

## Data policy for PSI research data (the “Policy”)

The Policy defines the rules for the following topics:

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

### 1 General Principles

- 1.1. The present Policy pertains to the ownership of, the curation of and access to experimental data and Metadata collected and/or stored by PSI Research Infrastructure.
- 1.2. Acceptance of this Policy is a condition for the award of access to Research Infrastructures and a binding part of each PSI employment agreement.
- 1.3. Users must not attempt to access, exploit or distribute 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 actions concerning the employment of PSI employees to denial of access to data or Metadata and/or denial of future beam time requests at PSI, as well as to legal actions.
- 1.5. All data and Metadata will be subject to Swiss law.

### 2 Definitions

For the purposes of this Policy the term:

- 2.1 **Custodian** refers to the head of the institute divisions (or a delegated person) storing, curating and providing access to Raw Data, Metadata and Results.
- 2.2 **Experimental Team** includes the PI and any other person to whom the PI designates the right to access data and associated Metadata related to this proposal.
- 2.3 **Long-term** means a minimum of five (5) years and PSI will strive for ten (10) years, depending on the type and volume of data concerned and the economic consequences associated with Long-term data storage. Thus, PSI reserves the right to restrict the storage periods or data sets in consultation with the respective communities.

- 2.4 **Metadata** means information pertaining to data collected from Research Infrastructures at PSI, including (but not limited to) the scientific and administrative context of the experiment, the Experimental Team and the experimental conditions.
- 2.5 **On-line Catalogue** pertains to a computer database of Metadata containing links to data files, that can be accessed by a variety of methods.
- 2.6 **Open Access** means belonging to the public at large, unprotected by most copyrights or patents and subject to appropriation by anyone. Those data will be made available under CC-BY-SA (<https://creativecommons.org/licenses/by-sa/4.0/>).
- 2.7 **Principle Investigator (PI)** means the main proposer identified on the experiment proposal for peer-reviewed experiments or the leader of the Experimental Team for non-peer-reviewed experiments, both at the PSI large-scale facilities, or at other Research Infrastructures.
- 2.8 **Proprietary** research refers to research done through commercial access.
- 2.9 **Public Research** refers to research done through peer reviewed experiments and experiments done during in-house research or management contingency beam time.
- 2.10 **Research Infrastructure** includes but is not limited to PSI facilities SLS, SINQ, SpS, SwissFEL, and other PSI research infrastructures.
- 2.11 **Raw Data** means data collected from experiments performed at PSI. This 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.
- 2.12 **Result** pertains to data, intellectual property, and outcomes arising from the analysis of Raw Data (i.e. derived data).

Unless there is something inconsistent in the subject or context, words denoting the singular number include the plural and vice versa; words denoting one gender include the other gender and the neuter.

### 3 Raw Data and Metadata

- 3.1 Data resulting from Public Research and Proprietary Research
  - 3.1.1. All Raw Data and Metadata obtained as a Result of Public 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. PSI is the Custodian of the Raw Data and Metadata.
  - 3.1.3. All Raw Data and Metadata obtained as a Result of Proprietary Research will be owned exclusively by the client who purchased the ac-

cess and is not covered by this Policy. Data from Proprietary Research will be removed after the experiment from PSI storage, unless otherwise agreed with the PSI management before the start of the experiment.

### 3.2 Curation of Raw Data and 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 PSI.
- 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. Raw Data and Metadata will be read-only for the duration of their life time.
- 3.2.4. Raw Data and Metadata will be migrated or copied to archival facilities for Long-term curation.
- 3.2.5. 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).
- 3.2.6. High level Metadata such as title, authors, abstract, specific Research Infrastructure 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 in addition to a direct access protected file based access within the PSI network.
- 3.3.2. Access to the On-line Catalogue of PSI will be restricted to registered users of the On-line Catalogue. PSI sets up the on-line procedure to become a registered user of the On-line Catalogue.
- 3.3.3. Access to Raw Data and Metadata obtained from an experiment is restricted to the Experimental Team for an embargo period of three (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 (3) years will have this possibility on a yearly basis on a maximum prolongation of two (2) years. For longer extension a written request shall be submitted, specifying the reasons for the proposed prolongation, to the head of the corresponding PSI division who decides on the request, all at their sole discretion. Data can always be made openly accessible earlier on simple request of the PI. In exceptional circumstances the head of the corresponding PSI division can grant access to official committees at any time for the purpose of verifying data integrity.
- 3.3.4. Raw Data and Metadata explicitly used for peer-reviewed publication will become Open Access at the time of such publication.

- 3.3.5. It is the responsibility of the PI to ensure that the experiment number and requirements concerning the confidentiality, integrity and availability of the data is correctly entered into the Metadata for each Raw Data set.
- 3.3.6. Authorized PSI staff (including but not limited to facility management, instrument scientists, computing group members) have access to any curated data or Metadata for facility related purposes. PSI will undertake that confidentiality of such data is preserved during the embargo period.
- 3.3.7. The On-line Catalogue will enable linking experimental data to experimental proposals. Access to the full proposal text will only be provided to the Experimental Team and appropriate facility staff, unless otherwise authorized by the PI.
- 3.3.8. The PI has the possibility to transfer any or all of his rights during the embargo period to another person. This transfer has to be documented.
- 3.3.9. The PI may to create and distribute copies of the Raw Data, without violating against the rules that apply during the embargo period.
- 3.3.10. To the extent permitted by law, PSI cannot be held liable in the case of unavailability or loss of Raw Data.

## 4 Results

### 4.1 Ownership of Results

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. PSI will provide curation of Results on a best effort basis, and acts as Custodian of Results in the long term.
- 4.2.2. PSI cannot be held liable in case of unavailability or loss of data or Results.
- 4.2.3. PSI cannot be held liable in case of unavailability or loss of data analysis software.

### 4.3 Access to Results

Access to the Results of analysis performed on Raw Data and Metadata, is restricted to the person performing the analysis, unless otherwise requested by that person. However, if the Raw Data being analysed is still restricted, access to the analysis results must be granted by the PI on request.

## 5 Leading practice for Metadata capture and Results storage

- 5.1. The Experimental Team shall 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. PSI 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 carry out analyses of Raw Data and Metadata which are openly accessible shall, to the extent practicable, contact the original PI to inform him and suggest a collaboration if required. Researchers must acknowledge the source of the data and cite its unique identifier as well as the original publication linked to the same Raw Data.
- 5.4. PI and researchers who carry out analyses of Raw Data and Metadata shall link the Results of these analyses to the Raw Data / Metadata using the facilities provided by the On-line Catalogue. Furthermore, PI and researchers shall make such Results openly accessible.

## 6 Publication information

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

### Implementation

The data policy for PSI research data will be implemented over time, depending foremost on our capacity to tag raw data with meaningful Metadata. A Metadata catalogue will be installed. A process document describes the details.

---

<sup>1</sup> <http://pan-data.eu/PaNdataEurope>