Change History

Version	Datum	Änderung	Autor
0.0	25.07.2016	Final Draft	Mirjam van Daalen
0.1	11.08.2016	Formated as PSI «Weisung»	Tanja Wagner, Gerd Mann
0.2	02.11.2016	Finished document	Mirjam van Daalen
0.3	16.11.2021	General revision and update of policy	Gerd Mann
0.4	22.11.2021-10.01.2022	Consolidation of input	Alex Amato, Alun Ashton, Oliver Bunk, Michèle Erat, Gerd Mann
0.5	04.02.2022	Final review	Alex Amato, Markus Ammann, Alun Ashton, Oliver Bunk, Gregor Cicchetti, Michèle Erat, Didier Gavillet, Michel Kenzelmann, Klaus Stefan Kirch, Joachim Kohlbrecher, Gerd Mann, Andreas Suter
0.5.u	26.04.2022	Final adaption to user facilities	Alex Amato, Alun Ashton, Oliver Bunk, Michèle Erat, Stefan Janssen

PSI Data Policy

public version

for the external users of the
PSI Large Scale Facilities

SLS, SwissFEL, SINQ, SμS, CHRISP

Table of Contents

1 Background, Purpose and General Principles

1.1 Background and Purpose

Data are a valuable and essential product and resource for research conducted at PSI. Especially the large research facilities operated at PSI are a source for large data sets relevant for national and international research teams. Therefore, research data management (RDM), i.e. organization, storage, preservation, and sharing of data collected and used in research projects is pivotal. Research Data Management contributes to the efficient use of resources, reproducibility of research, accessibility of data, good scientific practice and research integrity.
In addition, there are growing requirements imposed by funders and publishers to make research data available as Open Research Data (ORD) in accordance with FAIR (cf. Appendix 4.1) principles. PSI supports the Swiss National ORD strategy[1]. It collaborates with national and international initiatives and projects to further promote and develop RDM and ORD.

This document defines the general principles for research data management at PSI and its user facilities. A glossary with definitions can be found in the Appendix.

1.2 General Principles

a) This document pertains to the ownership, curation and access to research data and metadata generated, collected and/or processed and/or stored by PSI or at its facilities (unless exempt or covered by other agreements).
b) This document is binding for all PSI employees and external users of PSI facilities. It applies to all activities concerning supervision, coordination or execution of research with public funding (unless exempt or covered by other agreements).
c) Acceptance of this policy as far as applicable is a condition for the award of access to research infrastructures for internal and external users. The relevant parts of this document are published on the PSI internet and thus available to all users.
d) In addition to this policy, legal, regulatory or contractual requirements may be relevant (e.g. federal Act on Data Protection) and superseding. This includes requirements from national or international funding agencies (Swiss National Science Foundation, European Commission etc.).
e) Proprietary research activities (e.g. funded by commercial entities) must adhere to the contractual agreements entered into by PSI, with particular care being taken with regard to non-disclosure agreements and publication embargo clauses. In all other aspects, the terms of this document are mandatory. In case of doubt, the terms of the commercial agreement shall take priority.
f) Under the umbrella of this document, PSI organizational entities can define additional rules and requirements for RDM and ORD, addressing specific circumstances and needs.
g) No one may or attempt to access, exploit or distribute research data or metadata unless they are entitled to do so under the terms of this document.
h) Deliberate infringements of rules laid out herein may lead to denial of access to data or metadata and/or denial of access to research infrastructures at PSI, to legal actions as well as to disciplinary actions for PSI employees or external users.

[1] https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/ORD/Swiss_National_ORD_Strategy_en.pdf

2 Research Data Management

2.1 Data Ownership

a) By default, the PI is owner of Research Data. The PI can transfer the ownership to another person. Transfer of ownership must be documented. If a PI is not named, the main author of a proposal is considered as PI.
b) In case the data owner cannot be identified or contacted, the Data Steward decides on the ownership.

2.2 Data Management Plans

a) PSI encourages all its researchers and external users to include RDM in the planning of their activities. The definition and implementation of Data Management Plans (DMP) is recommended.
b) Researchers, in particular the PIs, are responsible for fully complying with the RDM obligations imposed by the funding institution when accepting a research grant or entering a contract. With regard to the present policy, this applies primarily to any obligation regarding the definition, maintenance and implementation of a DMP[2].

2.3 Data Acquisition and Processing

a) All activities concerning data management and processing must be in accordance with the PSI instruction on Research Integrity (VA-6213-302[3]).
b) It is the responsibility of the PI to ensure all relevant Research Data and Meta Data are recorded and stored in an appropriate repository. In case the repository is not based on application services provided or supported by the PSI Data Custodian, it is the obligation of the PI to ensure third party repositories fulfill all applicable requirements as defined in this document.
c) Especially for experiments generating large data sets the PI must ensure that it is defined which data are considered as Raw Data and which data are considered as Research Data and how these data should be managed to fulfill the applicable RDM and ORD requirements. To make best use of its infrastructure and to save cost, PSI reserves the right to apply Data Reduction on all Research Data generated at PSI. For data reduction, the focus is on data streams and data sets with high and very high data throughput and/or data volumes. PSI provides information to PIs, researchers and other interested parties about methods and tools used for Data Reduction.
d) All Research Data and Metadata should be stored and archived in well-defined and documented data formats to ensure data can be accessed and processed. The use of non-propriety data formats is recommended.
e) The acquisition and documentation of personal data shall adhere to the principles of data minimization (collect only data relevant for specific research questions, not as much as possible) and privacy by design (implement data protection at the earliest possible project stage).
f) Where applicable, the use of Electronic Lab Journals is recommended.

[2] https://www.snf.ch/en/FAiWVH4WvpKvohw9/topic/research-policies

[3] https://www.psi.ch/en/media/66107/download?attachment

g) Each data collection recorded in the PSI data catalogue will receive a unique persistent identifier when made publicly accessible. Anybody publishing results based on such data must quote the identifier (and related publications if available & applicable).
h) Research Data and Metadata that have been archived will be read-only for the duration of their lifetime.
i) For the large research facilities, PSI strives for automated processes to support acquisition and recording of Research Data and Metadata.
j) Processed data obtained by artificial intelligence and other methods that do not allow a tracing of the original data nor the exact replication of the calculation need to provide plausible explanations according to which underlying rules data was obtained.

2.4 Access to and Publication of Research Data

a) All Research 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. Exceptions are defined in paragraph f) of this chapter.
b) High level Metadata such as title, authors, abstract, specific Research Infrastructure used are made public as soon as the experiment has been carried out at PSI's large research facilities.
c) The embargo period is three (3) years, starting at the end of data collection as agreed between the Data Steward and PI. Thereafter, the data will become openly accessible. On written request of the PI the embargo period can be shortened or omitted.
d) The embargo period can be extended twice by a single year on request of the PI, up to a total period of five (5) years. Requests must be submitted to the Data Custodian and the responsible Data Steward. Extensions of the standard embargo period of three (3) years must not be in conflict with legal or regulatory requirements or the rules of the funding agency.
e) For embargo periods exceeding five (5) years, a written request must be submitted, specifying the reasons for the requested prolongation, to the Data Steward who decides on the request. The data custodian must be informed about the decision.
f) Access to data and code can be restricted beyond the initial embargo period in case of:
    i. ethical and legal constraints, in particular regarding personal data which cannot be anonymized to a sufficient degree and medical data,
    ii. security constraints including possible code abuse,
    iii. ongoing preparation of a patent application,
    iv. contractual obligations e.g. to industry partners,
    v. intended commercial use of a software linked to a data set,
    vi. intended use of data for commercial purposes or
    vii. other special circumstances (to be confirmed by data steward).
g) Researchers must not hand over exclusive rights to the publication and further use of data to repositories. PSI strongly recommends the use of Open Access Repositories.
h) If processes are not automated, it is the responsibility of the PI to ensure that a unique experiment ID and requirements concerning the confidentiality, integrity and availability of the data is correctly entered into the Metadata for each Raw Data set.
i) PIs may transfer any or all of their rights during the embargo period to another person. This transfer has to be documented.
j) When Program Code is required to derive published results from the associated data, it is recommended that the relevant source code and/or machine code is stored along with the data or referenced.
k) It is the obligation of the Data Custodian to ensure that appropriate identification and access control mechanisms are implemented to limit access to Research Data to authorized persons.

l) 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 and adherence to good scientific conduct.
m) Authorized PSI staff (including but not limited to facility management, instrument scientists, computing group members) have access to any Research Data or Metadata for facility related purposes (e.g. troubleshooting). PSI will ensure that confidentiality of such data is preserved during the embargo period.

2.5 Data Retention, Data Disposal and Cost for Data Archiving

a) Services for data archiving provided by PSI and/or its partners aim at data retention periods of 10 years.
b) On request of the PI or the Data Steward the Data Retention period can be prolonged for an additional period of 5 or 10 years.
c) On request of the PI or the Data Steward Research Data and Meta Data can be stored with redundancy (to minimize the risk of data loss due to technical failure, disasters etc.) or without redundancy.
d) After the end of the data retention period data will be disposed by the Data Custodian. Before data are disposed, the PI and the Data Steward have to be informed in advance and they can execute a veto right, i.e. prolong the data retention period for another period of 10 years.
e) Cost for data curation services will be covered by PSI IT as part of general IT services.
f) Cost related to data volumes, data retention periods, prolongation of data retention periods, and redundancy must either be covered by the Data Steward or the Data Steward must define the party bearing the cost (e.g. the PI owning the data, via the grant application).
g) Details on cost are published as appropriate.

3 Liability

a) All services related to the acquisition, transportation, access, processing, storage, archiving and disposal of Research Data, Metadata and/or Program Code provided by PSI and/or its partners are on a best effort basis.
b) To the extent permitted by law, PSI cannot be held liable in the case of unauthorized access, unavailability or loss of Research Data, Metadata and/or Program Code.

4 Appendix

4.1 Definitions

- **Data Custodian** denotes a role responsible for technical environment and IT services to ensure the safe custody, transport, storage, archiving of data and implementation of business rules (for details see chapters 2.4, 2.5).
- **Data Management Plan (DMP)** is used to plan the life cycle of research data. It is designed for the long term and sets out how the data should be produced, collected, documented, published and archived.
- **Data Reduction** refers to a set of methods and procedures aiming at reducing the capacity required to transfer, handle and to store data. Examples of data reduction methods are compression and

selection. Data reduction methods can be lossless (original data set can be reconstructed) and lossy (original data set cannot be reconstructed, with a possible degradation of data as a consequence).

- **Data Steward** denotes a role responsible for data content, context, and associated business rules (for details see chapters 2.4, 2.5).
- **Electronic Lab Journal** denotes a computer program used by scientists, engineers, and technicians to document research, experiments, and procedures performed in a laboratory.
- **FAIR Guiding Principles** for scientific data management and stewardship require that data (and other digital assets) are findable, accessible, interoperable, and reusable.
- **Metadata** is data that provides information about other data. Examples are the scientific and administrative context of an experiment, the experimental team or the experimental conditions.
- **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/).
- **On-line Catalogue** pertains to a computer database of Metadata containing links to data files that can be accessed by a variety of channels (e.g. European Open Science Cloud).
- **Principle Investigator** (PI) means the main proposer identified on a research proposal or the leader of the research team. A PI can be a PSI employee or an individual not employed by PSI.
- **Program Code** is defined as the instructions that are created in the course of software development for a specific computer program or a part thereof and that describe or represent its functionality in a specific programming language.
- **Proprietary Research** is defined as that for which users request confidentiality of proposal, data and results for a certain period of time. It is often linked, but not limited to research funded by commercial entities.
- **Public Research** refers to research funded or supported by public entities (e.g. research institutions, national and international research councils, public innovation agencies or government departments).
- **Raw Data** means data collected from experiments. This includes data created automatically or manually by software, staff or external users in order to facilitate subsequent analysis of the experimental data (see the note at end of this chapter).
- **Research Data** are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others (see the note at end of this chapter).
- **Research Data Management** refers to the set of all methodological, conceptual, organizational and technical measures and procedures for organization, storage, preservation, and sharing of data collected and used in research projects over its life cycle.
- **Research Infrastructure** includes but is not limited to PSI large research facilities, and other PSI research infrastructures.
- **Research Team** includes the PI and any other person to whom the PI designates the right to access data and associated Metadata related to the research project of the PI. Research Team members can be PSI employees or individuals not employed by PSI.
- **Result** pertains to data, intellectual property, and outcomes arising from the analysis of Research Data (see the note at end of this chapter).

Note: Definitions of Raw Data, Research Data and Results may depend on the research topic and the context. If the above definitions are not appropriate or sufficiently specific, then they can be adapted in project or topic-specific data management plans.