COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

WORLD
DATA
SYSTEM

Contribution ID: **218**   Type: **Poster**

Facilitating Cross-Domain Interoperability of X-Ray Absorption Spectroscopy (XAS) Data: Developing a CDIF Profile for the Galaxy Platform.

Monday 13 October 2025 18:00 (1h 30m)

The Cross Domain Interoperability Framework (CDIF) provides a set of implementation guidelines designed to lower the barriers to cross-domain research data reuse. CDIF provides standards and methodologies for addressing interoperability issues preventing cross-domain research data utilization. CDIF's initial version comprises five core profiles: Discovery, Access, Controlled Vocabularies, Data Description for Integration, and Universals, which collectively support the cross-disciplinary implementation of the FAIR principles.

The challenge of embedding FAIR principles into research outputs applies not only to data and metadata, but also to the methods used to analyse them. The metadata associated with raw data must be Interoperable support reusability. However, if the parameters used in the analysis of the data -and the corresponding metadata- are not recorded properly, Reusability will also be compromised.

Two European projects are addressing these challenges of cross domain interoperability and data analysis reproducibility in X-ray Absorption Spectroscopy (XAS). The first one is the OSCARS-funded CDIF-4-XAS project that is applying CDIF to enhance the interoperability and reusability of XAS data. The objective is to streamline data exchange between applications, databases, and institutions, making XAS data interoperable across multiple research disciplines. The second project is the EuroScienceGateway, which aims to leverage European computing infrastructures for data-intensive research guided by FAIR principles.

Recently, the CDIF-4-XAS project published its first deliverable: a comprehensive landscape analysis of standards, vocabularies, ontologies, data formats, and practices in XAS. For the EuroScienceGateway project, we have contributed a set of custom tools that can be used in the Galaxy platform for managing workflows associated to XAS data processing and analysis. Galaxy offers a number of features that ensure that the workflows' outputs retain all the metadata needed for them to be reproduced: histories store the data and parameter inputs associated to all output data; software tools are strictly versioned and run in containers, and the execution of workflows can be exported as Research Object Crates.

Building on this, the CDIF-4-XAS is developing semantic descriptions of two XAS community standards (NXxas for multi-spectra raw and processed data and XDI for single spectra data) that will produce a CDIF profile (XAS-CDIF). This includes: exploration of the use of the CDIF Discovery profile and related standards such as schema.org, DCAT, and PROV-O and of DDI-CDI for data description of variables; characterisation of the HDF5 data structure using DDI-CDI and mappings of key NXxas and XDI concepts. The XAS-CDIF profile will be used to extend, and update, the existing XAS Galaxy tools and workflows, which will facilitate the seamless integration of data from various sources into processing and analysis workflows.

This paper will briefly present the OSCARS and EuroSciencegateway projects and explain how the XAS-CDIF may facilitate seamless integration of XAS datasets from different beamlines and laboratories, promoting data reuse across diverse research domains. The prototype implementation of XAS-CDIF into Galaxy tools and workflows will be presented as an example of how the XAS-CDIF profile facilitates building advanced tools that take advantage of data interoperability.

Primary authors: NIEVA DE LA HIDALGA, Abraham (School of Computer Science and Informatics, Cardiff University, Wales, UK.); GREGORY, Arofan (CODATA, the Committee on Data of the International Science Council, France.); GÖRZIG, Heike (Helmholtz Zentrum Berlin für Materialien und Energie (HZB), and Helmholtz Metadata

Collaboration (HMC), Germany.); LIBORIO, Leandro (Scientific Computing Department, STFC, UKRI, UK); KUBIN, Markus (Helmholtz Zentrum Berlin für Materialien und Energie (HZB), and Helmholtz Metadata Collaboration (HMC), Germany.); AUSTIN, Patrick (Scientific Computing Department, STFC, UKRI, UK); KRAHL, Rolf (Helmholtz Zentrum Berlin für Materialien und Energie (HZB), Germany.); HODSON, Simon (CODATA, the Committee on Data of the International Science Council, France.)

Presenter: LIBORIO, Leandro (Scientific Computing Department, STFC, UKRI, UK)

Session Classification: Poster Session

Track Classification: SciDataCon Persistent Themes: Data Science and Data Analysis