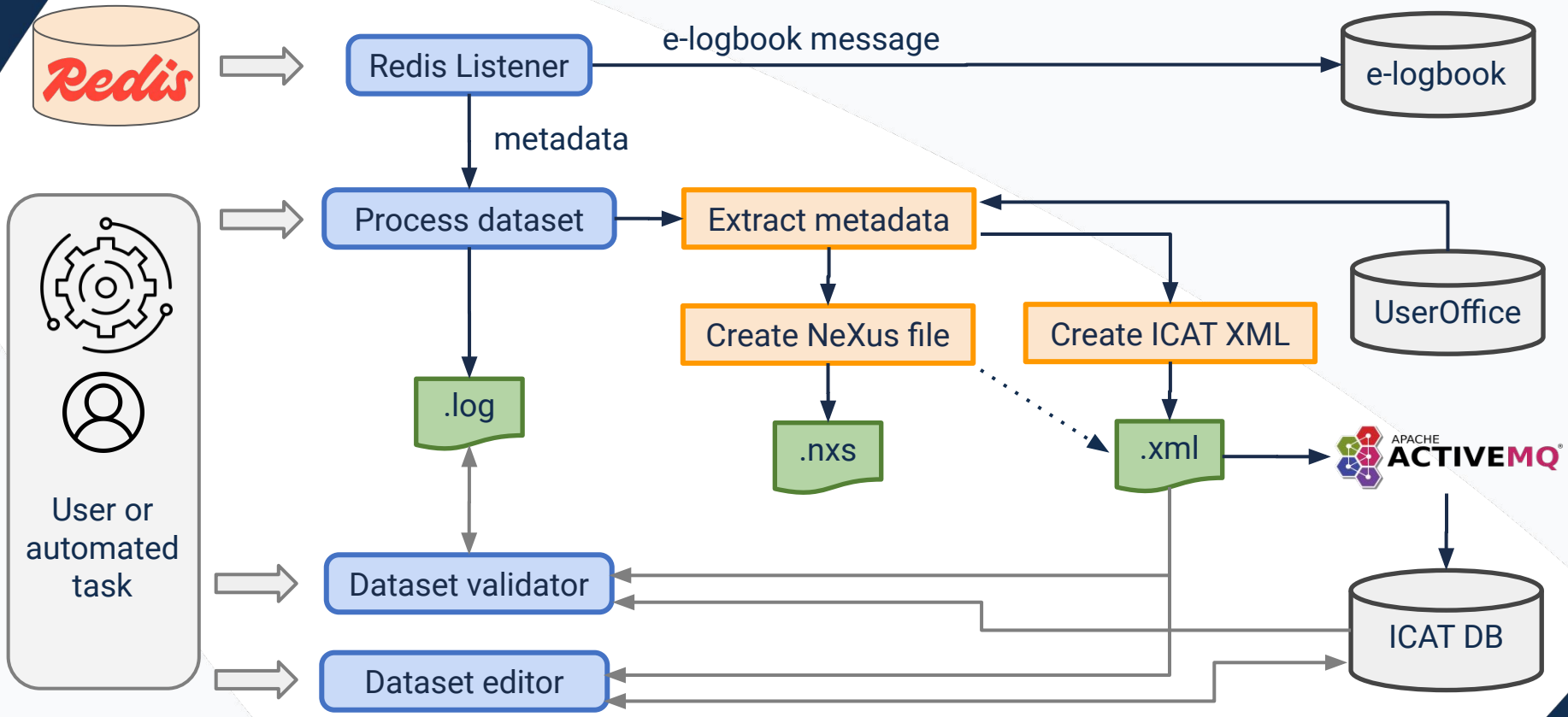
Design and implementation

Scientific Data Management



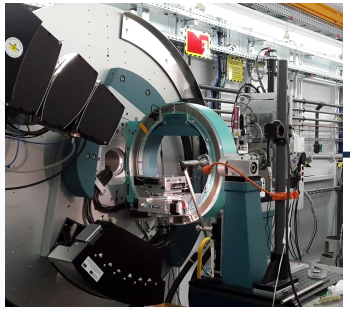
Design and implementation

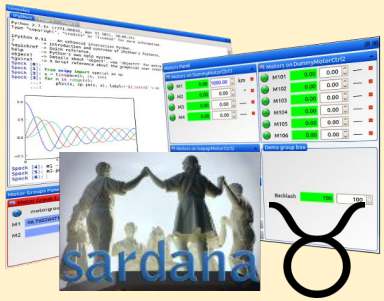
Scientific Data Management



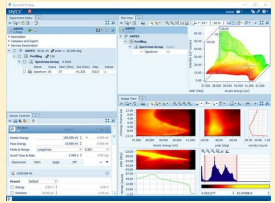
Design and implementation
the (meta)data flow

Beamline
Instrument





data/metadata



Proprietary endstation software

Redis

NeXus

EDF

SPEC

ESRF

...

Storage

Control System

TANGO

elogbook

ALBA Data Portal

Find, visualize and access data acquired at ALBA

metadata

NeXus

Storage

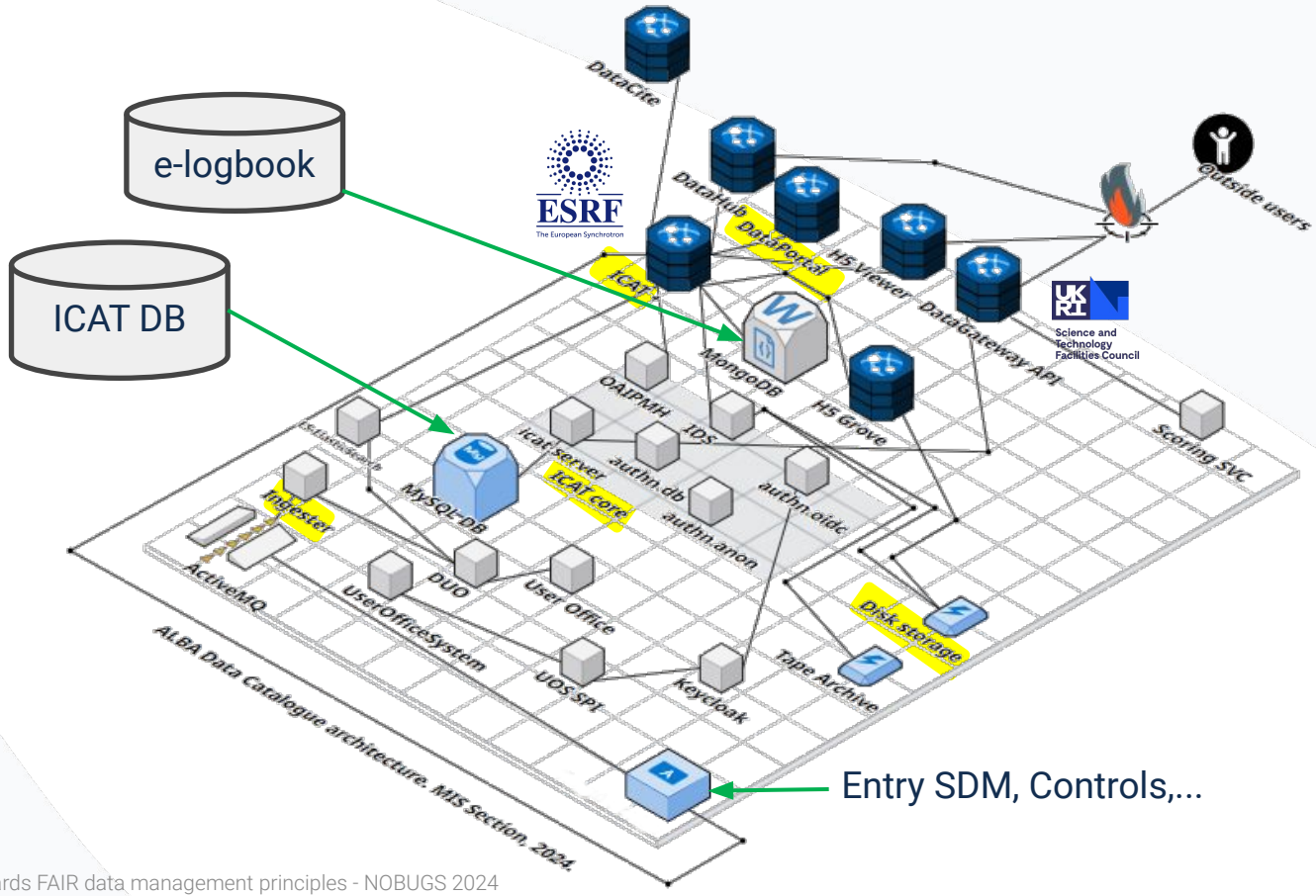
Other external sources

Scientific Data Management

MIS

Design and implementation

Management of Information Systems. ALBA Catalogue Architecture



Outline

- FAIR at ALBA
- Challenges
- Design and implementation
- **Remarks & future considerations**

Remarks & future considerations

- After several iterations on the design of how to implement Data Catalog on Beamlines, successful integration of first Beamline 😊
- Half of ALBA beamlines on *testing* ICAT DB. Close to move into production.
- Slow progress due to organizational challenges and operational constraints.

Remarks & future considerations

- After several iterations on the design of how to implement Data Catalog on Beamlines, successful integration of first Beamline 😊
- Half of ALBA beamlines on *testing* ICAT DB. Close to move into production.
- Slow progress due to organizational challenges and operational constraints.
- Reduce SDM processing layer:
  - Write metadata directly in Redis
  - Direct ingestion in ICAT from Control layer (MxCuBE)
  - Improve NeXus file writting from Sardana
- Add micro frontends for other techniques or processes
- Launching of data (re)processing/analysis directly from the catalog (VISA, JupyterLab,...). Save metadata of the processed datasets to ICAT.

Acknowledgements

SDM
Nicolas Soler
Fernan Saiz
Emilio Centeno
Albert Castellví

Controls
Fulvio Becheri
Zbigniew Reszela

ALBA BL Scientists

MIS
Marc Armenter
Rodrigo Cabezas

Systems
Gemma Rosas

Oscar Matilla

ALBA User Office

ESRF
The European Synchrotron

Alex de Maria
Maël Gaonach
...

ALBA SYNCHROTRON

Towards FAIR data management principles

Thank you!

questions?

Oriol Vallcorba (ovallcorba@cells.es)

on behalf of

Emilio Centeno, Fernan Saiz, Fulvio Becheri, Gemma Rosas, Marc Armenter, Nicolas Soler, Rodrigo Cabezas, Zbigniew Reszela

NOBUGS 2024, 23-27 September, ESRF and ILL, Grenoble