



## Overview of X-Ray Absorption Spectroscopy standards, vocabularies (and ontologies), data formats and practices.

### Landscape Analysis

Abraham Nieva de la Hidalga, Leandro Liborio, Patrick Austin, Markus Kubin, Heike Görzig,  
Simon Hodson, Arofan Gregory, Rolf Krahl.

## Executive summary

X-ray Absorption Spectroscopy (XAS) has become a fundamental tool in diverse scientific fields, including Physics, Chemistry, Surface Science, Nanoscale Science, Biology, and Environmental and Earth Sciences. However, the lack of standardized data formats and metadata poses challenges for interoperability and reuse. This document analyses current XAS data management practices, reviewing the efforts to improve data integration and usability.

Key areas covered include:

- Overview of XAS techniques and applications
- Development of custom data formats and analysis techniques
- Initiatives for interoperable standards.

Part of the CDIF-4-XAS project, this analysis is the first step towards developing a model for enhanced XAS data interoperability based on the Cross Domain Interoperability Framework (CDIF) developed by CODATA as part of the WorldFAIR project.



## CDIF-4-XAS project report metadata

<b>Project number</b>	01-314	
<b>Project acronym</b>	CDIF-4-XAS	
<b>Project name</b>	Describing X-ray spectroscopy data for cross domain use	
<b>Call</b>	1st OSCARS Open Cal	
<b>Topic</b>	OSCARS-PaNOSC	
<b>Type of action</b>	Task 1 Overview of standards, vocabularies (and ontologies), data formats and practices within the XAS area (landscape analysis)	
<b>Project starting date</b>	01/10/2024	
<b>Project duration</b>	24 months	
REPORTING PERIOD		
<b>Period covered</b>	from 01/10/2024 to 28/02/2025	
<b>Reporting period number</b>	1	
<b>Periodic report date and version</b>	28/02/2025, version 1	
<b>Authors</b>	Abraham Nieva de la Hidalga, Leandro Liborio, Patrick Austin, Markus Kubin, Heike Görzig, Simon Hodson, Arofan Gregory, Rolf Krahl.	
<b>Contributions</b>	ANH, LL, and PA contributed to the collection of examples, initial draft, editing, and revision of structure, MK, HG, SH contributed editing the draft during revision. AG and RK commented on the structure and coverage. All authors discussed and agreed on the coverage, content, and formatting of the document during the reported period.	
<b>CHANGE HISTORY</b>		
VERSION	PUBLICATION DATE	CHANGE
1.0	28/02/2025	Initial version (new MFF).

<b>1. Introduction.....</b>	<b>5</b>
<b>2. X-Ray Absorption Spectroscopy.....</b>	<b>6</b>
2.1. Brief introduction to XAS.....	6
2.2. XAS Measuring Methods.....	7
<b>3. File formats, schemas, and ontologies.....</b>	<b>10</b>
<b>4. Generation, Curation and Publishing of XAS Data.....</b>	<b>13</b>
4.1. Instrument Dependent XAS Schemas.....	13
4.1.1. Large-scale facilities.....	14
4.1.2. Laboratory-based XAFS.....	14
4.1.3. Summary.....	15
4.2. XAS Database Schemas.....	15
4.2.1. Materials Data Repository XAFS database (MDR XAFS DB).....	15
4.2.1.1. MDR XAFS Schema.....	17
4.2.1.2. MDR XAFS Metadata.....	18
4.2.1.3. MDR XAFS Data (9809 data format).....	21
4.2.2. Canadian Light Source XAS Database (CLSXAS DB).....	22
4.2.3. The SSHADE FAME Database (FAME DB).....	24
4.2.4. IXAS X-ray Absorption Data Library (XASLIB).....	26
4.2.5. LISA XAS Database.....	28
4.2.6. Experimental XAS database.....	30
4.2.7. XAS Reference Research Database (RefXAS DB).....	32
4.2.7.1. The RefXAS DB Schema.....	34
4.2.8. Summary.....	37
4.3. Software Dependent XAS Schemas.....	38
4.3.1. Demeter.....	39
4.3.1.1. Demeter Inputs.....	39
4.3.1.2. Demeter Outputs.....	40
4.3.2. X-ray Larch.....	41
4.3.2.1. Larch inputs.....	41
4.3.2.2. Larch outputs.....	42
4.3.3. DAWN.....	42
4.3.4. Summary.....	42
<b>5. Community Standardisation Efforts.....</b>	<b>43</b>
5.1. Community defined standards.....	43
5.1.1. The NeXus data format.....	44

5.1.1.1. NXxas application definition.....	46
5.1.1.2. NXxasproc application definition.....	48
5.1.2. XAFS Data Interchange (XDI).....	49
5.2. Other Interoperable Schemas for XAS Data.....	51
<b>6. Conclusions and outlook.....</b>	<b>52</b>
6.1. Community buy-in.....	52
6.2. Expected advantages.....	53
<b>7. References.....</b>	<b>54</b>

## 1. Introduction

X-ray Absorption Spectroscopy (XAS) research has expanded to become a set of widely used scientific methods with applications across Physics, Chemistry, Surface Science, Nanoscale Science, Biology, and Environmental and Earth Sciences. Over time, various scientific communities, research facilities, device providers, and software developers have created different formats and applications to store XAS data and describe it with metadata. These custom data formats serve their specific purposes, but they are not easily integrated or interoperable. Consequently, reusing XAS data from diverse sources—whether for further research, AI training, or the reproduction and replication of results—can be challenging. This is due to the diverse ways in which data is presented across domains and the limited availability of XAS data and metadata in public repositories. Using or combining these datasets often requires expert intervention and manual steps to map and process the data. Establishing commonly accepted standards for publishing XAS data is the first step in addressing these challenges.

This document provides a landscape analysis of current practices, standards, vocabularies, ontologies, schemas, and data formats used in the generation, curation, and publishing of X-Ray Absorption Spectroscopy (XAS) data, with a particular focus on efforts to create community standards that facilitate interoperability. We begin with an overview of current XAS techniques and their areas of application. Next, we discuss the development of custom formats for storing data and efforts to produce analysis techniques applicable regardless of data origin. We describe current efforts to generate standards that facilitate data interchange and integration, including various initiatives from research consortia aimed at creating interoperable formats, ontologies, and vocabularies. We conclude by observing the emerging consensus around using NXxas for multi-spectra raw and processed data and XDI for single spectra data.

This landscape analysis is part of the CDIF-4-XAS project, aimed at piloting an improved model for XAS data interoperability and reusability across scientific disciplines through the Cross Domain Interoperability Framework (CDIF) developed by CODATA as part of the WorldFAIR project [1]. The CDIF-4-XAS project seeks to enable seamless integration of XAS data into data catalogues and analysis frameworks in a universally interoperable manner, making it easier to reuse, compare, and incorporate into larger studies or used for training AI applications. Therefore, we conclude by introducing the next phase of work, which will be to use CDIF to develop interoperability profiles to enable the integration and mapping of NXxas and XDI data: the CDIF-4-XAS interoperability model.

## 2. X-Ray Absorption Spectroscopy

X-Ray Absorption Spectroscopy (XAS) also known as X-ray Absorption Fine-Structure Spectroscopy (XAFS) is a widely used experimental analysis technique for studying the electronic and structural properties of materials. XAS is observed when X-rays are absorbed by a material as the energy of the radiation is varied. The technique is sensitive to chemical elements contained in a material.

### 2.1. Brief introduction to XAS

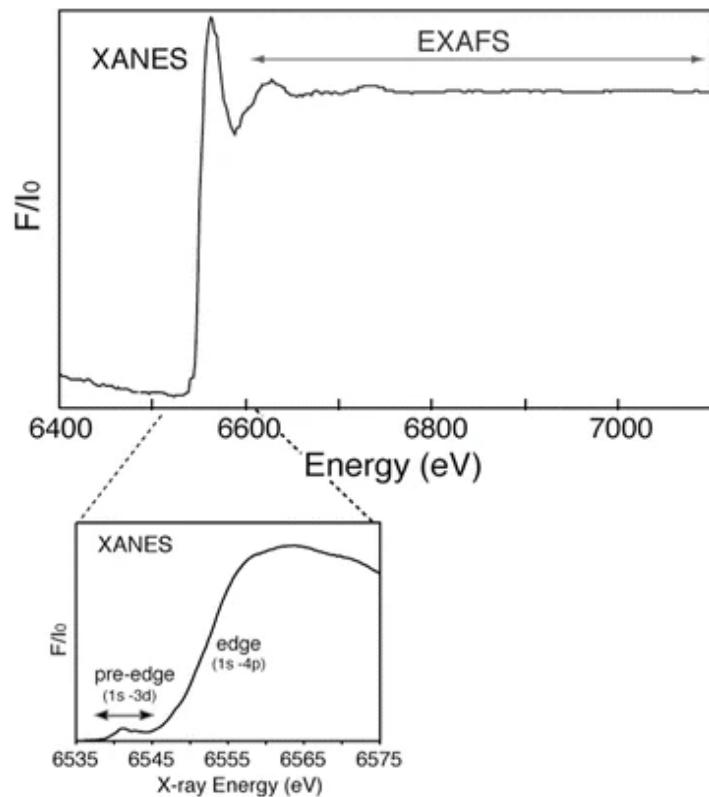
Depending on the chemical element and the wavelength of X-rays used (scientific communities distinguish soft, tender, and hard X-rays), XAS probes different absorption edges (K-edge, L-edge, M-edge etc.) of that element. The product of a XAS measurement is an X-ray absorption spectrum. It reveals essential information on the local electronic structure such as oxidation- and spin-states, as well as the local coordination- or ligand environment surrounding the probed element, thereby providing crucial insights into the chemical properties of materials in a variety of scientific disciplines like catalysis, materials science, and environmental chemistry.

The typical spectral regions of a XAS spectrum of the chemical element Manganese (Mn) are shown in [figure 1](#) (here measured at the K-edge). Two spectral regions typically used for two different types of analyses the X-ray absorption near edge structure (XANES), also known as Near-edge X-ray absorption fine structure (NEXAFS), reveals information about the local electronic structure, while analysis of the Extended X-ray absorption fine-structure spectroscopy (EXAFS) spectral region reveals information about the local atomic structure like coordination number and distance [2]. Researchers often combine XAS techniques with other spectroscopic techniques to provide a more detailed picture of the nature of the chemical bonds to better understand reaction mechanisms and intermediates [2].

X-ray spectroscopy is employed in a broad range of applications in technology and research [3]. To give just a few examples, applications range from fundamental research in chemistry, physics and material science, to environmental research, architecture, art, archaeology and industrial applications, geology, and astronomy Diamond [4], APS [5], BESSY [6], DESY [7], ESRF [8], and RCSLA [9]).

The most used XAS techniques are XANES and EXAFS, and they can be applied using soft X-rays -energies from 100 eV to 10 keV- or hard X-rays -energies from 10 keV to 100 keV-.

There are various methods to detect X-ray absorption spectra. In the following section we list the most common ones.



*Figure 1. X-ray absorption spectrum measured at the K-edge of the chemical element Manganese (Mn) located at the active site of a biological sample relevant for photosynthesis. The spectral regions for XANES- and EXAFS-analyses are assigned. (Figure adapted [3] Used under CC BY-NC 2.0 license).*

## 2.2. XAS Measuring Methods

Various methods for acquiring X-ray Absorption Spectroscopy (XAS) spectra are available due to differences in experimental setups, sample types, measurement objectives, and specific applications. XAS can be conducted in multiple configurations to meet the unique requirements of a study, including different sample environments, energy ranges, and sensitivity levels. Some of the existing methods and the motivation for their existence are listed below.

### Transmission mode XAS

In transmission mode XAS, the sample is placed in the X-ray beam and the X-ray flux transmitted through the sample is measured as a function of the X-ray energy. This method

uniquely allows for measuring X-ray absorption in absolute units [10]. For this method, samples need to be sufficiently transmissive, hence comparably thin.

### **Electron yield-detected XAS (EY-XAS)**

In electron yield-detected XAS, the sample is placed in the X-ray beam and the X-ray energy is varied. Shortly after X-ray absorption by the sample, electrons are emitted due to Auger decay. The current of electrons, measured as a function of the X-ray energy, is often used as a measure of the relative XAS amplitudes [11]. This method is surface-sensitive.

### **Fluorescence yield-detected XAS (FY-XAS)**

In fluorescence yield-detected XAS, the sample is placed in the X-ray beam and the X-ray energy is varied. Shortly after X-ray absorption by the sample, X-ray photons are emitted due to fluorescence decay. The relative intensity of emitted X-rays as a function of the X-ray energy is often used as a measure of the relative XAS amplitudes. In some cases, this method is biased by method-specific distortions of the X-ray absorption spectrum [12].

### **X-ray magnetic Circular Dichroism (XMCD)**

In X-ray magnetic circular dichroism extends XAS to provide information about magnetic properties of the probed sample or material. For XMCD, the sample is placed in a strong magnetic field and XMCD is obtained from the difference of XAS spectra measured once with left and once with right polarization of the X-ray beam hitting on the sample [13].

### **Energy Dispersive XAS (ED XAS)**

The principle of the Energy Dispersive XAS extends the scheme of transmission mode XAS. A polychromatic beam is focused on the sample and is partially transmitted towards a position-sensitive detector where the beam position is correlated to the X-ray energy. The dispersive configuration has the advantage of collecting the whole X-ray absorption spectrum simultaneously, which makes the technique especially useful for the study of fast processes. ED-XAS produces the same type of XAS spectrum of Figure 1, but with limited precision and sensitivity as compared to standard transmission mode XAS. A potential application of ED-XAS is, e.g., the rapid screening of biological/biomedical systems [14]. The technique can also be used for fast collection of the EXAFS spectral region [14], and to study catalysts under operating conditions [29], offering the possibility to study reaction processes on second to millisecond time scales.

### **High Energy Resolution Fluorescence Detected XAS (HERFD-XAS), or Partial fluorescence-yield detected XAS (PFY-XAS)**

High-energy-resolution fluorescence detection XAS, also known as partial fluorescence-yield detected XAS, restricts FY-XAS (see above) to monitoring the intensity of a single, element-specific fluorescence line, which can help to reduce experimental noise as compared to FY-XAS [15].

### **Imaging XAS**

By combining XAS with X-ray microscopic techniques, 2D imaging with spatially resolved chemical sensitivity can be obtained [16]. This can be applied to the mapping of 2D and 3D images to XAS data, which can be applied, for instance, to obtain new insights into the geometric and electronic configuration of particles in a catalyst [17].

### **Surface XAS**

Surface XAS is an in-situ technique, where surface sensitivity can be achieved by using a carefully controlled grazing incidence geometry. The main idea is that the beam scans a surface region, so that the atoms involved in the absorption are at the surface of the sample. This grazing incidence technique is also an ideal tool for determining the electronic structure and local coordination environment of novel, surface-engineered, materials in realistic operating conditions. It has been applied to the study of electrocatalysts [30].

This is just a brief overview, not an exhaustive list, as new methods and modifications are constantly being developed for different materials. Moreover, experimental techniques, including ex-situ, in-situ, and operando studies, can be combined with different acquisition methods [18], [19]. Normally, ex-situ refers to the analysis of samples after they have undergone some process in the lab. For instance, a study that aims to determine the changes in a sample that was treated under different atmospheric conditions -such as different gas mixes- by studying different samples that had been treated under different gas combinations. In-situ and operando experiments are more complex, requiring setting up reactors on the targeting station to facilitate acquiring spectra while conditions such as pressure, temperature or the atmospheric composition change. In-situ experiments usually involve fewer samples taken at specific intervals (of time, pressure, temperature or gas flow) to monitor the sample during the process; while operando experiments involve continuous or frequent data collection over longer periods, allowing for real-time monitoring of the sample's behaviour while also monitoring the environmental variables.

### 3. File formats, schemas, and ontologies.

When discussing the preservation of data, the terms *file formats*, *schemas*, and *ontologies* [20] can be used interchangeably. For this overview of the XAS data landscape, it is important to define their use in our discussion.

The **file format** refers to the structural organisation of a file. It specifies the encoding of data for storage, processing or transmission<sup>1</sup>. There are different layers of encoding. The character encoding can be distinguished between text (ASCII, UTF-8), and binary formats. Where ASCII and UTF-8 are specific forms of text encoding. The next layer is the organisation of data in the file such as key-value, tables or hierarchical and the combination of these in a file. For instance, CSV, JSON, and XML are formats for saving data encoded as ASCII and are defining a structure of how vocabulary is related to values.

The controlled **vocabulary** is the list of standardised terminology, words, or phrases, used for indexing or content analysis and information retrieval, usually in a defined information domain<sup>2</sup>. The terms used in the vocabulary are mostly related to a concept that might be described in an ontology (see ontology).

The **data schema** refers to the logical organisation of data that groups closely related entities together. Formally, the schema is the organisation or structure for a database. The activity of data modelling leads to a schema. (The plural form is schemata.) The term is used in discussing both relational databases and object-oriented databases. The term sometimes refers to a visualisation of a structure and sometimes to a formal text-oriented description<sup>3</sup>. Groups of data could be defined in terms of the information they represent. In a database, for instance, you could define tables that contain information as well as define the relationships between those tables: one table could define the components used in an experiment and another table could define the details of the experimenter. In text formats sometimes you divide the data into the heading section of key-values describing the contents and set values and the data section containing the values stored inside represented in a table. In a data schema the terms of a terminology are arranged in a specific way to comply with requirements for certain applications. There can be different schemata for the same application, but they will need to express the same things and therefore use the same concepts.

---

<sup>1</sup> <https://terms.codata.org/rdmt/data-file-format>

<sup>2</sup> <https://terms.codata.org/rdmt/controlled-vocabulary>

<sup>3</sup> <https://terms.codata.org/rdmt/schema-database>

An **ontology** is a higher level of knowledge representation that classifies the concepts in a domain and the relationships between those concepts. The ontology provides a framework and a vocabulary which can be used to assign meaning to data and enable reasoning about that meaning. The emergence of ontologies supports more sophisticated data uses, enabling for instance reasoning and generalisation<sup>4</sup>.

A visual example of the encoding, vocabulary, format, and schema concepts is provided in [Figure 2](#), the image shows a column data format, with detailed information about the data in the header lines (marked by a #). The data in the example is encoded as ASCII text. The schema used in the example is XDI. The vocabulary is aligned with the knowledge of the XAS community. In the example on [figure 2](#), the format for the data is XDI. The XDI format defines the way information is stored. The lines starting with a # are header lines and contain metadata, the remaining lines are data lines and represent the actual data values in columns.

```

# XDI/1.0 GSE/1.0
# Column.1: energy ev
# Column.2: i0
# Column.3: itrans
# Column.4: mutrans
# Element.edge: K
# Element.symbol: Cu
# Scan.edge_energy: 8980.0
# Mono.name: Si 111
# Mono.d_spacing: 3.13553
# Beamline.name: 13ID
# Beamline.collimation: none
# Beamline.focusing: yes
# Beamline.harmonic_rejection: rhodium-coated mirror
# Facility.name: APS
# Facility.energy: 7.00 GeV
# Facility.xray_source: APS Undulator A
# Scan.start_time: 2001-06-26T22:27:31
# Detector.I0: 10cm N2
# Detector.I1: 10cm N2
# Sample.name: Cu
# Sample.prep: Cu metal foil
# GSE.EXTRA: config 1
# /**
# Cu foil Room Temperature
# measured at beamline 13-ID
# -----
# energy i0 itrans mutrans
8779.0 149013.7 550643.089065 -1.3070486
8789.0 144864.7 531876.119084 -1.3006104
8799.0 132978.7 489591.10592 -1.3033816

```

*Figure 2. Example of a file formatted as column data and encoded in as ascii text (from [11]).*

For the example on [figure 2](#), an ontology would help in correctly interpreting all the information units and how they relate to each other. For instance, the fact that beamline

---

<sup>4</sup> <https://terms.codata.org/rdmt/ontology>



13ID is located at the APS facility is not evident from simply reading the data. In this case, the fact that beamlines are installed and operated within specific facilities is knowledge contained in an ontology.

This distinction is important because, to date, most of the XAS data published follows different data schemas. However, the knowledge for interpreting them, i.e. the ontology, is assumed to be common knowledge or self-evident from the context and contents of the data.

The following sections focus on two aspects of curation and publishing schemas for XAS data: first, providing an environmental perspective on the domain-dependent conditions that promoted the emergence of custom schemas ([Generation, Curation and Publishing of XAS Data](#)); and second, explaining the efforts to gather common knowledge and improve the overall situation through community-driven efforts towards interoperability ([Community Standardisation efforts](#)).



## 4. Generation, Curation and Publishing of XAS Data

There are many reasons for the diversity of XAS data schemas, including the configuration of acquisition devices, the software used for processing and analysis, the schemas of the databases that preserve the data, the research goals, and the publishers' requirements. Analysing the types of published XAS data available, four types of schemas for the publishing of XAS data are of relevance: instrument dependent, database, software dependent and community defined. In addition to these schemas, and because of the lack of a commonly agreed schema, the researchers/publishers of XAS data use various ad-hoc schemas which are not suitable for further use, such as custom text files, PDF documents, or custom software/hardware dependent schemas. The advantage of commonly agreed schemas is that their ontology, i.e. the knowledge about how the different parts of the information is related and interpreted, is documented and understood by a wide community. Moreover, the actual codification of the schema enables automated integration, validation, and interpretation of data, even when data is acquired from various sources. In the following sections, we will focus on the schemas which, over the years, have been adopted by different institutions and practitioners for the preservation and publishing of XAS data. The focus will be on public schemas which can be fully inspected and are associated with openly available public data. For this reason, we focus on the schemas sponsored by different synchrotron facilities, schemas from public databases, and the schemas supported by the most popular processing tools. Performing this classification of schemas will enable us to decide which are the best candidates to be integrated into CDIF-4-XAS. In this case, we are using the number of published data objects under a specific schema, as well as the availability of software which can identify and parse data stored in those schemas, as the criteria for deciding which are the most successful candidates for the next stage of the definition of CDIF-4-XAS.

### 4.1. Instrument Dependent XAS Schemas

The schemas for curation, storage, and publishing of XAS data are used to store XAS data at time of collection/generation in a format which is as close as possible to the raw data streams generated by the spectroscopy equipment. These formats need to be accessible to the researchers and their preferred processing and analysis tools. There are two primary sources for XAS type of data: large-scale facilities and institutional laboratories. This section explores some of the reasons why these environments contribute to the development of



custom data schemas and highlights some of the proposed solutions to address this variability.

#### 4.1.1. Large-scale facilities

Large-scale facilities like Diamond [4], APS [5], BESSY [6], DESY [7], ESRF [8], and RCSLA [9] allow access to cutting edge facilities for XAS experiments. Most of these facilities use some form of HDF5 structure. However, they also offer software that can preprocess and export the data for the beamline users. The diversity of XAS schemas stems from the fact that each beamline hosts highly customised XAS acquisition devices to capture XAS spectra. These devices require custom software to collect and store the data produced. As such, large-scale research facilities require the development of custom software to codify the data. This situation alone fosters the appearance of custom data schemas that register different details of experiment, spectra, beamline and devices involved in generation and recording of the data. One of the ways in which these facilities address the challenges of curating and publishing XAS data is the creation of public databases with open schemas which can be accessed by practitioners and software developers (see [XAS databases](#)).

In addition to this, large-scale facilities actively collaborate in the definition of standard schemas for the transmission and preservation of data (see [Community Standardisation Efforts](#)). A couple of these standards have matured to the point that they will be recommended for adoption in the IXAS XAFS2025 conference. In addition to these efforts, there are alternative proposals for developing custom extensions (plugins, mapping libraries) to correctly interpret data (see the section on [Software Dependent XAS Schemas](#)).

#### 4.1.2. Laboratory-based XAFS

In addition to large-scale facilities, there are several research groups that have implemented X-ray absorption spectroscopy (XAS) in the laboratory [21], [22]. These laboratories have either built custom developed spectrometers or use commercially available devices such as easyXAFS<sup>5</sup>, HP Spectroscopy<sup>6</sup>, and Sigray<sup>7</sup>. For instance, HarwellXPS<sup>8</sup>, The X-ray diffraction (XRD) RTP facility at the University of Warwick<sup>9</sup> and the Durham X-Ray Absorption Spectroscopy Facility (DXAF)<sup>10</sup>, operate easyXAFS systems offering XAFS as service to the

---

<sup>5</sup> <https://www.easyxafs.com>

<sup>6</sup> <https://www.hp-spectroscopy.com>

<sup>7</sup> <https://sigray.com>

<sup>8</sup> <https://portal.harwellxps.uk/xas>

<sup>9</sup> <https://warwick.ac.uk/research/rtp/xrd/instruments/xafs/>

<sup>10</sup> <https://www.durham.ac.uk/departments/academic/chemistry/about-us/services/dxaf/>

wider research community. In this case, these laboratory-based facilities also tend to create custom data schemas, which need additional processing and expert knowledge for integration with other data. In this case, laboratories are free to select the schema to use, for instance, DXAF uses Athena project files to transfer XAS data for their users.

#### 4.1.3. Summary

The file formats used by facilities and laboratories directly influence the creation of programs for processing and analysing XAS data. The lack of commonly accepted formats however is evident in the need to create and maintain various parsing tools which allow transforming data into a format suitable for processing and analysis.

### 4.2. XAS Database Schemas

This review includes only publicly available XAS databases that are accessible at the time of writing. The first place to look at is the International XAFS Database Portal<sup>11</sup> of the Japanese XAFS Society [23] that lists 6 XAS databases: MDR XAFS DB (Japan), CLS XAS Database (Canada), SSHADE/FAME DB (France), XASLIB (USA), Experimental XAS database (China), and LISA XAS Database (Italy). The following sections present an overview of these databases, plus the DAPHNE4NFDI XAS Reference Research Database (RefXAS DB).

#### 4.2.1. Materials Data Repository XAFS database (MDR XAFS DB)

The Materials Data Repository XAFS database is a collection of publicly available data compiled at the National Institute for Materials Science (NIMS) in Japan [24]. The MDR XAFS DB is a special collection of the MDR. The collection contains data from six facilities (Japan Synchrotron Radiation Research Institute [25]; Ritsumeikan SR Center [26]; Hokkaido University, Institute for Catalysis [27]; Photon Factory, KEK [28]; Aichi SR, Aichi Synchrotron Radiation Center [29]; and SAGA-LS, Kyushu Synchrotron Light Research Center [9]). MDR XAFS DB currently holds 2,264 XAS spectra.

The database can be accessed as a special collection of the Materials Data Repository<sup>12</sup> or using the search box in the landing page. The interface for MDR XAFS DB presents the information about the DB and the XAS spectra data browsing section in a single page ([Figure 3](#)). The list in the XAS spectra browsing section can be filtered and ordered. Additionally, there are four options for viewing the data (list, gallery, tiled, and slideshow).

---

<sup>11</sup> <https://ixdb.jxafs.org/>

<sup>12</sup> <https://mdr.nims.go.jp/collections/qz20st57x>

**a**

This collection of data has been compiled at the National Institute for Materials Science (NIMS) with the cooperation of related organizations for the purpose of wide use of X-ray Absorption Fine Structure (XAFS) spectra.

The data published in this collection are publicly known data, and the data providers have guaranteed that the publication of the data does not violate the Copyright Law, the Personal Information Protection Law, or any other laws and regulations by submitting the "Application for Data Publication in the Materials Data Repository (MDR)".

The data providers as of June 2023 are as follows.

- Japan Synchrotron Radiation Research Institute
- Ritsumeikan SR Center
- Hokkaido University, Institute for Catalysis

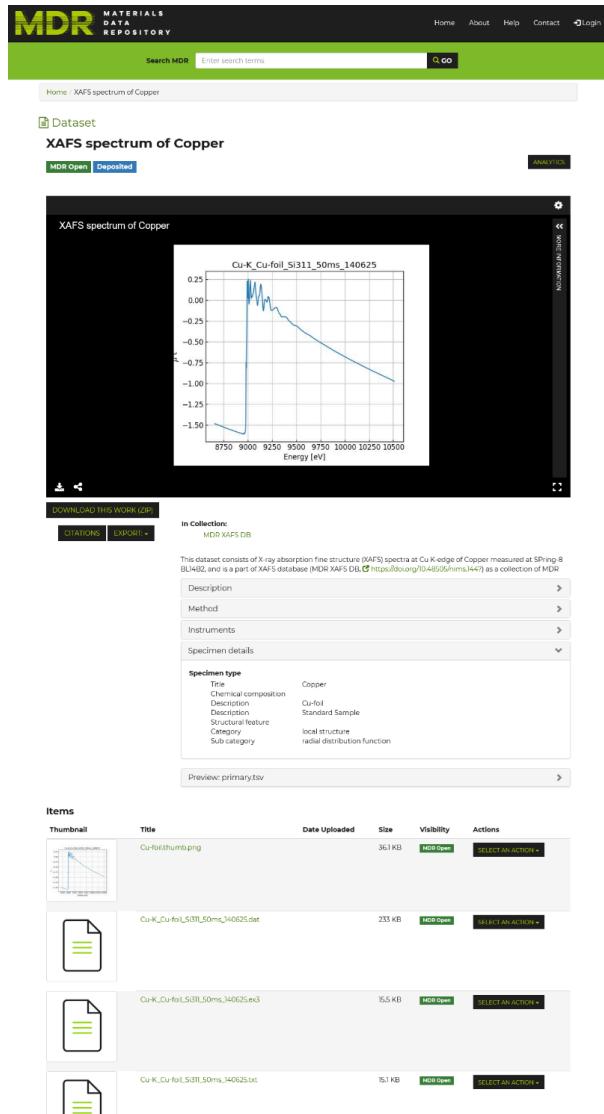
**b**

Title	Date Added	Visibility
XAFS spectrum of Terbium(III) chloride Anhydrous	24/07/2021	MDR Open
XAFS spectrum of Lithium nickel cobalt oxide	24/07/2021	MDR Open
XAFS spectrum of Barium aluminate	22/07/2021	MDR Open
XAFS spectrum of Nickel carbonate, hydrous	23/07/2021	MDR Open
XAFS spectrum of Lanthanum oxide	24/07/2021	MDR Open

Figure 3. Landing page for the MDR XAFS DB. The interface presents the information about the DB and the XAS spectra data browsing section in a single page. In the figure, both sections are cuts from the complete page, (a) presents the heading explaining the contents of the collection and (b) shows the data browsing section.

After searching and selecting a spectra, the spectra is presented in a single page ([Figure 4](#)) which shows a diagram of the spectra, the metadata corresponding to the description (name, identifiers, creator, keywords, licensing), measurement method, instrument details, and specimen details. Each record includes a collection of files which include the spectra in different formats (xml, txt), and the various representations of the metadata (JSON, YAML, XML), and an image of the spectra (PNG). Spectra data can be downloaded from this interface in text format as 9098 and column data. 9098 is a custom data schema developed in Japan and used widely for their data. The column data includes just two columns: energy and transmission, with no metadata. Each spectra has an independent license field, so licensing may vary.





*Figure 4. Spectra view of the MDR XAFS DB. The spectra is presented as a complex data object which includes several files for the spectra data, the corresponding metadata and a visualisation of the spectra.*

The full schema of the MDR XAFS DB is public and it is used to guide the formatting of data for ingestion [30]. The schema allows creating metadata files in YAML and JSON formats. Notice that these files contain metadata only, spectra are contained in separate files, e.g. simple text column data or 9098. The following elaborate on the metadata and data formats in more detail.

#### 4.2.1.1. MDR XAFS Schema

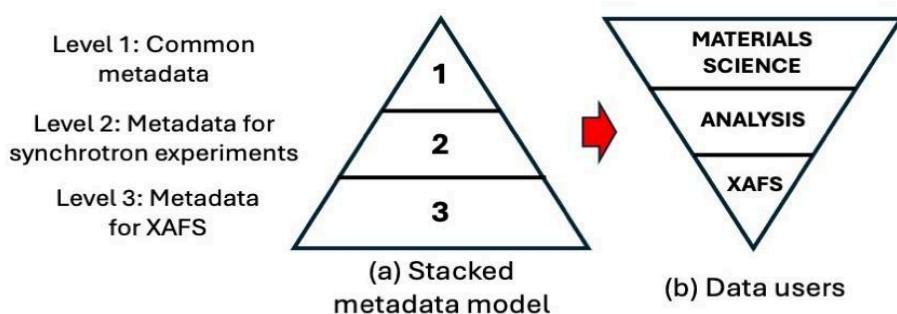
The Materials Data Repository XAFS data base (MDR XAFS DB) is a collection of publicly available data compiled at the National Institute for Materials Science (NIMS) in Japan [24].



The collection contains data from Japan Synchrotron Radiation Research Institute [25], Ritsumeikan SR Center [26], Hokkaido University, Institute for Catalysis [27], Photon Factory, KEK [28], Aichi SR, Aichi Synchrotron Radiation Center [29], SAGA-LS, and Kyushu Synchrotron Light Research Center [9].

Before the establishment of the MDR XAFS DB, there were four notable databases linked to various institutions. The BL14B2 XAFS Standard Sample Database at Spring8 [31], [32], operated by the Japan Synchrotron Radiation Research Institute, was the largest XAFS database in Japan, containing spectra on 1913 chemical substances. The Hokkaido University XAFS Database [27], the oldest in Japan, was developed in collaboration with the Japan XAFS Society and managed by Hokkaido University's Institute for Catalysis, housing approximately 300 spectra. The Ritsumeikan University Soft X-ray XAFS Database [26], accessible to the public at the university's soft X-ray synchrotron radiation facility and managed by the Ritsumeikan SR Center, includes 194 spectra from 98 samples. Lastly, the Photon Factory XAFS database [28], published by the Institute of Materials Structure Science, offers 148 publicly available spectral data.

The XAFS data formats produced in the above databases placed metadata in the header of the data file, providing some basic information necessary for data analysis. However, to have properly reusable XAFS data, a wider variety of metadata needs to be provided in an adequate format. The MDR XAFS DB adds a structured, YAML, metadata file to the corresponding data file. [Figure 5](#) [33] shows the concepts adopted for the metadata structure in the MDR XAFS DB.



*Figure 5. (a) Stacked metadata model with a hierarchy of keywords that increase in number as they become more specialized, and (b) the type of users at each level of the hierarchy.*

#### 4.2.1.2. MDR XAFS Metadata

The structure comprises a general metadata hierarchy called “stacked metadata model”, which is correlated with a hierarchical scale of users. The first level in the stacked metadata

model is metadata that is always present in any experiment, such as names, facilities and beamlines. This metadata does not require specialised knowledge, as its users are general materials scientists. The second metadata level comprises measurements specific to synchrotron radiation experiments and the particular samples used in them and can be interpreted by users with a certain level of specialized knowledge. Finally, the third metadata level is specific to XAFS experiments. It is highly specialized and has in-depth content that has little in common with the metadata in levels 1 and 2. The users of level 3 metadata are researchers specialized in XAFS.

In [figure 5](#) (a), the increasing size of each level's area indicates that the number of metadata keys increases as the hierarchy becomes deeper, because it is necessary to handle a wider variety of contents. Conversely, the decreasing size of each data user's area in [figure 5](#) (b) indicates that the number of users of each metadata level decreases with increasing metadata specificity.

There are many ways of using metadata, but accurate keywords must be defined if metadata will be used for meaningful searches. Cross searches among the member databases are one of the main purposes of the MDR XAFS DB. Hence, it is desirable that the keywords are useful for searches as wide and deep as possible.

[Table 1](#) shows how the MDR XAFS DB defines its metadata keywords [33]. The metadata associated with reproducibility of XAFS experiments and integration of XAFS spectrum are mostly necessary for making use of the data, but not when the data is being searched. Hence, the general information keywords are the ones relevant for searches in MDR XAFS DB.

*Table 1. Categories of keywords in MDR XAFS DB's metadata [31].*

Purpose	Typical keywords	Use case
General information	Date, scientist/s, facility, beamline, method, sample	Comparison with other experimental data, discover relevant data
Reproducibility and reliability of XAFS experiments	Monochromator, mirror, slit, energy, calibration, number of measurement points, step width, ion chamber gas, amplifier gain	Accuracy evaluation, detection limits, reproduction of experiments, precise analysis
Integration of XAFS spectrum data	Column name, unit, data format	Big data creation, statistical analysis, machine learning.

MDR XAFS DB uses the NIMS's materials vocabulary platform, MatVoc [34], to manage materials names linked to specific IDs that are, in turn, also used as Uniform Resource Identifiers (URI), which form a namespace of material-related lexicons [35]. In this space, the standardized name of materials, and their chemical formulas, can be found associated with their IDs.

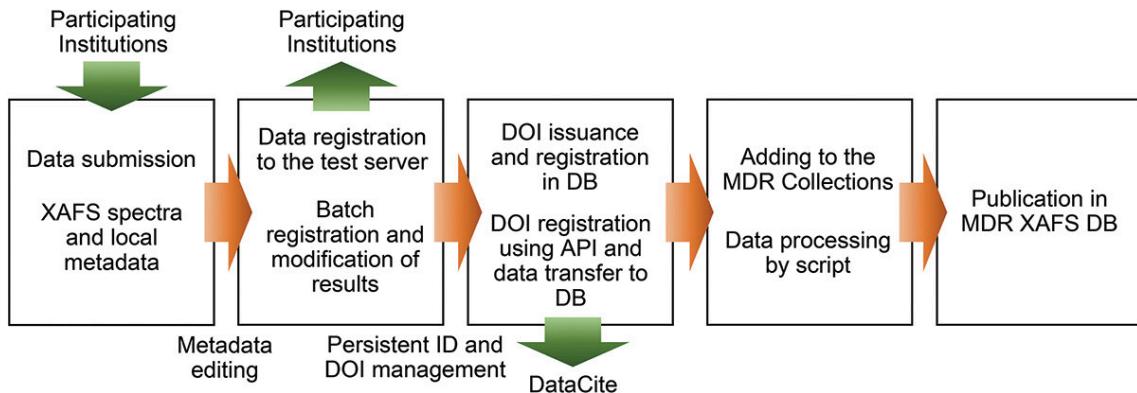
The YAML metadata files, containing the names of materials, chemical formulas, absorption edges, spectrometer crystals and other relevant information, are built using the metadata extracted from data that each participating institution provides. The MDR XAFS DB has its own schema [30]. [Figure 6](#) shows an example of a YAML metadata file created with keywords for cross searching extracted from the local metadata of each institution. The comment text after the # is for ease of understanding for the reader and the definition of the value. Here, the vocabulary used in the keywords is the nomenclature as described in MatVoc [34], [35].

```
- subject: Nickel # Material name
- subject: Ni # Chemical formula
- subject: Ni K-edge # Absorption edge
- subject: Pure metal # Material superordinate
- subject: Si(111) # Monochromator crystals
- subject: BL-12C # Beamline name
- subject: Photon Factory # Data provider
- subject: XAFS # Measurement method (fixed)
- subject: collection - MDR XAFS DB # Identification of collection (fixed)
```

*Figure 6. Spectra view of the MDR XAFS DB. The spectra are presented as a complex data object which includes several files for the spectra data, the corresponding metadata and a visualisation of the spectra.*

[Figure 7](#) [33] schematically shows how the MDR XAFS DB is managed. Data registration begins with the submission of spectral data and local metadata, including information such as data provider information and rights statements. The registration process finishes when the registration data is displayed correctly on the test server. Within MDR, after a DOI is issued electronically, the data is added to the MDR XAFS DB in the MDR Collection and eventually released to the public.

With regards to the registration of metadata, on the one hand, facility-specific metadata, such as storage ring current, and beamline-specific metadata, such as features of the monochromator crystals, are automatically defined by the institutions participating in the MDR XAFS DB. On the other hand, sample-specific metadata is input using a user interface provided by the facility, and then combined with the rest of the metadata. Hence, the MDR XAFS DB schema should incorporate keywords that can manage these diverse definitions.



*Figure 7. Registration workflow for publishing material in the MDR XAFS DB [31].*

For example, in the SPring-8 database, the metadata keywords focus on identifying individual samples for internal processing, rather than describing the data with a view of linking it with external databases. Therefore, information, such as supplier, model number, and lot number, is attached to almost all samples, and each study sample is easy to track and retain its provenance. In this database, there are examples of alloy and composites that have names but no chemical formula to identify them.

In the Hokkaido University database, the sample-specific metadata is extracted from sample names freely written by users. In many cases, these sample names incorporate experimental conditions in addition to the substance names and are written in a non-standardized way.

#### 4.2.1.3. MDR XAFS Data (9809 data format)

The screenshot in [figure 8](#) shows an example of how XAS spectra is stored in a 9809 file. These files are usually text files. At the top of the file, you have the metadata, which includes details of the sample and the instrument, and below you have columns with data. In this format, the first column corresponds to the spectrometer angle, and the fourth and fifth column correspond to the values of the incident and transmitted intensities. To plot a XAS spectra, the spectrometer angle values must be converted into energy values, and the absorption coefficient is calculated using the values of incident and transmitted light intensities.

```

9809      KEK-PF    BL9A
Fe 4um foil aft 19.02.20 10:22 - 19.02.20 10:30 Serial#KEKPF-BL9A_009665
Fe 4um foil
Ring : 2.5 GeV   416.6 mA - 413.7 mA
Mono : Si(111)      D= 3.13551 A   Initial angle= 16.14260 deg
BL9A      Transmission( 2)  Repetition= 1  Points= 301
Param file : Std-XANES      energy axis(2)  Block = 1

Block      Init-Eng  Final-Eng       Step/eV      Time/s      Num
      1        7076.20  7181.20          0.35        1.00       301
Ortec(-1)  NDCH = 3
Angle(c)  Angle(o)  time/s       2            3
Mode        0          0            1            2
Offset      0          0  4439.800  6782.000
16.22460  16.22460  1.00  2558083  3695919
16.22377  16.22380  1.00  2555481  3693252
16.22295  16.22310  1.00  2550781  3687075
16.22212  16.22210  1.00  2541422  3674535
16.22130  16.22120  1.00  2535577  3667245
16.22047  16.22040  1.00  2535750  3668245
16.21965  16.21970  1.00  2540868  3676290
16.21883  16.21890  1.00  2547400  3686295
16.21800  16.21810  1.00  2549831  3690594
16.21718  16.21710  1.00  2550312  3693064

```

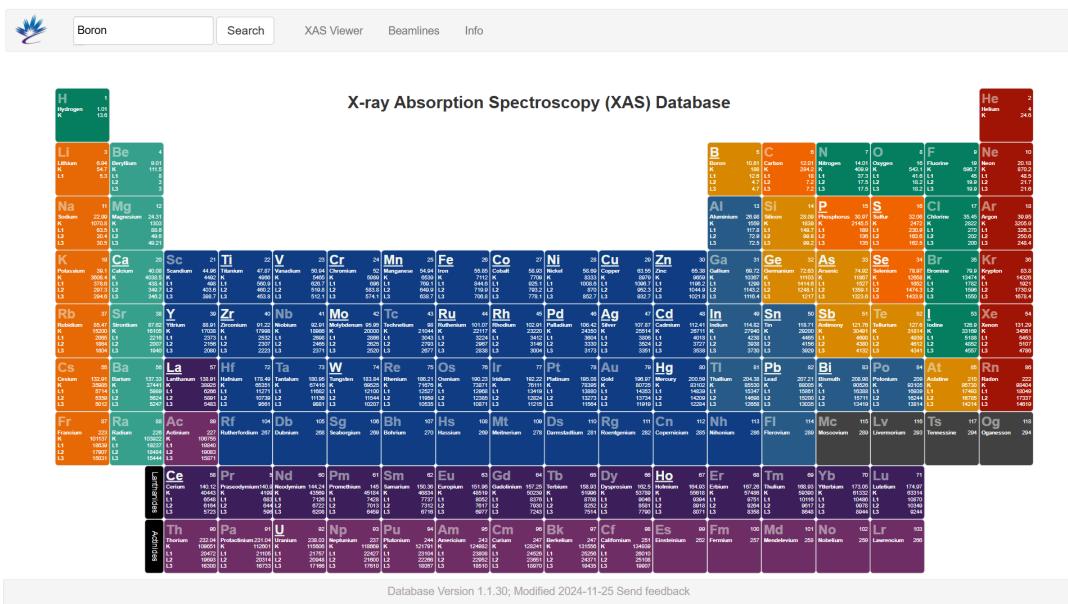
*Figure 8. An example of the 9809 data format*

#### 4.2.2. Canadian Light Source XAS Database (CLSXAS DB)

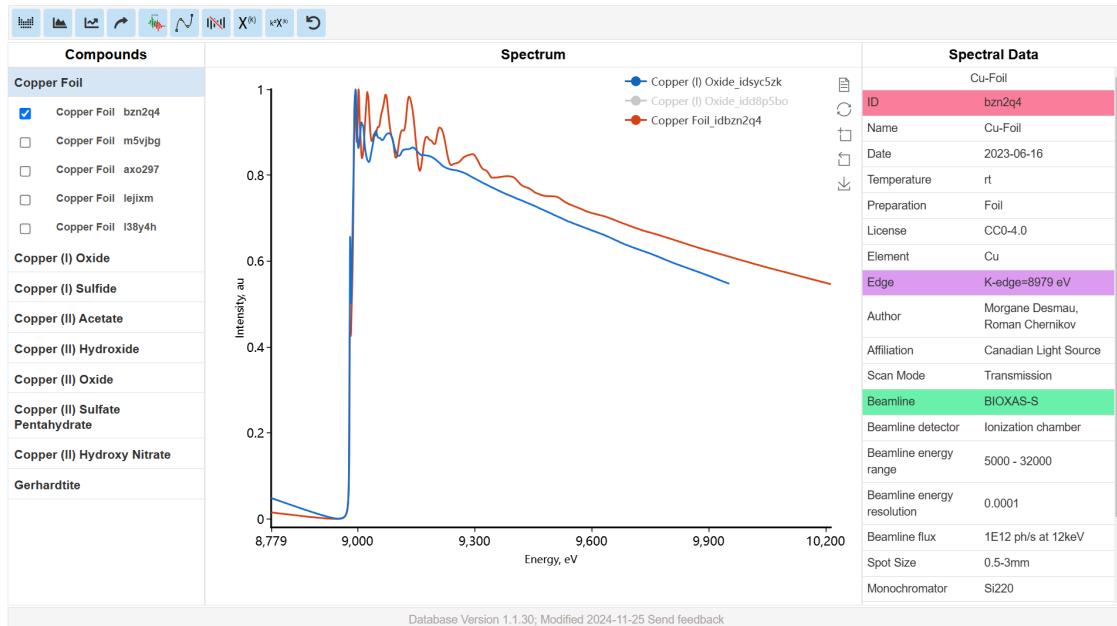
The Canadian Light Source XAS Database<sup>13</sup> currently holds 693 XAS spectra [23], [36]. This is a standalone database for XAS data. The data can be accessed through a periodic table interface or using the search box at the top ([Figure 9](#)). After selecting an element or entering a search term, the available spectra are shown in a plotting interface ([Figure 10](#)). Spectra data can be downloaded from this interface in text formats as raw and XDI. The raw data is stored as text data, with multiple columns for the detectors or fluorescence channels. This can vary depending on the scanning mode used. The XDI data is standardised and normalised to conform to XDI 1.0. The CLSXAS DB also offers a viewer which can be used to visualise XAS spectra with basic plotting options (normalised, background subtracted,  $X(k)$ ) ([Figure 11](#)). The data is licensed under CC BY 4.0. There is no information about the schema of the DB.

---

<sup>13</sup> <https://xasdb.lightsource.ca/>



**Figure 9.** CLSXAS DB Landing page. The top bar contains a search box and menu. The Periodic table allows accessing the spectra from different elements and their compounds (e.g. oxides). Only the bold and underlined elements have spectra.



**Figure 10.** CLSXAS DB Data view. After searching or selecting an element from the periodic table, the DB shows the available spectra in a plotting interface which allows viewing more than one spectrum. For instance, the image above shows the spectra for copper foil and copper oxide.



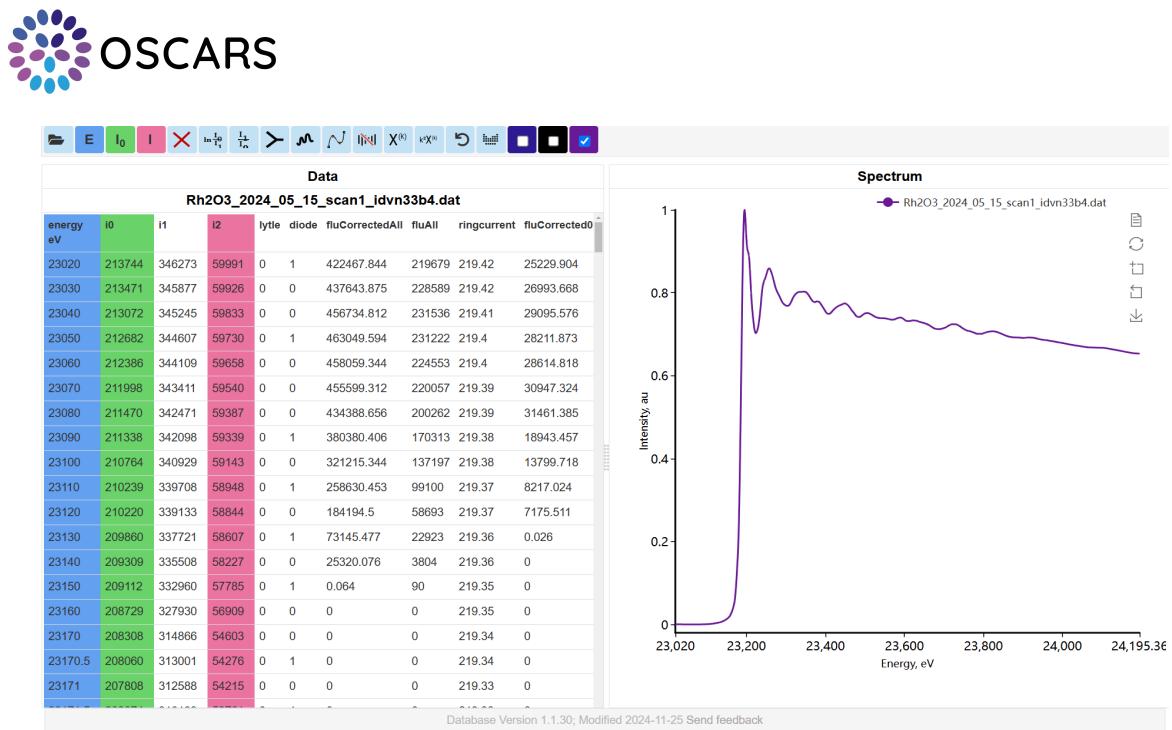


Figure 11. CLSXAS DB XAS Viewer. The top bar contains a toolbox which allows selecting the data columns ( $E$ ,  $I_0$ ,  $I_1$ ), and the plotting options.

#### 4.2.3. The SSHADE FAME Database (FAME DB)

The French Absorption spectroscopy beamline in Material and Environmental science database service<sup>14</sup> provides access to 660 spectra [37]. The data can be accessed through a search/browse interface (Figure 12)<sup>15</sup>.

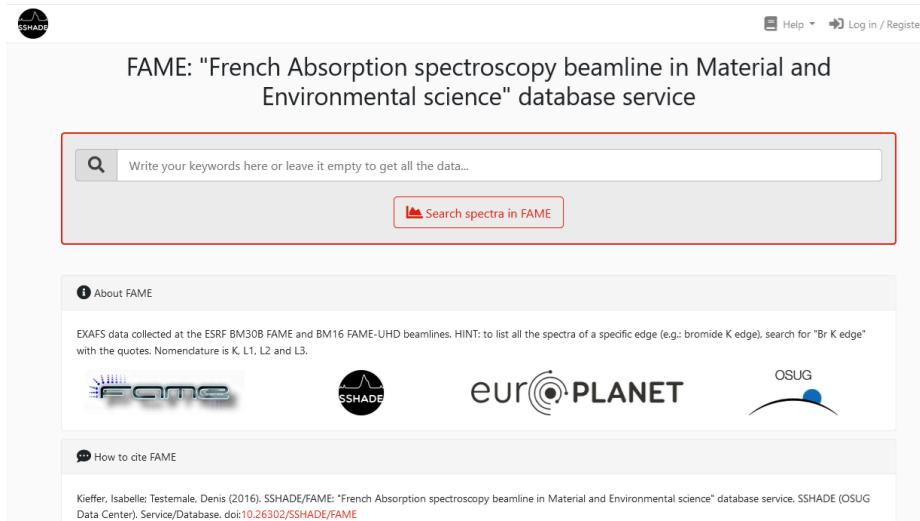


Figure 12. Landing page of the FAME DB. The search box can be used to filter the different spectra in the DB.

<sup>14</sup> <https://www.sshade.eu/doi/10.26302/SSHADE/FAME>

<sup>15</sup> <https://www.sshade.eu/db/fame>





After a search term, the available spectra are shown in a list interface ([Figure 13](#)). The list presents the spectra grouped by experiment and each experiment can include various spectra. The list interface allows downloading all the spectra from an experiment in a single step.

The screenshot shows a search results page for the FAME DB. At the top, there is a search bar with the placeholder "Write your keywords here" and a search button. Below the search bar are sections for "Spectra search" (containing a text input with "copper" and a "Search" button), "Filters" (with a "Reset all filters" button), and "Results". The results section displays a table of 17 spectra, each with a download icon. The columns in the table are: Spectra number, Title, Spectral range(s), Temperature, Type, Date created, and three small icons. The titles of the spectra include "Cu K edge XAS fluorescence of copper(II) solution at hydrothermal conditions", "Cu K edge XAS transmission of Cu metallic reference foil at ambient conditions", and "Cu K edge XAS HERFD (Kalpha1) of CuSO<sub>4</sub>.nH<sub>2</sub>O at ambient conditions".

*Figure 13. Search results page of FAME DB. The search results are presented as a list. Notice that spectra are grouped by experiment and each experiment may contain more than one spectrum record. If you click on the download button, you can download all the spectra of the experiment in a single step (requires an account registered in SSHADE).*

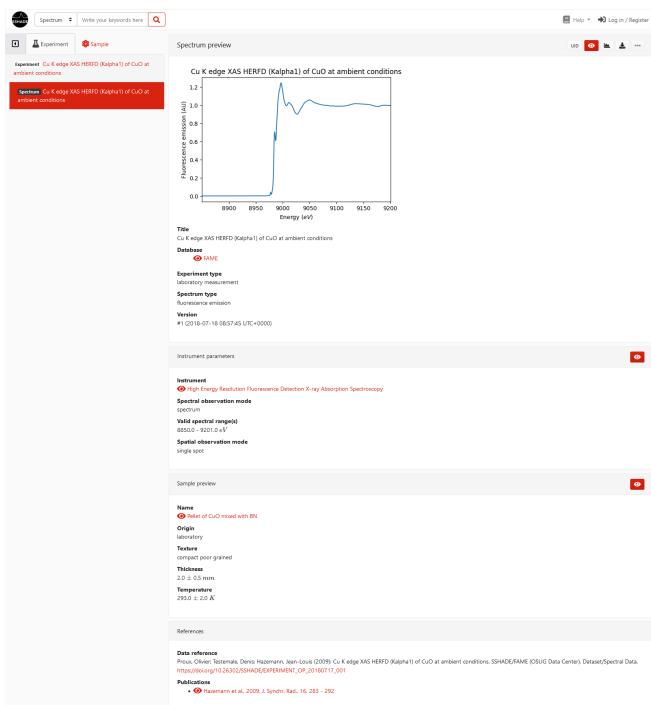
An individual spectrum can be selected from this list by clicking on their name. The selected spectra is presented in an interface ([Figure 14](#)) that includes an image of the spectrum, and shows its metadata, including name, sample, experiment, and instrument. The spectrum data can be downloaded from this interface; however, users need to create a user account with SSHADE before being able to download data. The data is exported as a complex object packaged as a zip file. The zip file contains the metadata in simple text and HTML formats. The file includes the spectrum data in a custom text in columns (energy and transmission). SSHADE does not provide any information about the licensing of the data but establishes that the intellectual property belongs to the persons listed as experimentalists or validators of the data.

The schema of the SSHADE-FAME DB is the Solid Spectroscopy Data Model (SSDM), defined by SSHADE<sup>16</sup>. SSDM is a relational data model for the description of spectral data of solid materials and it is shared among all the SSHADE databases. SSDM is device and technique

<sup>16</sup> <https://wiki.sshade.eu/sshade/documentation/ssdm>



independent by design to allow its use for different spectroscopy techniques. This model is used for guiding the preparation of data for ingestion into the SSHADE databases. The SSDM covers the description of solid samples, instruments, spectra and associated publications. Solid samples are described in detail including layers, materials, constituents and species. The description of instruments also includes techniques used for the measurements. Spectra data includes additional higher-level products (instrument specific) and band lists (band parameters and transition attributions). The SSDM also addresses the description of publications associated with the spectral data.



*Figure 14. FAME DB spectrum view showing the metadata for the spectrum including a plot. The data should be downloaded using the Download file button on the toolbar, but it requires a registered account with SSHADE.*

#### 4.2.4. IXAS X-ray Absorption Data Library (XASLIB)

The IXAS X-ray Absorption Data Library<sup>17</sup> provides access to 227 spectra [38]. XASLIB is a standalone XAS only DB. The data can be accessed through a periodic table interface or using the search and filter controls to the right of the periodic table ([Figure 15](#)). After selecting an element or entering a search term, the available spectra are shown in a list under the periodic table ([Figure 16](#)). The list shows the name of the spectra, the edge, the beamline of origin, rating, and a selection box. Various spectra data can be downloaded

<sup>17</sup> <https://xaslib.xrayabsorption.org/elem/>



from this interface by clicking on the select box and then on the Save Zip File button. The selected spectra will be grouped together in a single zip file that includes the spectra in XDI files. Clicking on the name of a spectrum will show its metadata, a plot of the spectrum and allow downloading its individual file ([Figure 17](#)), the data is available in XDI text file format. The XDI data can be in XDI 1.0 or XDI 1.1, depending on the age of the scan. The data is licensed under CC0, no rights reserved.

**IXAS X-ray Absorption Data Library**

[Log In](#) [Create Account](#)

[Elements](#) [Suites](#) [Samples](#) [Beamlines](#) [Upload Data](#) [About](#)

**Select Absorbing Element (277 spectra, 20 elements):**

H Li Be <a href="#">Show All Spectra</a> Na Mg K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe Cs Ba La Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn Fr Ra Ac  Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr	<b>He</b> Filter Spectra by Edge, Beamline, Measurement Mode, Rating, or Text in Name, Comment, or other Field: Edge: <input type="button" value="Any"/> Beamline: <input type="button" value="Any"/> Measurement Mode: <input type="button" value="Any"/> Rating: <input type="button" value="Any"/> Required Text: <input type="text"/> <input type="button" value="Search All Spectra"/>
---	---

*Figure 15. IXAS X-Ray Absorption Data Library XASLIB landing page. The main search interface is a periodic table. The spectra are provided for the element symbols indicated in blue. The right boxes allow searching and filtering the spectra by Edge, Beamline, Measurement mode, rating and keyword.*

**IXAS X-ray Absorption Data Library**

[Log In](#) [Create Account](#)

[Elements](#) [Suites](#) [Samples](#) [Beamlines](#) [Upload Data](#) [About](#)

**Select Absorbing Element (277 spectra, 20 elements):**

H Li Be <a href="#">Show All Spectra</a> Na Mg K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe Cs Ba La Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn Fr Ra Ac  Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr	<b>He</b> Filter Spectra by Edge, Beamline, Measurement Mode, Rating, or Text in Name, Comment, or other Field: Edge: <input type="button" value="Any"/> Beamline: <input type="button" value="Any"/> Measurement Mode: <input type="button" value="Any"/> Rating: <input type="button" value="Any"/> Required Text: <input type="text" value="copper"/> <input type="button" value="Search All Spectra"/> <input type="button" value="Filter Current Selections"/>
---	---

6 spectra for all (filtered from 277) - Click on heading labels to toggle sorting:

Name	Edge	Beamline	Rating	Select		
CuO_tenorite_13K_01	K	SSRL 4-3	unrated	<input type="checkbox"/>	Act on Selected Spectra:	
CuS_covellite_13K_01	K	SSRL 4-3	unrated	<input type="checkbox"/>	<a href="#">Select All</a>	<a href="#">Select None</a>
Se_CuSe_rt_02	K	APS 13-BM-D	unrated	<input type="checkbox"/>	<a href="#">Plot Spectra</a> (up to 20 spectra plotted)	
Se_Cu2Se_rt_01	K	APS 13-BM-D	unrated	<input type="checkbox"/>	<a href="#">Save Zip File</a>	

*Figure 16. IXAS X-Ray Absorption Data Library XASLIB search/filter results. The results appear under the periodic table and can be further refined using the filters. The data can also be downloaded and plotted from this view.*





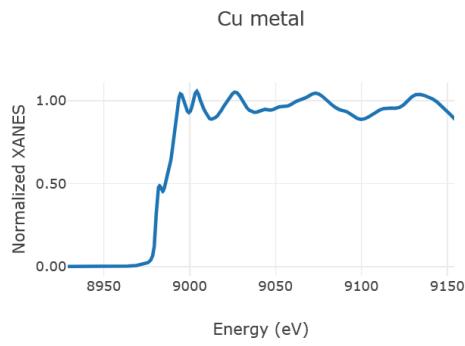
IXAS X-ray Absorption Data Library

[Log In](#) [Create Account](#)

[Elements](#) [Suites](#) [Samples](#) [Beamlines](#) [Upload Data](#) [About](#)

### Spectrum: Cu metal

Data File: Cu metal.xdi  
 Short Description: Cu metal foil  
 Ratings: No ratings  
 Suites: Mono Calibration Foils, APS-13-ID-E, 2013  
 Absorption Edge: Cu K edge  
 Measurement Mode: transmission  
 Beamline: APS 13-ID-E  
 Mono d-spacing: 3.134770 Å (None)  
 Energy Units: eV  
 Energy Resolution:  
 Sample: Cu  
 Cu, metal foil  
 Reference Sample: none  
 Date Measured: 2013-10-03 13:38:43-04:00  
 Date Uploaded: 2020-05-17 13:26:31.272626-04:00  
 Owner: Matt Newville (matt.newville@gmail.com)  
 Citation: unknown  
 User Comments: None



[Switch Plot to Raw XAFS](#)

```
unsorted file headers:
# Mono: d_spacing: 3.134770, name: None
# Beamline: name: APS 13-ID-E
# Data: upload_date: 2020-05-17 13:26:30, submitted_by: Matt
Newville
# Sample: notes: sample for 'Cu_Foil_rt_2016Foils_13IDE_01.xdi',
uploaded 2020-05-07 19:57:14, {"temperature": "room
temperature"}
# Reference: mode: no reference spectra
```

Figure 17. IXAS X-Ray Absorption Data Library XASLIB spectrum view showing the metadata for the spectrum including a plot. All data is directly taken from the stored XDI file. The spectrum can be downloaded by clicking on the file name.

The schema for the XASLIB is simple, as it accepts single spectra text files for ingestion. The instructions for formatting the data recommend the use of the XDI specification, but this is not strict. The only mandatory information, apart from the spectra data are Sample Name, Absorbing Element, Edge, Monochromator d-spacing (or nominal reflecting cut), Energy units, Beamline, and Measurement Mode. The database contains data from different USA facilities (APS, NSLS, SSRL).

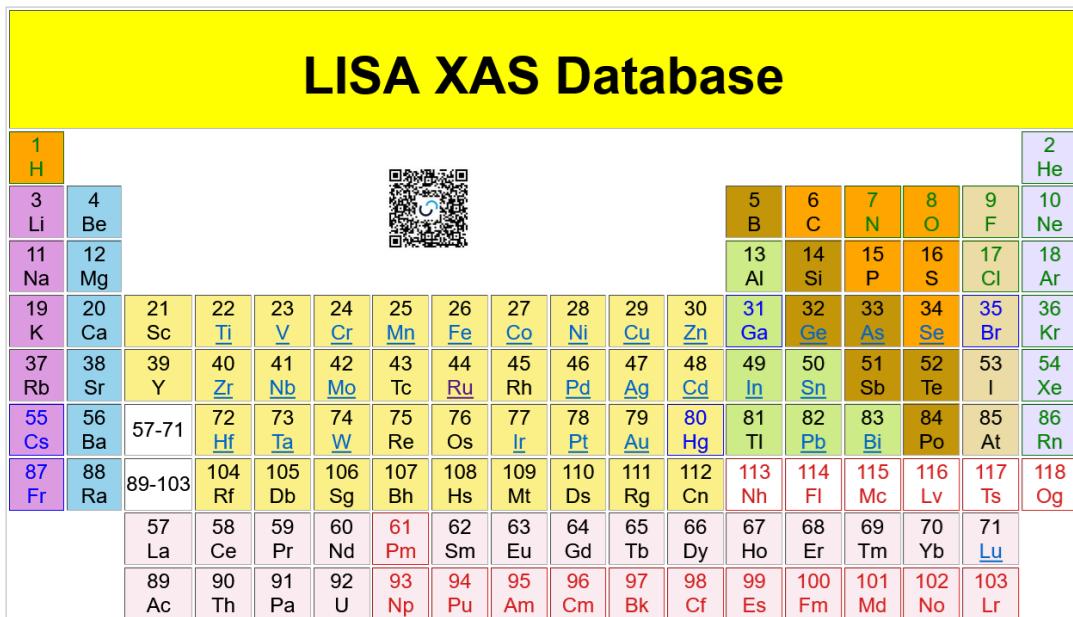
### 4.2.5. LISA XAS Database

The Italian Beamline for X-ray Absorption Spectroscopy (Linea Italiana per la Spettroscopia d'Assorbimento X - LISA) XAS database (LISA XAS DB)<sup>18</sup> ESRF-CNR. This database presents an interface for accessing 48 spectra [39], [40]. The data can be accessed through a periodic table interface (Figure 18). After selecting an element, the available spectra are shown in a simple list (Figure 19), each item of the list presented is a direct link to download the corresponding file for a spectrum, there is no preview. The spectra data is available in XDI

<sup>18</sup> <https://lisa.iom.cnr.it/xasdb/>



text file format. The XDI data is formatted to conform to XDI 1.0. The data is licensed under Creative Commons Attribution 4.0 International. In addition to the online DB, the full dataset can be downloaded from Zenodo [40]



© Copyright Public Domain. The Database may not work with safari

Created by Alessandro Puri - University of Bologna; CNR-IOM-OGG c/o ESRF LISA CRG. Periodic table template from [syed504](#)

To cite this Database DOI [10.5281/zenodo.1077809](https://doi.org/10.5281/zenodo.1077809)

*Figure 18. LISA XAS Database landing page. The main search interface is a periodic table. The spectra are provided for the elements displayed as links (underlined blue text).*



The screenshot shows a web browser displaying search results for 'Copper compounds'. The title 'Copper compounds' is in bold. Below it is a bulleted list of spectra names:

- [Cu\\_metal\\_foil\\_111](#)
- [Cu<sub>2</sub>O\\_111](#)
- [CuO\\_111](#)
- [CuFeS<sub>2</sub>\\_80K\\_111](#)

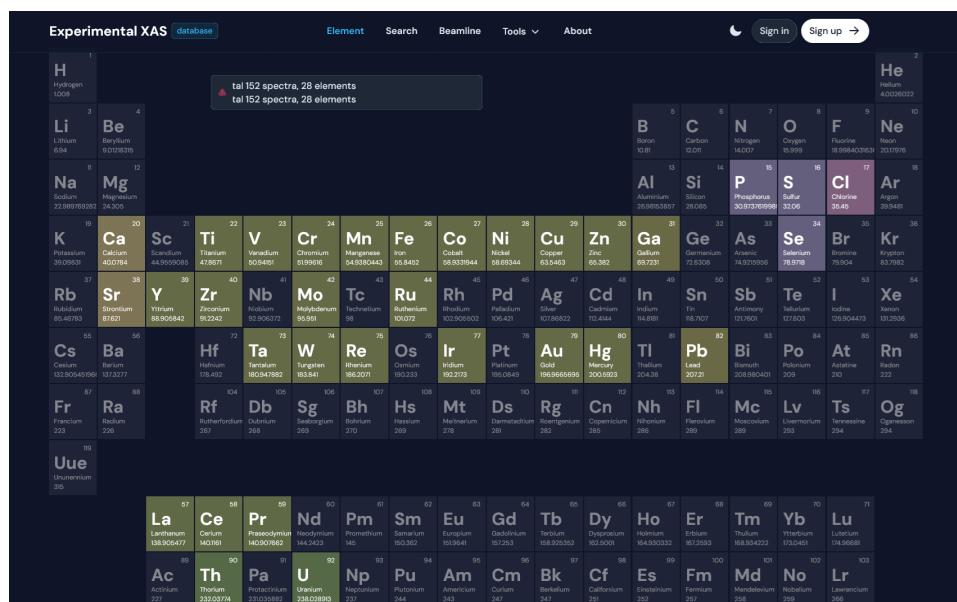
*Figure 19. LISA XAS Database search results. The results appear as a list in a single page. The data can be downloaded by clicking on the spectra names displayed as links. There are no previews of the spectra.*



There is no published schema for the LISA XAS DB. Looking at the Zenodo image [40], it is a simple file storage for XDI files. The linking and organisation of this data is manual using html pages.

#### 4.2.6. Experimental XAS database

The Experimental XAS database (E-XAS DB)<sup>19</sup> of the Institute of High Energy Physics, Chinese Academy of Sciences (IHEP-CAS) presents an interface for accessing 152 spectra [41].The data can be accessed through a periodic table interface ([Figure 20](#)). After selecting an element, the available spectra are shown in a list under the periodic table ([Figure 21](#)). The list shows the name of the spectra, the edge, the beamline of origin, collection data, and rating. Clicking on the name of a spectrum will show its metadata, a plot of the spectrum ([Figure 22](#)). At the moment of writing the download button is not operational or takes a long time to process. Only the images of the spectra can be downloaded for the moment. There is no information about licensing of the data. Additionally, the database schema is also not public.



*Figure 20. The periodic table search interface of the E-XAS DB. Spectra provided for green and greyish-blue coloured elements.*

<sup>19</sup> <http://xasdb.ihep.ac.cn/>



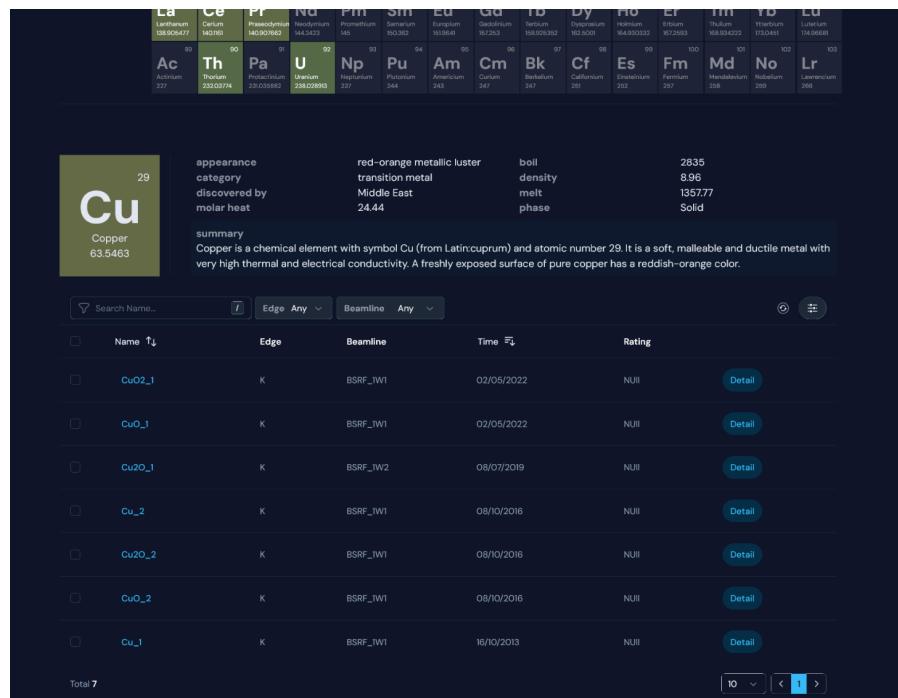


Figure 21. E-XAS DB search results appear under the periodic table. Spectra is accessed by clicking on names.

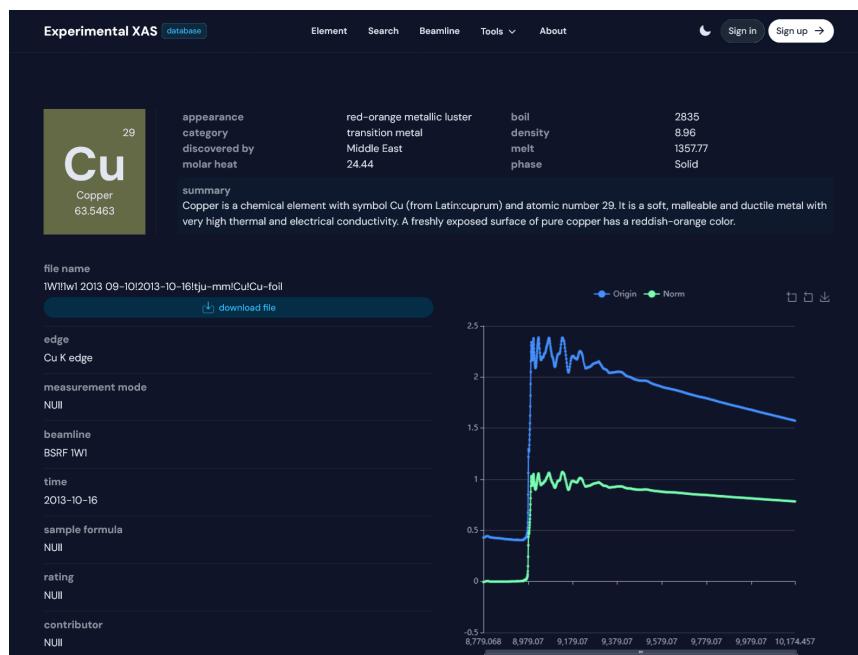


Figure 22. E-XAS DB spectrum view showing the metadata for the spectrum including a plot. The plot shows both the original spectrum and the normalised one. The “Download file” button appears non-functional at the moment.

#### 4.2.7. XAS Reference Research Database (RefXAS DB)

The XAS Reference Research Database<sup>20</sup> from the DAPHNE4NFDI project currently lists 42 spectra (each designated as a dataset) [42], [43]. The list indicates that it stores 63 but will only show 42 verified records. The data can be accessed through a direct search using the search box at the top, or by clicking the browse all datasets button in the centre of the page ([Figure 23](#)).

After entering a search term or clicking on search, the available spectra are shown in a list ([Figure 24](#)), the list can be filtered using the buttons on the left side of it.

Each spectra can then be accessed by clicking on the more button on the right side. The spectra view ([Figure 25](#)) shows views of the spectra in different modes (Raw, Normalised, Chi(k), and Chi(R)), an interactive plot of the spectra (can be scaled, copied, downloaded, etc), and the corresponding metadata (Sample, Instrument, Bibliography, Experiment). This page also allows downloading the data as a zip file. The zip file includes four versions of the data as: JSON, YAML, TXT, and HDF5. The schemas of the data appear to be custom, they do not indicate adhering to a known standard, but rather being different conversions of the DB schema [44]. The contents of the TXT format resemble the XDI schema, as the file includes a header section for the metadata, and the data section under it ([Figure 26](#)).



*Figure 23. RefXAS DB landing page. Access to spectra is through the search box at the top and the Browse all datasets button at the centre of the page.*

<sup>20</sup> <http://xafsdb.ddns.net/>

**≡ RefXAS**    Home    Upload    Search all DataSets (with filters)    Team/Contact    Try to find a spectrum... 

Element:    A total of **5** datasets match the search:  
(number of records displayed may be lower due to non-visible records that are not yet verified):

Nickel: 3    Zinc: 5    Gold: 6    More...    Filter applied: Zinc     Reset All Filters

id	Title	More
PID.SAMPLE.PREFIX2...	Zn_foil_Zn_K_300.00001.xdi 2024-06-19-13_34_56 Description: Zn foil - XAS CONTINUOUS MODE	
PID.SAMPLE.PREFIX0...	Zn_foil_Zn_K_0000.xdi 2024-11-25-10_02_08 Description: Zn foil at RT	

1 of 1  
 © 2023 - 2024 Sebastian Paripsa (University of Wuppertal),  
 Frank Förster (TU Berlin) and Abhishek Gour (KIT).

Figure 24. RefXAS DB search/filter results appear as a list. Spectra data can be accessed by clicking the "More" button on the right of each dataset.

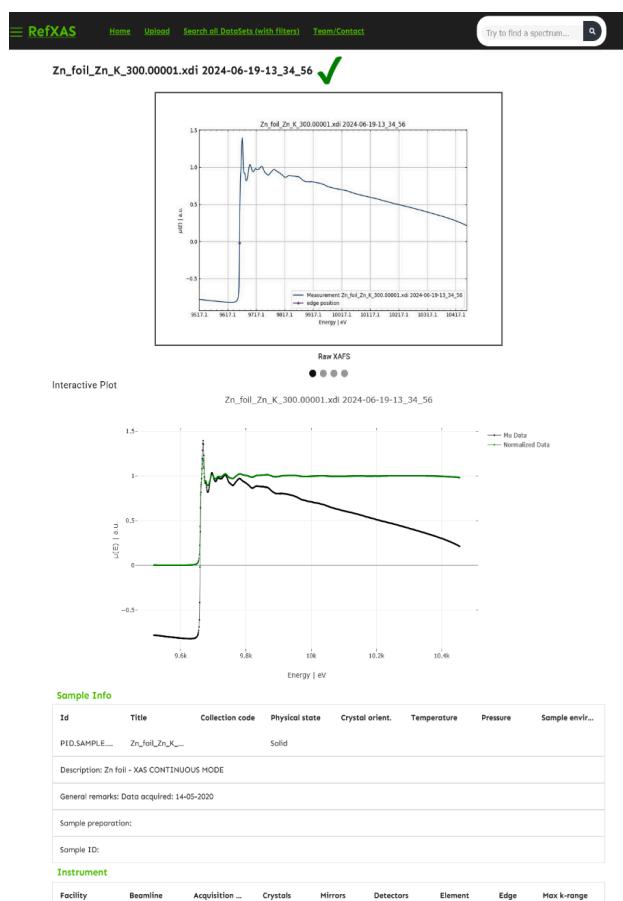


Figure 25. Spectra view of the RefXAS DB. The spectra is presented as in two plot views along with the corresponding metadata from the RefXAS schema. The download link is at the bottom of the page (not shown).

```

#   - Doi:
#   - Reference:
#   - Disclaimer: Yes
# -----
# *Instrument*:
#   - Facility: DESY Petra III
#   - Beamline: P65
#   - Aquisition Mode: XAS CONTINUOUS MODE
#   - Crystals: Si 111
#   - Mirrors:
#   - Detectors:
#   - Element Input: Zn
#   - Edge Input: K
#   - Max K Range:
# -----
# *Results quality control*:
#   - Edge step: 1.88 nm.
#   - K-max: 14.4 Å-1
#   - Energy resolution: 0.9 eV
#   - Edge energy: 9659 eV
# -----
# *Update / Verification Info*:
# (The following entries have been altered)
#   - General Remarks: Data acquired: 14-05-2020
#   - Max K Range: 14.4
#   - Updated by: Pacinsa, paripsa@uni-wuppertal.de, 06/19/2024
# -----
# *Data*:
# Energy, Mu, Normalized:
9517.132130384254 -0.7819263244053435 0.00053247778872702
9518.049182201748 -0.782282399569089 0.0005401872709240756
9518.972530182353 -0.7827563400713364 0.00048642440277403245

```

*Figure 26. View of the contents of the TXT file from the Zinc spectra downloaded from the RefXAS database. Metadata is contained in the top part of the file, those starting with a hash (#). The rest of the lines contain the data in columns. The title and contents of each column are indicated by the names in the last header line, they are Energy, Mu and Normalised.*

#### 4.2.7.1. The RefXAS DB Schema

The RefXAS DB schema is publicly available [26], and it is used to guide the quality control and ingestion processes for the database. An output of DAFNI4NFDI [45], RefXAS is a web based reference database for XAS spectra intended to provide quality, metadata-rich spectra for the comparison steps in XAS analysis pipelines. Lack of metadata around sample preparation and experimental conditions are noted as shortcomings with previous XAS databases. The schema for RefXAS is shown in [Table 2](#).

*Table 2. The schema for RefXAS*

Entity	(Meta)data field	Description/example
Sample	Sample ID	Provides a unique identifier for each sample and process
	Collection code	Chemical symbol of element with edge (K, L, M etc.) and scan number

<b>Entity</b>	<b>(Meta)data field</b>	<b>Description/example</b>
	CAS No. (if available)	Standardized identifier for chemical compounds
	Physical state	<i>E.g.</i> crystalline, powder, thin film, liquid, gas
	Structural parameters for crystalline samples	Can affect the diffraction pattern, for now only crystalline or single-phase samples can be used, in future we aim for more complex systems
	Crystal orientation	<i>E.g.</i> Miller indices
	X-ray or neutron diffractogram (if available)	Provides information about the crystal structure
	Temperature, pressure and environment	Can impact the XAS spectrum
	Remarks about sample preparation	<i>E.g.</i> foil, pellet, capillary, powder on tape
	Sample environment	<i>E.g.</i> cell/microreactor/batch, gases, solvents, potential
<b>Spectra</b>	Raw data file	Provides the original data acquired by the instrument
	Raw absorbance spectrum	Processed XAS spectrum
	Normalized absorbance spectrum	Normalized XAS spectrum/XANES
	$X(k)$	Represents the extended X-ray absorption fine structure (EXAFS) oscillations in $k$ -space (wavevector space)
	$X(R)$	Represents the EXAFS oscillations in $R$ -space (radial distance space), obtained by performing the Fourier transform of the EXAFS oscillations in $k$ -space without employing corrections for phase shifts
	Reference spectrum <sup>†</sup>	If available, used as a standard to check the energy scale calibration and for comparing beamline parameters
<b>Instrument</b>	Facility	Whether synchrotron or laboratory instrument

Entity	(Meta)data field	Description/example
	Beamline	<i>E.g.</i> P65, CAT-ACT <i>etc.</i> , preferentially a DOI
	Acquisition mode	Refers to the way in which data are collected from the sample ( <i>e.g.</i> 'continuous scan', 'step-scan' or 'energy-dispersive' measurement mode), absorption, fluorescence or total electron yield mode
	Crystals	<i>E.g.</i> Fe <sub>2</sub> O <sub>3</sub> (hematite), DCM, CCM, polychromator, specify materials and orientation [ <i>e.g.</i> Si (111), Ge(220), <i>etc.</i> ]
	Mirrors	Optical devices that are used to focus, collimate or deflect X-ray beams; specify mirror material, incident angle and form of the mirror (plane, bent)
	Detectors	Devices used to measure and record the intensity of X-rays that have interact with a sample during an experiment, <i>e.g.</i> gas-filled ionization chambers, solid-state detectors
	Element	<i>E.g.</i> Fe
	Absorption edge	<i>E.g.</i> K-edge
	Maximum <i>k</i> range (EXAFS)	Refers to the upper limit of the wavevector ( <i>k</i> ) values considered during the analysis of the EXAFS oscillations
	Resolution	Energy resolution, which is the ability of a spectrometer to distinguish between two closely spaced X-ray photon energies
<b>Bibliography</b>	DOI	Identifier
	Title	<i>e.g.</i> <sample-name> + <time-stamp>
	Author	<Full name>
	Literature reference	Provides more information about the data

Users can submit a data file and its associated metadata via a web form. TXT, ASCII, SPEC and XDI formats are supported, with HDF5 (and by extension NeXus) under development. To ensure a minimum standard of quality for the database, automated checks are applied to

uploaded data ([Table 3](#)). This is processed using the Larch library to extract absorption and energy values, find the position of the edge, normalise the spectra, subtract the background, calculate  $\chi(k)$  and finally  $\chi(R)$

From this process, the normalised spectra, edge step, maximum value of  $k$ , and  $\chi(k)$  are stored and displayed as quality criteria. This process is tailored to metal foils, and can have issues, for example due to data artifacts or multiple edges in the same spectra. To account for this, the user is shown the automated criteria and there is always a manual curation step performed before the data is finally accepted. In addition to reviewing the plots and values described above, quantitative limits on quality criteria are also applied. Spectra with all values outside these ranges are rejected outright without being manually reviewed.

*Table 3. RefXAS quality criteria*

Quality criteria	Range	Description
Edge step (transmission mode)	0.5–2.0	Refers to the steepness of the onset of the absorption edge and is an indicator of the quality of the sample and the measurement. A high edge step, characterized by a steep onset of the absorption edge, is preferred as it reduces background noise in the EXAFS data and improves the accuracy of the measurement results.
Maximum usable $k$ -range	15–20 Å <sup>-1</sup>	Refers to the range of wavevector ( $k$ ) values used to examine the EXAFS oscillations. The choice of $k$ -range directly impacts data quality, structural resolution and the accuracy of derived results.
Energy step or point density as calculated from spectral points	0.5–2 eV	Refers to sampling frequency in the energy domain in the XANES region.
Amplitude reduction factor ( $S_0^2$ ) <i>Specifically for metals foils where the structure is known</i>	0.7–1.0	Correction factor that accounts for the reduction in the EXAFS oscillation amplitude. Essential for accurately determining the coordination numbers from the EXAFS data.
Signal-to-noise ratio	0.1% of edge jump	Quantifies the proportion of meaningful signal to the background noise present in the data.

#### 4.2.8. Summary

The seven databases described in this section are all openly available and allow downloading data. [Table 4](#) shows an overview of these databases, indicating if the schema is public

(verifiable), the main uses of the schema, the type of format used for outputs (downloads) and the location of the metadata in these outputs.

All of the databases support at least an export format which is text based and that can be mapped and used in current processing software. None of these databases report providing data in complex formats such as Nexus (or HDF5). As things stand, combining data as provided by the different databases is complicated, since the formats vary. Depending on the origin, special processes may be needed to parse and transform the data. For instance, MDR XAFS DB, SSHADE/FAME and RefXAS provide data in custom text formats. The files can be interpreted by human users, but there is some processing needed to identify column contents correctly. These three databases are also the ones with the most complete data models. Their models are similar and cover different aspects of the generation of XAS data, including the facility, instrument, sample and experiment (scanning) details.

XDI is the only data schema which is not dependent on the database. XDI is used by CLS DB, XASLIB and LISA. This is not surprising as it is natively supported by the Demeter software suite [46], which is widely used for processing and analysis of XAS data (Athena, Artemis, Ephesus).

*Table 4. Overview of XAS Databases revised.*

Database	Data Schema	Use of Schema	Output	Metadata
MDR XAFS DB	Public	Guide data curation/ingestion	Text 9809	External
CLS DB	Not public	NA	XDI	In file
SSHADE/FAME	Public	Guide data curation/ingestion	Text	In file and external
XASLIB	Public (Simple)	Guide data curation/ingestion	XDI	In file
LISA DB	Public (Simple)	NA	XDI	In file
EX-XAS	Not public	NA	NA	
RefXAS	Public	Guide data curation/ingestion	Text	In file and external

### 4.3. Software Dependent XAS Schemas

Software dependent schemas are used to manage data and parameters for the processing of XAS data. The types of schemas used to feed data and output data from software can vary

also. In this case we also contemplate software which is supported and used openly for processing XAS data: Demeter, Larch, and DAWN.

### 4.3.1. Demeter

Demeter is one of the most popular and widely used desktop programs for processing and analysing X-ray Absorption Spectroscopy data [46]. Demeter is better known for its two processing and analysis interfaces Athena and Artemis. Athena allows basic processing of XAS data and XANES analyses, while Artemis is used for EXAFS analyses.

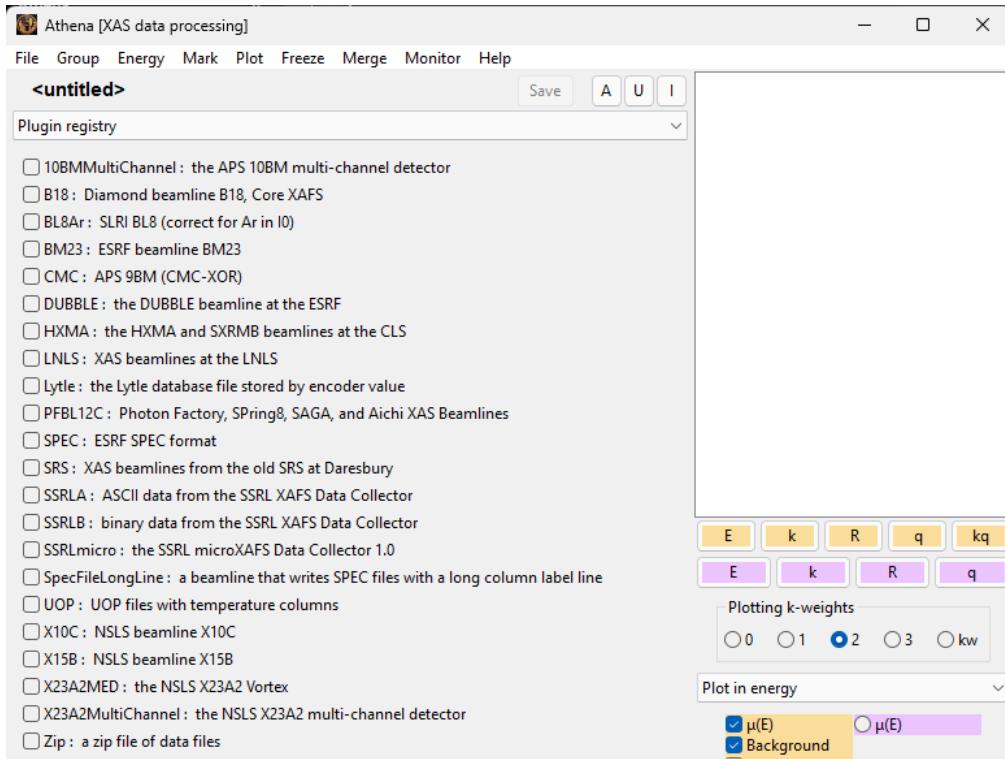
#### 4.3.1.1. Demeter Inputs

The main input of Athenea is text data. However, considering that many facilities have created custom formats, Demeter allows creating beamline specific plugins<sup>21</sup>. The plugins enable reading text data into Athena, taking into account a range of variations such as adding multiple columns. Version 0.9.26 has 22 plugins for reading text data from different beamlines ([Figure 27](#)). The list indicates that even within a facility, different beamlines will present the information using different data schemas. For instance, the BM23 and DUBBLE from ESRF require independent plugins in order to correctly import data into Athena, and there are at least three alternative specifications for importing BM23 data. The Athena plugin registry lists 21 plugins from 12 facilities, some with more than one beamline, and even more than one schema per beamline. Some plugins are kept to guarantee access to data from facilities no longer in operation, such as the Lytle database, the SRS, and the NSLS old beamlines.

The issues addressed by developing plugins are recurrent, especially after the upgrade or commission of new equipment which may be seen as an opportunity to implement changes to data management practices. In this case, a standardised data schema could prevent the appearance (need for) of custom formats.

---

<sup>21</sup> A plugin is a software component that adds new features or extends the functionality of a program.



*Figure 27. Custom plugins for reading data from different beamlines. Notice that in some cases data schemas vary even within facilities. For instance, the Stanford Synchrotron SSRL has three plugins, because of variations between beamlines.*

#### 4.3.1.2. Demeter Outputs

The main outputs from Athena (.prj) and Artemis (.fpj) are project files. These project files contain the data imported, the parameters associated with each data file, the content of the journal, and several other collections of important data. All of this gets saved in a single file. Demeter project files, especially Athena (.prj) files, are popular in the research community for processing, analysing, and sharing XAS data. Additionally, Athena allows exporting individual spectra groups as XDI including data for absorption ( $\mu(E)$ ), normalised absorption ( $norm(E)$ ), shown in [Figure 28](#)), normalised absorption coefficient ( $\chi(k)$ ), Fourier transform ( $\chi(R)$ ), or ( $\chi(q)$ ).

```

# XDI/1.0 Athena/0.9.26
# Element.edge: K
# Element.symbol: Rh
# Column.1: energy eV
# Column.2: norm
# Column.3: nbkg
# Column.4: flat
# Column.5: fbkg
# Column.6: nder
# Column.7: nsec
# Athena.e0: 23224.387
# Athena.eshift: 0
# Athena.rbkg: 1.0
# Athena.importance: 1
# Athena.standard: None
# Athena.bkg_kweight: 2
# Athena.edge_step: 1.0109109
# Athena.fixed_step: no
# Athena.pre_edge_range: -124.709 -60.000
# Athena.pre_edge_line: 2.844356 - 0.000122994 * E
# Athena.normalization_range: 150.000 774.827
# Athena.post_edge_polynomial: -3.613226 + 0.000390185 * E - 8.24990365e-009 * E^2 + 0 * E^3
# Athena.spline_range_energy: 0.000 874.823
# Athena.clamps: 0 24
# Athena.spline_range_k: 0.000 15.153
# Athena.kweight: 2
# Athena.window: hanning
# Athena.phase_correction: no
# Athena.k_range: 3.000 13.153
# Athena.dk: 1
# Athena.r_range: 1 3
# Athena.dr: 0.0
# Athena.window: hanning
# Athena.plot_multiplier: 1
# Athena.y_offset: 0
# /**
# -----
#   e      norm      nbkg      flat      fbkg      nder      nsec
23099.678  0.27981433E-02  0.27981433E-02  0.27981432E-02  0.27981433E-02  -0.16993422E-02  0.58264162E-03
23100.353  0.16510874E-02  0.16510874E-02  0.16510873E-02  0.16510874E-02  -0.13060591E-02  0.10795664E-02
23101.028  0.10349636E-02  0.10349637E-02  0.10349636E-02  0.10349637E-02  -0.24678753E-03  0.13769165E-02
23101.704  0.13176774E-02  0.13176775E-02  0.13176774E-02  0.13176775E-02  0.55415508E-03  0.37939266E-03

```

*Figure 28. Example of normalised spectra data exported from Athena as XDI.*

#### 4.3.2. X-ray Larch

X-ray Larch (or simply Larch) is a suite of open-source Python libraries and applications for analysing and processing X-ray spectra data. It is used for processing and analysing X-ray absorption and fluorescence spectroscopy (XAFS and XRF) data [47].

##### 4.3.2.1. Larch inputs

There are various schemas which can be used to provide XAS data including: text column data, XDI, HDF5, NeXus, Athena, NetCDF, and Epics. Additionally, Larch includes special libraries for reading custom data from APS beamlines, and ESRF Spec/BLISS HDF5 files. The Larch user guide<sup>22</sup> states that the preferred data format for larch is the XAFS Data Interchange (XDI). In addition to XDI, Larch also allows reading simple plain text (ASCII or UTF-8) column data files, which are a straightforward way to store small amounts of numerical data widely used in the XAFS community. Due to the popularity of Athena for processing XAS data (see Demeter above), Larch also allows reading data from Athena projects. Finally, Larch also allows reading its own data format, larch session files (.larix). Larch does not offer the functionality for defining import schemas as Demeter, however, they could be created as Larch (python) scripts.

<sup>22</sup> <https://xraypy.github.io/xraylight/data.html>





#### 4.3.2.2. Larch outputs

Larch also allows saving data in several formats including simple text columns, Athena project files, HDF5 and Larch session files (.larix). There are several special functions to convert Athena files to HDF5 or adding XDI data to HDF5 groups. A Larch Session File (.larix) stores all user-generated data from a Larch session, including input data arrays, processed arrays, dictionaries, Journals, processing parameters, analysis results, fit histories, and a list of executed Larch commands. It also contains session configuration details (e.g., Larch and Python versions, operating system). These files allow saving sessions as "Projects", making it easy to share or resume analysis later. They are designed to be portable across different computers and versions.

#### 4.3.3. DAWN

The Data Analysis Workbench (DAWN) [47] is an open-source software (licence) for the visualisation and processing of scientific data. DAWN was specifically developed for data from synchrotron-based techniques by Diamond Light Source. DAWN loads data from many common formats (text files, tiffs, hdf5...), as well as those specific to X-ray techniques (NeXus, EDF, MAR...). DAWN supports reading NeXus data by allowing the exploration of the entire structure of the HDF5 file. DAWN can also be used to export processed data both in NeXus (.nxs) and text format (.dat).

#### 4.3.4. Summary

The data file formats accepted by popular processing and analysis software are relevant because they reflect commonly used data formats from the source and common outputs. Demeter is particularly significant as it is currently the only software suite producing XDI files as output. Additionally, due to its popularity for processing and analysing XAS data, Demeter project files (especially Athena .prj files) have become a popular medium for sharing processed XAS data. For instance, researchers often share them as supporting data for published articles.



## 5. Community Standardisation Efforts

Historically there have been several attempts to produce standards for XAS data. The main idea behind these efforts is to allow the retrieval of data produced by different devices. Without a standard format, the longevity of the data is reduced to the lifetime of the associated programs and those users that know how to interpret the data. Community defined XAS schemas are used by various facilities and research groups. Currently, the most popular schemas for XAS data are the XDI, NeXus [48] and the 9809 [13, 14] format, which can be transformed into Athena format. Moreover, measurements that require a variety of experimental environments, such as operando experiments, accumulate data that was produced under different experimental environments. Given the diverse range of users involved, the integration of XAS data from different sources requires the cooperation of researchers, and the development of a set of requirements for data integration and the functionality of the infrastructure used to store and make this data browsable. Examples of such requirements were those proposed by Ishii et. al [33]:

- The benefits of data integration should be not only in data-driven science but also in everyday research.
- The data and metadata should be in as few formats as possible (ideally following and agreed data schema).
- The publication infrastructure should be prepared as a repository with policies for data utilization, such as the FAIR Principles.
- The database infrastructure should have search functionality and not just storage online.

### 5.1. Community defined standards

The International Workshop on Improving Data Quality and Quantity in XAFS Spectroscopy (Q2EXAFS), is the result of a joint initiative from IUCr and IXAS. The interest group within IUCr and IXAS organises periodic workshops focused on X-ray Absorption Spectroscopy (XAS) standards and criteria. The first workshop Q2EXAFS 2011 [51], was held during the 22nd IUCr conference. From then, three additional in-person workshops have been held: 2nd Q2EXAFS<sup>23</sup>, 3rd Q2EXAFS<sup>24</sup>, and 4th Q2EXAFS<sup>25</sup>. Additionally, during 2025 a virtual workshop was held from July to October<sup>26</sup>. The workshops intended to establish new standards and

<sup>23</sup> <https://indico.desy.de/event/10169/>

<sup>24</sup> <https://www.diamond.ac.uk/Conference/Q2XAFS2017/Scientific-Programme.html>

<sup>25</sup> <https://www.iucr.org/resources/commissions/xafs/iucr-2023-satellite-q2xafs2023>

<sup>26</sup> <https://sites.google.com/view/q2xafs/>

criteria for XAFS experiments and analyses as well as to discuss new data formats, databases and ideas for data deposition. Among the proposals from the Q2EXAFS working group, there is a plan to propose NXxas and XDI as standards for XAS data [52] during the International XAFS Conference 2025 (Chicago, USA, July 14 to July 19)<sup>27</sup>. NXxas will be recommended as the preferred standard for multispectral and XDI for single spectra. The idea is to be able to homogenise the creation and reading of data from different sources, minimising the need for specialised plugins or mappers for each XAS instrument.

### 5.1.1. The NeXus data format

The main motivation for the development of the NeXus format in the 1990s was to prevent scientists from needing to write software to manage the custom file formats in use locally at different instruments and facilities [48]. In response, a self-describing and extensible format was proposed. HDF5 became the default recommended storage format for NeXus data due to increased performance requirements at facilities, its presence in the public domain, and increased support for HDF5 elsewhere such as in third party software packages. In addition to being designed to support both raw and processed data, NeXus also defines standards to facilitate the exchange of data between applications. These include both generic and domain specific specifications (see NeXus Standards).

A NeXus file can contain one or more entries, each of which correspond to one measurement (either under fixed conditions, or a scan) and include both data and metadata (such as information about the beamline and parameter logs). The same file can be modified in place or extended with more information (such as the outcomes of processing steps) without the need to write a completely new file. Another feature of the format is the use of hard links, so that information somewhere within the NeXus file structure can be accessed at other (more accessible and intuitive) places in the hierarchy. It is also possible to link to locations within other files, allowing (for example) data written by commercial detector software in one file to be referenced in a second file containing additional metadata written by the acquisition software.

The NeXus standards for structuring data can be split into base classes and application definitions. All NeXus classes begin with NX by convention, and when saved in HDF5 format will correspond to the underlying format's storage elements: groups, data fields and attributes. [Figure 29](#) shows an example of the nexus hierarchy using the base classes.

---

<sup>27</sup> <https://xrayabsorption.org/events/xafs-2025/>

root of HDF5 file Root level of a NeXus data file	
<b>NXentry</b>	required
All data belonging to one scan or run. A given NeXus file can contain multiple related scans or runs	
<b>NXinstrument</b>	
The data needed to describe an instrument. Contains groups for each relevant instrument component.	
<b>NXsource</b>	
<b>NXcollimator</b>	
<b>NXattenuator</b>	
<b>NXdetector</b>	
...	
<b>NXsample</b>	
All the information about the sample	
<b>NXmonitor</b>	
Incident intensity monitor	
<b>NXuser</b>	
User information	
<b>NXdata</b>	required
Links to plottable data in the NXdetector group -- one instance for each detector bank. Provides support for generating a view of the data automatically.	
...	
<b>NXentry</b>	
Additional NXentry groups are optional	

*Figure 29. Basic example of a NeXus hierarchy using only base classes. [48]*

In addition to base classes, application definitions describe the content required for a particular use case. There are two existing NeXus application definitions specific to XAS NXxas [53] and NXxasproc [54], as well as a third under development [55].

### 5.1.1.1. NXxas application definition

The NXxas application definition is designed to store raw data from an X-ray absorption spectroscopy experiment [53]. This is essentially a scan on energy versus incoming/absorbed beam. The structure of NXxas is shown in figures 30, 31, 32 and 33. Figure 30 shows the high level classes, while figures 31, 32 and 33 show the detail of instrument, experiment, and data classes respectively.

```

ENTRY: (required) NXentry
  @entry: (required) NX\_CHAR
    NeXus convention is to use "entry1", "entry2", for analysis software to locate entry.
  title: (required) NX\_CHAR  $\Leftarrow$ 
  start_time: (required) NX\_DATE\_TIME  $\Leftarrow$ 
  definition: (required) NX\_CHAR  $\Leftarrow$ 
    Official NeXus NXDL schema to which this file conforms
    Obligatory value: NXxas
  INSTRUMENT: (required)...
  SAMPLE: (required)...
  MONITOR: (required)...
  DATA: (required)...

```

Figure 30. NXxas application definition first level classes. [53]

```

INSTRUMENT: (required) NXinstrument  $\Leftarrow$ 
  SOURCE: (required) Nxsource  $\Leftarrow$ 
    type: (required) NX\_CHAR  $\Leftarrow$ 
    name: (required) NX\_CHAR  $\Leftarrow$ 
    probe: (required) NX\_CHAR  $\Leftarrow$ 
    Obligatory value: X-ray
  monochromator: (required) NXmonochromator  $\Leftarrow$ 
    energy: (required) NX\_FLOAT (Rank: 1,
      Dimensions: [nP])  $\Leftarrow$ 
  incoming_beam: (required) NXdetector  $\Leftarrow$ 
    data: (required) NX\_NUMBER (Rank: 1,
      Dimensions: [nP])  $\Leftarrow$ 
  absorbed_beam: (required) NXdetector  $\Leftarrow$ 
    data: (required) NX\_NUMBER (Rank: 1,
      Dimensions: [nP])  $\Leftarrow$ 
    This data corresponds to the sample
    signal.

```

Figure 31. NXxas application definition instrument description classes [53].

**SAMPLE:** (required) [NXsample](#)

**name:** (required) [NX\\_CHAR](#)

Descriptive name of sample

**MONITOR:** (required) [NXmonitor](#)

**mode:** (required) [NX\\_CHAR](#)

Count to a preset value based on either clock time (timer) or received monitor counts (monitor).

Any of these values: [monitor](#) | [timer](#)

**preset:** (required) [NX\\_FLOAT](#)

preset value for time or monitor

**data:** (required) [NX\\_NUMBER](#) (Rank: 1, Dimensions: [nP])

This field could be a link to  
[/NXentry/NXinstrument/incoming\\_beam:NXdetector/data](#)

Figure 32. NXxas application definition experiment description classes [53].

**DATA:** (required) [NXdata](#)

**mode:** (required) [NX\\_CHAR](#)

Detection method used for observing the sample absorption (pick one from the enumerated list and spell exactly)

Any of these values:

- [Total Electron Yield](#)
- [Partial Electron Yield](#)
- [Auger Electron Yield](#)
- [Fluorescence Yield](#)
- [Transmission](#)

**energy:** [link](#) (suggested target: [/NXentry/NXinstrument/monochromator:NXmonochromator/energy](#))

**absorbed\_beam:** [link](#) (suggested target: [/NXentry/NXinstrument/absorbed\\_beam:NXdetector/data](#))

Figure 33. NXxas application definition data description classes [53].

In the NXxas application definition, much metadata about the experiment is required, such as details about the instrument. The NXdata class is also used in a similar manner to [Figure 29](#) the energy and absorbed\_beams are singled out as being useful for simple plotting and linked to by NXdata. It also requires the mode of detection to be specified.

### 5.1.1.2. NXxasproc application definition.

The NXxasproc application definition is designed for processed data from X-ray absorption spectroscopy [54]. The structure of NXxas is shown in [figure 34](#). The expected data is energy versus I(incoming)/I(absorbed). NXxasproc is intended for processed data, consequently, it includes classes for the details of the program(s) used for reduction and processing. Additionally, the processing class (XAS\_data\_reduction) includes attributes that link to the raw data from which the NXxasproc data was derived.

```

ENTRY: (required) NXentry
  @entry: (required) NX\_CHAR
    NeXus convention is to use "entry1", "entry2", ... for analysis
    software to locate each entry.
  title: (required) NX\_CHAR
  definition: (required) NX\_CHAR
    Official NeXus NXDL schema to which this file conforms
    Obligatory value: NXxasproc
  SAMPLE: (required) NXsample
    name: (required) NX\_CHAR
      Descriptive name of sample
  XAS_data_reduction: (required) NXprocess
    program: (required) NX\_CHAR
      Name of the program used for reconstruction
    version: (required) NX\_CHAR
      Version of the program used
    date: (required) NX\_DATE\_TIME
      Date and time of reconstruction processing.
    parameters: (required) NXparameters
    raw_file: (required) NX\_CHAR
      Original raw data file this data was derived from
  DATA: (required) NXdata
    energy: (required) NX\_CHAR (Rank: 1, Dimensions: [nP])
    data: (required) NX\_FLOAT (Rank: 1, Dimensions: [nP])
      This is corrected and calibrated
      I(incoming)/I(absorbed). So it is the absorption.
      Expect attribute signal=1

```

*Figure 34. NXxasproc application definition structure.*

The SAMPLE entry appears in both definitions, and because they use the same underlying base class (NXsample) the structure of this portion of the metadata will be the same. By contrast, the NXxasproc application definition also requires information about the XAS\_data\_reduction program, which is only relevant in the context of processed data, and no information about the INSTRUMENT. One final point of note is that despite being present and serving the role of exposing the high value “plottable” data in both, the DATA class will

link to different places in the overall structure of the entry. All of this shows how different application definitions within the same overall domain can serve different needs.

It is also possible to use the NeXus format without following a specific application definition. While it is entirely possible to include the same information in this way, by not following an application definition there is no longer a “contract” between the producer and possible consumers of the data; in principle those attempting to use the data do not have any guarantees about how it may be formatted or what it contains.

Application definitions must follow an XML Schema known as the NeXus Definition Language (NXDL). New application definitions, written in NXDL, can be submitted to the NeXus International Advisory Committee (NIAC). Applications receive feedback and are required to be in use for a year before being officially adopted to allow further refinement to take place.

More broadly, the NIAC maintains the underlying NeXus format, ensures it is well documented, and coordinates with other organisations in the interest of compatibility between data formats.

### 5.1.2. XAFS Data Interchange (XDI)

The XAFS Data Interchange (XDI) standard is designed to encapsulate a single spectrum of XAFS along with relevant metadata [57]. XDI is a text-based format with a syntax which separates metadata from the data table in a way interpretable programmatically and by humans. The metadata header contains metadata names and values as an associative array. The data table is represented as columns of numbers. This format can be imported as is into most existing XAFS data analysis tools, spreadsheet, or data visualization programs.

The main idea behind the development of XDI is that a processed  $\mu(E)$  spectrum is the basic unit of currency for an XAFS analysis problem and XDI designed to encode this basic unit of currency. An XDI file encodes a single processed  $\mu(E)$  spectrum [Not raw data]. This processed data might be a single scan, or the average of repeated scans on a sample, or the consolidation of the channels of a multi-element detector after dead-time corrections have been applied.

The standard defines 8 semantic namespaces for 8 types of metadata encoded in the header.

- **Column:** groups an ordered list of column labels. Each column is numbered, and the order is ascending, and it follows the order of the columns from left to right.

- **Element**: contains the edge and symbol of the absorbing element
- **Scan**: description of the scanning event, edge energy and step size
- **Mono**: description of the monochromator including name and d-spacing
- **Beamline**: beamline description, including name and optics
- **Facility**: description of the synchrotron radiation facility (name)
- **Detector**: type and atmosphere in the targeting station.
- **Sample**: name and preparation of the sample under study.

The example in [Figure 35](#) shows the use of these namespaces in a XDI file. Notice that the ordering and grouping of the metadata namespaces in the file can be mixed, there is no strict ordering, except for ensuring that the data columns follow the metadata section. Notice also that the format can be extended to include custom metadata, in additional namespaces. For instance, encoding special parameters when data has been processed.

1 # XDI/1.0 GSE/1.0	Define data columns
2 # Column.1: energy eV	
3 # Column.2: mutrans	
4 # Column.3: i0	
5 # Element.edge: K	Experiment
6 # Element.symbol: Fe	
7 # Scan.edge_energy: 7112.0 eV	
8 # Mono.name: Si 111	
9 # Mono.d_spacing: 3.13550	
10 # Beamline.name: 13-BM-D	Beamline
11 # Beamline.collimation: none	
12 # Beamline.focusing: no	
13 # Beamline.harmonic_rejection: detuned 50%	
14 # Facility.name: APS	Facility
15 # Facility.energy: 7.00 GeV	
16 # Facility.xray_source: bending magnet	
17 # Scan.start_time: 2002-03-16T17:50:01	Experiment
18 # Detector.I0: 15cm N2	
19 # Detector.II: 15cm N2	
20 # Sample.name: FeN	Sample details
21 # Sample.prep: powder on tape, 5 layers	
22 # GSE.extra: some configuration parameter	Experiment
23 # ///	
24 # room temperature	
25 # measured at beamline 13-BM-D	
26 # vert slits = 2mm (at 45m)	
27 #	
28 # energy mutrans i0	Data
29 6962.0 -0.39452033 302897.20	
30 6972.0 -0.41004339 304569.20	
31 6982.0 -0.42398716 304504.20	
32 6992.0 -0.43863027 304514.20	
33 7002.0 -0.45328045 304407.20	

*Figure 35. Example XDI file (from [11]). The colours highlight the types of metadata encoded in the header.*

## 5.2. Other Interoperable Schemas for XAS Data

Among the databases described above (see [databases](#)), MDR XAFS DB, and RefXAS DB stand out for having the most comprehensive and well-documented schemas for XAS data curation, and for being targeted at accepting data from various sources. These schemas extend beyond basic data formatting and metadata, offering detailed information about experimental conditions, devices, personnel, software tools, and data cataloguing. This section will examine these schemas in more detail, as they reflect the expertise of practitioners in addressing the practical needs of XAS data preservation. EXASLIB and LISA DB are not included, because they are mainly XDI repositories, and they are covered in the [XDI section](#).

## 6. Conclusions and outlook

The efforts to foster the greater publishing and reuse of XAS data in the XAS research community include the creation of online databases, open formats and schemas, and community standardisation initiatives. Despite the considerable number of XAS experiments conducted globally, the interoperability of available XAS data remains limited due to challenges with varying levels of metadata description. Ensuring data and metadata quality, designing a unified metadata format, and establishing a coherent policy for its use are crucial steps moving forward.

The establishment of interoperable XAS schemas will facilitate the effective use of diverse data sources in larger studies, promoting FAIRness (Findable, Accessible, Interoperable, and Reusable) from dataset creation onward, and supporting the development of processing and analysis tools. The IUCr-IXAS group on Q2EXAFS has reached a consensus on the recommendation of two schemas which can complement each other: NXxas for multi-spectra raw and processed data and XDI for single spectra data.

Given the existing challenges for XAS data preservation and sharing, the agreement on recommending NXxas and XDI as the ideal formats and standards marks a significant milestone. The adoption of these standards is expected to enable practices that support XAS data interoperability, more effective data integration, the creation of larger XAS databases, and improved data usability including through the development of software.

### 6.1. Community buy-in

The adoption of NXxas and XDI as community standards in practice has high possibilities of success. Firstly, most of the synchrotron facilities already implement some form of HDF5 encoding, and NXxas is a customised HDF5 format. Similarly, processing and analysis software also support reading and writing HDF5 data, so adoption would take the form of creating mapping libraries/packages for them. The case of XDI is similar, as most facilities and laboratories already support the creation of text-based tabular data, with some text metadata. In fact, three of the databases described above have already adopted XDI. The arguments for processing and analysis software consuming and producing XDI are similar, Demeter being a popular package that consumes and produces XDI data.



## 6.2. Expected advantages

Adopting NXxas and XDI as community standards will enable the seamless integration of XAS data from diverse sources into larger, more comprehensive studies. By reducing barriers to data combination, this approach will not only expand the size and diversity of XAS datasets but also ensure that they are aligned with FAIR (Findable, Accessible, Interoperable, Reusable) principles from the outset. This, in turn, can accelerate the development and enhancement of processing and analysis tools, fostering a more robust and collaborative research environment.

The next stage of the project will be to systematically apply the Cross-Domain Interoperability Framework (CDIF) recommendations [1] to create an interoperability profile that allows for the integration and mapping of NXxas and XDI data, known as the CDIF-4-XAS interoperability model. CDIF-4-XAS will enable the interoperability of data from different databases and repositories, enhancing cross-domain collaboration and supporting the efficient, FAIR-compliant use of data across XAS research communities.



## 7. References

- [1] A. Gregory *et al.*, ‘WorldFAIR (D2.3) Cross-Domain Interoperability Framework (CDIF) (Report Synthesising Recommendations for Disciplines and Cross-Disciplinary Research Areas)’, May 2024, doi: 10.5281/ZENODO.11236871.
- [2] A. Iglesias-Juez, G. L. Chiarello, M. Patience, and M. O. Guerrero-Pérez, ‘Experimental methods in chemical engineering: X-ray absorption spectroscopy—XAS, XANES, EXAFS.’, *Can. J. Chem. Eng.*, vol. 100, pp. 3–22, 2022.
- [3] P. Zimmermann *et al.*, ‘Modern X-ray spectroscopy: XAS and XES in the laboratory’, *Coord. Chem. Rev.*, vol. 423, p. 213466, Nov. 2020, doi: 10.1016/j.ccr.2020.213466.
- [4] Diamond Light Source, ‘Techniques Available, About XAS’, Diamond Light Source. Accessed: Oct. 01, 2024. [Online]. Available: <https://www.diamond.ac.uk/industry/Techniques-Available/Spectroscopy/X-ray-spectroscopy-XAS/About-XAS.html>
- [5] Advanced Photon Source, ‘XSD Groups’, X-ray Science Division Groups Advanced Photon Source. Accessed: Jan. 06, 2025. [Online]. Available: <https://www.aps.anl.gov/XSD-Groups>
- [6] Helmholtz-Zentrum Berlin für Materialien und Energie, ‘BESSY II Light Source’, Helmholtz-Zentrum Berlin. [Online]. Available: [https://www.helmholtz-berlin.de/forschung/quellen/bessy/index\\_en.html](https://www.helmholtz-berlin.de/forschung/quellen/bessy/index_en.html)
- [7] Deutsches Elektronen-Synchrotron DESY, ‘X-ray Absorption Spectroscopy’, X-ray Absorption Spectroscopy DESY. Accessed: Jan. 06, 2025. [Online]. Available: [https://photon-science.desy.de/users\\_area/industrial\\_users/methods\\_techniques/x-ray\\_absorption\\_spectroscopy/index\\_eng.html](https://photon-science.desy.de/users_area/industrial_users/methods_techniques/x-ray_absorption_spectroscopy/index_eng.html)
- [8] European Synchrotron Radiation Facility ESRF, ‘BM23 XAS Beamline’, BM23 XAS Beamline.
- [9] Kyushu University, ‘Research Center for Synchrotron Light Applications’, Research Center for Synchrotron Light Applications, Kyushu University. Accessed: Jan. 08, 2025. [Online]. Available: <https://www.rcsla.kyushu-u.ac.jp/en/index.html>
- [10] M. Kubin *et al.*, ‘Direct Determination of Absolute Absorption Cross Sections at the L-Edge of Dilute Mn Complexes in Solution Using a Transmission Flatjet’, *Inorg. Chem.*, vol. 57, no. 9, pp. 5449–5462, May 2018, doi: 10.1021/acs.inorgchem.8b00419.
- [11] S. P. Cramer *et al.*, ‘Ligand field strengths and oxidation states from manganese L-edge spectroscopy’, *J. Am. Chem. Soc.*, vol. 113, no. 21, pp. 7937–7940, Oct. 1991, doi: 10.1021/ja00021a018.
- [12] F. M. F. De Groot, M. A. Arrio, Ph. Saintcavit, Ch. Cartier, and C. T. Chen, ‘Fluorescence yield detection: Why it does not measure the X-ray absorption cross section’, *Solid State Commun.*, vol. 92, no. 12, pp. 991–995, Dec. 1994, doi: 10.1016/0038-1098(94)90027-2.
- [13] R. D. Peacock and B. Stewart, ‘Natural Circular Dichroism in X-ray Spectroscopy’, *J.*

*Phys. Chem. B*, vol. 105, no. 2, pp. 351–360, Jan. 2001, doi: 10.1021/jp001946y.

- [14] M. A. Newton and A. J. Dent, ‘Energy-Dispersive EXAFS: Principles and Application in Heterogeneous Catalysis’, in *In-situ Characterization of Heterogeneous Catalysts*, 1st ed., J. A. Rodriguez, J. C. Hanson, and P. J. Chupas, Eds., Wiley, 2013, pp. 75–119. doi: 10.1002/9781118355923.ch3.
- [15] R. Mitzner *et al.*, ‘L-Edge X-ray Absorption Spectroscopy of Dilute Systems Relevant to Metalloproteins Using an X-ray Free-Electron Laser’, *J. Phys. Chem. Lett.*, vol. 4, no. 21, pp. 3641–3647, Nov. 2013, doi: 10.1021/jz401837f.
- [16] Z. Li *et al.*, ‘Highly sensitive 2D X-ray absorption spectroscopy via physics informed machine learning’, *Npj Comput. Mater.*, vol. 10, no. 1, p. 128, Jun. 2024, doi: 10.1038/s41524-024-01313-7.
- [17] S. W. T. Price *et al.*, ‘Chemical imaging of single catalyst particles with scanning  $\mu$ -XANES-CT and  $\mu$ -XRF-CT’, *Phys. Chem. Chem. Phys.*, vol. 17, no. 1, pp. 521–529, 2015, doi: 10.1039/C4CP04488F.
- [18] X. Liu *et al.*, ‘Distinct charge dynamics in battery electrodes revealed by in situ and operando soft X-ray spectroscopy’, *Nat. Commun.*, vol. 4, no. 1, p. 2568, Oct. 2013, doi: 10.1038/ncomms3568.
- [19] I. López, J. Morey, J. B. Ledeuil, L. Madec, and H. Martinez, ‘A critical discussion on the analysis of buried interfaces in Li solid-state batteries. *Ex situ* and *in situ* / *operando* studies’, *J. Mater. Chem. A*, vol. 9, no. 45, pp. 25341–25368, 2021, doi: 10.1039/D1TA04532F.
- [20] M. Pidd, ‘A Practical Guide to Using Data Ontologies in the Arts, Humanities and Social Sciences’, A Practical Guide to Using Data Ontologies in the Arts, Humanities and Social Sciences. [Online]. Available: <https://www.dhi.ac.uk/books/ontology-guide>
- [21] W. Malzer, C. Schlesiger, and B. Kanngießer, ‘A century of laboratory X-ray absorption spectroscopy – A review and an optimistic outlook’, *Spectrochim. Acta Part B At. Spectrosc.*, vol. 177, p. 106101, Mar. 2021, doi: 10.1016/j.sab.2021.106101.
- [22] N. S. Genz, A. Kallio, F. Meirer, S. Huotari, and B. M. Weckhuysen, ‘Operando Laboratory-based X-ray Absorption Spectroscopy: Guidelines for Newcomers in the Field’, *Chemistry—Methods*, vol. 4, no. 1, p. e202300027, Jan. 2024, doi: 10.1002/cmtd.202300027.
- [23] M. Ishii, ‘International XAFS DB Portal’, International XAFS DB Portal. Accessed: Jan. 28, 2024. [Online]. Available: <https://ixdb.jxafs.org/>
- [24] National Institute for Materials Science (NIMS), ‘Materials Data Repository’, Search MDR. Accessed: Dec. 06, 2024. [Online]. Available: [https://mdr.nims.go.jp/catalog?f%5Bmember\\_of\\_collections\\_ssim%5D%5B%5D=MDR+XAFS+DB](https://mdr.nims.go.jp/catalog?f%5Bmember_of_collections_ssim%5D%5B%5D=MDR+XAFS+DB)
- [25] Japan Synchrotron Radiation Research Institute (JASRI), ‘JASRI’, JASRI Registered Institution for Facilities Use Promotion. Accessed: Dec. 05, 2024. [Online]. Available: <https://www.jasri.jp/en/index.html>
- [26] Ritsumeikan SR Center, ‘Ritsumeikan SR Center Soft X-Ray XAFS Database’, Ritsumeikan SR Center Soft X-Ray XAFS Database. Accessed: Jan. 12, 2025. [Online]. Available:

[https://www.ritsumei.ac.jp/acd/re/src/sx\\_xafs\\_db/](https://www.ritsumei.ac.jp/acd/re/src/sx_xafs_db/)

- [27] Hokkaido University, ‘Institute for Catalysis, Hokkaido University’, Institute for Catalysis, Hokkaido University. Accessed: Dec. 05, 2024. [Online]. Available: <https://www.cat.hokudai.ac.jp/index-e.html>
- [28] Institute of Materials Structure Science, ‘Photon Factory’, Photon Factory - KEKK IMSS. Accessed: Jan. 08, 2025. [Online]. Available: <https://www2.kek.jp/imss/pf/eng/>
- [29] Aichi Synchrotron Radiation Center, ‘Aichi Synchrotron Radiation Center’, Aichi Synchrotron Radiation Center. Accessed: Jan. 08, 2025. [Online]. Available: <https://aichisr.jp/en/index.html>
- [30] Materials Data Platform Center, National Institute for Materials Science Japan, ‘MDR Schema’. Accessed: Jan. 16, 2025. [Online]. Available: <https://github.com/nims-dpfc/mdr-schema/>
- [31] Spring-8, ‘Spring-8 Industrial Use Promotion Office’, Spring-8 Industrial Use Promotion Office. [Online]. Available: <https://support.spring8.or.jp/BL/bl14b2/xafs/standardDB/>
- [32] J. Yano and V. K. Yachandra, ‘X-ray absorption spectroscopy’, *Photosynth. Res.*, vol. 102, no. 2–3, pp. 241–254, Dec. 2009, doi: 10.1007/s11120-009-9473-8.
- [33] M. Ishii *et al.*, ‘Integration of X-ray absorption fine structure databases for data-driven materials science’, *Sci. Technol. Adv. Mater. Methods*, vol. 3, no. 1, p. 2197518, Dec. 2023, doi: 10.1080/27660400.2023.2197518.
- [34] Masashi Ishii, ‘MatVoc vocabulary (subset)’. Accessed: Jan. 18, 2025. [Online]. Available: <https://dice.nims.go.jp/ontology/mdr-xafs-ont/Item>
- [35] National Institute for Materials Science - Japan, ‘MatVoc Explorer Dictionary’, MatVoc Explorer Materials Vocabulary. Accessed: Jan. 19, 2025. [Online]. Available: <https://matvoc.nims.go.jp/explore/en/dictionary/Q713>
- [36] Canadian Light Source, ‘XAS Database’, XAS Database. Accessed: Jan. 28, 2025. [Online]. Available: <https://xasdb.lightsource.ca/>
- [37] I. Kieffer and D. Testemale, ‘SSHADe/FAME: “French Absorption spectroscopy beamline in Material and Environmental science” database service’, 2017, *SSHADe (OSUG Data Center)*. doi: 10.26302/SSHADe/FAME.
- [38] International X-ray Absorption Society (IXAS), ‘IXAS X-ray Absorption Data Library’, XASLIB: X-ray Absorption Data Library.
- [39] A. Puri, ‘LISA XAS Database’, LISA XAS Database. Accessed: Feb. 03, 2025. [Online]. Available: <https://lisa.iom.cnr.it/xasdb/>
- [40] A. Puri, ‘LISA XAS Database, Full image’. Zenodo, Mar. 04, 2024. doi: 10.5281/ZENODO.10778068.
- [41] Institute of High Energy Physics, Chinese Academy of Sciences (IHEP-CAS), ‘Experimental XAS database’, Experimental XAS database. Accessed: Feb. 02, 2025. [Online]. Available: <http://xasdb.ihep.ac.cn/>
- [42] A. Gaur *et al.*, ‘Metadata Fields and Quality Criteria - XAS Reference Database under DAPHNE4NFDI’, in *Proceedings of the Conference on Research Data Infrastructure*, Sep. 2023. doi: 10.52825/cordi.v1i.258.
- [43] A. Gaur *et al.*, ‘Poster: Metadata Fields and Quality Criteria - XAS Reference Database

under DAPHNE4NFDI', Oct. 2023, doi: 10.5281/ZENODO.8412918.

- [44] S. Paripsa *et al.*, 'RefXAS: an open access database of X-ray absorption spectra', *J. Synchrotron Radiat.*, vol. 31, no. 5, pp. 1105–1117, Sep. 2024, doi: 10.1107/S1600577524006751.
- [45] A. Barty *et al.*, 'DAPHNE4NFDI - Consortium Proposal', Jun. 2023, doi: 10.5281/ZENODO.8040606.
- [46] B. Ravel and M. Newville, 'ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT', *J. Synchrotron Radiat.*, vol. 12, no. 4, pp. 537–541, Jul. 2005, doi: 10.1107/S0909049505012719.
- [47] M. Newville, 'Larch: An Analysis Package for XAFS and Related Spectroscopies', *J. Phys. Conf. Ser.*, vol. 430, p. 012007, Apr. 2013, doi: 10.1088/1742-6596/430/1/012007.
- [48] M. Könnecke *et al.*, 'The NeXus data format', *J. Appl. Crystallogr.*, vol. 48, no. 1, pp. 301–305, Feb. 2015.
- [49] Materials Informatics Group, 'The 9809 format Raw Data'. [Online]. Available: [https://github.com/Materials-Informatics-Group/XAFS/blob/main/World/02-2\\_9809format.md](https://github.com/Materials-Informatics-Group/XAFS/blob/main/World/02-2_9809format.md)
- [50] HX-XAFS group of Nagoya Univ. SR-Center and Aichi SR, 'tabuchi:9809format', tabuchi:9809format. Accessed: Jan. 31, 2025. [Online]. Available: <https://titan.nusr.nagoya-u.ac.jp/Tabuchi/BL5S1/doku.php/tabuchi/9809format>
- [51] I. Ascone, K. Asakura, G. N. George, and S. Wakatsuki, 'International Workshop on Improving Data Quality and Quantity for XAFS Experiments (Q2XAFS 2011)', *J. Synchrotron Radiat.*, vol. 19, no. 6, pp. 849–850, Nov. 2012, doi: 10.1107/S0909049512043506.
- [52] M. Newville, 'Progress Report: on Data Format Working Group', presented at the Q2XAFS2024, Jul. 26, 2024. Accessed: Dec. 10, 2024. [Online]. Available: <https://www.youtube.com/watch?v=56jxb2bHGKs>
- [54] NeXus International Advisory Committee, P. R. Jemian, R. Berg, T. Richter, and J. Wuttke, 'NXxasproc documentation', 3.3.2.26. NXxasproc - nexus v2024.02 documentation. Accessed: Nov. 02, 2024. [Online]. Available: <https://manual.nexusformat.org/classes/applications/NXxasproc.html>
- [57] M. Newville, B. Ravel, V. A. Solé, and Wellenreuther, 'XAS Data Interchange Format Draft Specification, version 1.0', XAS Data Interchange Format. Accessed: Oct. 01, 2024. [Online]. Available: <https://github.com/XraySpectroscopy/XAS-Data-Interchange/blob/master/specification/spec.md>