Chapter 7.3

Related chapters

The Open Source Database of the Japanese XAFS Society

Kiyotaka Asakura*

Institute for Catalysis, Hokkaido University, Kita-21, Nishi-10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan. *Correspondence e-mail: askr@cat.hokudai.ac.jp

The concept and activities of the Japanese XAFS Society database and its future are briefly reviewed.

1. Introduction

The Japanese XAFS Society (JXS) is an academic society for the development of X-ray absorption fine-structure (XAFS) and related spectroscopies. The JXS was founded in 2001 just after XAFS Conference XI in Ako, and nearly 200 scientists and technicians have since joined the JXS. The JXS has an annual meeting once a year to discuss the science and collaborations in XAFS and organizes XAFS schools and lectures for younger generations. It has also published XAFS textbooks and white papers on XAFS and light sources. The JXS co-organized the XAFS Tutorial for Crystallographers in 2008 and the Q2XAFS workshop in 2012 (Ascone *et al.*, 2012) together with the IUCr XAFS Commission. The XAFS database was started in 2013 in collaboration with the Institute for Catalysis (ICAT), Hokkaido University, which provides the data server and a database manager. ICAT is a unique Joint Usage/Joint Research Center in the field of catalysis, with a mission to promote collaboration among all fields of science and engineering in order to develop new catalysts that are necessary for a sustainable society. XAFS is an important characterization technique for catalysts and ICAT thus decided to develop an XAFS database. Here, we discuss the concept and structure of the XAFS database. In the final part, we discuss the future direction of the database.

2. Concept of the database

For a long time, the XAFS community has desired to know the validity of the XAFS that they have measured and to compare the spectra of unknown compounds with those of standard compounds with known structure in order to obtain hints to the structures of unknown compounds. In the discussion during the annual meeting of JXS in 2011 held at the Institute for Molecular Science (IMS), Japan, we decided to construct a database for XAFS, in particular X-ray absorption near-edge structure (XANES), of standard compounds under the following concept: 'All XAFS data can be deposited with no restriction through the website. The validity of the data is confirmed by the majority rule. It is completely open for all and available free of charge.'

2.1. Who deposits

Although it is possible that synchrotron-radiation facilities and facility members may deposit data that they are responsible for, or a company related to XAFS might collect data to create a database that is offered for sale, we decided on an open-access database in both download and upload directions.
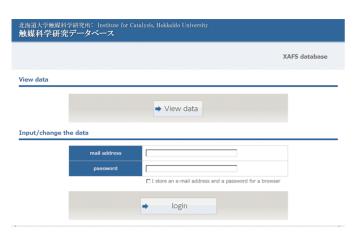
Figure 1
Login window.

Thus, anyone involved in XAFS spectroscopy and interested in the database can deposit their data voluntarily and anyone who requires data can download the data from the database freely. This idea requires only a data server and a web site. In order to prevent illegal access to the database, anyone who wishes to deposit data has to be registered in advance. All of the necessary cost is now covered by ICAT, Hokkaido University.

2.2. How to and what to deposit

We only accumulate the raw data in text format. The data should include metadata as shown later. The data format is simple. Each line should contain energy and absorbance. The deposited data are mainly those for standard compounds which have well-defined structures and properties. In order to obtain energy calibration, the foil XAFS data measured at the same beamline and the same time are deposited.

2.3. Where is the database?

ICAT provides the server computer and manages the database. The login page can be found at https://www.cat.hokudai.ac.jp/catdb/index.php?action=xafs_login_form&opnid=2.

2.4. How to confirm, and who confirms, the validity of the data

A peer-review system is one way to confirm the validity of the data. However, it requires a great cost, human resources and time. Here, we apply the majority rule. We accept any data for standard compounds even if data for the same standard compounds have already been uploaded. This means that we have a large amount of data for the same standard compounds measured on different occasions, on different beamlines and by different experimentalists. If these data for the same compounds agree well with each other, the reliability of the data should be high. However, if the data for the same compound contradict each other, the data reliability becomes low and users should use the data with care. The database may function as a round robin for the beamline by comparing data for the same compounds.

3. The structure of the database

Fig. 1 shows the login page for the database at https://www.cat.hokudai.ac.jp/catdb/index.php?action=xafs_login_form&opnid=2.

Without a login ID (email address) and password, anyone can access the database to search it, see the data and download the data by pressing the View data button described in Section 3.1.

A person who wishes to deposit data first has to register by sending the necessary information to a manager. After obtaining a password they can log in and the deposition window appears as described in Section 3.2.

3.1. Search, see and download

When the View data buttom is pressed, the search window with keywords appears (Fig. 2). When one inputs search words such as the absorbing atom, edge and so on, a list of files appears at the bottom of the window (scrolling down may be necessary). The data can then be downloaded in *ATHENA* or *REX*2000 format, both of which contain energy and μt data with different headers.
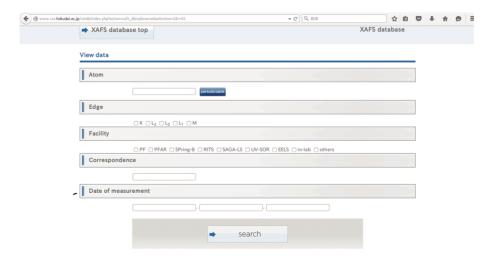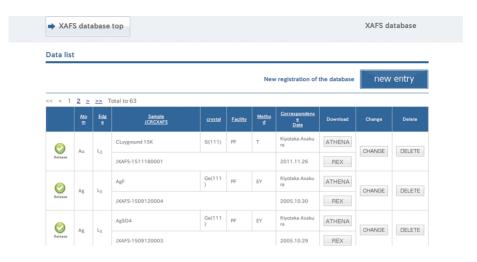

Figure 2
Data-search window.
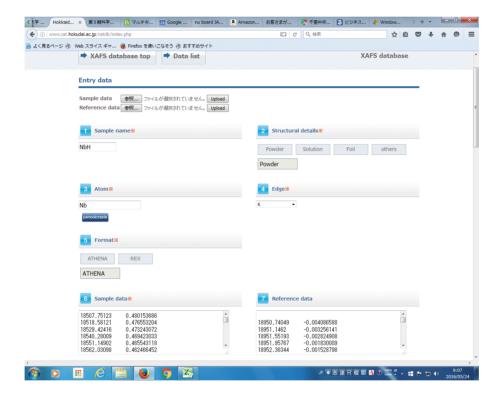
Figure 3
Deposition window.

Figure 4
Data-input window.

3.2. Deposition window

To deposit data, click on the 'New entry' button shown in Fig. 3 when the deposition window is open. The data-deposition buttons are at the top of the window (Fig. 4). The 9806 format used at Photon Factory and SPring-8 can directly be accepted using the upload button. If the data are not in the 9806 format they can be entered manually. The header is composed of the following items: sample name, sample details, absorbing atom, edge, format (*ATHENA* or *REX*), sample data (energy and μt are separated by a space on one line; the data can be copied and pasted), reference data (energy and μt in the same format) and reference name, date of measurement, mono-chromator, facility, beamline, beam current, beam energy, method of rejection of higher harmonics, method of focusing, method of measurement, detection gases, name of the experimentalist and comments.

4. The future

4.1. What is the problem with the database?

The problem is the small amount of data in the database. In November 2022, there were 1014 data sets in the database. The majority rule has been adopted to guarantee the data. However, it does not work well with such a small number of data. The only way to solve this is the automatic deposition of

all data to one storage computer just after measurement, followed by categorizing the data as an open or a secret file. Consequently, we may obtain a larger database in a few years and the majority rule will work to confirm the data quality. The question then is who will draw useful information from such a large database?

4.2. Artificial intelligence

The recent rapid developments in artificial intelligence (AI) will open new possibilities if such a large database is obtained. AI will search through the database and pick up XAFS spectra of the same standard samples. It will judge whether the data are valid or not by comparing them based on the majority rule, but minor spectra should be kept with a tag 'minor or less reliable'. In the future, each data file may have its own reliability factor based on the majority rule and other pieces of information (facility, signal-to-noise ratio, operator and so on) which can be determined by AI after deep learning. AI may then create a new database with reliability factors. If such a large database is obtained and is open to all, AI will search through the database and pick up similar XAFS spectra of standard compounds to the unknown compound that one has measured. AI will provide some hints about the structure of the material. We can then further analyse the results. The XAFS data files are preferably connected to other data sets of physical and chemical properties, synthesis details and so on (Ishii *et al.*, 2023), and a hierarchical data format (such as HDF5) will need to be developed for storage of these big data.

5. Summary

The JXS Open Source Database is an open database that uses the majority rule to confirm the validity of the data. The majority rule works better as the database becomes larger. In the future, AI should work to evaluate the reliability of standard XAFS data and provide hints to the structure of unknowns just by comparison with XAFS data for standard compounds. From this point of view, we need to construct a large XAFS database immediately, which should preferably be connected to other large materials databases. The automatic accumulation of measured XAFS data will be discussed in the international community (Asakura *et al.*, 2018).

References

Asakura, K., Abe, H. & Kimura, M. (2018). *J. Synchrotron Rad.* **25**, 967–971.
Ascone, I., Asakura, K., George, G. N. & Wakatsuki, S. (2012). *J. Synchrotron Rad.* **19**, 849–850.
Ishii, M., Tanabe, K., Matsuda, A., Ofuchi, H., Matsumoto, T., Yaji, T., Inada, Y., Nitani, H., Kimura, M. & Asakura, K. (2023). *Sci. Technol. Adv. Mater. Methods*, **3**, 2197518.