

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

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

- **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>



ALBA Data Portal

Find, visualize and access data acquired at ALBA

Public data is accessible to anyone with an ALBA User Office account. You need to be logged-in to visualize your data when it is under embargo. See [ALBA data policy](#) for more details.

Start searching data

Experiment title, abstract, beamline, DOI...

Search

Or browse experiment sessions:

All data Public data Embargoed data

My data

You need to log-in to see your experiment sessions.

Continue where you left off

Nothing yet. Start browsing!

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.

Experiment sessions

Search: name, title, abstract, DOI...

Filter by: Public data Embargoed data

My data

Beamline: any

Start date: 24/08/2024

User: Select a user...

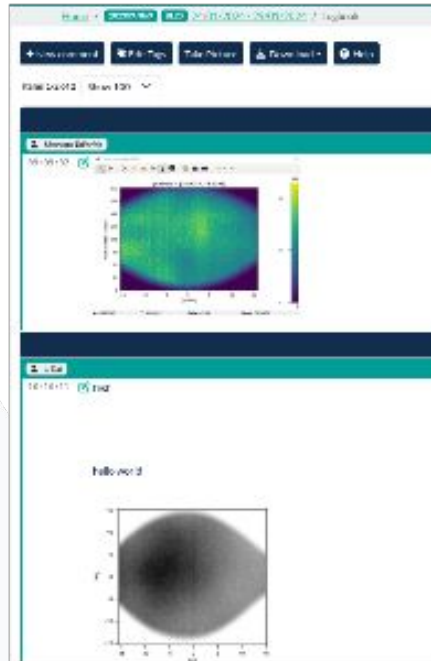
Beamline	Start	Title	A-Form	Samples	Datasets	Files	Release	DOI	Logbook
2024028206	BL11	18/07/2024	Operando SAXS/WAXS Investigation of Silicon Anode Morphology evolution upon cycling, and effect of binder selection for High-Energy Li-Ion Batteries				21/07/2022		
2024028317	BL01	16/07/2024	Evaluation of lipid and protein alterations in the secondary brain lesion of a preclinical model of intracerebral hemorrhage by µFTIR				20/07/2022		
2024028339	BL20	16/07/2024	Spin and Angle-Resolved photoemission investigations on the spin texture of magnetically modified topological surface states	14 265.58 MB	14 265.58 MB	14	20/07/2022	DOI: 10.57710/ALBA-E5-2024028339	
2024028174	BL11	11/07/2024	Unveiling the mechanism of hydrometallurgical recycling of neodymium from permanent magnets using phosphonium ionic liquids				11/07/2022		
2024028232	BL20	10/07/2024	Atomic mechanisms of electron doping in perovskite nickelates	254 3.98 GB	270 3.98 GB	270	14/07/2022	DOI: 10.57710/ALBA-E5-2024028232	
2024028286	BL01	09/07/2024	Structural changes in bio-sourced thermoplastics during aging studies in the marine environment				11/07/2022		
2024068477	BL01	05/07/2024	Feasibility test for "Investigating the interaction between microstructure and mechanical performance in MEW Soft Robotic Actuators I (MIRAS part)"				13/07/2022		
2024028081	BL11	03/07/2024	Investigating coiled coil-based self-assembled structures				07/07/2022		
2024028144	BL11	03/07/2024	Correlation between nature, loading and spatial confinement of biomolecules within Metal-Organic-Frameworks				05/07/2022		

ICAT: ALBA's data catalogue

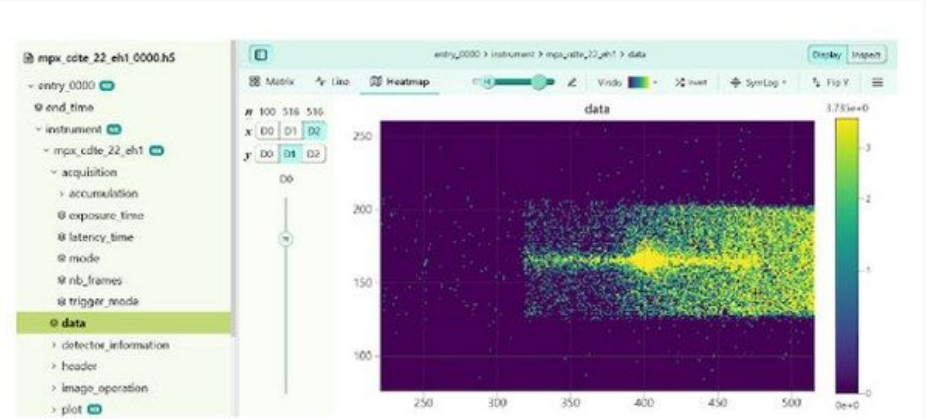
integrated tools



e-logbook to take notes during experiments



NeXus data visualizer (h5web)



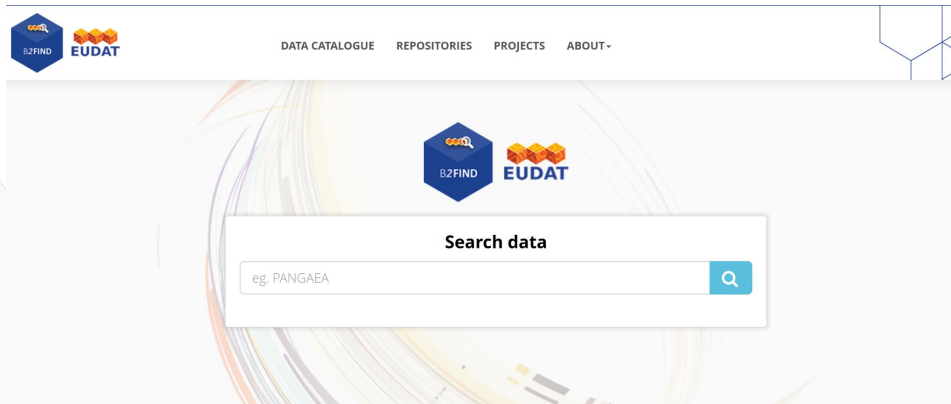
LIMS for MX



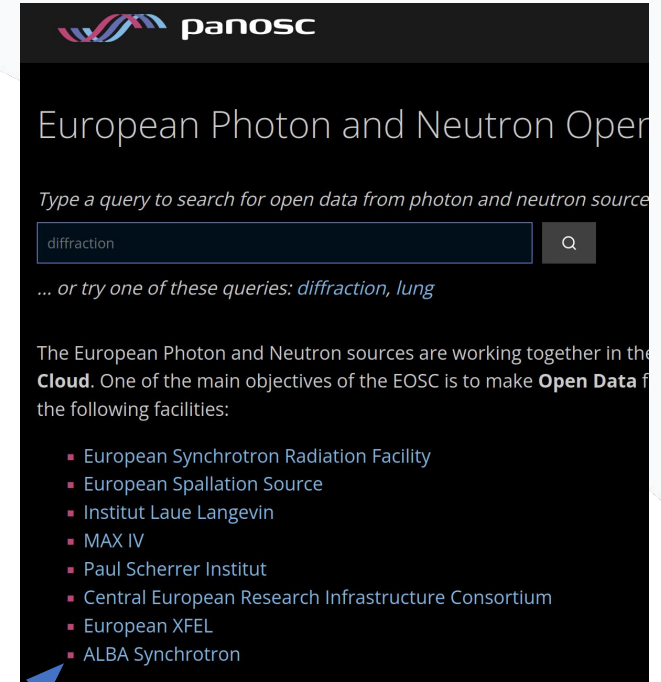
ICAT: ALBA's data catalogue



- 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**



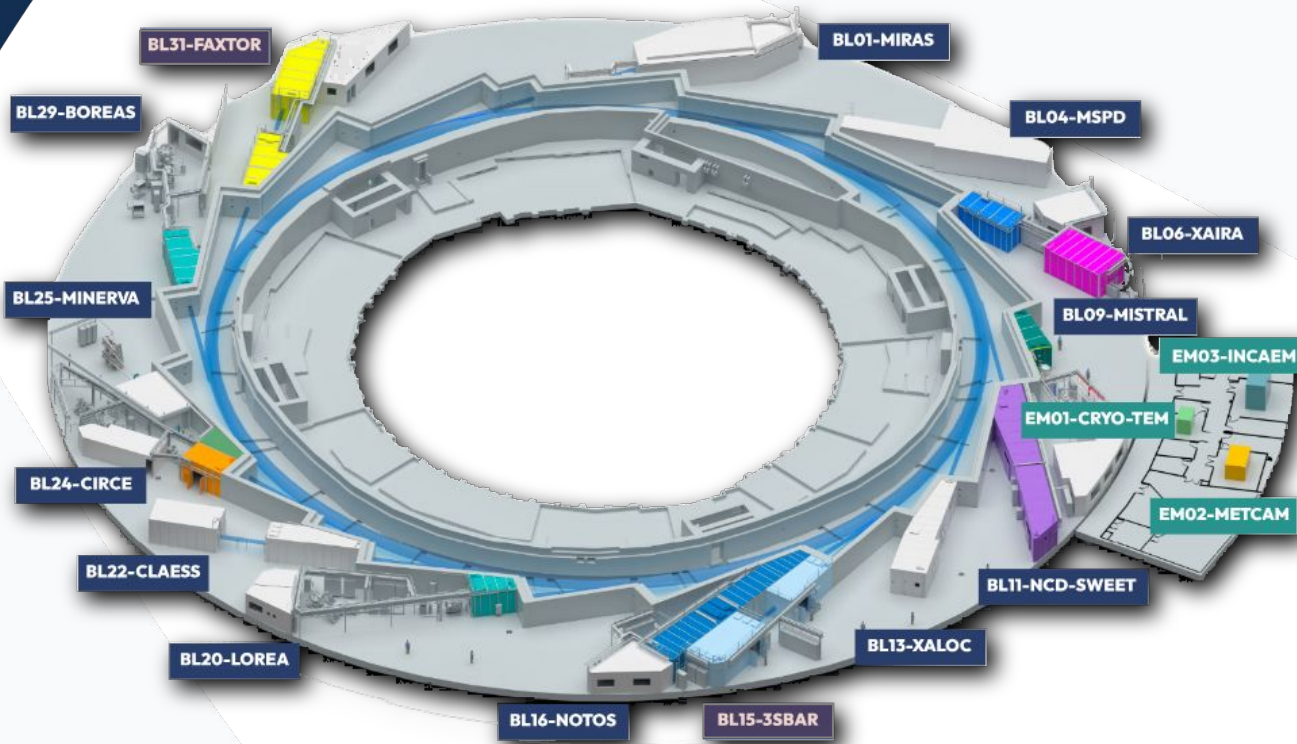
<https://b2find.eudat.eu>



<https://data.panosc.eu/>

ICAT: ALBA's data catalogue

Status and Roadmap

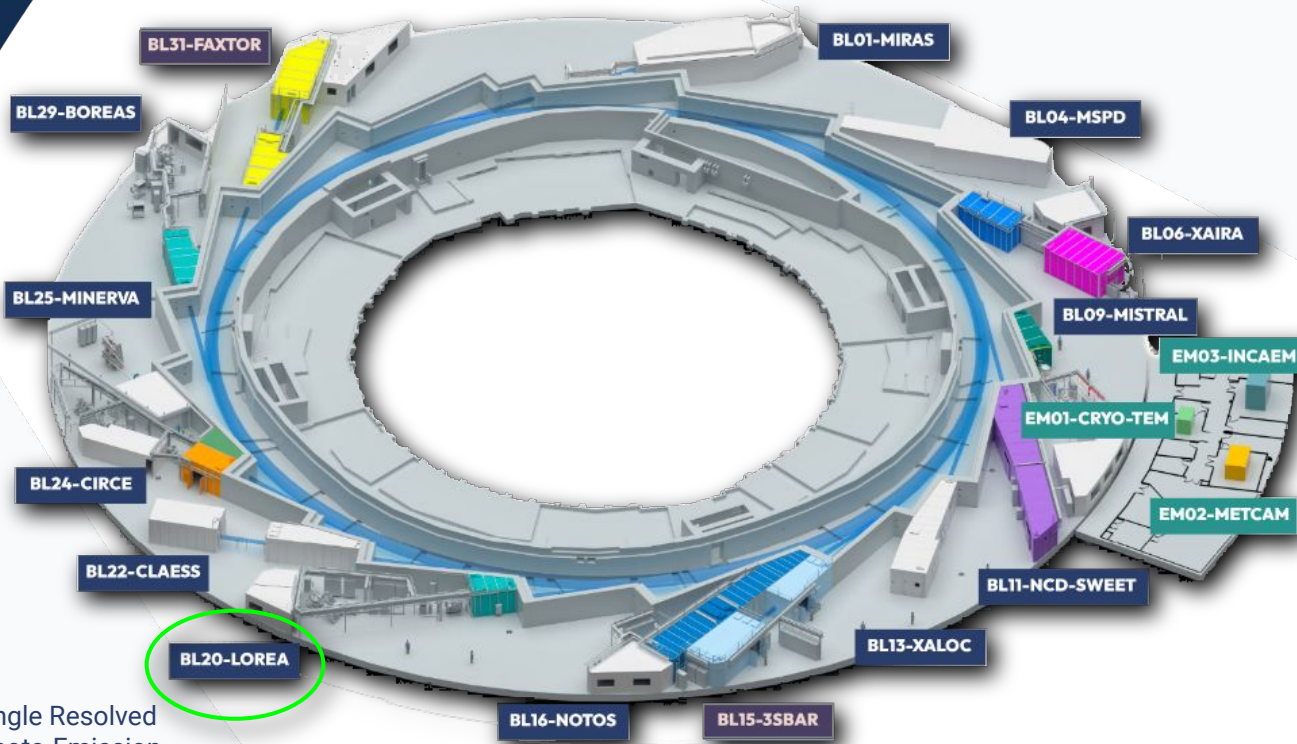


12 BLs in operation

2 BLs in construction

ICAT: ALBA's data catalogue

Status and Roadmap



12 BLs in operation

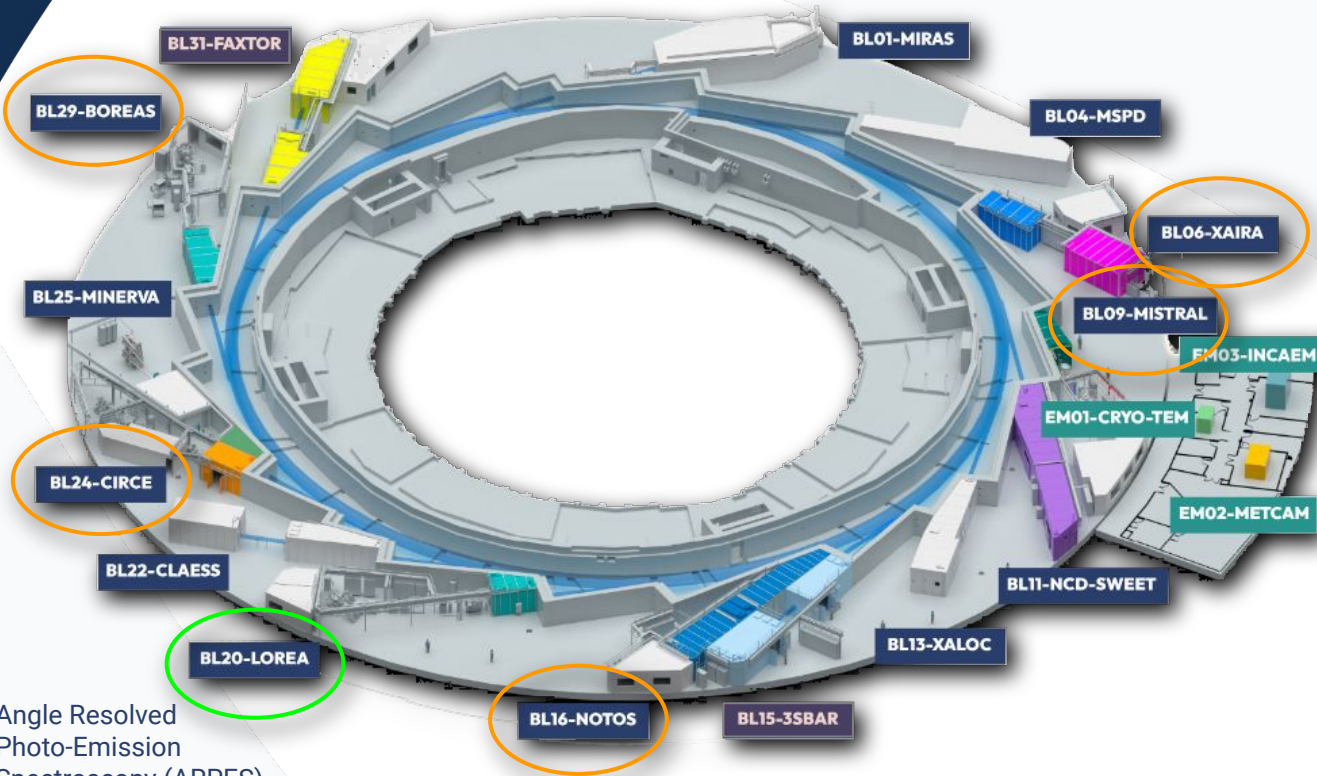
2 BLs in construction

1 BL integrated in ICAT

Angle Resolved
Photo-Emission
Spectroscopy (ARPES)

ICAT: ALBA's data catalogue

Status and Roadmap



12 BLs in operation

2 BLs in construction

1 BL integrated in ICAT

5 in process (end 2024)

All other shortly after (2025)

Angle Resolved
Photo-Emission
Spectroscopy (ARPES)

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

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

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

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. **Maintainability**.

Coordination with **Scientists**

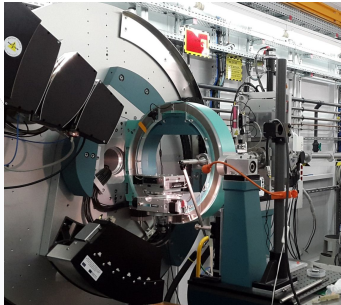
- Balance FAIR requirements with existing workflows, aiming to **minimize interference** and additional workload. Point out the benefits to the users.
- Instrumentation context. Software to exploit the data.
- Identify relevant metadata to publish in ICAT for searches (findables).
- **Standardize** solutions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

Coordination with **Users**

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

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

Beamline Instrument



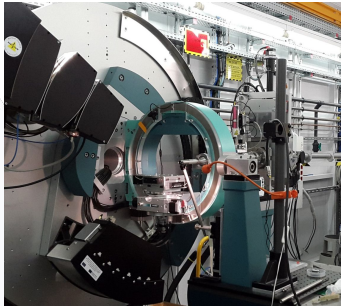
ALBA Data Portal

*Find, visualize and access
data acquired at ALBA*

Design and implementation the (meta)data flow



Beamline
Instrument



data/metadata

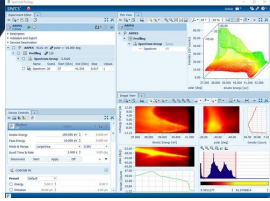


Control
System

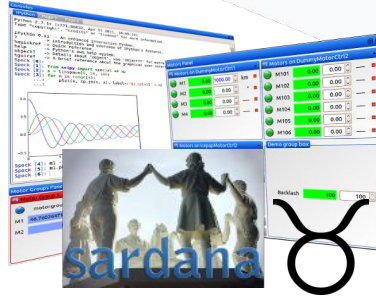


ALBA Data Portal

*Find, visualize and access
data acquired at ALBA*



Proprietary endstation software



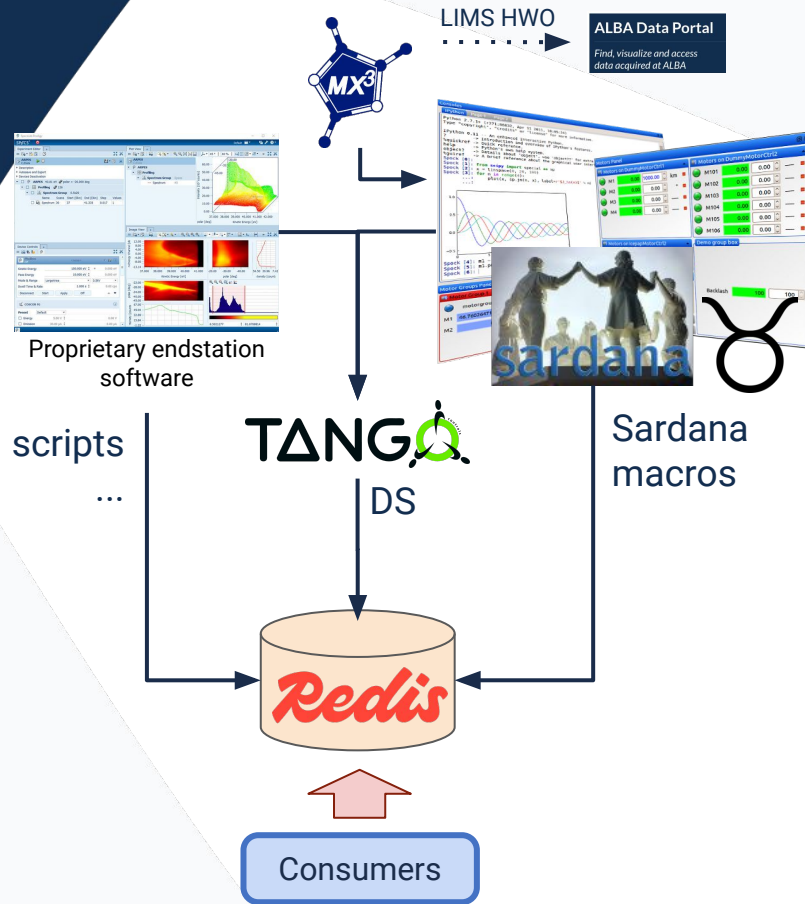
Consumers

- 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

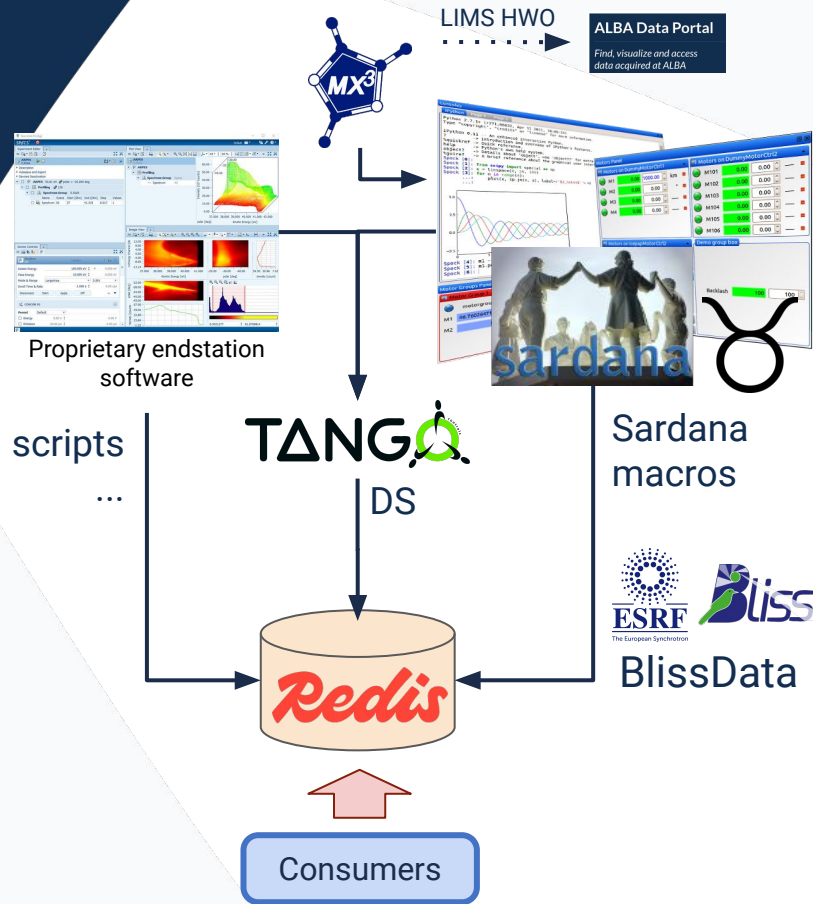
Controls



- 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 file.
 - Extended data definition with all metadata.
- Store from Sardana or from Tango device server.

Design and implementation

Controls

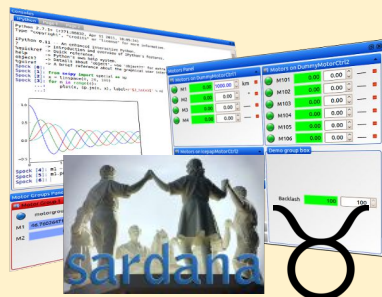
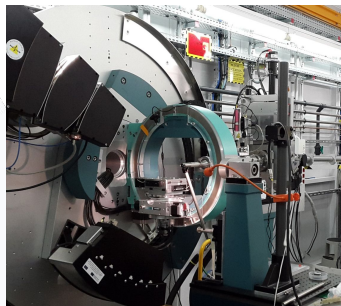


- Store metadata in **Redis** Streams.
 - Append-only data structure
 - Consumers act as processing queue.
 - Stream keys (e.g. *elogbook* or *meta*)
 - No data is lost. Consumers can retake processing from specified timestamps.
- Stream may contain:
 - Minimum info and point to NeXus fle.
 - Extended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

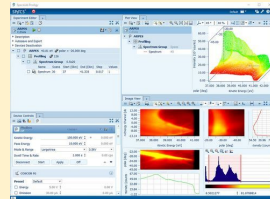
Design and implementation the (meta)data flow



Beamline
Instrument



data/metadata

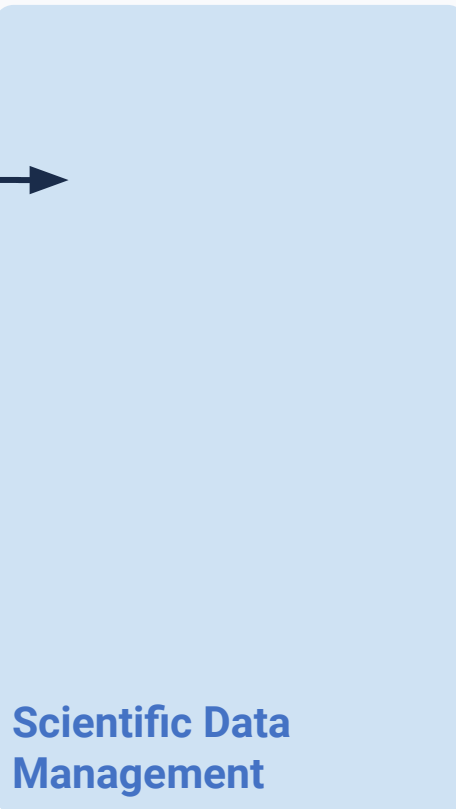


Proprietary endstation
software

Control
System



...

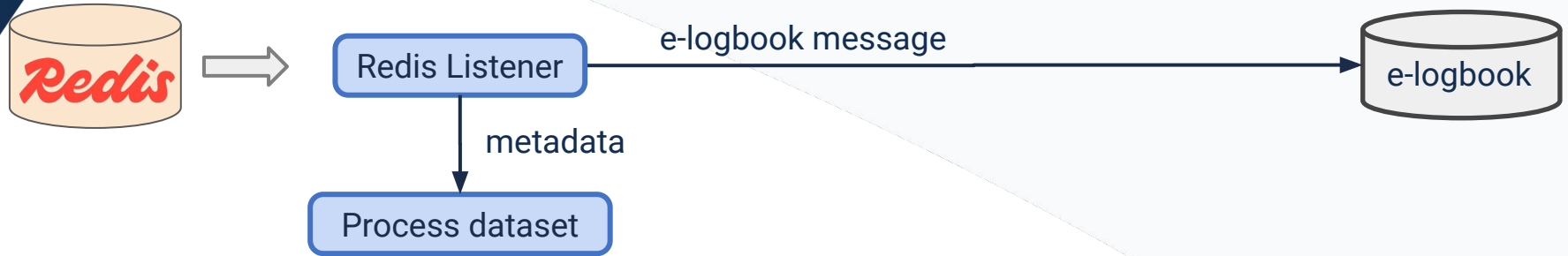


ALBA Data Portal

Find, visualize and access
data acquired at ALBA

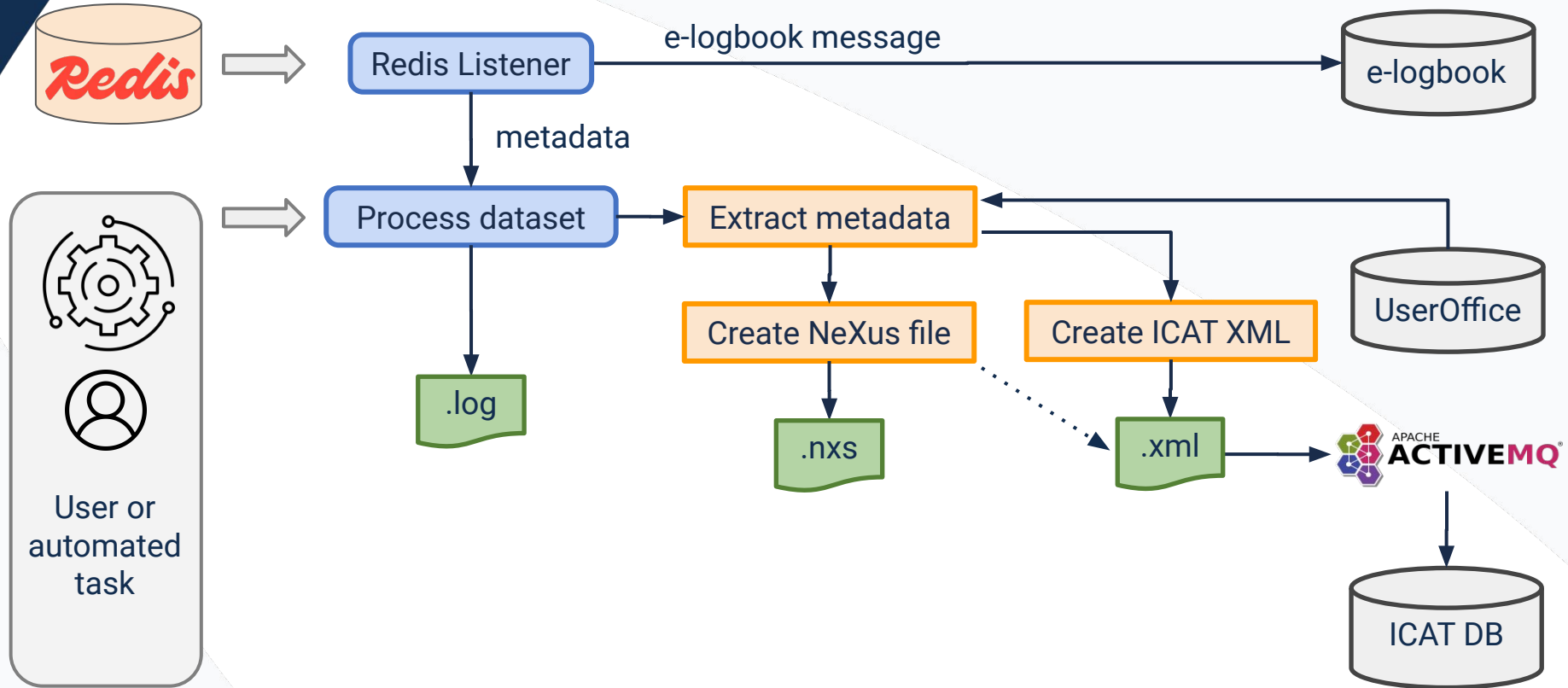
Design and implementation

Scientific Data Management



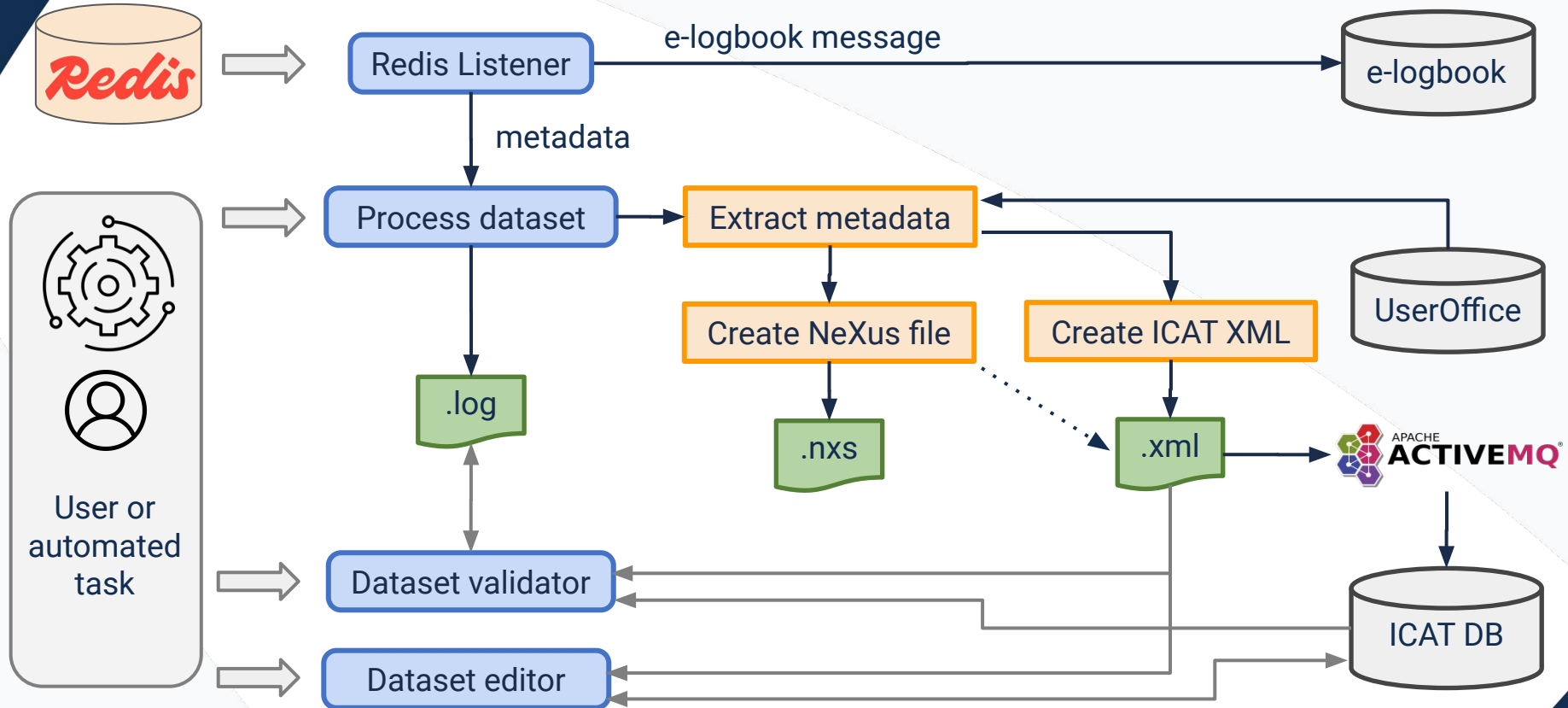
Design and implementation

Scientific Data Management



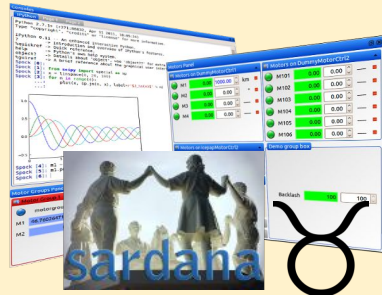
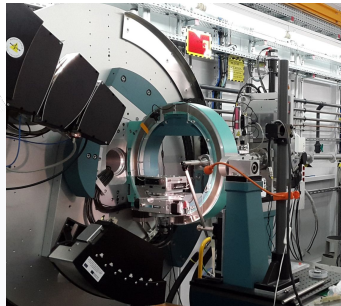
Design and implementation

Scientific Data Management

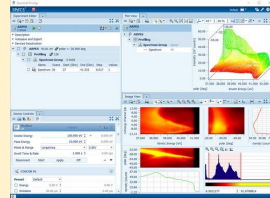


Design and implementation the (meta)data flow

Beamline
Instrument

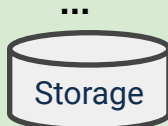


data/metadata



Proprietary endstation
software

Control
System



Scientific Data
Management

eelogbook

ALBA Data Portal

Find, visualize and access
data acquired at ALBA

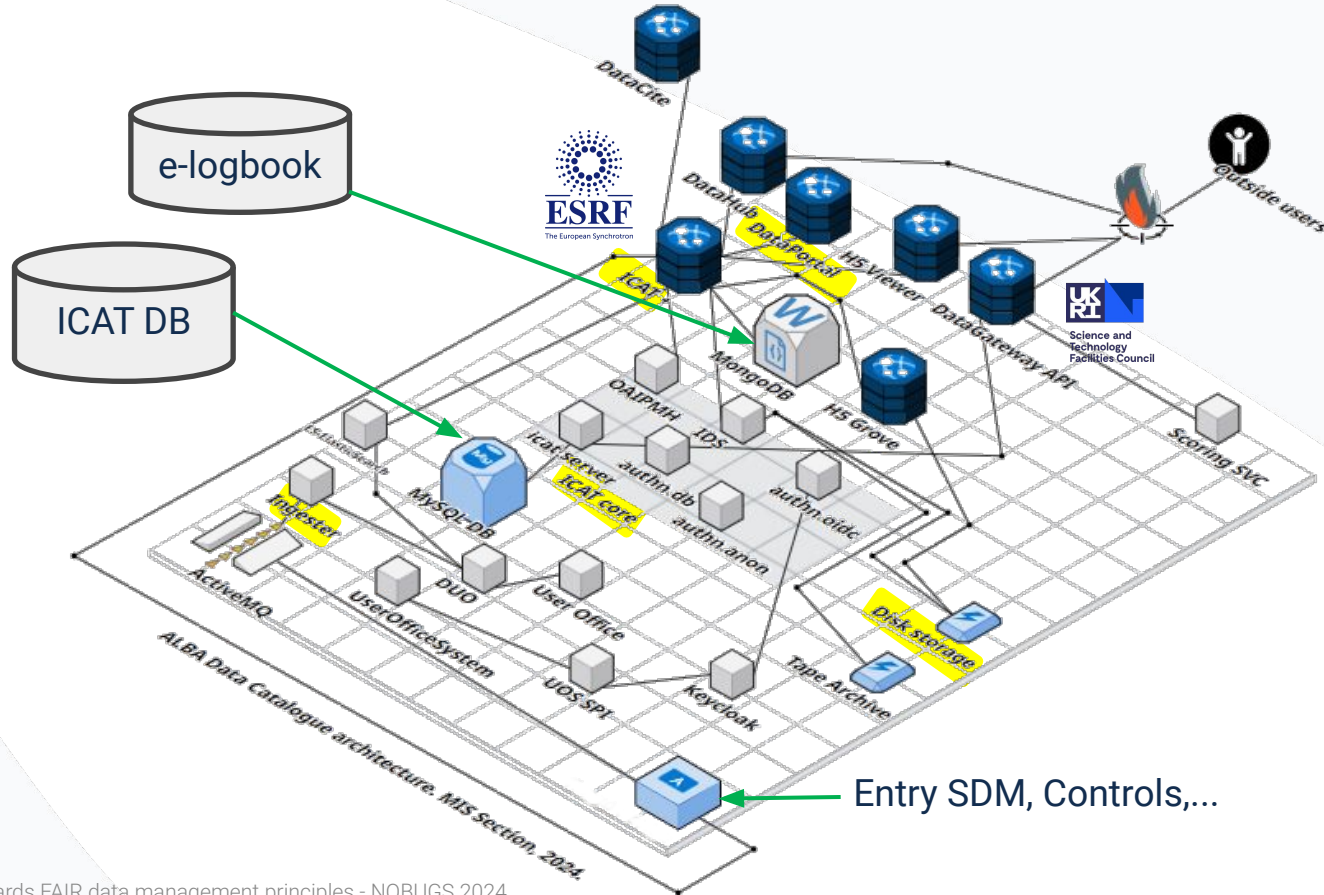
metadata

Other
external
sources

MIS

Design and implementation

Management of Information Systems. ALBA Catalogue Architecture



- 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 writing 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

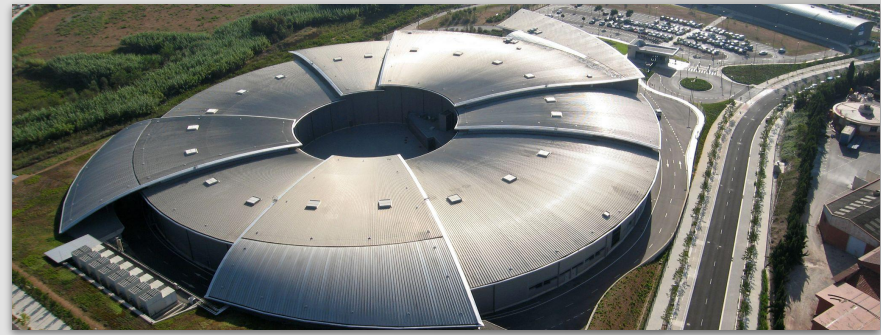
Oscar Matilla

ALBA User Office



Alex de Maria
Maël Gaonach

...



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