

SOLEIL DATA MANAGEMENT POLICY

Date of circulation	Editor	Auditors	Approver	Changes	Index
02/10/2018	B.Gagey	SOLEIL Management SOLEIL Legalist Users' Representatives	J.Daillant		0

PUBLIC

Circulation: For action / For information

Recipients: SOLEIL Users, SOLEIL Staff

SOLEIL DATA MANAGEMENT POLICY

TABLE OF CONTENTS

1. GENERAL PRINCIPLES	3
2. DEFINITIONS.....	4
3. OWNERSHIP	6
4. ROLES AND RESPONSIBILITIES	6
5. DATA STORAGE AND CURATION	7
6. DATA ACCESS AND REUSE	8
7. PUBLICATION INFORMATION.....	9
8. TERMINATION OF CUSTODIANSHIP OR ON-LINE CATALOGUE	9

SOLEIL Data Management Policy

2 October 2018

As « Modern Science builds on extensive scientific dialogue and advances by improving earlier work »¹, SOLEIL is encouraging its Users and Scientists to produce FAIR (Findable, Accessible, Interoperable, Reusable) data. SOLEIL aims to facilitate this with data management services for storage and retrieval of experimental data and associated metadata collected and/or stored at SOLEIL. This also implies the definition of a framework to access these data and metadata.

This is the purpose of the present data management policy. It is based on a model for scientific data management at photon and neutron facilities formulated by the PaNdata project², co-funded by the European Commission under the 7th Framework Program, and on its application at other facilities. Adapted to SOLEIL with the help of science data experts from CNRS, it will be implemented starting 2018.

1. GENERAL PRINCIPLES

- 1.1. The present data management policy pertains to the ownership of, the curation of, and access to experimental data and associated metadata collected, reduced, processed and/or stored at SOLEIL.
- 1.2. SOLEIL may use subcontractors to perform its obligations under this data management policy. These subcontractors act as data processors in case personal data are concerned.
- 1.3. Acceptance of this policy is a condition for beamtime allocation.
- 1.4. 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.
- 1.5. Deliberate infringements of the policy may lead to denial of access to data or metadata and/or denial of future beamtime requests at SOLEIL.
- 1.6. If, and to the extent that, data and metadata include personal data, the data protection legislation of France and the European Union, respectively, will be applicable. This data management policy will be governed by and constructed in accordance with the French law.
- 1.7. Users of the SOLEIL facility must be personally registered via the SOLEIL user portal ([SUN set](#)). Each User performing an experiment at SOLEIL must agree and follow the SOLEIL User Charter Rules.

¹ Cf. [Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020](#), European Commission, March 2017

² Cf. [PaNdata Europe](#). Common Data Policy Framework on Scientific Data, December 2010

2. DEFINITIONS

For the purposes of the present data management policy:

- 2.1. The term **experimental data**, see Figure 1, pertains to data collected from experiments performed on SOLEIL instruments. This definition includes (but is not limited to) data that are created automatically or manually by facility specific software and/or facility staff expertise to facilitate subsequent analysis of the experimental data.
- 2.2. The term **raw data**, see Figure 1, pertains to the experimental data that is recorded during experiments, as produced by the detection system, and cannot be derived from other persistent data.
- 2.3. The term **reduced data**, see Figure 1, pertains to the experimental data that is derived from raw data through pre-processing during experiments including (but not limited to) formatting and qualifying raw data and helping to decide on the continuation of the experiment.
- 2.4. The term **processed data**, see Figure 1, pertains to the experimental data that is derived from raw data along the analysis steps.
- 2.5. The term **results**, see Figure 1, pertains to a subset of processed data and other outcomes arising from the analysis of experimental data, excluding publications based on such analysis and intellectual property (IP) rights.
- 2.6. The term **metadata**, see Figure 1, describes information pertaining to data collected from SOLEIL instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.

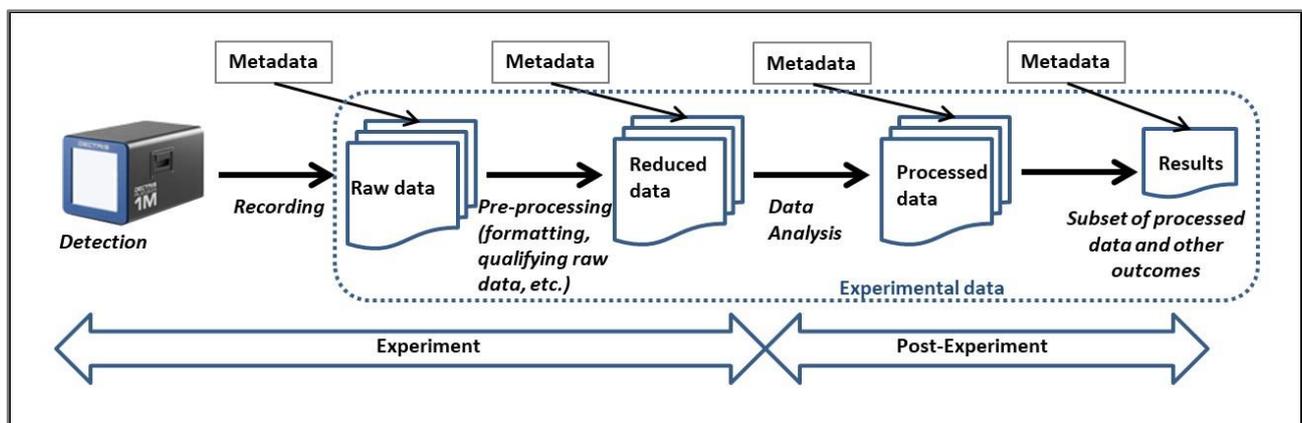


Figure 1 : Simplified illustration of Data Classes

- 2.7. The term **beamtime** means the time period, allocated by SOLEIL to the experimental team, for performing scientific experiments on a SOLEIL instrument. Every beamtime has a well-defined start date, end date, and experimental team.
- 2.8. The term **cycle of beamtime allocation** pertains to the time period starting from one call of proposals to the next one, typically six months.
- 2.9. The term **standard proposal** pertains to a proposal for an experiment running on one cycle of beamtime allocation.
- 2.10. The term **long-term proposal** pertains to a proposal for an experiment running over several cycles of beamtime allocation.
- 2.11. The term **block allocation group proposal** (BAG) pertains to a proposal for multiple experiments submitted by local, national or international consortia and running over two or

more cycles of beamtime allocation. Each BAG is composed of a BAG coordinator (drawn from the team members of the BAG) and several teams, each led by a **principal investigator (or PI)**.

- 2.12. The term **main proposer (MP)** pertains to the main proposer identified on a standard or long-term proposal or to the BAG coordinator identified on a BAG proposal. The MP is the person who submits the proposal and who is the manager of the proposal for any matter. He/she is also the contact person to SOLEIL for any administrative issue related to the proposal.
- 2.13. The term **experimental team** includes the MP and all co-proposers of a given proposal, all participants in the experiment(s) performed for the proposal (including the SOLEIL local contact(s)), and any other person to whom the MP designates the right to access resultant experimental data and associated metadata.
In the case of a BAG, each experimental team includes the PI and all co-proposers of the experiment, all participants in this experiment (including the SOLEIL local contact(s)), and any other person to whom the PI designates the right to access resultant experimental data and associated metadata.
- 2.14. The term **proprietary research** refers to research done through purchased (commercial) access.
- 2.15. The term **public research** refers to research done through peer reviewed access (e.g. the SOLEIL proposal system), or via in-house research beamtime, or use of Rapid Access beamtime but excluding proprietary research and some specific agreements.
- 2.16. The term **on-line catalog** 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.
- 2.17. A **registered user** refers to any person who has been granted a personal [SUN set](#) account, which can be requested via the login web page of the SOLEIL User portal.
- 2.18. The term **custodian** refers to the Institute storing, curating and providing access to raw data, metadata and results.
- 2.19. The term **long-term** means up to five years and SOLEIL will strive for 10 years starting with the end of the respective beamtime. The precise duration will depend on the beamline used, the type and volume of data concerned, and the economic consequences associated with long-term data storage. Thus, SOLEIL reserves the right to restrict the storage periods or data sets in consultation with the respective communities for high data rate instruments. Long-term durations applicable at each beamline or end station will be communicated on the SOLEIL user portal ([SUN set](#)), and reminded to the MP on acceptance of the proposal and before end of long-term duration.
- 2.20. The term **open access** means belonging to the community at large, unprotected by copyright or patent and subject to appropriation by anyone. Data held in long-term storage will be made available under the CC-BY license ³.
- 2.21. The term **embargo period** means the period, starting from the end of the respective beamtime, during which access to the collected raw data and associated metadata is restricted to the experimental team. This period will be 3 (three) years by default with the possibility of up to 5 (five) years, as detailed in 6.5.
- 2.22. The term **Data Management Plan**, or DMP, refers to the document describing the data management life cycle for the data to be collected, processed and/or generated by a research project: “the handling of research data during and after the end of the project, what data will be collected, processed and/or generated, which methodology and standards will be applied,

³ Cf. (<https://creativecommons.org/licenses/by/4.0/>).

whether data will be shared/made open access, and how data will be curated and preserved (including after the end of the project)⁴. Such a document may be required by external funders, research institutions and other organizations.

3. OWNERSHIP

- 3.1. All raw data and the associated metadata obtained via public research conducted at SOLEIL will be open access after an embargo period.
- 3.2. All raw data and the associated metadata obtained via proprietary research conducted at SOLEIL will be owned exclusively by the client who purchased the access and is not covered by the present data management policy. Data from proprietary research will be removed after the experiments from SOLEIL storage, unless otherwise agreed with SOLEIL management before the start of the experiment.
- 3.3. In some specific cases (as a national security related experiment, a classified experiment carried out by a ZRR⁵ laboratory, etc.), SOLEIL and the organization carrying out the experiment may agree on a specific data management policy. The MP is invited to mention it when submitting the proposal, and to contact SOLEIL Management at the acceptance of the proposal.
- 3.4. Ownership, including any IP rights that might arise, of all results derived from the analysis of the experimental data is determined by the contractual obligations of the person(s) performing the data analysis and/or data interpretation respectively the applicable law.

4. ROLES AND RESPONSIBILITIES

- 4.1. Unless otherwise specified, SOLEIL will act as a custodian for raw data and associated metadata.
- 4.2. SOLEIL aims at providing means for reduction and/or processing of raw data.
- 4.3. The MP is required to establish a Data Management Plan. SOLEIL can bring help by providing a template or by sending out the available information to the MP if the funder or research institution or other organization has particular data management requirements.
- 4.4. It is the responsibility of the MP (and/or each PI in a BAG) to ensure that the experiment number is correctly entered into the metadata for each raw data set, in order to correctly associate each data set with the MP and, if relevant, with the PI. If this is not done, the experimental team will not be able to access the data via the on-line catalog (6.1) or other users may inadvertently be given access rights to the data. Normally this will be done by simply entering the experiment number via the data acquisition software.
- 4.5. The MP (or each PI in a BAG) has to ensure sample descriptions are included in the metadata.
- 4.6. The experimental team is encouraged to ensure that 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.
- 4.7. SOLEIL undertakes to provide means for the capture of such metadata items that are not automatically captured by an instrument, including the Data Management Plan, to facilitate recording the fullest possible description of the raw data.

⁴ Cf. [Guidelines on FAIR Data Management in Horizon 2020](#), European Commission, July 2016, p. 4

⁵ ZRR = « Zone à Régime restrictif », cf. <http://www.sgdsn.gouv.fr/missions/protection-du-potentiel-scientifique-et-technique-de-la-nation/>

- 4.8. It is the responsibility of the MP (or each PI in a BAG) to ensure that any information, including personal data, added by the experimental team into the metadata, comply with the French and European data protection regulations. SOLEIL is not responsible for their compliance with data protection legislation. SOLEIL can bring help by providing available information.
- 4.9. It is the responsibility of the MP to ensure that the experimental team agrees on any change of the embargo period (6.5).
- 4.10. It is the responsibility of the MP (and/or each PI in a BAG), to manage possible authorizations of access to the data, granted to persons from outside the designated experimental team, during this embargo period (6.4) and in conformity with the legislation.
- 4.11. The MP (and/or each PI in a BAG) has the possibility to transfer or grant parts of all his/her rights during the embargo period to another registered person.
- 4.12. The MP (or each PI in a BAG) has the obligation to recover and to save all the data issued from the experiment within the duration specified at the acceptance of the proposal. SOLEIL can advise him/her to find an appropriate data repository for very long-term preservation.
- 4.13. Should a MP (or a PI) no longer be able to ensure his/her role, he/she may inform SOLEIL management and the experimental team, who will together organize a replacement, if possible.
- 4.14. SOLEIL will, at its own discretion, make its best to ensure an accurate storage and curation as well as an uninterrupted access to experimental data and metadata, during the duration specified at the acceptance of the proposal. However, failures caused by technical or human mistakes cannot be ruled out regarding any data processing. SOLEIL cannot guarantee an absolutely accurate storing and curating. Also, access might be temporarily limited or impossible, especially due to necessary maintenance or overhaul of services or failure of third-party service providers.
- 4.15. SOLEIL and MP (or each PI in a BAG) cannot be made liable in case of loss or defect of data, metadata or results, as well as for access being limited or unavailable.
- 4.16. SOLEIL and MP (or each PI in a BAG) cannot be made liable in case of unavailability or loss of data analysis software.
- 4.17. SOLEIL cannot be made liable for the consequences of any interpretation of the data.
- 4.18. As a matter of precaution (and without prejudice to the question of ownership) all members of the experimental team grant SOLEIL the unlimited and unrestricted right to use the experimental data and metadata to the extent necessary to curate them and make them available in accordance with this data management policy.

5. DATA STORAGE AND CURATION

- 5.1. All raw data will be curated in well-defined formats, for which the means of reading the data will be made available by SOLEIL.
- 5.2. Metadata, whether automatically captured by instruments or manually recorded with SOLEIL tools, will be curated either within the raw data files, within an associated on-line catalog, or within both.
- 5.3. Raw data and metadata will be read-only for the duration of their life-time.
- 5.4. Raw data and metadata will be migrated or copied to long-term storage facilities that could be sub-contracted, for long-term curation.
- 5.5. It is planned that each experiment and data set will have a unique persistent identifier, so that they will be uniquely attributable.

- 5.6. Unless otherwise specified, reduced data and the associated metadata will not be curated for long-term by SOLEIL. The tools for regenerating them from raw data will be made available by SOLEIL on a best effort basis (via recent versions of supported tools running on recent operating systems).
- 5.7. Unless otherwise specified, the processed data from the interim analysis steps and the associated metadata will not be curated long-term by SOLEIL.
- 5.8. Results issued from analyses performed on raw data and metadata using SOLEIL means will be stored long-term by SOLEIL in case of in-house research, and only on a best effort basis in other cases. It will not be the responsibility of SOLEIL to fully curate this data e.g. to ensure the semantic of this data or that software to read / manipulate this data is available.

6. DATA ACCESS AND REUSE

- 6.1. Access to raw data, metadata and results stored by SOLEIL will be available via a searchable on-line catalog, under the conditions described below.
- 6.2. Access to the on-line catalog will be restricted to registered users.
- 6.3. Download of data may be subject to restrictions of the volume which SOLEIL can provide. Download of data will be logged and the information made available to the MP (or PI in a BAG), within the limits of the legislation.
- 6.4. During the embargo period, access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team, except specific transfer or grant of rights is granted by the MP (or PI in a BAG).
If people from outside the designated experimental team request access to the data, SOLEIL will transfer the request to the MP (or PI in a BAG), who will examine the request in coordination with the experimental team.
- 6.5. After the embargo period, all raw data and the associated metadata will become openly accessible.
At least two months before end of embargo period, SOLEIL will remind the deadline to the MP. Any MP that wishes data to retain restricted access for a period longer than three (3) years (for example to align with the practices of the scientific disciplines to which the proposal relates) will have this possibility on a yearly basis, with justification (via the [SUN set](#) tool), up to a maximum prolongation of two (2) years. For longer extensions, a written request shall be submitted to SOLEIL Scientific Directors, specifying the reasons for the proposed prolongation, who decide on the request. In exceptional circumstances, data can be made openly accessible earlier than the initial embargo period if the MP requests SOLEIL to do so.
- 6.6. Access to results issued from analyses 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 analyzed is still restricted, access to results must be granted by the MP (or PI in a BAG) on request.
- 6.7. Authorized SOLEIL staff (e.g. instrument scientists, IT staff) have access to any curated data or metadata for facility related purposes. SOLEIL will undertake that confidentiality of data with restricted access is preserved.
- 6.8. The on-line catalog 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 requested by the MP (or PI in the case of a BAG).
- 6.9. A limited subset of metadata will be made public immediately on completion of experiments. This information will be available via the unique persistent identifier landing page on the web.

- 6.10. SOLEIL strongly recommends that researchers, who aim to carry out analyses of raw data and metadata which are openly accessible, should, when possible, contact the original MP (or PI in a BAG), to inform him/her/them, suggest collaboration and propose to be co-author in any future publication arising from the data, if appropriate.
- 6.11. Anybody publishing results based on open access data must acknowledge the source of the data, cite its unique persistent identifier and any publications linked to the same raw data or the origin of the raw data. Furthermore, SOLEIL encourages making such results openly accessible.
- 6.12. Allocation of SOLEIL beamtime commits the MP (and/or each PI in a BAG) to respond to requests about data obtained from the experiment once openly accessible.

7. PUBLICATION INFORMATION

- 7.1. Publications related to data from experiments carried out at SOLEIL must cite the unique persistent identifier of the experiment and data in their publication.
- 7.2. References for publications related to experiments carried out at the facilities must be deposited in the publications database (function "Register Publications" of the User menu in the [SUN set](#)) as soon as the article is published. Failure in providing publication reference can lead to a refusal of new beamtime allocation.

8. TERMINATION OF CUSTODIANSHIP OR ON-LINE CATALOGUE

- 8.1. If SOLEIL decides to not continue to act as custodian and/or to maintain and provide the on-line catalogue, SOLEIL will inform the MPs (and PIs) concerned in a timely manner and provide them with effective means to make a copy of the respective raw data, metadata, and results, provided SOLEIL is aware of the correct email address of the MP (or PI in a BAG) at that time.