

Workpackage 7

Data Analysis and Management

SESAME Data Policy document

D 7.2

August 2020



PROJECT DETAILS

PROJECT ACRONYM

BEATS

PROJECT TITLE

BEAMline for Tomography at SESAME

GRANT AGREEMENT NO:

822535

THEME

START DATE

01/01/2019

DELIVERABLE DETAILS

WORK PACKAGE: 07

EXPECTED DATE: 31/08/2020

WORK PACKAGE TITLE: DATA HANDLING AND MANAGEMENT

DELIVERABLE TITLE: SESAME DATA POLICY DOCUMENT

WORK PACKAGE LEADER: THE CYPRUS INSTITUTE

DELIVERABLE DESCRIPTION: REPORT

DELIVERABLE ID: D7.2

**PERSON RESPONSIBLE FOR THE DELIVERABLE:
CHARALAMBOS CHRYSOSTOMOU**

NATURE

R - Report P - Prototype D - Demonstrator O - Other

DISSEMINATION LEVEL

P - Public
 PP - Restricted to other programme participants & EC:
 RE - Restricted to a group
 CO - Confidential, only for members of the consortium

REPORT DETAILS

VERSION: 1

DATE: 31/08/2020

NUMBER OF PAGES: 6

DELIVERABLE REPORT AUTHOR(S): CH. CHRYSOTOMOU FOR MORE INFO PLEASE CONTACT: C.CHRYSOSTOMOU@CYI.AC.CY

STATUS

Template Draft
 Final Released to the EC

CONTENTS

Description of the deliverable4

ATTACHMENT A: SESAME DATA POLICY DOCUMENTS

DESCRIPTION OF THE DELIVERABLE

Work Package 7 (WP7) is responsible for the choice and design of hardware and software solutions implementing the workflow for BEATS, SESAME's tomography beamline. This includes data acquisition, processing, analysis, treatment, and storage. The tasks of this work package are challenging since X-ray tomography is among the most demanding techniques in terms of computing power and data storage. The approach followed for this work package is based on the replication and adaptation of hardware and software solutions already implemented at established tomography beamlines of European based partner light sources, and in particular, the Paul Scherrer Institut (PSI) and the European Synchrotron Radiation Facility (ESRF). The design of a computing infrastructure must consider the prospected scientific case of the beamline as well as the anticipated number of experiments per year.

More specifically, WP7 is responsible for creating the first version of SESAME data policy and for its endorsement by the SESAME Council. The purpose of the SESAME data policy is to provide users researching at SESAME with information and guidance on subjects such as experimental data ownership, storage, access and management, and to ensure that experimental data is managed and used in ways that maximize public benefit. Furthermore, the purpose of the document is to provide a Policy regarding the management of in-house experiments. The data policy has been designed to follow SESAME's existing policies along with other data policies implemented in neutron scattering and synchrotron X-ray facilities, particularly those in the European region, and so contribute to the fostering of a common framework for the management of experimental data.

Task 1 of the BEATS WP7 (Months 1 - 16) was dedicated to the preparation and endorsement of the SESAME data policy. The result of this exercise is the SESAME Data policy document contained in this deliverable, which was endorsed in its final version by the SESAME Council on the 8th of June 2020.

ATTACHMENT A: SESAME DATA
POLICY DOCUMENT



SESAME

The Synchrotron-Light for Experimental Science and Applications in the Middle East

SESAME/TELECONFERENCE/36-C/20-03

SESAME EXPERIMENTAL DATA MANAGEMENT POLICY

CONTENTS

1.	Policy Statement	1
2.	General Principles	1
3.	Definitions	1
4.	Raw data and associated metadata	3
5.	Results	4
6.	Good practice for metadata capture and the storage of results	5
7.	Publication information	5

1. Policy statement

SESAME is committed to ensuring that there is transparency in the manner in which Experimental Data produced using its facilities is owned, stored, accessed and managed.

The purpose of this Policy is to provide users conducting peer-reviewed and Director Discretionary time research with information and guidance on experimental data ownership, storage, access and management, and to ensure that experimental data is managed and used in ways that maximise public benefit. It is also to lay down the policy insofar as the management of in-house experiments is concerned. It is designed to align SESAME's Policy with that of an ever-increasing number of neutron scattering and synchrotron X-ray facilities in the world and so contribute to the fostering of a common framework for the management of experimental data.

Experimental Data arising from proprietary and industrial research and research carried out by Participatory Research Teams (PRTs) is not covered by this Policy and is subject to separate contractual arrangements.

The Policy defines the rules governing:

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

2. General principles

- 2.1 The present Experimental Data Management Policy pertains to the ownership of, the curation of, and access to experimental raw data and metadata collected and/or stored at SESAME.
- 2.2 Acceptance of this Policy is a precondition for the award of beam time for peer-reviewed and Director Discretionary time research proposals and in-house experiments.
- 2.3 No person will have the right to access, exploit or distribute raw data or metadata unless otherwise entitled to do so under the terms of this Policy.
- 2.4 Deliberate infringement of this Policy may lead to denial of access to raw data or metadata and/or rejection of future applications for beam time at SESAME.

3. Definitions

For the purposes of this policy:

- 3.1 The term **experimental data** is the umbrella term for raw data and all associated metadata.
- 3.2 The term **raw data** pertains to data collected from peer-reviewed, Director Discretionary time and in-house experiments performed on SESAME and Participatory Research Team (PRT) instruments. This definition includes data that are created automatically or manually by SESAME-specific software and/or SESAME staff expertise in order to facilitate subsequent analysis of the experimental data.

- 3.3 The term **metadata** describes information pertaining to data collected from SESAME instruments, including (but not limited to) the context of the experiment, experimental conditions and other logistical information. It also describes the same information pertaining to data collected from Participatory Research Team (PRT) instruments when this data relates to raw data as defined in this Policy. In both cases, irrespective of the instruments used, no personal data are recorded in the metadata.
- 3.4 The term **Director Discretionary time** refers to beam time allocated by the Director of SESAME at his/her discretion, typically for short trials deemed desirable for the development of SESAME's scientific programme, very promising beam time requests received after a proposal submission deadline, and beam time that may be allocated on an individual user basis in the case of lost beam time.
- 3.5 The term **Participatory Research Team (PRTs)** refers to a research team that has contributed to the funding, building and/or operation of a SESAME beamline, and has exclusive use of a certain percentage (negotiated with the Director of SESAME) of the available time at the beam to which it contributed.
- 3.6 The term **principal investigator** (PI) pertains to the main proposer identified in the experiment proposal (for peer-reviewed experiments and in experiments granted Director Discretionary time), or the Safety Approval Form (for in-house experiments).
- 3.7 The term **experimental team** consists of the PI and any other person to whom the PI designates the right to access resultant raw data and associated metadata.
- 3.8 The term **proprietary research** refers to research done through purchased (commercial) access to SESAME.
- 3.9 The term **online 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.
- 3.10 The term **results** pertains to data, intellectual property, and outcomes arising from the analysis of raw data. It does not include publications.
- 3.11 The term **custodian** refers to SESAME storing, curating and providing access to raw data, metadata and results.
- 3.12 The term **long-term** normally means a minimum of 5 years, and SESAME will strive for 10 years. The actual number of years will depend on the type and volume of data concerned and the economic consequences associated with long-term data storage. Thus, SESAME reserves the right to shorten the storage periods and/or reduce the number of data sets in consultation with the pertinent communities for high data rate instruments.
- 3.13 The term **open access** means belonging to the community at large, unprotected by copyright or patent, and subject to appropriation by any person. Open access to research data from public funding should be easy, timely, user-friendly and preferably internet-based. The SESAME data archive will be made available under CC-BY (Creative Commons – By Attribution).
- 3.14 The term **landing page** identifies a stand-alone web page specifically created to direct online search requests to a predetermined place.

4. Raw data and associated metadata

4.1 Access to raw data and associated metadata

4.1.1 All the raw data and associated metadata obtained as a result of peer-reviewed access to SESAME, use of Director Discretionary time and in-house research will be open access after an initial embargo period of 3 years, during which access will be restricted to the experimental team, represented by the PI.

4.1.2 SESAME will be the custodian of the raw data and associated metadata.

4.2 Curation of raw data and associated metadata

4.2.1 All raw data and metadata will be curated in well-defined formats, for which the means to read the data will be made available by SESAME.

4.2.2 Metadata that are automatically captured by instruments will be curated either within the raw data files, within an associated online catalogue, or both.

4.2.3 Only data with metadata generated by SESAME software will be archived.

4.2.4 Raw data and metadata will be read-only for the duration of their lifetime.

4.2.5 Raw data and metadata will be migrated or copied to archival facilities for long-term curation.

4.2.6 Each experiment and data set will eventually have a unique persistent identifier. Any person providing data with the same identifier must make sure that the data in the copy is identical to that in the Facility's database. Any person publishing results based on open access data will quote the same identifier (and related publications if available and appropriate).

4.2.7 High-level metadata such as Title, Authors, Abstract, and 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.

4.3 Access to raw data and metadata

4.3.1 Access to raw data and metadata at SESAME will be via a searchable online catalogue.

4.3.2 Access to the online catalogue of SESAME will be restricted to users registered on the catalogue. SESAME will set up the online procedure to become a registered user of its online catalogue.

4.3.3 Access to the raw data and associated metadata obtained from an experiment will be 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 who wishes data to retain *restricted access* for a period longer than 3 years will have this possibility by submitting a written request, specifying the reasons for the proposed prolongation, to the SESAME Director who decides on the request. Data may be made openly accessible earlier than 3 years on simple request of the PI. A limited subset of metadata will be available on the immediate

completion of experiments. If data cannot be stored at the Facility for 3 years, the PI will have exclusive access to it for the full duration of the storage period.

- 4.3.4 It will be the responsibility of the PI to ensure that the experiment number is correctly entered into the metadata for each raw data set in order to associate each data set with him/her. Failing this, the experimental team will be unable to access the data via the online catalogue, and other users may inadvertently be given access rights to the data.
- 4.3.5 Authorised SESAME staff (e.g. instrument scientists, computing group members) will have access to any SESAME-curated data or metadata for Facility-related purposes. SESAME will use reasonable endeavours to preserve the confidentiality of such data during the embargo period.
- 4.3.6 The online catalogue will enable the linking of experimental data to experimental proposals. Unless otherwise authorised by the PI, access to proposals will only be provided to the experimental team and appropriate Facility staff.
- 4.3.7 The PI will have the possibility of transferring parts, or the totality, of his/her rights during the embargo period to another person registered on the online catalogue.
- 4.3.8 The PI will have the possibility of creating and distributing copies of his/her raw data.

5. Results

5.1 Ownership of results

Ownership of all results (intellectual property) derived from the analysis of the raw data will be determined by the contractual obligations of the person(s) performing the analysis. These contractual obligations may be to the users' establishment(s), to other users on the experimental proposal, to grant-funding agencies or to other third parties, or it may be a pre-existing contractual obligation to SESAME.

5.2 Curation of results

- 5.2.1 SESAME will provide curation of the results on a best effort basis and will act as custodian of results in the long term.
- 5.2.2 SESAME will not be held liable in the case of the unavailability or loss of data or results.
- 5.2.3 SESAME will not be held liable in the case of the unavailability or loss of data analysis software.

5.3 Access to results

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

6. Good practice for metadata capture and the storage of results

- 6.1 The experimental team will be encouraged to ensure that the metadata of the experiments are as complete as possible as, in the long term, this will enhance the possibilities for open search, retrieval and interpretation of the data.
- 6.2 SESAME will provide 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.
- 6.3 Researchers who aim to carry out analyses of raw data and metadata which are openly accessible should, when possible, contact the original PI or his/her designate team member to inform him/her of this and suggest a collaboration if appropriate. Researchers must acknowledge the source of the data and cite its unique identifier and any publication(s) linked to the same raw data.
- 6.4 PIs and researchers who carry out analyses of raw data and metadata will be encouraged to link the results of these analyses to the raw data/metadata using the facilities provided by the online catalogue. They will be further encouraged to make such results openly accessible.

7. Publication information

- 7.1 In publications related to data from experiments carried out at SESAME the persistent identifier of the experiment and data must be cited.
- 7.2 The references of publications related to experiments carried out at SESAME must be deposited in a publications database within 3 months of the publication date, or during any new application for beam time, whichever is the earlier.