OSCARS

Open Science Clusters' Action for Research & Society

Funded Project

CDIF-4-XAS: Describing X-Ray Spectroscopy Data for Cross-Domain Use

Principal Investigator: S. Hodson, Codata

Project team members: A. Gregory; M. Heikkurinen; L. Liborio; P. Austin; A. Nieva de la Hidalga; H. Görzig; R. Krahl; M. Kubin

Implemented by

UKRI Science and Technology Facilities Council

Cardiff University Prifysgol Caerdydd

CODATA Committee on Data International Science Council

HZB Helmholtz Zentrum Berlin

HMC

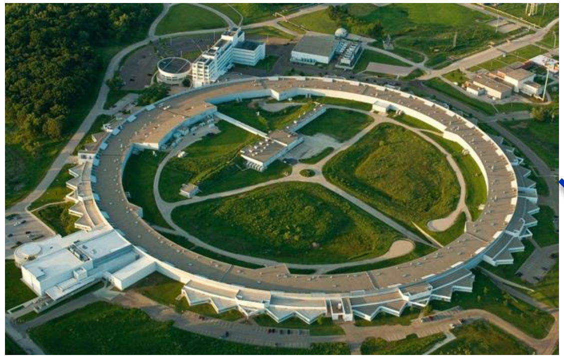
Funded by the European Union

What problem(s) are you going to solve?

- Related to (meta)data obtained from X-ray Absorption Spectroscopy (XAS) experiments.
- Over time, scientific communities, research facilities and software developers have created different formats and applications to store XAS data and describe it with metadata. This resulted in the data not being easily integrated or interoperable.
- XAS experiments are performed all over the world, and reusing XAS data from different sources—whether for further research, AI training, or the reproduction and replication of results—can be challenging. Or simply impossible.

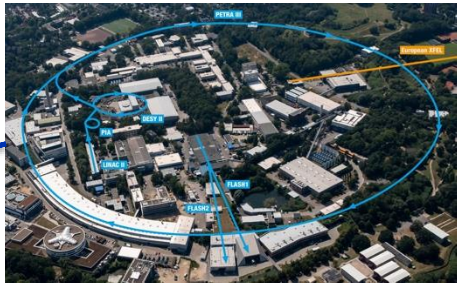
CHALLENGE ADDRESSED

CDIF-4-XAS



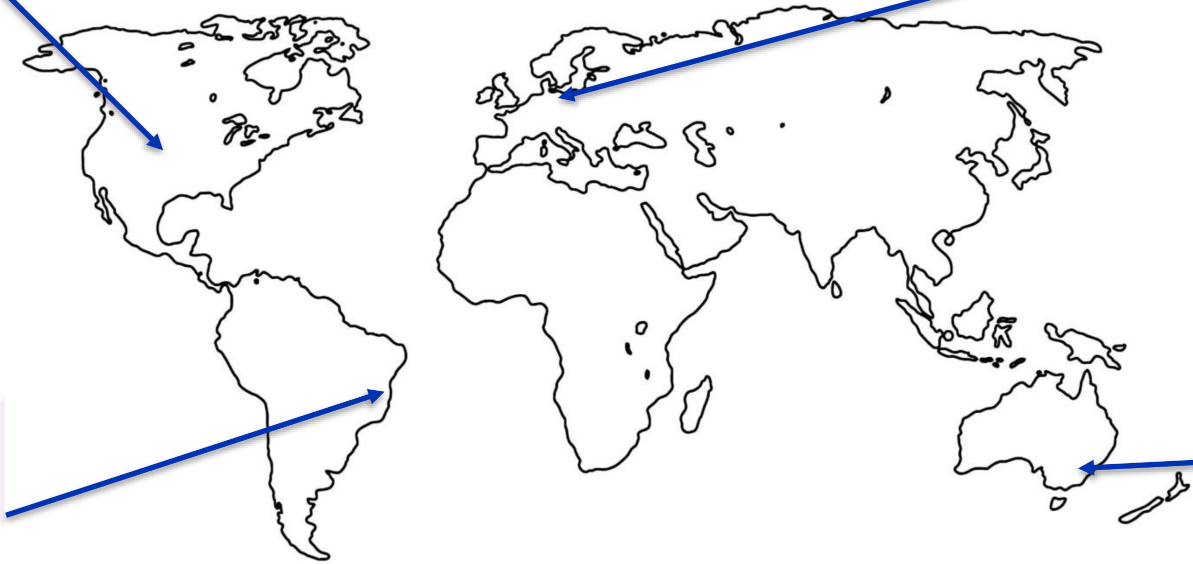
Argonne NATIONAL LABORATORY | Advanced Photon Source

Chicago, U.S.A



DESY.

Hamburg
Germany



LNLS

Brazilian Synchrotron Light Laboratory. Campinas, Brazil.

ANSTO

Australian Synchrotron. Clayton, Australia.

- Improving the transfer, interoperability and reusability of X-ray spectroscopy data and metadata.
- Enabling new science through combining X-ray spectroscopy data at scale across infrastructures.

What are you planning to do to solve the problem?

- We want to develop a metadata framework aligned the **Cross-Domain Interoperability Framework (CDIF)** to tackle this problem.
- The **CDIF** will serve as the scaffold for preserving the relevant metadata needed to map the published results to their supporting XAS data, which may be available in various formats.
- The **CDIF** does this by establishing domain-agnostic standards to support data sharing across institutional and disciplinary boundaries.
- Through CDIF-4-XAS, metadata and provenance for XAS data are standardised, enhancing the integration of XAS datasets. This promotes data reuse across different research domains, making XAS data more accessible, valuable and interoperable.

What will be the results and how do you plan to make them available to the broader community?

- The first step will be a **landscape analysis** of standards, ontologies, data formats and practices within the XAS community.
- Then we will perform semantic descriptions of at least two chosen XAS community standards using a CDIF profile.
- The CDIF profile will facilitate integration with a set of Galaxy software tools that we have developed, as part of the EU's EuroScienceGateway project, for processing XAS data.
- Galaxy is a web-based workflow management platform, and our tools are installed in European and STFC servers and are available to the XAS community.
- We will also write recommendations and guidelines for the use of CDIF for XAS data (XAS-CDIF) and plan to write scientific articles to describe and promote this work on the community.

OSCARS

Landscape Analysis: Delivered

- Reviewed over 50 x-ray facilities around the world to learn about the current practices, standards, vocabularies, ontologies, schemas, and data formats used in the generation, curation, and publishing of X-Ray Absorption Spectroscopy (XAS) data and metadata.
- Small percentage of XAS (meta)data is made public.
- Despite the community standardisation efforts, there are several facility-scale databases and schemas, but there is no common schema for XAS (meta)data and no defined ontology.
- Currently, XAS (meta)data difficult to integrate or make interoperable.
- At the 2025 International XAFS Conference in July, NeXuS and XDI are likely to be proposed as reference standards. We will perform semantic description of these.
- Landscape analysis published here: https://doi.org/10.5281/zenodo.14920226

What risks could limit the success of the project, and how can they be mitigated

- For implementing the use of XAS-CDIF, and for developing prototypes for using XAS-CDIF, it would be useful for the CDIF-4-XAS project to have access to XAS data and metadata from different research facilities around the world. We are in contact with scientists working at the Diamond Synchrotron Source in the UK, the ESRF synchrotron in Grenoble, and the BESSY synchrotron in Germany. However, there may be a small risk of not getting data/metadata from all these sources. We are planning to contact scientists in other synchrotrons.

- Community uptake of new standards risk. But we plan to apply XAS-CDIF in diverse cases to mitigate this risk.

Who is doing it?

- **CODATA, the Committee on Data of the International Science Council, France.**
  Simon Hodson (PI), Arofan Gregory, Matti Heikkurinen

- **Rutherford Appleton Laboratory, STFC, UKRI, UK.**
  Leandro Liborio, Patrick Austin

- **School of Computer Science and Informatics, Cardiff University, Wales, UK.**
  Abraham Nieva De La Hidalga

- **Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, HZB, Germany.**
  Heike Görzig, Rolf Krahl

- **Helmholtz Metadata Collaboration (HMC) Hub Matter, Germany.**
  Markus Kubin