ESRF Data Policy 2024

Version 14/10/2023

The ESRF data policy covers the following topics:

- **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[1] and the PaNOSC data policy framework[2]. The main objective of this policy is to make ESRF open data FAIR: Findable, Accessible, Interoperable and Reusable.

The policy defined in this document applies to all data collected at ESRF in the frame of publicly funded research starting from 01/01/2024. A separate policy document is available for commercial access.

1. General Principles

1.1. This data policy governs the curation of and access to experiment data and metadata collected, uploaded and/or stored at the ESRF. This includes raw, processed and auxiliary data which have been curated at the ESRF.

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

1.3. Users shall 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 the initiation of legal actions by the ESRF in the courts of law.

1.5. All data and metadata curated at the ESRF will be subject to the data protection legislation applicable in France.

1.6. Users shall ensure that all data are collected with accurate metadata to adhere to FAIR principles. The ESRF will define a minimum subset of metadata which is collected automatically during the experiment.

1.7. Users shall endeavour to include auxiliary data to augment the experiment data.

[1] https://doi.org/10.5281/zenodo.3738497
[2] https://doi.org/10.5281/zenodo.3826039

1.8. The ESRF cannot be held liable in case of unavailability or loss of data or data analysis software.

2. Definitions

For the purposes of this policy, the following definitions apply:

2.1. The term **raw data** refers to the primary data collected from peer-reviewed and in-house experiments performed on ESRF's instruments, including 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 experiment 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 experiment team, experiment conditions, the sample, electronic logbooks and other logistical information.

2.3. The term **auxiliary data** refers to data that provide contextual information regarding the experiment, such as information regarding the sample (images, provenance and preparation) or the software used for data processing.

2.4. The term **processed data** refers to the data obtained by processing raw data.

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

2.6. The term **experiment team** includes the PI and any other person to whom the PI grants the right to access resultant the data and associated metadata.

2.7. The term **beam time** refers to the period of time when the experiment team has access to the facility resources to conduct an experiment.

2.8. The term **users** refers to anyone who has been a member of an experiment team at the ESRF, which make use of the ESRF's instruments.

2.9. The term **embargo** period refers to the period before the data is made publicly accessible.

2.10. The term **public research** refers to research conducted on ESRF's instruments through peer-reviewed or in-house research beam time.

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

2.12. The term **on-line catalogue** pertains to a computer database of metadata containing links to raw data files, which can be accessed by a variety of methods, including (but not limited to) web-based browsers.

2.13. The term **data** pertains to raw data (see 2.1), metadata (see 2.2), auxiliary data (see 2.3), and processed data (see 2.4). This does not include publications.

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

2.15. The term **open access** means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone.

2.16. The term **Digital Object Identifier** (DOI) refers to the persistent identifier system managed by the International DOI Foundation for unique digital identifiers of a wide range of objects including publications and data

3. Data and associated metadata

3.1 Access to raw data and associated metadata

3.1.1. The ESRF is the custodian of the data stored at the ESRF.

3.1.2. All data obtained as a result of peer-reviewed access to the ESRF, in-house research and use of Management Contingency beam time, excluding proprietary research, will be open access after an initial embargo period of three years after the end of the experiment, during which time access is restricted to the experiment team, represented by the PI. Thereafter, the data will become openly accessible.

3.1.3 The experiment report submitted by the user after each beam time session will be included as part of the auxiliary data, subject to the same embargo rules and three-year embargo period.

3.1.4. The embargo period begins at the end of the experiment session i.e. at the end of the scheduled beam time on the beamline.

3.1.5. Any PI wishing to restrict access to data for a period longer than three years is entitled to submit a written request, which should specify the reasons for the proposed extension. The request will be considered and decided upon by the ESRF Directors of Research.

3.1.6. Data can be made openly accessible earlier than three years by the PI.

3.1.7. ESRF will release open data under an open licence.

3.1.8. Any raw, processed, and auxiliary data stored by the ESRF will be archived for a duration of at least five years, and up to a maximum of ten years if possible. Metadata will be stored indefinitely. The ESRF reserves the right to restrict the storage period or data sets in consultation with the respective beamline responsible for high data rate instruments.

3.1.9. ESRF encourages users to triage all data (raw and processed) acquired during their experiments, by keeping only scientifically useful datasets at the end of the experiment, in order to limit the stored data volume. Automated triage can be put in place for high data rate instruments.

3.1.10. For high data volume experiments, lossy compression may be used to store data, if required because of raw data storage issues. Prior approval by the PI is encouraged.

3.1.11. Access to data, results and the associated metadata stored at the ESRF is via the ESRF data portal[3].

3.1.12. Authorized ESRF staff (e.g. beamlines scientists, computing group members, software engineers) have access to any curated raw data or metadata if required for IT infrastructure maintenance or upgrade. The ESRF reserves the right to use data still under embargo to improve facility processes and performance. ESRF will undertake that confidentiality of such data is preserved during the embargo period.

3.2 Curation of raw data and associated metadata

3.2.1. All raw data and metadata will be curated in well-known formats.

3.2.2. Experiment datasets are referenced by a unique persistent identifier based on the DOI system. Anybody publishing results based on data collected at ESRF must quote the DOI identifier.

3.2.3. 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 DOI landing page on the web.

4. Curation of processed data

4.1 Curation and access to processed data

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

4.1.2 When appropriate, the ESRF will store only processed data and not the raw data.

4.1.3 The ESRF will provide a means for the PI to upload processed data and associated metadata to the ESRF and enable them to associate these data with raw data collected at the ESRF.

4.1.4. The ESRF cannot be held liable in case of unavailability or loss of data.

4.1.5. The ESRF cannot be held liable in case of unavailability or loss of data analysis software.

4.2 Access to processed data

4.2.1. Access to the processed data of analysis performed on raw data and metadata uploaded to the ESRF data portal is governed in the same manner as raw data.

[3] https://data.esrf.fr