

X線吸収分光の統合型データベース MDR XAFS DBの構築と公開

石井真史

国立研究開発法人物質・材料研究機構 〒305-0047 茨城県つくば市千現 1-2-1

要旨

研究開発において、「第四の科学」と呼ばれるデータ中心の研究手法（データ駆動型研究）が一般化しつつある。X線吸収分光（X-ray Absorption Fine Structure, XAFS）のデータ共有・再利用の国際的な議論が活発化する中で、日本の放射光コミュニティも方針を示す時期が来ていると思われる。本稿では、国内のXAFSに関係する機関が、データを一か所にまとめて公開するMDR XAFS DB（Materials Data Repository XAFS Database）プロジェクトの活動を概観し、構築・公開されたデータベースについて、利用法を含めて実用的視点で紹介する。またデータ共有・再利用に欠かせないメタデータの統一に関して、現状で主軸となっているXDI（XAS Data Interchange）の考え方を述べるとともに、これに対して、本プロジェクトでも採用している日本国内の共通メタデータの位置づけを示す。特に他国にはない「横断検索」に焦点を当て、メタデータの値をマネージする独自の戦略と、具体的な活動を明らかにする。すなわちMDR XAFS DBの強みである、すべてのデータを等しく検索できる機能の根幹となっている材料辞書の構築と、そこでの目的志向の語彙編纂方針を述べる。更に、本データベースにおける課題をまとめ、今後の放射光データの共有・再利用に向けて、コミュニティとして必要なことを論ずる。

1. 序

X線吸収微細構造（X-ray Absorption Fine Structure, XAFS）^{1,2)}はスペクトルを参照しあうことで解釈が深まることから、いわゆるデータ駆動型の研究が盛んになる以前から「データ共有」に対する理解があり、施設やコミュニティの中で、データベースの構築や、どのようにすれば広くデータを共有できるかの議論が進んできた³⁾。一方で、最近のデータ駆動型研究における「データ再利用」の動きは、これまでの議論の範囲を越えて、機械でデータを扱うための更に踏み込んだ要件を求めている。日本の主要なXAFS関係機関がデータを持ち寄り、物質・材料研究機構（National Institute for Materials Science, NIMS）の材料データリポジトリ（Materials Data Repository, MDR）で統合し公開するプロジェクトMDR XAFS DB⁴⁾は、取り組みの開始時期が、データ駆動型研究の高まりとちょうど重なることから、比較的新しい視点でデータの共有と再利用に取り組んできた。旧来の利用者の声や最近の科学の状況を考えると、共有と再利用の両方を念頭にプロジェクトを進める必要があることは間違いない。

- GUI（Graphical User Interface）での操作を前提としたwebサービスでの利便性や、そもそもXAFSのデータ共有に必要な要件
- API（Application Programming Interface）でのデータ取得や、そのあとのバッチ処理、機械学習などでのデータ再利用に必要な要件

を、大括りに「人可読」と「機械可読」の要件と読み替えたとき、MDR XAFS DBはそれぞれに対してどのようにサービスを展開しているかを、これまでの経緯を振り返りつつ、実用的視点で述べるのが本稿の目的の一つである。一方、本稿で頻出するにもかかわらず「機械可読」と「人可読」の定義や境界は明確ではなく、特に「機械可読」については「読める」だけから「厳密に理解する」までの広い幅がある。更に、実験・理論・計算に続くデータ中心の科学「第四の科学」⁵⁾の時代に突入し、データの資産化が進み、生成AIが世界的なブームとなっている現在において、データベースの在り方は、「人可読」と「機械可読」の考え方が混じりあいながら、刻々と変化していくであろう。放射光データを、個人の研究のみならず科学全体や社会の発展につなげるために何をすべきか、読者と共に考え、日々続く活発な議論とデータベースの構築に多くの方に加わっていただくきっかけにしたい、と考えている。

2. MDR XAFS DBの歴史

XAFSコミュニティには、よく知られているFarrel Lytle Databaseのように長い歴史を持つ大きなデータベースがある⁶⁾。LytleのデータベースとMDR XAFS DB⁷⁾を見比べると、記録や眺めるためのデータベースから、誰にでもわかり使われるデータベースへ移行し、まさに「人可読」から「機械可読」への動きが確実に進んでいることがわかる。ここでは国内での動きを追ってみることにする。

2011年に北海道大学 触媒化学研究所で構築が始まった XAFS データベース⁸⁾をはじめとして、各所で様々な形態のデータベースが公開されはじめた。施設を越えた XAFS データを俯瞰する、光ビームプラットフォーム (科学技術振興機構 先端研究施設共用促進事業 2016-2020 年)⁹⁾が進む中で、データベース統合の観点での大きな転機は 2018 年に北海道大学で開催された XAFS Database Workshop であったであろう。以降、ユーザーの利便を考えたデータ共有のための会合が年 1, 2 回の頻度で開催され、データの公開の方針、権利、データ提供者へのインセンティブなど、データベース統合における様々な課題の議論が続いている。多くの模索・試行が続く中で、2020 年には MDR でのデータベース統合が決まり、2021 年 7 月にはファーストデータが公開され、以降参画機関を増やし、2023 年には 6 機関がデータを登録し、施設を越えた横断検索が可能になった。現在は更に参画機関が増える見込みであり、ウラン化合物などアクチノイド系への拡張も検討が進んでいる。2024 年 9 月現在、合計 2,263 の XAFS スペクトルが公開され、吸収端元素数は 56, K 吸収端数 50, L 吸収端数 24 が収録されるに至っている。データの統計情報を含む詳細は

<https://mdr.nims.go.jp/concern/datasets/vh53wz94c> にまとめてある。後述の通り、データ共有・再利用において必要不可欠なメタデータ (研究データの素性を明示する「データのデータ」) のマネージメントに関する取り組みも進み、放射光科学における共有データ基盤作成の先行事例となっているといえる。

3. MDR XAFS DB の人可読な使い方

ここで GUI をベースにした、MDR XAFS DB の人可読な使い方を簡単に述べておく。ここでは硬 X 線の XAFS の標準試料としてよく用いられる銅を検索例として添えておく (スペクトル数などは 2024 年 9 月現在)。

(1) トップページへのアクセス

<https://doi.org/10.48505/nims.1447>

MDR XAFS DB にはこの DOI が付与されており、ユーザー登録などの手続きをしなくてもどこからでも、だれでもオンラインでアクセス可能である。ここには MDR XAFS DB の概要や、論文などで参照する際の例文などが記載されている。トップページのスクリーンショットを Fig. 1 に示す。

MDR XAFS DB

User Collection MDR Open

2264 items · Created by: ISHII, Masashi · Last Updated: 19/06/2023

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
- Photon Factory, KEK
- Aichi SR, Aichi Synchrotron Radiation Center
- SAGA-LS, Kyushu Synchrotron Light Research Center

The data published in this repository can be used to determine the electronic state and local structure of materials. Please refer to the readme for file formats and other details. For details of analysis methods, please refer to the many references available.

Information on substances and measurements is compiled from metadata provided by relevant organizations into a format suitable for data linkage on the Data Platform (DPF), and used for display and search. For substances, IDs (Q numbers) are assigned, and the substances are classified and managed in a dictionary. This dictionary can be browsed as "NIMS XAFS DB Project Materials Dictionary" on NIMS's materials vocabulary website MatVoc Explorer. For measurements, information from each organization is included as is, so it can be used for new measurements and analysis.

The metadata provided by the related organizations can be downloaded in three different formats depending on the purpose of use.

- tsv format: Converted to a flat Key-Value format that is easy to display in MDRs and spreadsheets.
- json format: A hierarchical structure expressed in a format suitable for organizing data in a machine.
- yaml format: A hierarchical structure expressed in paragraphs, etc., that is easier for people to read than json format.

Collection Details	
Total items	2264
Size	unknown
Resource type	Dataset
Creator	ISHII, Masashi
Contributor	MATSUDA, Asahiko TANABE, Kosuke NAGAO, Hiroko YOSHIKAWA, Hideki
Keyword	x-ray local structure absorption spectroscopy
License	Creative Commons BY-NC-SA Attribution-NonCommercial-ShareAlike 4.0 International
Publisher	National Institute for Materials Science
Date Created	2021-07-21
Language	English
Identifier	https://doi.org/10.48505/nims.1447

Related URL

- http://www.ritsumei.ac.jp/acd/re/src/sx_xafs_db/
- <https://www.saga-ls.jp/main/2026.html#gsc.tab=0>
- <http://ttan.nusr.nagoya-u.ac.jp/ASR-XAFS-DB/>
- https://www.cat.hokudai.ac.jp/catdb/index.php?action=xafs_login_form&opnid=2
- <https://support.spring8.or.jp/BL/bll4b2/xafs/standardDB/>
- <https://user.spring8.or.jp/sp8info/?p=2751>
- <https://pfxafs.kek.jp/xafsdata/list.php>

Fig. 1 (Color online) MDR XAFS DB Top Page (<https://doi.org/10.48505/nims.1447>).

(2) データベースへのアクセス

(1) トップページでの「Access MDR XAFS DB」または「MDR XAFS DB 閲覧」をクリックすることで、データベースへアクセスできる。ここでは全スペクトルが、データのタイトルと Keywords (次に詳述) などがサムネイルとともに一覧で表示される。

(3) スペクトルのスクリーニング

所望のスペクトルに迅速に行き着くためには、「Keywords」によるスクリーニングが最も有用である。ここで Keywords として、全機関の全スペクトルに共通する約 10 個の項目を定め、各スペクトルにその内容をタグ付けしている。この共通項目の中には、物質名、化学式、吸収端、分光結晶方位、ビームライン名、施設名を含んでいる。Keywords はアルファベット順 (A-Z SORT) か、登録数順 (NUMERICAL SORT) を選ぶことができ、表示項目に対して完全一致の条件でスクリーニングされる。銅の例では、「Copper」を使うのがよい。現在、各機関から提供された 13 のスペクトルの一覧に絞り込まれるはずである。この一覧に表示されているサムネイルを見れば、広域 X 線吸収微細構造 (EXAFS) か X 線吸収端近傍微細構造 (XANES) かの区別が付き、また提供施設や分光結晶

の情報から、利用目的に合ったスペクトルが選択できる。

(4) スペクトルデータの詳細

希望のスペクトルが見つければ、タイトル「XAFS spectrum of X」(X は材料名) をクリックすれば、スペクトルデータに行き着く。例えば 13 ある銅のデータ「XAFS spectrum of Copper」の内、Aichi SR の BL5S1 を選択した場合を Fig. 2 に示す¹⁰⁾。MDR XAFS DB では、この詳細にある全ファイルおよび MDR メタデータ (本稿では、GUI で表示されている MDR 共通のメタデータを「MDR メタデータ」と呼ぶことにする。) のセットを「Work」と定義して、データ処理の単位としている。現在、MDR XAFS DB に収録しているデータは、多くが標準試料の単一スペクトルであるが、一つの試料に対して複数のスペクトルを収集している場合、参照スペクトルを添えてある場合、より複雑な補遺データがある場合などについても、Work に格納することで一括管理できる。また、サムネイルや後述のメタデータファイルについても、Work を単位にして作成して格納している。

この画面では、(3) よりも大きくサムネイルが表示され、また XAFS メタデータ (本稿では、データ提供機関が作成した、XAFS の専門家のための詳細なメタデータを

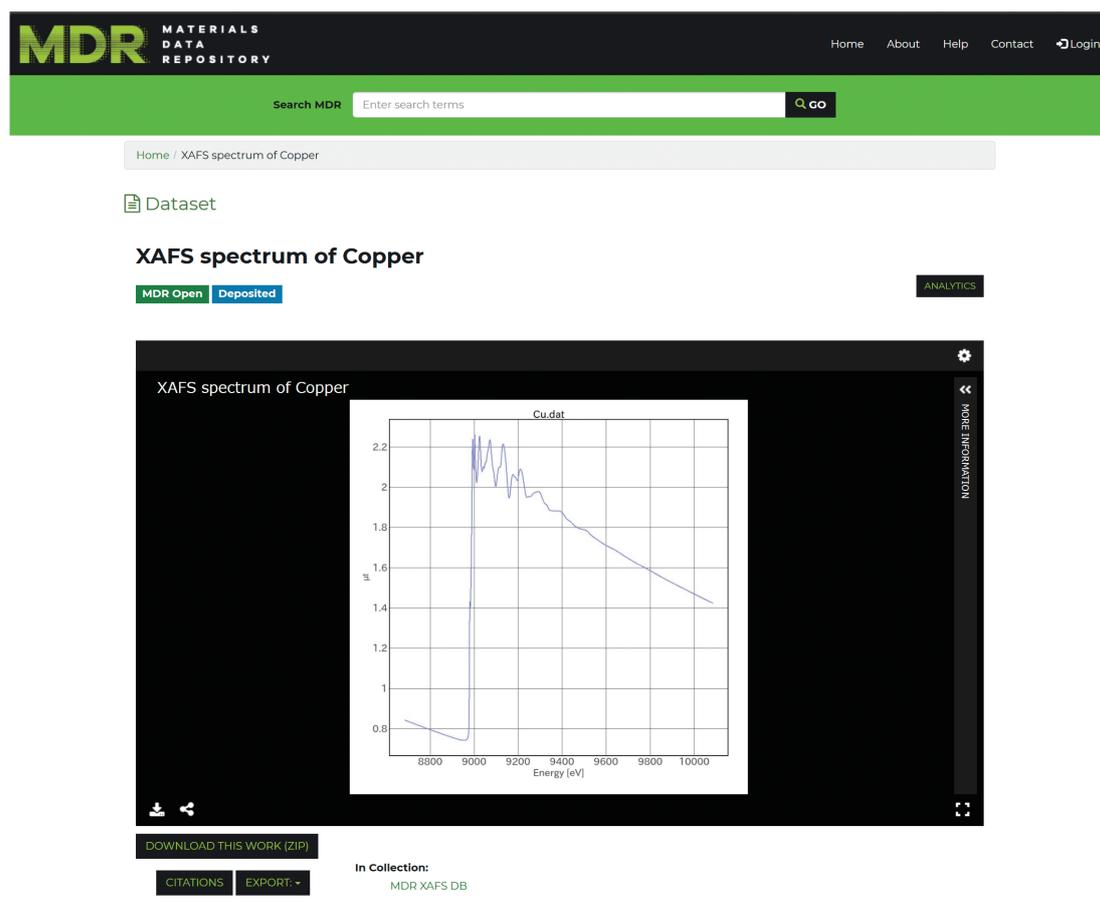


Fig. 2 (Color online) Example of access to Work in MDR XAFS DB (<https://doi.org/10.48505/nims.3899>).

「XAFS メタデータ」と呼ぶことにする) が、Preview 機能で表示される。Fig. 2 の Aichi SR の例では、XAFS メタデータは 277 項目あり、そのうち 200 項目が Preview で簡易的に確認できる。

(5) データのダウンロード

Work に含まれる全ファイル一覧は、(4) の Preview 機能の画面の下に表示され、個々のファイル名をクリックすれば、ファイル形式やサイズなどの情報が表示される。あるいはファイル名の後にある Actions ボタンを使ってローカル環境にダウンロードすることが可能である。上記の Aichi SR の例では Table 1 のファイルがダウンロード可能である。この表にファイルの内容も記しておく。本データベースでは、これらと MDR メタデータがセットで Work を形成している。XAFS メタデータは、内容が同一の三つの形式が揃っており、目的に応じて使い分けることができる。現在、データ登録時によく提出されるメタデータ形式は YAML (Yet Another Markup Language) であるが、表現方法が直感的な JSON (JavaScript Object Notation) と、TSV (Tab Separated Values) が YAML から変換生成され、提供されている。TSV が最も構造がシンプルで人可読性が高く、実際、これが (4) の Preview 機能で表示されている。

4. MDR XAFS DB のシステム的特点

「3. MDR XAFS DB の人可読な使い方」で、Work をデータ処理の単位としていることを述べた。実際、MDR XAFS DB では全 Work に DOI (Digital Object Identifier) を付与しており、実質的に一つの実験結果に対して一つの DOI が対応していることになる。論文一報に対して、一つの DOI がつくことは多いが、データを DOI の単位にすることは、機械可読なデータ共有の特徴の一つと言えよう。この「データの web (web データ空間)」とも言える先端的扱いはセマンティック web 技術として知られており、その詳細は web 技術の標準化を進めている W3C (World Wide Web Consortium) で示されている¹¹⁾。実際は、こうしたデータ流通に関する技術的背景は認識しなくとも、

ユーザーは一つ一つのデータに DOI によってどこからでもアクセスでき、MDR XAFS DB の重要な特徴の一つになっている。

Work の各データを、手でダウンロードする方法は、「3. MDR XAFS DB の人可読な使い方」で紹介したが、一括ダウンロードのための API が別途提供されている。この API の特長は、GUI 上でスクリーニングした結果の URL を、出力形式を指定する数文字を加える程度で、ほぼそのまま API のクエリパラメータに転用できることである。一般的に API のクエリパラメータは、習熟が難しい場合が多いが、MDR XAFS DB の場合は、その障壁を極めて低くしてある。こうした GUI と API の関係は、「人可読」から「機械可読」へのシームレスな橋渡しといえる。API 利用方法の詳細は

<https://dice.nims.go.jp/services/MDR/manual/html/api.html>

に記載されている。ここでは、Python で書かれたサンプルプログラムも示されており、目的に応じたアプリケーションの開発が容易に可能である。

5. MDR XAFS DB プロジェクトのメタデータ

5.1 メタデータ項目の整備によるデータ再利用

MDR XAFS DB プロジェクトは、最近の機械可読の視点でのデータの再利用を目標の一つとしてきたが、一方で旧来の人可読な視点でのデータ共有の取り組みも並行して進めている。この両立では、メタデータのマネージメントが極めて重要になる。メタデータについては、百人百様の考え方があるが、これをできるだけ総意に近く客観的にながめるために、国際・国内の XAFS メタデータの統一に関する動きを振り返ってみる。

2011 年の The International Workshop on Improving Data Quality and Quantity in XAFS Spectroscopy (Q2XAFS2011)において、既に IXAS (International X-ray Absorption Society) のワーキンググループによって、XAFS の推奨フォーマットとして、XDI (XAS Data

Table 1 Example of files that can be downloaded at Work in MDR XAFS DB.

ファイル名	内容
Cu-AichiSR-BL5S1-20220927.dat	スペクトルのオリジナルテキストデータ
Cu-AichiSR-BL5S1-20220927.info	Aichi SR 固有の測定に関する補遺データ
Cu-AichiSR-BL5S1-20220927.png	サムネイルの画像ファイル
metadata.yml	yml 形式の XAFS メタデータ
metadata.json	json 形式の XAFS メタデータ (metadata.yml からファイル形式を変換したもの)
primary.tsv	tsv 形式の XAFS メタデータ (metadata.yml からファイル形式を変換したもの)

Interchange)¹²⁾ と HDF5 (Hierarchical Data Format version 5) が提示されている³⁾。XDI は、スペクトルの素性を知るための必要最小限のメタデータを付与することを前提としており、一方で特殊環境下の高度な実験のように複雑なファイル構造とメタデータを必要とするものは、HDF5 にすることを推奨している。現在 XDI では、

Family.Field: Value

の形式により、大分類 (Family) と小分類 (Field) の組み合わせで、項目名に階層構造を持たせ、それに対して値を与える形式になっている。これらの分類の定義は

<https://github.com/XraySpectroscopy/XAS-Data-Interchange/blob/master/specification/dictionary.md>

で与えられている。大分類は Table 2-1 の 8 つであるが、ここで括弧内の数は、各大分類に属する小分類の数を表す。この表に示す通り、XDI の小分類をすべて書いたとしても 29 項目しかなく、実験の再現性や品質を担保するメタデータは、スコープ外といえる。更に必須メタデータは以下の三つのみである。

- **Element.symbol:** 吸収元素。興味深い点として、これを必須メタデータに定めた理由として、スペクトルを見ただけでは、吸収端が偶然一致する別の元素との区別がつかないためとしている。例えば Cr K 吸収端と Ba L1 吸収端はいずれも 5989 eV であり、ちょうどこれに相当する。翻って考えると、もしこうした偶然の一致がなければ、吸収元素はハンドブックなどを調べればわかることから、必須メタデータにならなかったことを示している。一般知識を前提としたメタデータは、明らかに人可読といえ、XDI の位置づけが垣間見える。

- **Element.edge:** 吸収端。ここでも、IUPAC (International Union of Pure and Applied Chemistry) による表記は示しつつ、L1-L3 など吸収端をまたぐ測定の場合は、L と記載することを容認している。しかしこれは機械可読の観点では、表記の揺らぎに他ならない。

- **Mono.d_spacing:** モノクロメータ結晶の間隔。これは、結晶に対する X 線の入射角を光子エネルギーに換算する際に必要なメタデータである。ただし、換算式は XAFS ドメインには常識ではあるものの、一般にデータ再

利用を促すのであれば、むしろ換算後のデータを併記することが望ましいであろう。

XDI の推奨メタデータは以下のとおりである：

- **Facility.name:** 放射光やほかの X 線施設の名前
- **Facility.xray_source:** 偏向磁石やアンジュレータなどの光源名
- **Beamline.name:** ビームライン名
- **Scan.start_time:** 測定開始時間
- **Column.1:** 測定配列の横軸

一般的に、施設名、光源名、ビームライン名は、エネルギー分解能や光の強度を知り、スペクトルの概要を知るのに役に立つと考えられるが、XDI ではこの記載がないことも容認している。必須項目の **Element.symbol** と同様に、スペクトルが全ての情報を持っているという考え方に基いているといえる。

メタデータ項目やデータ構造が、個別の実験内容に強く依存する HDF5 については、ここでは全体的特徴を述べるに留める。階層構造によって様々なファイルを整理よく格納することができ、圧縮ファイルの取り扱いが容易であり、チャンクと呼ばれる小さな単位で読み書きを行うため処理速度が早いなど、大規模なデータを再利用するのに優れている。ファイルの構造を与える自己記述型のフォーマットであり、アプリケーションがまさに「機械可読」に内容を理解し、目的に合った使い方ができる。しかし、XAFS に関するスキーマを別途定義する必要があり、NeXus と呼ばれる放射光やミュオン施設のためのスキーマからの援用が有望視されている¹³⁾。ここに、XAFS 固有の知識を導入し、実験データ説明するのに十分なメタデータの項目をリストアップし、分類し構造化する必要がある。この試みは既に NXxas として一部が文書化されている¹⁴⁾。NeXus の狙いは、様々なオリジナルデータ (時系列データを含む) を構造化して保存するとともに、データプロセスに関してトレーサブルな状況を作り出し、一方で統一形式により様々なアプリケーションの迅速な開発や共有を実現し、大量データゆえに難しくなるデータ循環を活性化することにある。こうしたトップダウンの動きにもかかわらず、XDI や HDF5 が現場で真に活用されるために

Table 2-1 Major categories of XDI metadata, their definitions and number of subcategories.

項番	大分類	定義と小分類数
1	Facility	測定が行われた施設 (4)
2	Beamline	ビームラインの構造と光学系の情報 (4)
3	Mono	分光器の情報 (2)
4	Detector	検出系の詳細という情報 (4)
5	Sample	試料の調合と測定の情報 (6)
6	Scan	測定スキャンに関する情報 (3)
7	Element	吸収元素の情報 (4)
8	Column	データ列とその単位に関する情報 (2)

は、これらのデータ形式そのものか、互換性が高いデータがビームラインで使われる必要があり、多くの施設でそれが整備されるまでには、強い動機と具体的な行動が必要である。

XDIの29のメタデータは、項目数を必要最小限に減らすことで普及が期待できる一方で、スペクトルの再現性や信頼性を検証するためには十分とは言い難い。これに対して、日本でのメタデータ統一の動きは、NeXusほどの機械可読性はないものの、メタデータの項目数はXDIよりもずっと多く、スペクトル取得の正確な記録を目指している点で独自色が強い。例えば2017年に、光ビームプラットフォームで検討されていたメタデータの大分類はTable 2-2のとおりである。括弧内は、XDIの小分類とほぼ同じレベルで考えたときの項目数を示す。この表を見ると、小分類でも50を超え（うち必須項目16）、更に下位の細かな項目を考えると、記載すべき内容はXDIよりずっと多いことが分かる。必須項目は、リングの蓄積電流の他、ほとんどが装置のID（Identification）であり、施設におけるデータ管理を重視していた。すなわち、この時点では機械可読によるデータ再利用は主目的ではなかったようである。一方で名前空間の定義など、国際ルールに基づいてメタデータをXML（Extensible Markup Language）で設計しており、拡張性は高い。

これに対し、2022年から2023年にかけて、放射光学会のデータ構造化諮問委員会¹⁵⁾の分光分科会が作成した共通メタデータ¹⁶⁾は、サンプル関係の項目が増え、データ

の広い再利用を考えた構造になっている。大分類はTable 2-3のとおりであり、他と同様に括弧内に小分類項目数を示す。この表に示した通り、小分類は250項目を超え、ほかのメタデータに比べて圧倒的に数が多い。特に、サンプルに関してよく知られているID（グローバルID）の記載項目を設けるなど、外部の情報との連携を考慮していることが特徴である。一方で、「5. measurement」、「6. sample」、「7. reference」は測定者自身が入力する必要があり、その合計156項目を完全に記載することは、ほぼ期待できないであろう。むしろ全項目の記載というよりは、予め準備された項目を必要に応じて使って、データを系統的に整理できることが、このメタデータ設計の思想といえる。実際、項目数が多いにもかかわらず、必須項目は、データ提供者の情報、施設名など9つに過ぎず、ユーザーの入力項目の中では、サンプル名のみが唯一の必須項目になっている。また実験そのものより、むしろデータに対する責任を明確にする項目が必須とされており、翻ってデータの権利が前面に出ていることを意味している。この共通メタデータは、作成された時期が「2. MDR XAFS DBの歴史」で述べた、参画機関が増加した時期と一致していることからわかる通り、MDR XAFS DBプロジェクトで使われることを想定しており、実際新しい参画機関から提供されるデータには、この適用が始まっている。「3. MDR XAFS DBの人可読な使い方」で紹介した、Aichi SRのmetadata.ymlの内容は、まさにこの共通メタデータに準拠しており、その内容が277項目に及んだ理由は、メタデー

Table 2-2 Major categories of Optical Beam Platform metadata, their definitions and number of subcategories.

項番	大分類	定義と小分類数
1	lightsource	施設名やビームライン名などの光源情報 (3)
2	optics	モノクロメータやミラーなどの光学素子の情報 (17)
3	detector	イオンチェンバーやSSDなど検出器に関する情報 (29)
4	measurement	実験開始日や試料名など測定に関する情報 (8)
5	comment	コメント (1)
6	research_subject	実験者や所属の情報 (3)

Table 2-3 Major categories of MDR XAFS DB metadata, their definitions and number of subcategories.

項番	大分類	定義と小分類数
1	data_info	データ登録者や組織、データ作成日などの情報 (27)
2	facility	施設名、蓄積電流やフィリングパターンなどの施設に関する情報 (12)
3	files	データファイル名、フォーマット、コラムの定義などの情報 (14)
4	instrument	光学系（モノクロ、ミラー、スリットなど）の配置などのビームラインにおける光導入に関する情報 (45)
5	measurement	吸収端、実験開始日時、エネルギー校正、エネルギースキャン、検出器、計測器などに関する情報 (101)
6	sample	サンプル名、化学式、グローバルID、形状、測定温度などの情報 (33)
7	reference	参照試料名、化学式、グローバルIDなどの情報 (22)

タ項目をできるだけ埋めることで記載例を示すことを試行したためである。また、「3. MDR XAFS DB の人可読な使い方」の (5) データのダウンロードで述べた通り、この共通メタデータは Work ごとに YAML などのファイルで提供されており、XDI のように、スペクトルデータのヘッダとして提供される形式とは異なる。大雑把な言い方をすると、日本のメタデータやデータ構造は、XDI と HDF5 の中間にあると考えられる。現在のデータ共有 / 再利用の状況から考えると、この形式で大半のデータに対応できると考えられるが、将来においてもデータベースの登録対象をこの範囲に留めるか、HDF5 に転向し、より網羅的にスペクトルを収集再利用してゆくかは、今後の検討事項である。現在のデータ再利用を考えた日本の共通メタデータが、国際標準にどのように織り込まれてゆくかは、2025 年の Q2XAFS で決まると考えられる。

5.2 メタデータ値の整備（語彙辞書作成）によるデータ再利用

メタデータはそもそも、項目名と値の組み合わせで与えられる。5.1 はメタデータ全体を議論しているように捉えられがちであるが、正確にはメタデータの項目名を議論したに過ぎない。MDR XAFS DB プロジェクトでは、項目名にとどまらず値の統一を重視している。それは項目名を統一するだけでは、値を処理する際に、結局は「人可読」にならざるを得ず、現代のデータ再利用の「機械可読」な要件に十分応えられないためである。これが如実に現れるのは、試料名であろう。試料名の表記の揺らぎを「人可読」に処理することは限界がある。特に大量の XAFS のデータの再利用のためには、まず試料名・吸収端名でスクリーニングされ、目的のデータが見つかる (Findable¹⁷⁾) ことが第一であろう。MDR XAFS DB における、施設を超えた横断検索を最重視した取り組みでは、まさにこの観点が根幹をなしている。「3. MDR XAFS DB の人可読な使い方」の (3) スペクトルのスクリーニングで紹介した、Keywords の背後にある語彙辞書について、ここで述べておく。

施設名、ビームライン名、吸収端名については、表記にそれほど多くの揺らぎがあるわけではないので、横断検索のための語彙統制にさほど労力は必要としない。一方で、データ提供機関から提示される様々な試料名を、名寄せして管理することが、すべてのデータを等しく俯瞰する上で必須となることは、MDR XAFS DB プロジェクト開始当初から予想していた。そこで NIMS で運用している、語彙管理基盤 MatVoc に、「NIMS XAFS DB プロジェクト材料辞書」を整備してゆくことにした⁴⁾。MatVoc では、MDR XAFS DB 以外にもプロジェクトごとに、それぞれのドメイン知識に整合するように語彙を定義し、階層構造を作っている。「NIMS XAFS DB プロジェクト材料辞書」の中の化学物質名の管理に関する階層構造を Fig. 3 に示してお

く。紙面の都合で、全てを書き出すことはできないが、以下のサイトで容易に確認することが可能である。

<https://matvoc.nims.go.jp/explore/ja/results/Q713>

世界には膨大な材料が存在するにもかかわらず、XAFS で良く扱われる材料はそれほど多くはないため、階層の定義や粒度は辞書内で不均一ではある。本来は一般的な材料体系と合致することが理想的であろうが、MDR XAFS DB に登録される試料のみ、横断検索の観点で、大きすぎず小さすぎないグループにまとめて登録している状況にある。一方で、MatVoc では階層間で概念を継承しているの、別の辞書があれば齟齬なく連携できることを考えると、単体では材料科学として理想的ではないにせよ、語彙管理としては妥当と考えている。むしろ今後、複合材料・デバイスなどの構造体が登録される場合など、概念継承が難しい物質の辞書登録は、答えのない課題として残っている。現状はそこに至る前の、様々な試料名を代表名に紐付け、一意の番号 (MatVoc 内では Q + 番号で ID 管理するため、QID と呼ぶ) を付番することで、「人可読」から「機械可読」への語彙変換の辞書として十分な役を果たしている。例えば、「3. MDR XAFS DB の人可読な使い方」で紹介した通り、銅は 13 スペクトルが登録されているが、データ提供機関によって、Cu と表記する場合もあれば、Copper や Cu-foil などと書かれる場合もある。これらの材料を全て、Q1426 の QID (代表名 Copper, <https://matvoc.nims.go.jp/explore/ja/results/Q1426>) に名寄せすることで、横断検索が実現できる。もちろん、厳密には 13 の銅は同一物

```

化学物質 (Q2828)
  有機材料 (Q714)
    酢酸塩 (Q717)
      酢酸コバルト(II) (Q1349)
    ...
    フタロシアニン (Q741)
      鉄(II)フタロシアニン (Q1458)
    ...
  無機材料 (Q715)
    臭化物 (Q721)
      臭化金(III) (Q1305)
    ...
    炭化物 (Q722)
      炭化モリブデン(II) (Q1573)
    ...
  生体材料 (Q3735)
    システイン (Q3736)
      コバルトシステイン (Q3765)
    ...
    グルタチオン (Q3737)
      コバルトグルタチオン (Q3766)
    ...
  ...

```

Fig. 3 Hierarchical structure of “Chemicals” in NIMS XAFS DB Project Materials Dictionary (<https://matvoc.nims.go.jp/explore/ja/results/Q713>).

質ではないし、研究・開発目的によっては分けるべき時もある。しかし、XAFSの横断検索という目的志向で辞書を編纂し、関係する可能性が高いデータをFindableにし、比較できるようにすることがデータベースの役割と考えている。現在本辞書には、材料名 955、化学式 748 が登録されている（2024年9月現在）。

6. MDR XAFS DB の課題

6.1 エネルギー校正

MDR XAFS DBに限ったことではないが、そもそもXAFSスペクトルの絶対エネルギーを確定することは、モノクロメータの精度を考えても、事実上不可能であろう。考え方としては、EXAFSのフーリエ変換をベースにした解析目的に特化する、理論計算に合わせる、といった方向付けはあり得る。MDR XAFS DBでは、各データ提供機関のエネルギー校正のルールをまとめて、

<https://mdr.nims.go.jp/concern/datasets/vh53wz94c>

に掲載している。よく知られたCu K吸収端のプリエッジピークのエネルギーの各機関の基準値はTable 3に示す通りとなっており、これらの値には1 eVを超える幅がある。ここでどれが正しい、間違えているという議論は意味がないであろう。原理的にはMDR XAFS DBの標準を定めて、それに合わせてデータを補正することは、あり得るかもしれないが、他の元素を含めた全試料の補正をすることは、不可能である。コミュニティとして統一見解を出すには幾多の困難が予想されるが、比較的網羅的に定められたSPring-8のエネルギー校正ルール（上記URLを参照）は、議論の土台として有用かもしれない。また別の考え方として、参照用の標準試料のスペクトルをWorkに格納することも、現実的な解決法として有望かもしれない。しかし、標準試料を何にするか、時として結晶性など、試料依存性が高い要素が含まれた場合の扱いを考えると、統一したルールにすることは容易ではなく、データの再利用を保証する、RTU (Ready-to-use) を標榜するには、長い時間が必要であろう。

6.2 権利表記

MDR XAFS DBの共通メタデータでは、保有者など、データの帰属や権利を明確にする設計思想があることは「5.1 メタデータ項目の整備によるデータ再利用」で述べ

た。現在のデータの利用ライセンスは、クリエイティブ・コモンズ (CC) の指針を取り入れ、CC BY-NC-SA（表示-非営利-継承）としている¹⁸⁾。つまり、データをダウンロードする場合は、データの出典を表示し、利用目的は非営利に限り、この権利が継承されることが、利用条件となっている。しかし、そもそもXAFSのスペクトル相互参照による益を思い出し、かつCCの指針を検証すると、以下のような本質的な部分で、CC BY-NC-SAの課題が見えてきた。

- データ提供機関のほぼすべてで有料のXAFS解析サービスを実施している

- 「営利」の定義が不明確で、人に依り解釈が異なる

これらに加えて、データ提供者のインセンティブを考えると、営利目的で使える機関を明確にすることが課題解決につながる可能性がある。すなわち、データ提供機関に限ってCC BY¹⁹⁾として、出典の表示のみで営利目的の利用を可能にする案が浮上している。これが実現すれば、企業も含めて多くの機関が、MDR XAFS DBプロジェクトに加わり、参画者はインセンティブを得、いずれは全ての人がデータを提供し、自由にデータを使えるようになる、データ共有の文化が育つ可能性がある。しかし、こうした特権的システムの導入は、データの管理を複雑にし、機関間の合議など、より組織的な運用が必須になる。MDR XAFS DBは、参画機関による共同運用の形態を取っており、組織力が乏しく、統制を伴う取り組みは明らかに困難を伴う。こうした状況を考えると、データ共有・再利用のボトムアップ的な取り組みは、スタート時には重要でインパクトがあると思うが、最終的にはコミュニティの中で作られる組織の下で、安定運用してゆくことが望ましいであろう。これをコミュニティの負担と考えるうちは有志による活動にならざるを得ないが、大多数の人が必要と感じるようになるかが、プロジェクト継続のステージゲートになる。現在、放射光学会ではデータ標準化委員会が立ち上がろうとしている。大きな方針の議論と、それを実装する取り組みが両輪となり、コミュニティ全体が動くことを期待している。

7. まとめ

国内の主なX線吸収分光法に関する機関が協力して構築している、MDR XAFS DBを広く俯瞰した。国際的

Table 3 Photon energy settings for the pre-edge peak of the Cu K-edge for several participating institutions.

施設名	Cu K 吸収端のエネルギー (基準値)	各機関の基準値の平均からの差
SPring-8	8980.23 eV	-0.453 eV
Aichi SR	8980.52 eV	-0.163 eV
SAGA-LS	8981.3 eV	+0.617 eV
平均値	8980.683 eV	

なメタデータの検討など、データ統合に向けた動きが高まる中で、メタデータの考え方をまとめるとともに、材料辞書を使ってすべてのスペクトルを等しく見つけられるようにする、横断検索に注力した取り組みを紹介した。今後のデータ共有と再利用がどのような仕組みで出来上がるにしても、材料語彙の辞書化は必要なマイルストーンであると考えている。一方で、現在もなおMDR XAFS DBは多くの課題を抱え動的に変化している。この活動は、今後の放射光データ全般をコミュニティとしてどのように扱ってゆかかを考える、試金石になると考えている。

謝辞

本取り組みは、MDR XAFS DBの参画機関の多くの方のお力の上に成り立っています。本プロジェクトの中心メンバーは以下の通りです（敬称略）：朝倉清高，木村正雄，田渕雅夫，稲田康宏，松本崇博，小林英一。皆様に感謝申し上げます。

本研究の一部は、国立情報学研究所公募型共同研究「競合分析を使った産学共有可能な放射光データ基盤の制度設計」の助成をいただきました。

参考文献

- 1) Brain M. Kincaid and P. Eisenberger: Phys. Rev. Lett. **34**, 1361 (1975).
- 2) J. J. Rehr and R. C. Albers: Reviews of Modern Physics **72**, 621 (2000).
- 3) B. Ravel, J. R. Hester, V. A. Solé and M. Newville: J. Synchrotron Rad. **19**, 869 (2012).
- 4) M. Ishii, K. Tanabe, A. Matsuda, H. Ofuchi, T. Matsumoto, T. Yaji, Y. Inada, H. Nitani, M. Kimura and K. Asakura: STAM Methods **3**, 2197518 (2023).
- 5) T. Hey, S. Tansley, K. Tolle and J. Gray: The Fourth Paradigm: Data-Intensive Scientific Discovery (Microsoft Research, 2009).
- 6) Farrel Lytle Database, http://ixs.iit.edu/database/data/Farrel_Lytle_data/
- 7) MDR XAFS DB, <https://doi.org/10.48505/nims.1447>
- 8) K. Asakura, H. Abe and M. Kimura: J. Synchrotron Rad. **25**, 967(2018).
- 9) 科学技術振興機構 先端研究施設共用促進事業 光ビームプ

ラットフォーム, <https://www.jst.go.jp/shincho/program/pfkeisei.html>

- 10) XAFS spectrum of Copper, <https://doi.org/10.48505/nims.3899>
- 11) W3C Semantic Web Standards, https://www.w3.org/2001/sw/wiki/Main_Page
- 12) B. Ravel and M. Newville: J. Physics: Conf Series. **712**, 012148 (2016).
- 13) XDI and NeXus formats for XAS data, <https://tinyurl.com/nxxas2023>
- 14) NeXus User Manual and Reference Documentation, 3.3.2.26. NXxas, <https://manual.nexusformat.org/classes/applications/NXxas.html>
- 15) 放射光学会 データ構造化諮問委員会, <https://jssrr.smoosy.atlas.jp/ja/datakouzouka>
- 16) XAFS 共通メタデータ, <https://github.com/xafs-db/xafs-schema>
- 17) FAIR Principles - GO FAIR, <https://www.go-fair.org/fair-principles/>
- 18) CC BY-NC-SA 4.0, 表示-非営利-継承 4.0 国際, コモンズ証, <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.ja>
- 19) CC BY 4.0 表示 4.0 国際, コモンズ証, <https://creativecommons.org/licenses/by/4.0/deed.ja>

著者紹介



石井真史

国立研究開発法人物質・材料研究機構 マテリアル基盤研究センター 主席研究員
E-mail: ISHII.Masashi@nims.go.jp
専門: オントロジー, データベース, データ科学

【略歴】

1995年3月大阪大学基礎工学研究科後期博士課程修了。博士(工学)。1995年4月理化学研究所入所, 1997年4月高輝度光科学研究センター入所, 2006年10月物質・材料研究機構入所, 同量子ビームセンター, 表界面構造・物性ユニット, 材料データプラットフォームセンターを経て, 2024年4月より現職。この間, University of Manchester Institute of Science and Technology, University of Marseille, The University of Manchesterなどで在外研究。

Construction and release of MDR XAFS DB, an integrated database of X-ray absorption spectroscopy

Masashi ISHII

National Institute for Materials Science
1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047 Japan

Abstract In research and development, data-intensive research methods (data-driven research), known as the “The Fourth Paradigm,” are becoming more prevalent. As international discussions on data sharing and reuse of X-ray absorption fine structure (XAFS) data become more active, it seems that the time has come for the Japanese synchrotron radiation community to set a policy. In this paper, we review the activities of the MDR XAFS DB project, in which domestic institutions involved in XAFS are collecting and releasing their data in one place, and introduce the constructed database from a practical viewpoint. In addition, regarding the unification of metadata, which is indispensable for data sharing and reuse, we will describe the concept of XDI, which is currently the mainstay of metadata, and show the position of common metadata in Japan. In particular, we will focus on “cross-search,” which is not found in other countries, and clarify our original strategy and specific activities to manage metadata values. In other words, the development of a materials lexicon, which is the basis of the MDR XAFS DB’s strength, i.e., the ability to search all data equally, and the purpose-oriented lexicon compilation policy are described. Furthermore, we summarize the issues of this database and discuss what is necessary for our community to share and reuse synchrotron radiation data in the future.