30 November 2015

The ESRF Data Policy

The ESRF aims to implement a Data Policy starting as soon as possible in 2016. The main elements of this policy comprise:

- **Data ownership**
- **Data curation**
- **Data archiving**
- **Open access to data**

This policy follows largely the recommendations of the PaN-data Europe Strategic Working Group laying out a common framework for scientific data management at photon and neutron facilities (Deliverable D2.1, PaN-data Europe, co-funded by the European Commission under the 7th Framework Programme)

1. General Principles

1.1. The present data management policy pertains to the ownership of, the curation of and access to experimental data and metadata collected and/or stored at the ESRF.

1.2. Acceptance of this policy is a condition for the award of beam time.

1.3. Users must not attempt to access, exploit or distribute raw data or metadata unless they are entitled to do so under the terms of this policy.

1.4. Deliberate infringements of the policy may lead to denial of access to raw data or metadata and/or denial of future beam time requests at the ESRF, as well as actions of the ESRF in the court of law.

1.5. All data and metadata will be subject to the data protection legislation of France.

2. Definitions

For the purposes of this policy:

2.1. The term **raw data** pertains to data collected from peer-reviewed and in-house experiments performed on ESRF's instruments and includes data collected from peer-reviewed experiments performed on CRG beamlines. This definition includes data that are created automatically or manually by facility specific software and/or facility staff expertise in order to facilitate subsequent analysis of the experimental data.

2.2. The term **metadata** describes information pertaining to data collected from ESRF instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.

2.3. The term **principle investigator** (PI) pertains to the PI identified on the Experiment Proposal (for peer-reviewed experiments) or the Safety Approval Form (for in-house experiments).

2.4. 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.

2.5. The term **public research** refers to research done through peer review or access via in-house research beam time.

2.6. The term **proprietary research** refers to research done through purchased (commercial) access.

2.7. The term **on-line catalogue** pertains 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) web-based browsers.

2.8. The term **result** pertains to data, intellectual property, and outcomes arising from the analysis of raw data. This does not include publications.

2.9 The term **custodian** refers to the Institute storing, curating and providing access to raw data, metadata and results.

2.10. The term **long-term** means a minimum of 5 years and the ESRF will strive for 10 years. This will depend on the type and volume of data concerned and the economical consequences associated with long-term data storage. Thus the ESRF reserves the right to restrict the storage periods or data sets in consultation with the respective communities of high data rate instruments.

2.11. The term **open access** means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone. The ESRF 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 peer reviewed access to the ESRF, in-house research and use of Management Contingency beamtime excluding proprietary research 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 ESRF 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 research will be owned exclusively by the client who purchased the access and is not covered by the ESRF data policy. Data from proprietary research will be removed after the experiment from ESRF disk storage, unless otherwise agreed with ESRF management before the start of the experiment.

3.2 Curation of raw data and associated metadata

3.2.1. All raw data and metadata will be curated in well-defined formats, for which the means of reading the data will be made available by the ESRF.

3.2.2. Metadata that are 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 ESRF software 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 & 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 is foreseen to be via a searchable on-line catalogue.

3.3.2. Access to the on-line catalogue of the ESRF will be restricted to registered users of the on-line catalogue. The ESRF sets up the on-line procedure to become a registered user of the on-line catalogue.

3.3.3. Access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team for an embargo period of 3 years after the end of the experiment. Thereafter, the data will become openly accessible. Any PI that wishes data to retain *restricted access* for a period longer than three years will have this possibility by submitting a written request, specifying the reasons for the proposed prolongation, to the ESRF Directors of Research who decide on the request. In exceptional circumstances, data can be made openly accessible earlier than 3 years if the PI or the ESRF Directors of Research inform the ESRF to do so.

3.3.4. It is the responsibility of the PI to ensure that the experiment number is correctly entered into the metadata for each raw data set.

3.3.5. Authorized ESRF staff (e.g. instrument scientists, computing group members) have access to any curated data or metadata for facility related purposes. ESRF will undertake that confidentiality of such data is preserved during the embargo period.

3.3.6 The on-line catalogue will enable linking experimental data to experimental proposals. Access to proposals will only be provided to the experimental team and appropriate facility staff, unless otherwise authorized by the PI.

3.3.7. The PI has the possibility to transfer parts or 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.

4.2 Curation of results

4.2.1 The ESRF will provide curation of results on a best effort basis, and acts as custodian of results in the long term.

4.2.2. The ESRF cannot be made liable in case of unavailability or loss of data or results.

4.2.3. The ESRF cannot be made liable in case of unavailability or loss of data analysis software.

4.3 Access to results

4.3.1. Access to the results of analysis performed on raw data and metadata is restricted to the person or persons performing the analysis, unless otherwise requested by those persons. However, if the raw data being analysed is still restricted, access to the analysis results must be granted by the PI on request.

5. Good practice for metadata capture and results storage

5.1. The experimental team is encouraged to ensure that experiments metadata are as complete as possible, as this will enhance the possibilities for everybody to search for, retrieve and interpret the data in the long term.

5.2. ESRF provides means 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 to the raw data / metadata using the facilities provided by the on-line catalogue. Furthermore, they are encouraged to make such results openly accessible.

6. Publication information

6.1. Publications related to data from experiments carried out at ESRF must cite the persistent identifier of the experiment and data in their publication.