



Generic data management policy at the ALBA Synchrotron and the JEMCA

This document describes the scientific data management policy that the CELLS consortium applies to the ALBA synchrotron and the Joint Electron Microscopy Center at ALBA (referred to hereinafter as the JEMCA and composed, on 1st of March of 2023, of two instruments: EM01-Cryo-TEM and EM02-METCAM).

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 by the CELLS.

1.2 Raw data collected from instruments included in the JEMCA are subjected to special regulations, as indicated in the following sections wherever the case applies.

1.3 Acceptance of this policy is mandatory for the awarding of public beam time.

1.4 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.

1.5 Deliberate infringements of the policy may lead to the denial of access to data and metadata and the denial of participation in future calls for proposals.

1.6 All data and metadata will be managed according 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 the ALBA Synchrotron and the JEMCA can be classified as public access (including 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), access derived from specific partnership agreements and proprietary access.

Public access and specific partnership experiments are covered by this policy.

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

For the purposes of this policy:

2.1 The Alba Synchrotron and the JEMCA are considered as two distinct facilities.

2.2 The term “raw data” refers to data collected from experiments performed on beamlines or other instruments. This definition includes data that are created automatically or manually by facility-specific software and/or 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 the 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) refers to the PI identified on the academic proposal (for peer-reviewed experiments) or the Safety approval form for any other proposals

2.5 The term “experimental team” includes the PI and any other person to whom the PI grants 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 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 The CELLS provides custody of the raw data and associated metadata.

3.1.3 All raw data and the associated metadata obtained as a result of proprietary access, 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 in order to guarantee confidentiality. Data from proprietary research will be removed after the experiment from the CELLS storage systems, unless otherwise agreed with the 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 the 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 the 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 from the ALBA Synchrotron will be archived. Data and metadata generated from the JEMCA facility will not 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 from the ALBA Synchrotron will be migrated or copied to archival facilities for long-term curation.

3.2.6 It is planned that each experiment from the ALBA synchrotron will have a unique persistent identifier. The plan includes that users will also be able to mint persistent identifiers to ensembles of datasets selected for publication. Anybody publishing results based on open access data must quote the same identifier (and related publications if available and required).

3.2.7 The ALBA synchrotron proposals' 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 provided through suitable list/search tools.

3.3.2 Access to data will be restricted to people who are registered as users. Any retrieval of data may require access to permanent storage, implying possible latencies..

3.3.3 The CELLS shall custody the data from the ALBA Synchrotron for a minimum of 5 years. On a best effort basis, the CELLS will keep the data accessible online for one year, after which it can be archived on a tape library or a warehouse, implying consequent latencies. In the case of the JEMCA facility, the corresponding data and metadata will be guaranteed to remain stored and accessible for a maximum period of sixty days.

3.3.4 Access to raw data and the associated metadata obtained from an experiment in the ALBA Synchrotron 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 the CELLS.

3.3.5 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. 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.

3.3.6 Authorized CELLS staff (e.g. instrument scientists, Computing group members) have access to any facility (the ALBA Synchrotron or the JEMCA) curated data or metadata for facility related purposes. The CELLS staff guarantees that it will preserve the confidentiality of data from the ALBA Synchrotron during the embargo period.

3.3.7 The on-line catalogue will enable linking experimental data from the ALBA Synchrotron to experimental proposals. Access to proposals will only ever be provided to the experimental team and appropriate CELLS staff, unless otherwise authorized by the PI during the 3-year embargo period.

3.3.8 The PI has the possibility to transfer parts of the totality of her/his rights during the embargo period to another registered person.

3.3.9 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.

4.2 Curation of results

4.2.1 The results of the analysis performed can be stored at the CELLS premises if the PI decides so. It will not be the responsibility of the CELLS to fully curate this data e.g. to ensure that software to read/manipulate this data is available.

4.2.2 The CELLS cannot be made liable in case of unavailability or loss of data, results or data analysis software.

4.3 Access to results

4.3.1 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.

4.3.2 Appropriate facility staff (e.g. instrument scientists, computing group members) has access to any facility-curated data or metadata for facility related purposes. The CELLS will guarantee that confidentiality of 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 The 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 the ALBA Synchrotron and the JEMCA must cite the persistent identifier of the experiment and data in their publication.

6.2 References for publications related to experiments carried out at the ALBA Synchrotron and the JEMCA must be deposited in the CELLS publications database within six months of the publication date, or during any new application for beamtime, whichever is earlier. Non-compliance to this rule may result on not getting further proposal time allocation.