

CERIC – ERIC

Scientific Data Policy

Contents

1.	Introduction	3
2.	Purpose.....	3
3.	Field of application	3
4.	General principles.....	4
5.	Definitions.....	4
6.	Raw Data and Metadata	5
6.1	Access to raw data and associated metadata.....	5
6.2	Curation of raw data and associated metadata	5
6.3	Access to raw data and metadata	5
7.	Results	6
7.1	Ownership of the results	6
7.2	Curation of the results.....	6
7.3	Access to results	6
8.	Good practice for metadata capture and results storage	6
9.	Publication information	6
10.	Sources	7

1. Introduction

The Central European Research Infrastructure Consortium, CERIC-ERIC (www.ceric-eric.eu), is a multidisciplinary Research Infrastructure for basic and applied research in all fields of Materials, Biomaterials and Nanotechnology. CERIC has been built by integrating leading national research facilities based in 8 countries (Austria, Croatia, Czech Republic, Hungary, Italy, Poland, Romania and Slovenia) into a unique European entity supporting the production of basic knowledge and technology transfer, and promoting the mobility of researchers in an international multicultural scientific environment. Based on this definition of CERIC – ERIC, the purpose of this document is to provide a common Scientific Data Policy for the all PFs in harmony with the National Data Policy of each country, if present.

The Scientific Data Policy will be published by CERIC and accepted by the users during the submission of their proposals.

2. Purpose

The purpose of the CERIC-ERIC Scientific Data Policy is to define the time, the way and the place where data, metadata and raw data are stored. This document describes the common framework for scientific data management at CERIC-ERIC Partners Facilities. The participating facilities are used by researchers in universities, publicly funded research entities, and industry. The general process includes the generation of raw data from each experiment, which is then analysed by the research team. The data format recommended by CERIC-ERIC for the raw data is NEXUS/HDF5, which in addition to the detector data includes administrative metadata, instrument metadata and scientific metadata.

The data framework presented in this document focuses on this “raw data” stage. It must be stressed that the full strength of this digital approach is only reached when all datasets, from detector data to final publication are included, giving full advantage to the experimenting team and the scientific community. The framework has been strongly influenced by the OECD’s “Principles and guidelines for access to research data from public funding”. It strives for a careful balance between aspects of competition and collaboration in science. Having an open access data policy with data in well-defined formats has many benefits:

- Raw data becomes open to scrutiny by other researchers which helps to uncover cases of scientific fraud. Thus open access policies foster scientific integrity.
- It makes previously measured data available for further analysis without the necessity to remeasure the same sample.
- It promotes interdisciplinary research.
- Scientists can mine data in previously unknown ways or reapply new methods to existing data.

3. Field of application

This CERIC-ERIC Scientific Data Policy applies to all CERIC-ERIC Partners Facilities. Having an identical approach to the management of scientific data will ease the life of scientists using more than one facility and add to the overall transparency of the scientific process.

4. General principles

This data management policy pertains to the ownership of, the curation of and access to experimental primary data and metadata collected and/or stored by CERIC-ERIC Partners Facilities.

- Acceptance of this policy is a condition of the award of beamtime.
- 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.
- Deliberate infringements of the policy may lead to denial of access to raw data or metadata and/or denial of future access to CERIC Partner facility.
- All data and metadata will be subject to the data protection legislation of the country in which the data and metadata are stored.

5. Definitions

Term	Definition
Partner Facilities (PFs)	Facilities ('Partner Facility') which provide access to users on behalf of the ERIC members and which have the scientific and technical capability to contribute to the common strategic objectives, purposes and access capabilities
Metadata	Metadata is defined as the data providing information about the CERIC – ERIC instrument associated to a set of data, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistic information
Row Data	The term Row Data refers to data collected from measurements performed at CERIC – ERIC Partners Facilities. This includes data created automatically or manually by software and/or staff expertise in order to facilitate the subsequent analysis of the experimental data
On-line Catalogue	On-line catalogue refers to a computer database of metadata containing links to raw data files, that can be accessed by a variety of methods
Result	The term result refers to data, intellectual property and outcomes arising from the analysis of raw data (publication NOT included)
Open Access	Open Access means belonging to the community at large, unprotected by copyright or patent and subject to appropriation to anyone
Principal Investigator	The term principal investigator (PI) pertains to the PI identified on the experiment proposal. For experiments outside of the facilities proposal system, the PI is the person initiating or performing the experiment
Experimental Team	The term experimental team includes the PI and any other person to whom the PI designates the right to access resultant raw data and associated metadata
Public Research	The term public research refers to research done through peer review and leading to publication(s)
Proprietary Research	The term proprietary research refers to research done through purchased (commercial) access to the research facility

Table 1: definitions

6. Raw Data and Metadata

6.1 Access to raw data and associated metadata

All raw data and the associated metadata obtained as a result of publically funded access to the research facility will become available through open or restricted access after a reasonable embargo period, with CERIC-ERIC and the Partner Facilities in which the experiments/measurements took place acting as the custodian.

All raw data and the associated metadata obtained as a result of proprietary research will be owned exclusively by the user paying for the instrument time. Proprietary users shall agree with CERIC-ERIC management on how they wish their raw data and metadata to be managed before the start of any experiment.

6.2 Curation of raw data and associated metadata

All raw data will be curated in well-defined formats, for which the means of reading the data will be made available by CERIC-ERIC and the Partner Facilities in which the experiments/measurements took place. Metadata that is automatically captured by instruments will be curated either within the raw data files, within an associated on-line catalogue, or within both. Data will be read-only for the duration of its life-time. Data may be migrated or copied to archival facilities for long-term curation. Each data set will have a unique identifier. Anybody providing data with the same identifier must make sure that the copy is identical to the data in the facility data-base. Anybody publishing results based on open access data must quote the same identifier (and related publications if available & appropriate).

6.3 Access to raw data and metadata

Access to raw data and metadata in CERIC-ERIC and in every Partners Facilities is foreseen to happen via a searchable on-line catalogue. Access to the on-line catalogue of the facility will be either open access or restricted to those who are registered users of the on-line catalogue (e.g. for large data sets). Access to raw data and the associated metadata obtained from an experiment is restricted to the experimental team for a period of 3 years after the end of the experiment. Thereafter, it shall become openly accessible. Any PI that wishes their data to remain restricted for a longer period will be required to address a justified request to CERIC-ERIC. Data can always be made openly accessible earlier on simple request of the PI. It is 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 correctly associate each data set with the PI. If this is not done, the experimental team will not be able to access the data via the on-line catalogue or other users may inadvertently be given access rights to the data.

Appropriate facility staff (e.g. instrument scientists, computing group members) has access to any facility curated data or metadata for facility related purposes. Every facility will undertake that they will preserve the confidentiality of such data. The on-line catalogue will enable the linking of experimental data to experimental proposals. Access to proposals will only ever be provided to the experimental team and appropriate facility staff, unless otherwise authorized by the PI. The PI has the right to transfer or grant parts or all of his rights to another registered person. The PI has the right to create and distribute copies of his raw data.

7. Results

7.1 Ownership of the 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.

7.2 Curation of the results

Each Partner Facility will provide a means for users to upload results and associated metadata to the facility and enable them to associate these results with raw data collected from the facility. The upload of results and associated metadata may be subject to volume restrictions. These results will be stored for at least five years by CERIC-ERIC or, in general, by the Partner Facilities in which the experiments/measurements took place. It will not be the responsibility of CERIC-ERIC or its partner facilities to fully curate this data e.g. to ensure that software to read / manipulate this data is available. CERIC-ERIC or its partner facilities shall not be liable in case of unavailability or loss of data or data analysis software.

7.3 Access to results

Access to the results of analyses performed on raw data and metadata is restricted to the person or to the once performing the analyses. Nevertheless, access rights can be granted to externals by the analyst. However, if the raw data are restricted, even access to the results of the analysis needs an additional permission by the PI.

8. Good practice for metadata capture and results storage

The experimental team is encouraged to ensure that experiments metadata are as complete as possible, as this will enhance the possibilities for them to search for, retrieve and interpret their own data in the future. Each partner facility undertakes to 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. Researchers who aim to carry out analyses of raw data and metadata which are openly accessible should, where possible, contact the original PI to inform them and suggest a collaboration if appropriate. Researchers must acknowledge the source of the data and cite its unique identifier and any publications linked to the same raw data. PIs and researchers who carry out analyses of raw data and metadata are encouraged to link the results of these analyses with the raw data / metadata using the facilities provided by the on-line catalogue. Furthermore, they are encouraged to make such results openly accessible.

9. Publication information

References for publications related to experiments carried out at the facilities must be deposited in a publications database within 3 months of the publication date, or during any new application for beamtime, whichever is the earlier.

10. Sources

This document has been written taking into account:

- The PaN data guideline for a common Scientific Data Policy:
<http://pan-data.eu/sites/pan-data.eu/files/PaN-data-D2-1.pdf>
- The Elettra Sincrotrone Trieste Scientific Data Policy:
<https://www.elettra.trieste.it/userarea/scientific-data-policy.html>
- The ALBA Synchrotron Data Policy
https://www.cells.es/en/users/call-information-1/bases/2017_07_data_policy_alba_approved-cr.pdf/view