試料PID付与に関する動向

高輝度光科学研究センター (JASRI)
松本 崇博

第25回XAFS討論会 XAFSデータベース セッション 2022/08/02

Outline

- はじめに
- RDUF小委員会「研究資料・実験機器へのPID付与検討」のご紹介
- XAFS試料メタデータ付与に関する課題整理

はじめに

- 今後、データ流通・利活用を進めていく上で、試料メタは重要
  - 実験測定で用いた試料のさまざまな情報のトレース（品質保証)
    - 試料の提供メーカー, 化学式, 試料前処理方法など
  - 既存の他DBの情報とリンクさせて物理解析を柔軟に行いたい

- 現況
  - 試料情報は実験ノートには記述されるがデジタル化は今後の課題
  - コミュニティ依存, 系統的にはなかなか進めづらい

XAFS試料メタの課題整理や、外部活動 RDUF小委員会紹介
→ 今後の進め方について皆さまのご意見を伺いたい

RDUF(研究データ利活用協議会)とは

研究データ利活用協議会
RDUF
Research Data Utilization Forum

- 日本のボランティアベースでのオープンサイエンス活動
  - 参加は無料,　JaLC, JST, NIMSなどがサポート
  - 2016より活動,　参加:　8 institutions,　201 members　(2022.6現在)
  - https://japanlinkcenter.org/rduf/

- 研究データ利活用に関する課題について小委員会や部会を設置
  - 小委員会: 1年間の活動,　8 小委員会　(2019~)
  - 部会:　パーマネントな活動,　2 部会　(2021~)
- 成果物
  - ガイドライン作成や報告物 → 学術機関やSocial Communityに提言

- ほか, JOSSやWorkshopのイベント開催

RDUF小委員会「研究資料・実験機器へのPID付与検討」2022.04~

- 目的
  - 研究DXの推進には, 研究データに加え, 試料,史資料,機材など有体物に関する情報もサイバー空間で参照できるようにすることが不可欠
  - 研究資源に関するPIDとメタデータ管理・利活用に関する調査を実施
- 成果物
  - 研究資源管理の効率化に向けたデジタル技術・運用制度の課題整理
    - PID付番やその実践方法を提示する
  - 研究資源へのPID・メタデータ付与・管理の世界動向をまとめる
    - 世界のPID体系: ARK, ePIC, PIDINST など
    - 実験機器IDのメタデータに関する提言: https://doi.org/10.15497/RDA00070
- 現況
  - 参加者からの事例紹介を実施, 今後, 調査・報告書作成を進める

XAFS試料メタデータ付与に関する課題

- 現況
  - NIMS MDRに各機関からのデータが集まってきた
  - 特に試料メタデータに関してはさまざまな課題が把握できてきた

  → コミュニティ・NIMS・各施設が一体となった取り組みが必要
- 課題
  - データ利活用における客観的な指標づくり
  - メタデータの書き方の統一
  - ユーザーフレンドリーなメタデータ入力インターフェース
  - 辞書化, 外部DB(CASなど)との連携
- 取り組み
  - NIMS材料辞書 (MatVoc)

データ利活用における客観的な指標づくり

- 各機関でのXAFS試料メタデータ項目の比較

SPring-8

key	value個数
name	1844
chemical_formula	1771
supplier	1838
model_number	1822
lot_number	1716
tag	174
総数	9165
平均	5.216/work

北大

key	value個数
chemical_formula	75
name	75
CAS_number	68
manufacturer	31
Product_number	24
sample_lot_numb	16
additional_data	75
総数	364
平均	4.853/work

立命館大SR

key	value個数
chemical_formula	206
name	206
CAS_number	169
additional_metad	121
総数	702
平均	3.408/work

- 標準試料でのメタデータ項目例
  - サプライヤなどの製品情報 (SP8)
  - CAS-RNなど連携ID (北大,立命館)
  - 形状・混合物など付加情報

- 手入力でメタデータ生成(3.4~5.2/work) → 手間がかかる

- 現在は標準試料でわかりやすいケースがほとんど
  ユーザー持ち込み試料(北大, KEK)のケースではその表記に課題が多い

試料DB活用によるメタデータ付与の効率化

NEDO燃料電池データベースの例

NEDO 燃料電池等利用の飛躍的拡大に向けた共通課題解決型産学官連携研究開発事業
参考URL:　https://fc-www.jasri.jp/mission/

```
27  sample:
28  - name: [redacted]
29    id: [redacted]
30    structure:
31      form: powder
32    catalyst:
33      pretreatment_id: H
34      pretreatment: 水素ガス還元処理
35      form: '2. 合金 '
36      carrying_rate: '45'
37      composition: Pt 40.4at% Co 4.6at%
38      uniformity: 合金触媒　Ptスキン層
39      particle_size: '3.4'
40      particle_size_unit: nm
```

- 事前に試料にローカルIDを付与し，連係情報から以下メタデータを抽出している
  - 試料責任者名
  - 触媒情報
  - 試料前処理方法
  - など

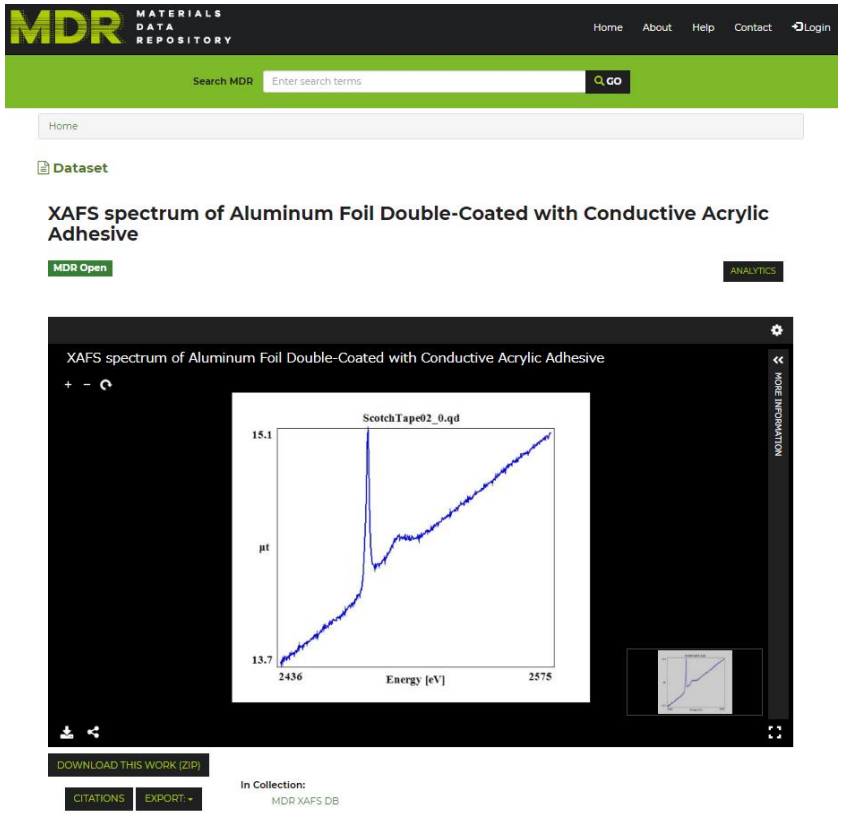
→ 試料DB構築を行い対応

- 今後，汎用&簡易利用に向けた取り組みが必要になりそう

標準試料以外のケース
KEK PF ユーザーデータの例

導電テープのXAFS測定

https://doi.org/10.48505/nims.3604



試料に関するメタデータ

Preview: primary.tsv

Show 10 entries Search:

key	value
@facility@ring@start_current_unit	mA
@facility@ring@end_current	449.947
@facility@ring@end_current_unit	mA
@sample_0@name	Aluminum Foil Double-Coated with Conductive Acrylic Adhesive
@sample_0@manufacturer	3M
@sample_0@product_number	AL-25DC
@measurement@method@name	XAFS
@measurement@method@detection	Fluorescence
@measurement@method@absorption_edge	S K
@measurement@date@start_time	2021-11-19 12:03:09

Showing 21 to 30 of 51 entries Previous 1 2 3 4 5 6 Next

このような複合体の場合
IDづけをどうしていくか?

メタデータの書き方の統一
検索性向上のための必須事項

オリジナルの試料名を
NIMS側でクレンジング → 大変な作業

例)

“LiFePO4 detune(80%)”→

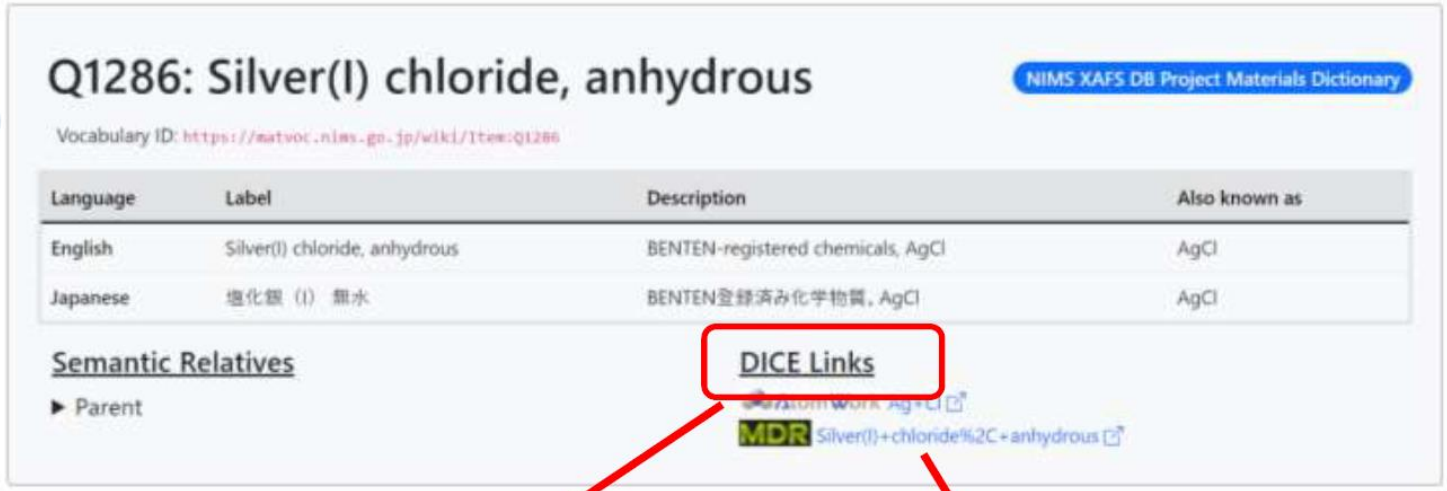
- sample:chemical_formula = “LiFePO4”
- sample:name = “Lithium iron(II) phosphate”
- instrument:additional_metadata = “detune(80%)”

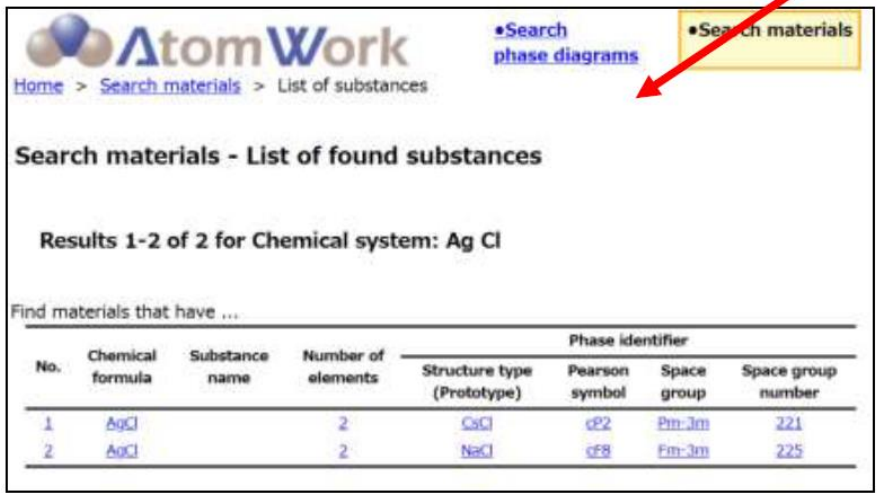
“SUS316L Ni K-edge 18.2 K” →

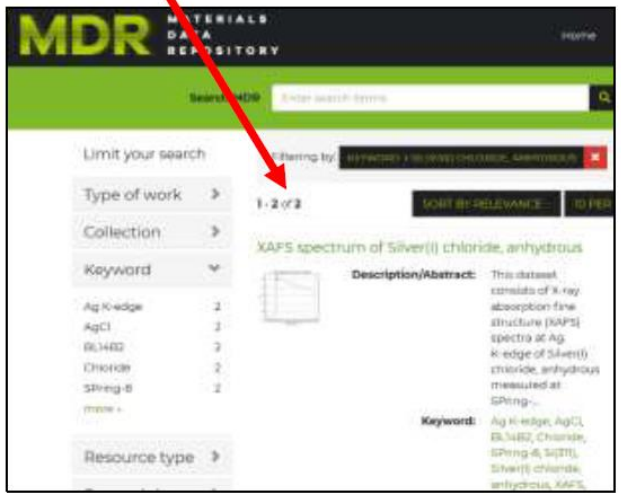
- sample:name = “SUS316L”
- measurement:method:absorption_edge = “Ni K-edge”
- measurement:atmosphere:sample:start_temperature = “18.2 K”

- 測定者しかわからない表記→ データ価値の低下
- 他人によるクレンジング後処理 → 限界, 不正確

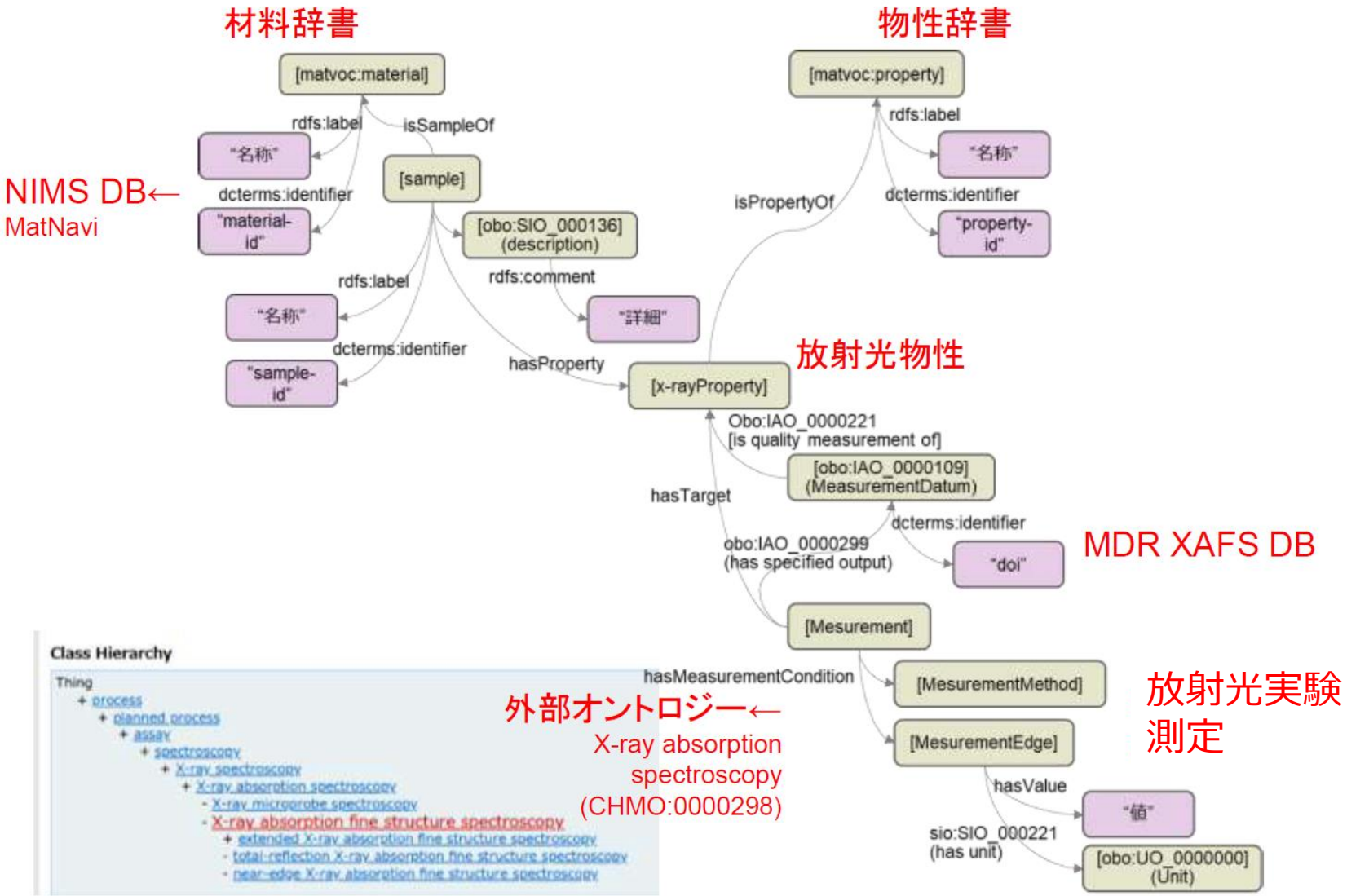
NIMS材料辞書（MatVoc）
材料データ連携の具体像







MDRの連携像


材料辞書
物性辞書
[matvoc:material]
[matvoc:property]
rdfs:label
isSampleOf
rdfs:label
"名称"
[sample]
"名称"
NIMS DB←
MatNavi
dcterms:identifier
isPropertyOf
dcterms:identifier
"material-id"
[obo:SIO_000136] (description)
"property-id"
rdfs:label
rdfs:comment
"名称"
"詳細"
dcterms:identifier
hasProperty
放射光物性
"sample-id"
[x-rayProperty]
Obo:IAO_0000221
[is quality measurement of]
[obo:IAO_0000109] (MeasurementDatum)
hasTarget
dcterms:identifier
obo:IAO_0000299
(has specified output)
"doi"
MDR XAFS DB
[Mesurement]
Class Hierarchy
hasMeasurementCondition
[MesurementMethod]
放射光実験
測定
Thing
process
planned process
assay
spectroscopy
X-ray spectroscopy
X-ray absorption spectroscopy
X-ray microprobe spectroscopy
X-ray absorption fine structure spectroscopy
extended X-ray absorption fine structure spectroscopy
total-reflection X-ray absorption fine structure spectroscopy
near-edge X-ray absorption fine structure spectroscopy
外部オントロジー←
X-ray absorption spectroscopy (CHMO:0000298)
[MesurementEdge]
hasValue
"値"
sio:SIO_000221
(has unit)
[obo:UO_0000000] (Unit)

まとめ

- NIMS MDRによるXAFSデータベース多機関連係が本格スタート
- データ流通・利活用の推進 → 試料メタデータ課題への対応要
  - どのパラメータを指標として記録していけばよいのか
  - メタデータの書き方の統一
  - メタデータ付与の手間軽減 → 試料DBの利活用
  - ID付番や他DB連係の実践をどう進めていくか？
- 今後の方向性
  - TIAや RDUF小委員会の活動で提言や課題解決を進めていくのがよいか
  - コミュニティ・各施設の意見も取り入れながら活動を進めていきたい