XAFS Database in Japan

Kiyotaka Asakura
Institute for Catalysis
ICAT, Hokkaido University
Introduction

Why do we need database?

1. We can guess the structure faster and more correctly.
   Strong demand from beginners and non-professionals

2. To create a new concept.

3. AI and DEEP LEARNING will give us directly the structure from the spectra based on database in near future.

Data-driven analysis

\[
\text{FeCl}_3 \text{ in PA} \quad \text{FeCl}_4 = \text{Co(NH}_3\text{)_6}[\text{FeCl}_6] \\
[\text{NEt}_4][\text{FeCl}_4] \\
[\text{NEt}_4]_2[\text{FeCl}_4]
\]

Introduction

Why do we need database?

1. We can guess the structure faster and more correctly.
   Strong demand from beginners and non-professionals

2. To create a new concept.

3. AI and DEEP LEARNING will give us directly the structure from the spectra based on database in near future.
   Data-driven analysis

Covalency and L3 edge peak height of Ag(I) are in proportion.

Introduction

Why do we need database?
1. We can guess the structure faster and more correctly.
   Strong demand from beginners and non-professionals

2. To create a new concept.
3. AI and DEEP LEARNING will give us directly the structure from the spectra based on database in near future.
   Data-driven analysis

http://cars.uchicago.edu/xaslib/search
   Graphical

http://ixs.iit.edu/database/
   Text base
SPring-8 BL14 Data repository https://sp8dr.spring8.or.jp/portal/dspace

   Sulfur database


XAFS SOCIETY OF JAPAN DATABASE
**XAFS database in Japan**

**History**
Catalysis Research Center (now ICAT) started to build the database of XAFS for future usage of data science in 2011, collaborating with XAFS Society of Japan (XSJ).

**System**
Web is prepared by ICAT and XSJ members upload their data.

**Fundamental principles**
- Volunteer base = anyone who is registered can deposit his/her data.
- Open database = anyone in the world can see and use it with citation.
- Low cost = no one manage the data because of no budget.
- Simple dataset = text base. Only energy and \( \mu t \) of raw spectrum.
- Compatibility = Foil data simultaneously measured should be deposited.
- Target = Mainly standard compounds. But some repository for the published data.
- Guarantee for accuracy = majority rule. We allowed the many data depositions of the same samples.
If we have 60 data with same quality (S/N ratio) of the same standard compounds. 

59 data sets are the same spectra and one is different from the other. The one should be rejected.

It requires number of data of the same compounds.

This one should be wrong.
Inside database


Search

Download

Click

The one where forgets a password.
Inside database


View data

Atom

Edge

Facility

Correspondence

Date of measurement

search
Inside database


View data

Input/change the data

Search

Download

Click

The one where forgets a password.
<table>
<thead>
<tr>
<th>Atom</th>
<th>Edge</th>
<th>Sample JCRCXAFS</th>
<th>crystal</th>
<th>Facility</th>
<th>Method</th>
<th>Correspondence</th>
<th>Date</th>
<th>Download</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nb</td>
<td>K</td>
<td>NbO2</td>
<td>Si(311)</td>
<td>PF</td>
<td>T</td>
<td>Kiyotaka Asakura</td>
<td>1984.06.20</td>
<td>ATHENA</td>
<td>DETAIL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JXAFS-1703150002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nb</td>
<td>K</td>
<td>Nb2O5</td>
<td>Si(311)</td>
<td>PF</td>
<td>T</td>
<td>Kiyotaka Asakura</td>
<td>1984.06.20</td>
<td>ATHENA</td>
<td>DETAIL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JXAFS-1412230001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FE</td>
<td>K</td>
<td>Fe2O3 XANES</td>
<td>Si(111)</td>
<td>PF</td>
<td>T</td>
<td>Kiyotaka Asakura</td>
<td>2017.02.25</td>
<td>ATHENA</td>
<td>DETAIL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JXAFS-1705190002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co</td>
<td>K</td>
<td>CoTPP_01</td>
<td>Si(111)</td>
<td>PF</td>
<td>T</td>
<td>Kiyotaka Asakura</td>
<td>2017.02.24</td>
<td>ATHENA</td>
<td>DETAIL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JXAFS-1705190001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rh</td>
<td>K</td>
<td>Rh foil</td>
<td>Si(311)</td>
<td>PF</td>
<td>T</td>
<td>Kiyotaka Asakura</td>
<td>2017.05.03</td>
<td>ATHENA</td>
<td>DETAIL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JXAFS-1705080004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RhCl3</td>
<td>Si(311)</td>
<td>PF</td>
<td>T</td>
<td>Kiyotaka Asakura</td>
<td>2017.05.0</td>
<td>ATHENA</td>
<td></td>
</tr>
</tbody>
</table>
Downloaded data

```
DATE=84.06.20
PLACE=PF
BEAMCURRENT(mA)=100
BEAMENERGY(GeV)=2.5
MONOCHROMATOR=Si(311)
OPTICS=
METHOD=Transmission(T)
\*GAS 10/1 = Ar(50)N2(50)/Ar(100)
MEASUREMENT TEMPERATURE=RT
REFERENCE=K_Nb205_Si311_19840620.txt
E(eV)  \mu t
18507.21312 -1.281775946
18518.04279 -1.284004024
18528.88541 -1.285567492
18539.74101 -1.288068686
18550.60962 -1.288701973
18561.49126 -1.290779827
18572.38504 -1.292854447
18583.2937 -1.294284416
18594.21456 -1.295542402
18605.14854 -1.297676177
18616.09667 -1.299018371
18627.05996 -1.30056751
18638.02944 -1.302029982
18649.01614 -1.304040825
18660.01608 -1.306237246
18671.02929 -1.307505888
```

```
GNUPLOT
Version 4.8 patchlevel 6  last modified September 2014
Build System: MS-Windows 32 bit
Thomas Williams, Colin Kelley and many others

gnuplot home:  http://www.gnuplot.info
faq, bugs, etc:  type "help FAQ"

Terminal type set to 'xt'
gnuplot> load 'C:\Users\Asakura Kiyo\Downloads\K_Nb205_Si311_19840620.txt'
gnuplot> plot 'C:\Users\Asakura Kiyo\Downloads\K_Nb205_Si311_19840620.txt'
```

![Graph showing data points with a downward trend](image)

Institute for Catalysis
Hokkaido University
### Entry data

**Template Download**

**Sample data**
- ファイルを選択
- Upload

**Reference data**
- ファイルを選択
- Upload

### 1. Sample name※

<table>
<thead>
<tr>
<th>Sample name ※</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

### 2. Structural details※

- Powder
- Solution
- Foil
- others

### 3. Atom※

<table>
<thead>
<tr>
<th>Atom ※</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Periodic table**

### 4. Edge※

```
▼ Please select ▼
```

### 5. Format※

- ATHENA
- REX

---

**Direct download from the original data (9801 format)**
Uploaded or copy \((E, \mu t)\) of sample

Uploaded or copy \((E, \mu t)\) of foil measured simultaneously.
### Web site and how to

#### 19 I gas

<table>
<thead>
<tr>
<th></th>
<th>He</th>
<th>N2</th>
<th>Ar</th>
<th>Kr</th>
<th>others</th>
<th>Ar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>50</td>
<td>70</td>
<td>75</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>others</td>
<td>clear</td>
<td>others</td>
<td>clear</td>
<td></td>
</tr>
</tbody>
</table>

#### 20 Measurement Temp.(°C)

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>others</th>
<th>clear</th>
</tr>
</thead>
</table>

#### 21 COMMENTS

[Blank field for comments]
What are the problems?

- Now August 12 2017. We have 194 datasets uploaded.
  - Too small and too biased.
  - Duplication and multiplication must be necessary.

- Sample name must be unified for easy link to other Material data base.

- Local data file name but no universal one. (Data DOI)
What are the problems?

• Now August 12 2017. We have 194 datasets uploaded.
  – Too small and too biased.
  – Duplication and multiplication must be necessary.
• Please help!!

• Sample name must be unified for easy link to other Material data base.
• Local data file name but no universal one. **(Data DOI)**
How to get more data = automatic depository and open after publication with DOI

Data are automatically uploaded to central computer and add their DOI but strongly protected by PWD

Users can access the data with PWD
How to get more data = automatic depository and open after publication with DOI

Users can access the data with PWD

Data are automatically uploaded to central computer and add their DOI but strongly protected by PWD

Submission

Journals (publisher)
How to get more data = automatic depository and open after publication with DOI

Users can access the data with PWD.

Data are automatically uploaded to central computer and add their DOI but strongly protected by PWD.

Open data requirement when the paper are accepted.

Open for public with DOI.
How to get more data = automatic depository and open after publication with DOI

- Data are automatically uploaded to central computer and add their DOI but strongly protected by PWD.
- Users can access the data with PWD.
- Journals (publisher) are required to open data when the paper is accepted.

- Open for public with DOI.
- All data should have DOI and should be cited by others.
- But all data published can freely be reproduced with DOI.
Proposals

To facilities stuffs. Please take action to store all data measured at your facility in your central storage with strong security with temporal local DOI.

To users. Please allow the facility to store your data.

To communities and Q2XAFS. Please discuss the rule about the data format, DOI and rule for open the data.

To journal editors or publishers. Please require the authors to open their data when accepted.

To authors. Please open the data in paper and please cite all data DOI used.
But what format.

Any data are OK if they include the sample chemical formula(+phase), Date, Facility name, and raw data (E, µt). We can make data convertors.

DOI

FacilityBeamLineYDTContactPersonName_BranchNo. is given automatically.

For example. PFARNW10A20170505235959KiyotakaAsakura_1
I propose following as the Q2XAFS missions.

1. Decide rule for the data file name or DOI and start the discussion for data opening.
2. Ask all facilities to start the collection of XAFS data.
3. Decide the sample name rule to link other database.
4. Propose the action plan and send advice to IXAS and XAFS Commission in IUCr.
5. The final decision in 2021(XAFS2021)
The presentation is based on the discussion with following people

H. Abe(PF), M.Kimura(PF), Y.Niwa(PF), Y.Nitani(PF), Y.Tamenori(SP8),
T.Honma(SP8), S.Ito(NIMS), Tanaka(NIMS),
K.Takahashi(NIMS),T.Chigyo(NIMS), K.Funatsu(Univ. Tokyo),
J.Hasegawa(ICAT), M.Nishida(ICAT), J.Lauterbach(USC)
K.Shimizu(ICAT), W.Ueda(Kanagawa),H.Takaya(Kyoto U)
XAFS Society of Japan
What next?

AI supported analysis or data driven analysis. Compare the data and give the conclusions.

But the number of experiments are finite.

We still need more data.
We need more data.

Automated calculations using some theoretical analysis programs

(FEFF, GNXAS, EXCURVE and so on.)

Theory Database should have input file and output chi(k) which both are automatically deposited.

Provide the cloud service or hub for analysis programs, where people freely access programs and calculate their own structure but automatically results are downloaded and stored. (Like nanoHub)
How to construct and control the database

Company(-ies) run(s) the hub and database.

Government(s) control(s).

A NPO or A Universities Alliance.

“All knowledge is the knowledge of all of us, controlled by all of us and used for happiness of all of us”
All data should be deposited on crowd and linked

All data have data name and DOI.

Analysis Data
EXAFS, XRD, XPS, and so on
Synthesis data
IoT;Robotics

Web
Published data

Public Files in the one
Crowd service
AI technology Utilization

Notebook
Internet LaboNote
Metafile to connect all dataset.

Cite the data name or DATA DOI.