Radiation Physics and Chemistry 175 (2020) 108479

Contents lists available at ScienceDirect

Radiation Physics and Chemistry

journal homepage: www.elsevier.com/locate/radphyschem

An open access, integrated XAS data repository at Diamond Light Source

Giannantonio Cibin[a,*], Diego Gianolio[a], Stephen A. Parry[a], Tom Schoonjans[a], Oliver Moore[b], Rachael Draper[c], Laura A. Miller[d], Alexander Thoma[e], Claire L. Doswell[f], Abigail Graham[g]

[a] Diamond Light Source Ltd., Harwell Science and Innovation Campus, OX11 0DE Didcot, United Kingdom
[b] School of Earth and Environment, University of Leeds, Leeds LS2 9JT, United Kingdom
[c] University of Manchester, Oxford Rd, Manchester M13 9PL, United Kingdom
[d] Research School of Earth Sciences, Australian National University, 142 Mills Rd, Acton ACT, 0200, Australia
[e] Queens' College, Cambridge, Cambridge, CB3 9ET, United Kingdom
[f] University of Birmingham, Edgbaston Birmingham B15 2TT, United Kingdom
[g] University of Nottingham, University Park Nottingham, NG7 2RD, United Kingdom

A B S T R A C T

The analysis of reference materials is a fundamental part of the data analysis process, in particular for XAS experiments. The beamline users and more generally the XAS community can greatly benefit from the availability of a reliable and wide base of reference sample spectra, acquired in standard and well-characterized experimental conditions. On B18, the Core EXAFS beamline at the Diamond Light Source, in the past years we have collected a series of XAS data on well characterized compounds. This work constitutes the base for a reference sample database, available as a data analysis tool to the general XAS community. This data repository aims to complement the bare spectroscopic information with characterisation, preparation, provenance, analysis and bibliographic references, so improving the traceability of the deposited information. This integrated approach is the base of success and wide distribution of data repositories in other fields, and we hope it will provide on one side a precious facility for the training of students and researchers new to the technique, and at the same time encourage the discussion of best practices in the data analysis process. The database will be open to the contribution of experimental data from the user community, and will provide bibliographic reference information and access control.

1. Introduction

X-Ray Absorption Spectroscopy is increasingly used as a scientific and diagnostic tool, as materials' technology and research is increasingly moving to the investigation of quasi-crystalline, amorphous and biologic materials in real use conditions. The characterisation of nanostructured systems in particular relies on non-crystallographic methods such as diffuse scattering, direct imaging or spectroscopic analysis. Among the latter, XAS has the important benefit of providing element-specific electronic and structural information, and if the element and absorption edge studied falls in the hard X-ray regime, in-situ and operando experiments are possible, replicating full-scale process conditions.

The analysis of XAS results presents however significant barriers to the newcomer. EXAFS analysis is an established technique, while getting robust and quantitative information from XANES simulations is still difficult. The interpretation of XANES still requires significant preparation work, typically including the collection of a substantial set of reference data on materials of known structure for direct comparison with the experimental unknowns. The daily experience as staff on the large scale facilities show that the preliminary selection of relevant compounds is often insufficient to provide a set of reference data wide enough to cover for unexpected experimental findings, in particular where quantitative analysis (e.g. via linear combination fitting of known model systems) is required. Missing key reference data can mean delays in the acquisition of precious information through successive beamtime access requests, or the search for datasets not necessarily compatible with the acquisition configuration used during the main experiment.

At the same time, initiatives for supporting the data analysis of XANES and EXAFS data based on deep learning algorithms are being proposed (Zheng et al., 2018). This approach relies for quantitative analysis, more than on the ability to accurately model the XAS signal, on the availability of large spectral libraries.

There is therefore a clear need for access to well-characterized reference XAS data, acquired on large families of reference compounds, covering multiple absorption edges and elements on most part of the periodic table.

2. Data sources

This large dataset collection cannot realistically be compiled by the

* Corresponding author.
E-mail address: giannantonio.cibin@diamond.ac.uk (G. Cibin).

https://doi.org/10.1016/j.radphyschem.2019.108479
Received 30 September 2018; Received in revised form 22 July 2019; Accepted 6 September 2019
Available online 08 September 2019
0969-806X/

facilities' staff alone, and contributions from the user community are necessary. While a core number of common compounds and a collection of pure element foils are available for routine calibration at most of the facilities, these sets are clearly not sufficient to cover the needs for reference measurements during the experiments. Sample preparation requires relevant time and often important materials' handling requires special care (as for samples highly sensitive to the exposure to ambient temperature, oxygen or moisture).

At the same time, an overwhelming number of reference measurements are routinely acquired during users' beamtime, providing potentially an ideal base of reference materials for the same community. Availability of these data would be at the same time a precious instrument for the statistical analysis of beamline performance and could form a base to build on-line data validation system - to confirm with the running experimentalists that the data being acquired are conforming to defined quality standards.

Databases for XAS are already available. We refer for brevity to the recent work of H. Asakura et al., (2018) which reports such resources, following an open or controlled access model.

In all cases, the data collection is relying on voluntary donations. The number of datasets deposited - also in collections open by several years - is however not comparable with what is reported e.g. on well-known crystallographic data sites as the ICSD (Bergerhoff et al., 1987), the American Mineralogist Crystal Structure Database (Downs and Hall-Wallace, 2003), the RRUFF™ Project for Raman data (Lafuente et al., 2015) or several FTIR libraries. While XAS diffusion is certainly small because of the restricted availability of synchrotron radiation-based instruments, data released by the user community is probably the limiting factor. Other routes are unlikely. While the automatic publication in the open domain of data acquired at publicly funded facilities data is being considered, at the same time it is unrealistic to consider this as a future source for reliable reference data of practical use, because of the intrinsic lack of metadata embedded with raw data collection methods.

We find indeed that the usability of the present database information suffers from limited ancillary information, regarding not only the data collection conditions but also – and more importantly - the materials provenance, characterisation and appropriate bibliographic references. The historical Farrell Lytle database for example, contains an ample collection of spectra, but supporting documentation is limited to the text comments entered by the user at the time of data acquisition. This suggests that useable repositories should aim at collecting information, in lieu of the bare spectra alone.

Other aspects to be considered are related to data ownership, data release policy and intellectual property management. Ensuring the intellectual property is correctly protected at the moment of release in the public domain, relies on the agreement on appropriate licensing schemes. Data owners could however feel comfortable with different degrees of openness for use of their data. This means that to ensure the widest participation, not necessarily a single licence model would cover most data owner preferences.

3. Database design

3.1. Web site

The website has been built using Django 2.0 (https://djangoproject.com). The information entered by the user will be stored in an SQL database: during development, SQLite (https://www.sqlite.org) was used as backend implementation but this will be replaced with MariaDB in production mode when it will be integrated into the Diamond Light Source computing infrastructure. The graphical output of the prototype and a high level block diagram of the database structure are presented in Figs. 1 and 2. The web site will be hosted on Diamond domain and reference provided through B18 and the Spectroscopy group information pages (https://www.diamond.ac.uk/Instruments/Spectroscopy/B18.html).

3.2. Data format

Data will be by default available and users are expected to upload XAS spectra as files that adhere to the XAS Data Interchange (XDI) format specification (Ravel and Newville 2015; 2016). XDI was designed to guarantee easy access and provide sufficient metadata information to describe single spectra and experimental conditions. Only datasets containing a reference measurement will be accepted, for energy calibration and evaluation of the experimental resolution. Such files will, after successful validation, be parsed and the information contained within will be added as a new entry to the SQL database, along with the additional metadata posted through the submission form. We will provide an automatic conversion tool to for XDI for data generated at Diamond.

The database will allow in addition the storage of raw datasets. For example, full emission spectra can be acquired on dilute systems, and analysis needs accurate processing to separate the fluorescence signal from nearby contributions. The availability of this information will allow reviewing the preliminary treatment process, and evaluate the extraction algorithms and noise properties. In addition, raw data availability allows in the future for reprocessing if changes to the data format used for the main interface will be required, as to guarantee compatibility for data exchange with the development of international methods to exchange experimental results (Asakura et al., 2018).

Individual datasets could be not sufficient to provide complete information. For example, presence of a strong pleochroism in anisotropic crystals requires knowledge of the sample orientation and access to a full set of oriented measurements. For this reason, data series can be grouped using user-defined reference tags, as adopted in the XAS data library hosted at CARS (M. Newville, XasDataLibrary 2016).

Supporting information

Supporting datasets will be hosted as well. As mentioned, aim is to provide information as complete as possible on the materials, experimental conditions and analysis processes. The importance of this additional information is clear is for example in the earth sciences, where whole families of minerals present very variable degrees of crystallinity and composition, that can strongly affect the XAS results. For this reason, we have included in our initial dataset examples of XRF and powder XRD data on natural samples, acquired with the conventional instrumentation available in Diamond laboratories.

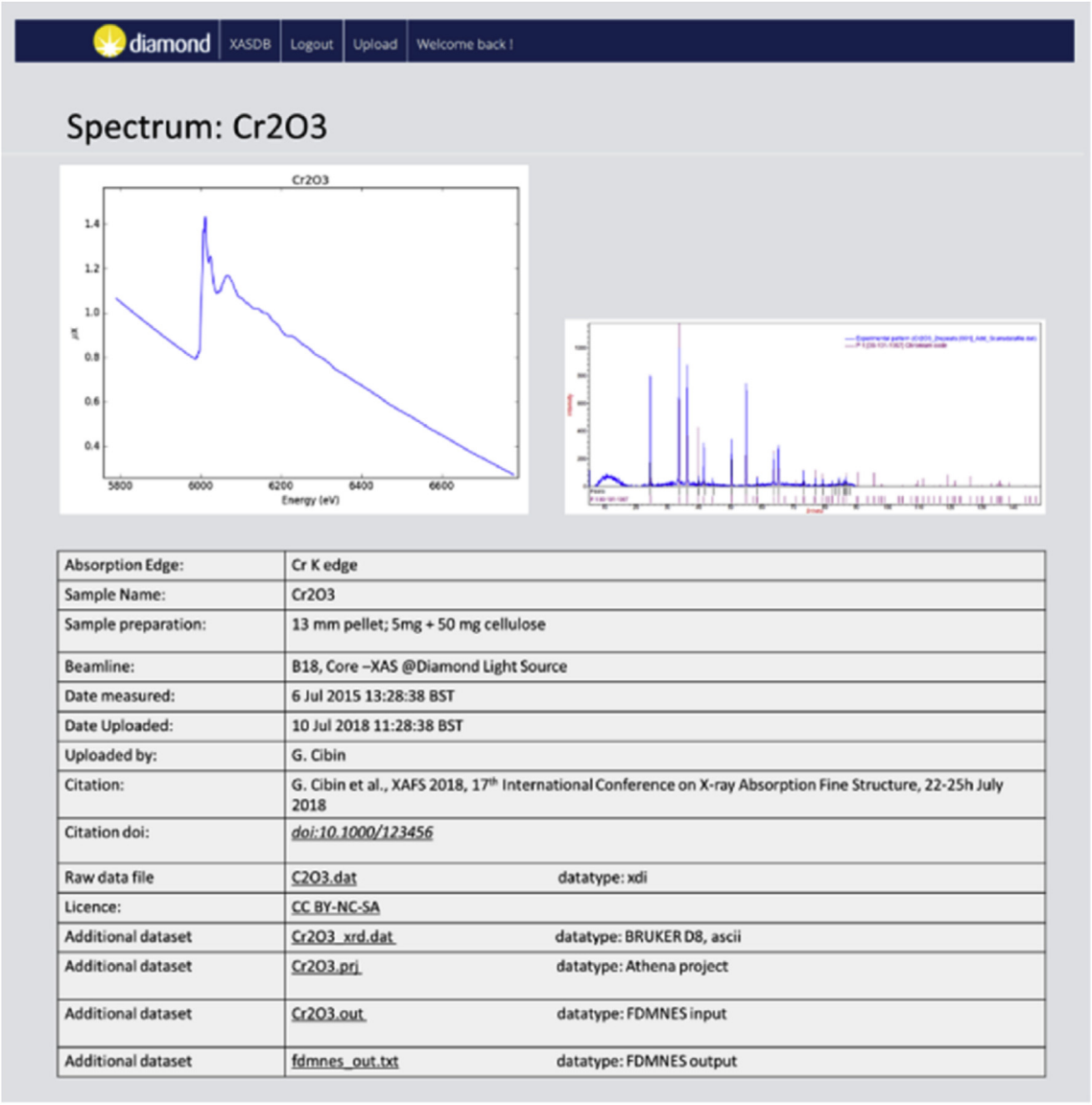
Managing data formats for these additional sets is beyond the scope of a XAS database. Basic metadata will need to contain information on the acquisition parameters, but user will be required only to attach a text description that should, as a minimum, to enable identification of the data format. The upload system will have a set of data conversion routines for known formats, to allow generating an image to present the data in the graphical interface.

3.3. Data deposition and access policy

Uploading new spectra will only be available to registered users. Before the newly added spectra will become publicly visible, a verification will be performed to ensure that the data is of acceptable quality.

We suggest restricting data donations to datasets accompanied by a publication reference. This ensures a full description of the sample characterisation follows the datasets as reported in the referenced publication. At the same time, use of these data will be tied to the acknowledgement of the appropriate bibliographic references, hopefully encouraging the data deposition.

To help with the control of access to those data, we expect to have a 2-tier stage. Among the mandatory information fields, the user will be



Absorption Edge:	Cr K edge	
Sample Name:	Cr2O3	
Sample preparation:	13 mm pellet; 5mg + 50 mg cellulose	
Beamline:	B18, Core –XAS @Diamond Light Source	
Date measured:	6 Jul 2015 13:28:38 BST	
Date Uploaded:	10 Jul 2018 11:28:38 BST	
Uploaded by:	G. Cibin	
Citation:	G. Cibin et al., XAFS 2018, 17th International Conference on X-ray Absorption Fine Structure, 22-25h July 2018	
Citation doi:	doi:10.1000/123456	
Raw data file	C2O3.dat	datatype: xdi
Licence:	CC BY-NC-SA	
Additional dataset	Cr2O3_xrd.dat	datatype: BRUKER D8, ascii
Additional dataset	Cr2O3.prj	datatype: Athena project
Additional dataset	Cr2O3.out	datatype: FDMNES input
Additional dataset	fdmnes_out.txt	datatype: FDMNES output

Fig. 1. A window output of the prototype database user interface. The XAS dataset will be always plotted by default, accompanied by an appropriate render of the supplementary datasets deposited, if a recognised interpreter and plotting method is available in the format conversion library for supplementary information.

required to select a licensing model (e.g. the different levels provided by the Commons Creative licence). The depositing user can then choose to release the full information in the open access, and data will be available to all. Alternatively, with a restricted access option, only the graphical output will be presented to the general public, while data download would be allowed to registered users. This will guarantee the donating user that access to their data sets is monitored.

3.4. Initial dataset

The initial dataset will contain about 200 spectra obtained on B18 (A. Dent et al., 2009) during a number of studentship projects. XAS Data at S, Ca, V, Ti, Cr, Mn, Fe K edges were acquired on B18, during in-house research time.

Part of samples were acquired from chemical suppliers, but some were extracted from mineral collections of known provenance. Because of the intrinsic variability of those natural samples, these required a characterisation work to confirm structure and composition of the samples. Data in these cases were acquired with lab-based XRF and XRD measurements and results are included with the XAS datasets in the supplementary information sections.

Most data were analysed using Athena and Artemis, the analysis tools provided in the Demeter processing package (Ravel and Newville 2005). The data analysis projects will be included in the dataset collection. For most of our set, we also completed the analysis by simulating the XANES region with FEFF 9.0 (Rehr et al., 2010) and FDMNES (Bunau and Joly, 2009). In these cases, we did not perform any optimisation of the simulation parameters, with the purpose of getting a direct comparison of the results obtainable with the two packages on a broad, cross-element sample base.

4. Conclusions

We are presenting a new XAS data repository based on an initial set of experimental measurements taken at Diamond Light Source. Our design strategy is aimed at encouraging the contribution and the sharing of data from the community of XAS researchers. The data base structure allows to integrate to the basic XAS spectra with rich information (as results of analysis processes and complementary characterisation measurements) in a flexible way, trying to maximise the final usability of the data deposited, adding traceability with bibliographic reference information and a clear licence deposition approach.

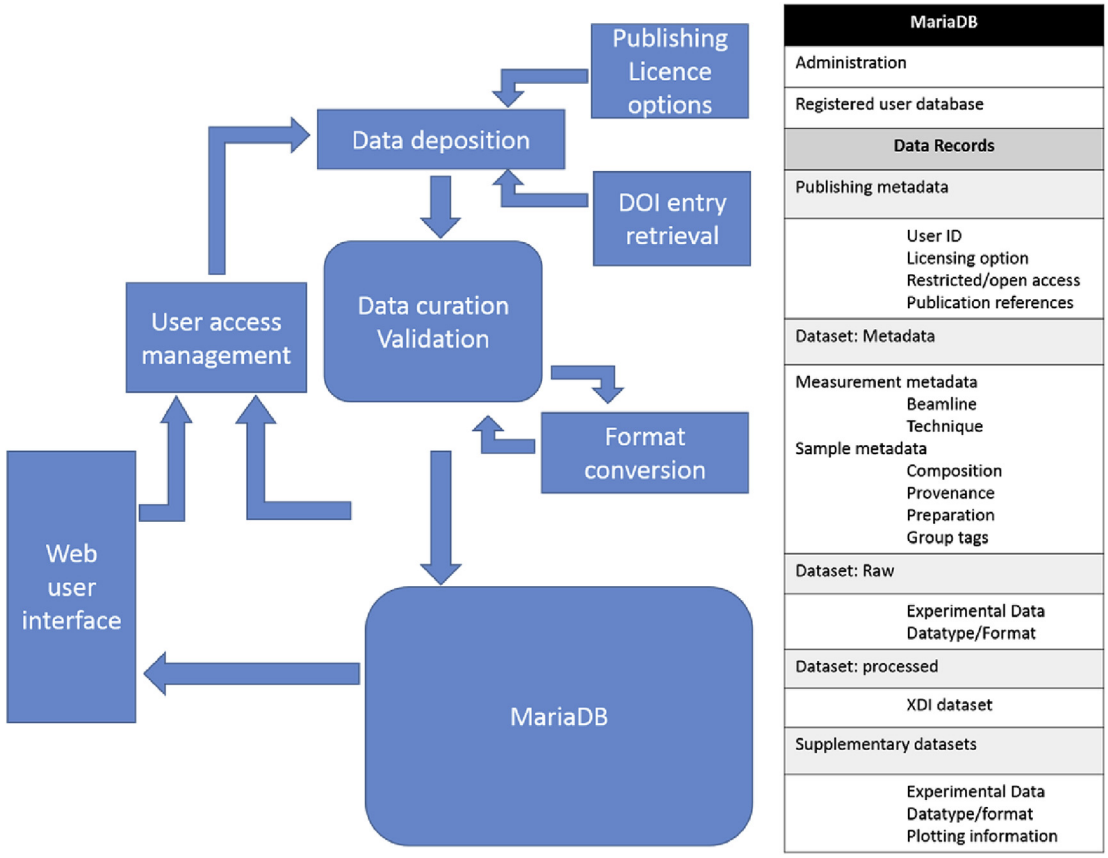


Fig. 2. High-level block diagram depicting the general structure of the database. Left panel: overall structure of the Maria DB information content.

Our attempt to capture as much metadata with the stored records as possible tries also to cover for future developments, in the perspective of an international network for data repositories that would be certainly a powerful tool in support of the fast growing XAS user community.

Acknowledgements

Data were collected on B18, the Core EXAFS beamline at Diamond Light Source within the in-house research program. O. Moore, R. Draper, L. A. Miller, A. Thoma, C. L. Doswell, A. Graham acknowledge support from Diamond Light Source for their studentships. Authors declare no competing interest.

References

Asakura, K., Abe, H., Kimura, M., 2018. The challenge of constructing an international XAFS database. J. Synchrotron Radiat. 25, 2–5.
Bergerhoff, G., Brown, I.D., 1987. In: Allen, F.H. (Ed.), „Crystallographic Databases“. International Union of Crystallography (Hrsg.) Chester.
Bunau, O., Joly, Y., 2009. Self-consistent aspects of x-ray absorption calculations. J. Phys. Condens. Matter 21, 345501.
Dent, A.J., Cibin, G., Ramos, S., Smith, A.D., Scott, S.M., Varandas, L., Pearson, M., Krumpa, N., Jones, C., 2009. B18: a core XAS spectroscopy beamline for Diamond. J. Phys.: Conf. Ser. 190, 012039.
Downs, R.T., Hall-Wallace, M., 2003. The American mineralogist crystal structure database. Am. Mineral. 88, 247–250.
Lafuente, B., Downs, R.T., Yang, H., Stone, N., 2015. The power of databases: the RRUFF project. In: Armbruster, T., Danisi, R.M. (Eds.), Highlights in Mineralogical Crystallography. W. De Gruyter, Berlin, Germany, pp. 1–30.
Newville, M., 2016. XasDataLibrary. (accessed 01 December 2018).
Ravel, B., Newville, M., 2005. ATHENA, artemis, hephaestus: data analysis for X-ray absorption spectroscopy using IFEFFIT. J. Synchrotron Radiat. 12, 537–541. https://doi.org/10.1107/S0909049505012719.
Ravel, B., Newville, M., 2016. XAFS Data Interchange: a single spectrum XAFS data file format. J. Phys. Conf. Ser. 712, 012148 2016.
Ravel, B., Newville, M., 2015. XAS Data Interchange Format Draft Specification. Retrieved from. https://github.com/XraySpectroscopy/XAS-Data-Interchange/blob/master/specification/spec.md.
Rehr, J.J., Kas, J.J., Vila, F.D., Prange, M.P., Jorissen, K., 2010. Parameter-free calculations of x-ray spectra with FEFF9. Phys. Chem. Chem. Phys. 12, 5503–5513.
Zheng, C., Mathew, K.C., Chen, Y., Tang, H., Dozier, A., Kas, J.J., Vila, F.D., Rehr, J.J., Piper, L.F.J., Persson, K.A., Ping Ong, S., 2018. *Automated generation and ensemble-learned matching of X-ray sabsorption spectra* npj. Comput. Mater. 4, 12. https://doi.org/10.1038/s41524-018-0067-x.