

# Generic data management policy at ALBA CELLS.

This document describes the scientific data management policy at the synchrotron ALBA at the CELLS consortium.

## 1 General principles

- 1.1 This data management policy covers the access and ownership of the experimental raw data and metadata collected and/or stored at ALBA-CELLS.
- 1.2 Acceptance of this policy is a condition of the award of public beam time.
- 1.3 Users must not attempt to access, exploit or distribute data or metadata under the terms of this policy unless they are entitled to do so.
- 1.4 Deliberate infringements of the policy may lead to denial of access to data and metadata and the denial of future beam time proposals.
- 1.5 All data and metadata will be managed subject to the Spanish Data Protection legislation (Ley Orgánica de Protección de Datos LOPD).

<http://www.cells.es/en/users/applying-for-beam-time>

## 2 Definitions

Experiments carried out at ALBA can be classified as public access (including the regular user calls, friendly users and training experiments, in-house research, access derived from formal collaboration agreements and the use of management contingency beam-time) and proprietary access.

Public access experiments are covered by this policy.

Proprietary access experiments are carried out keeping the proposal, raw data and results confidential and therefore are not covered by this policy.

For the purposes of this policy:

- 2.1 The term facility refers to the ALBA Synchrotron at CELLS.
- 2.2 The term **raw data** refers to data collected from experiments performed on Beamlines or other instruments at ALBA-CELLS. This definition includes data that are created automatically or manually by facility specific software and/or ALBA-CELLS staff expertise in order to facilitate subsequent analysis of the experimental data, unless otherwise agreed.
- 2.3 The term **metadata** describes information referring to data collected from ALBA-CELLS instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.
- 2.4 The term **principal investigator** (PI) pertains to the PI identified on the ALBA-CELLS academic proposal (for peer-reviewed experiments) or the Safety approval form for any other proposals (like in-house, training and user-friendly experiments, special international access, etc.).
- 2.5 The term **experimental team** includes the PI and any other person to whom the PI designates the right to access resultant raw data and associated metadata.

- 2.6 The term *on-line catalogue* refers to a computer database of metadata containing links to raw data files, that can be accessed by a variety of methods, including (but not limited to) www-based browsers.
- 2.7 The term *result* refers to data, intellectual property, and outcomes arising from the analysis of raw data. This does not include publications.
- 2.8 The term *open access* means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone. It also means unrestricted (but not anonymous) and free-of-charge access. The ALBA-CELLS data archive will be made available under CC-BY

(Creative Commons BY, <http://creativecommons.org/licenses/by/4.0/legalcode>)

### 3 Raw data and associated metadata

#### 3.1 Access to raw data and associated metadata

- 3.1.1 All raw data and the associated metadata obtained as a result of public access experiments, as defined in section 2 above, will be open access after an initial embargo period during which access is restricted to the experimental team, represented by the PI.
- 3.1.2 ALBA-CELLS is the custodian of the raw data and associated metadata.
- 3.1.3 All raw data and the associated metadata obtained as a result of proprietary access to ALBA-CELLS, will be owned exclusively by the client who purchased the access and is not covered by this data policy. Permissions and relevant procedures are in place to guarantee the confidentiality. Data from proprietary research will be removed after the experiment from the ALBA-CELLS storage systems, unless otherwise agreed with ALBA-CELLS management before the start of any experiment.

#### 3.2 Curation of raw data and associated metadata

- 3.2.1 Raw data may or may not be curated at ALBA-CELLS premises. The raw data will be curated in well-defined formats, for which the means of reading the data will be made available by ALBA-CELLS.
- 3.2.2 Metadata that is automatically captured by instruments will be curated either within the raw data files, within an associated on-line catalogue, or within both.
- 3.2.3 Only data with metadata generated by ALBA-CELLS will be archived.
- 3.2.4 Raw data and metadata will be read-only for the duration of their life time.
- 3.2.5 Raw data and metadata will be migrated or copied to archival facilities for long-term curation.
- 3.2.6 It is planned that each experiment and data set will have a unique persistent identifier. Anybody publishing results based on open access data must quote the same identifier (and related publications if available and required).
- 3.2.7 High level metadata such as Title, Authors, Abstract, Beamline will be made public as soon as the experiment has been carried out. This information will be available via the persistent identifier landing page on the web.

#### 3.3 Access to raw data and metadata

- 3.3.1 Access to raw data and metadata will be via suitable list/search tools.

- 3.3.2 Access to data will be restricted to those who are registered as users at ALBA-CELLS. Any retrieval of data may require access to permanent store with the consequent latencies and supplementary waiting times.
- 3.3.3 ALBA-CELLS shall custody the data for a minimum of 5 years. On a best effort basis, ALBA-CELLS will keep the data accessible online for one year, after which it can be archived on a tape library or a warehouse with the consequent latencies and waiting times. Access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team for 3 years. After the embargo period is over the data and metadata will become publicly accessible. Any PI that wishes their data not to become “publicly accessible” for a longer period will be required to make a special case to the Director of ALBA-CELLS.
- 3.3.4 It is the responsibility of the PI to ensure that the data is stored into the directories enabled to that purpose and that the experiment number is correctly entered into the metadata for each raw data set, in order to correctly associate each data set with the PI. If this is not done, the experimental team will not be able to access the data, or other users may inadvertently be given access rights to the data.
- 3.3.5 Authorized ALBA-CELLS staff (e.g. instrument scientists, computing group members) have access to any facility curated data or metadata for facility related purposes. ALBA-CELLS undertakes the compromise that they will preserve the confidentiality of such data during the embargo period.
- 3.3.6 The on-line catalogue will enable linking experimental data to experimental proposals. Access to proposals will only ever be provided to the experimental team and appropriate ALBA-CELLS staff, unless otherwise authorized by the PI during the 3-year embargo period.
- 3.3.7 The PI has the possibility to transfer parts of the totality of her/his rights during the embargo period to another registered person.
- 3.3.8 The PI has the possibility to create and distribute copies of the raw data.

## 4 Results

### 4.1 Ownership of results

- 4.1.1 Ownership of all results (intellectual property) derived from the analysis of the raw data is determined by the contractual obligations of the person(s) performing the analysis, as well as by the copyright agreements assumed upon publication.

### 4.1 Curation of results

- 4.1.1 The results of the analysis performed at ALBA-CELLS can be stored at ALBA-CELLS premises if the PI decides so. It will not be the responsibility of ALBA-CELLS to fully curate this data e.g. to ensure that software to read / manipulate this data is available.
- 4.1.2 ALBA-CELLS can not be made liable in case of unavailability or loss of data, results or data analysis software.

### 4.2 Access to results

- 4.2.1 Access to the results of analysis performed on raw data and metadata during the embargo period is restricted to the experimental team in charge of the analysis, unless otherwise requested by the members of that team.
- 4.2.2 Appropriate facility staff (e.g. instrument scientists, computing group members) has access to any facility-curated data or metadata for facility related purposes. ALBA-CELLS will undertake that confidentiality of such data is preserved during the embargo period.

## **5 Good practice for metadata capture and results storage**

- 5.1 The experimental team is encouraged to ensure that experimental metadata are as complete as possible, as this will enhance the possibilities for them to search for, retrieve and interpret their own data in the future.
- 5.2 ALBA-CELLS undertakes to provide means, on a best effort basis, for the capture of such metadata items that are not automatically captured by an instrument, in order to facilitate recording the fullest possible description of the raw data.
- 5.3 Researchers who aim to carry out analyses of raw data and metadata which are openly accessible should, where possible, contact the original PI to inform her/him and suggest a collaboration if required. Researchers must acknowledge the source of the data and cite its unique identifier as well as any publications linked to the same raw data.
- 5.4 PIs and researchers who carry out analyses of raw data and metadata are encouraged to link the results of these analyses with the raw data / metadata using the facilities provided by the on-line catalogue. Furthermore, they are encouraged to make such results publicly accessible.

## **6 Publication information**

- 6.1 Publications related to data from experiments carried out at ALBA-CELLS must cite the persistent identifier of the experiment and data in their publication.
- 6.2 References for publications related to experiments carried out at ALBA-CELLS must be deposited in the ALBA-CELLS publications database within six months of the publication date, or during any new application for beamtime, whichever is the earlier. Non-compliance to this rule may result on not getting further beamtime allocation.