Plan Overview

A Data Management Plan created using DMPTuuli

Title: Center for X-ray Spectroscopy

Creator: Simo Huotari

Principal Investigator: Simo Huotari, René Bes

Data Manager: René Bes

Affiliation: University of Helsinki

Funder: The Research Council of Finland (former The Academy of Finland)

Template: FIRI Data Management Policy

ORCID iD: 0000-0003-4506-8722

ORCID iD: 0000-0003-4206-1525

Project abstract:

The Center for X-ray Spectroscopy was established in 2017 to offer novel element-specific x-ray spectroscopy analytical services for national and international academic, educational and industrial users as well as members of state research institutes. The analytic services provide atomic scale information on the chemical state of elements in materials, such as chemical composition, valence states and information on atomic neighbours around a selected element atom or ion. This leads to unique information on chemical and structural processes also where in-situ and operando conditions needed, e.g., for batteries and catalysis. With this information the users find solutions for research questions in materials science, physics, chemistry, biology, geosciences and environmental sciences, radiochemistry, biophysics as well as cultural heritage. There are no such laboratories available elsewhere in the Fennoscandian region, and even globally they are very rare. Hence the demand for the services from national and international users has been overwhelming.

The data policy of the project will adhere to the internationally accepted PaNdata (http://pan-data.eu/) as well as Photon and Neutron Open Science Cloud (PaNOSC) (https://www.panosc.eu/) and European Open Science Cloud Photon and Neutron Data Servives (ExPaNDS) https://www.panosc.eu/related-projects/expands/ formed together by the European large experimental photon and neutron sources.

ID: 18941

Start date: 01-01-2023

End date: 11-03-2025

Last modified: 12-05-2022

Center for X-ray Spectroscopy

Data management policy

General description of data

For a full description of the PaNOSC data policy that the project adheres to, please see the document PaNOSC Photon and Neutron Open Science Cloud, H2020-INFRAEOSC-04-2018 (10.5281/zenodo.3826039). The data policy is also published on the research Infrastructure web pages (www.helsinki.fi/en/infrastructures/center-for-x-ray-spectroscopy).

The data is numerical data collected from experiments performed on facility instru-ments. This definition 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. The data format is Nexus (HDF5), that has been adopted by a majority of photon and neutron sources, and is supported by some detector suppliers and more and more data analysis software. In addition to the detector data the Nexus/HDF5 data includes sample, instrument and scientific metadata. The full strength of this digital approach will be reached when all
data from the detector to the final publication are included in a digital object which is
machine readable, giving full advantage to the experimental team and the scientific
community.

The data that is not governed by industrial users will be managed through the principles of being Findable, Accessible, Interoperable and Re-usable, i.e., the data policy framework is strongly aligned with the FAIR principles.

The amount of data is expected to be 1 TB per year of numeric data. Data quality is monitored during the experiments both by automated pre-analysis routines that can alert the user if experiment is not working out properly. The consistency and quality of the data will be controlled by regular calibration of the instruments by the RI staff. Standardised protocols for data acqusition, quality control as well as calibration of the instruments will be in place. Eventually the user is responsible for the quality of the data while all effort is done by the RI to ensure the data quality.

Agreements on rights of use and license

All data collected, stored, used and managed within the RI will comply with the ethical guidelines by the Finnish Advisory Board on Research Integrity and complies witht he EU General Data Protection Regulation (GDPR) and the Finnish data protection act. Agreements with all partners / researchers will be made in the beginning of the project to define issues of ownership and sharing of the data.

The RI does not store or generate sensitive personal information other than a user database (name, e-mail address, and affiliation). However it may generate confidential information, such as data for novel IPR and patents. In addition, discussions and knowledge exchange under separate NDA's might occur with collaboration partners as companies. Commercial users will thus have an option to keep their data non-open.

The PI has the right to transfer or grant parts or all of his rights to another person. The PI has the right to create and distribute copies of their raw data and metadata. 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.

All raw data and the associated metadata obtained as a result of public funded access to the research facilities are open access, with the research facility acting as the custodian. All raw data and the associated metadata obtained as a result of proprietary research will be owned exclusively by the client who purchased the access. Proprietary users must agree with the facility management how they wish their raw data and metadata to be managed before the start of any experiment.

Citing the infrastructure: The users are required to acknowledge the RI in all publications where data from the RI is used. This applies to all forms of publication including scientific and popularized articles, presentations and theses.

Opening or sharing data

The RI follows the Declaration of Open Science and Research by the Finnish research community. We will be committed to open-access publishing. Gold or green route open access will be encouraged.

Access to raw data and metadata collected at the facility is foreseen to be via a searchable on-line catalogue through a web portal. 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 will become openly accessible. Any PI that wishes their data to remain restricted accessfor a longer period will be required to make a special case to the respective facility management. If data can only be stored at the facility for less than three years, then access is exclusive to the PI up to the end of the storage period. Data can always be made openly accessible earlier on simple request of the PI. Appropriate facility staff (e.g. instrument scientists, computing group members) has access to any facility curated data or metadata for facility related purposes. The 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.

Documentation & metadata

Metadata is stored together with the data as completely as possible by the facility. The experimental team is encouraged to ensure that further 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. The 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 use the data collected by others, 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 online catalogue. Furthermore, they are encouraged to make such results openly accessible.

Storage, backup & access control to data

The RI 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 RI offers currently 120 TB RAID-5 data server, which is maintained by the UH IT group and it will be upgraded periodically. The RI does not collect nor store sensitive data except for commercial contract research, in which case the storage and transfer of data will be agreed with the commercial partner in the beginning of the project. These data will be stored long-term by the originating facility. It will be responsible to curate this data e.g. to ensure that software to read/ manipulate this data is available. The facility cannot be made liable in case of unavailability or loss of data. The facility cannot be made liable in case of unavailability or loss of data analysis software.

Long-term storage of data: We will use the CSC IDA service and/or the EUDAT service and will store the data for at least 10 years, and longer if technically feasible. A persistent identifier will be acquired for the data.