DataCiteを選んだ理由

Materials Data RepositoryにおけるDataCite DOIの活用

田辺浩介 TANABE.Kosuke@nims.go.jp

物質・材料研究機構

技術開発・共用部門 材料データプラットフォーム データ基盤ユニット 主幹エンジニア

https://orcid.org/0000-0002-9986-7223

2023年6月22日 JAPAN OPEN SCIENCE SUMMIT 2023 セッションD3

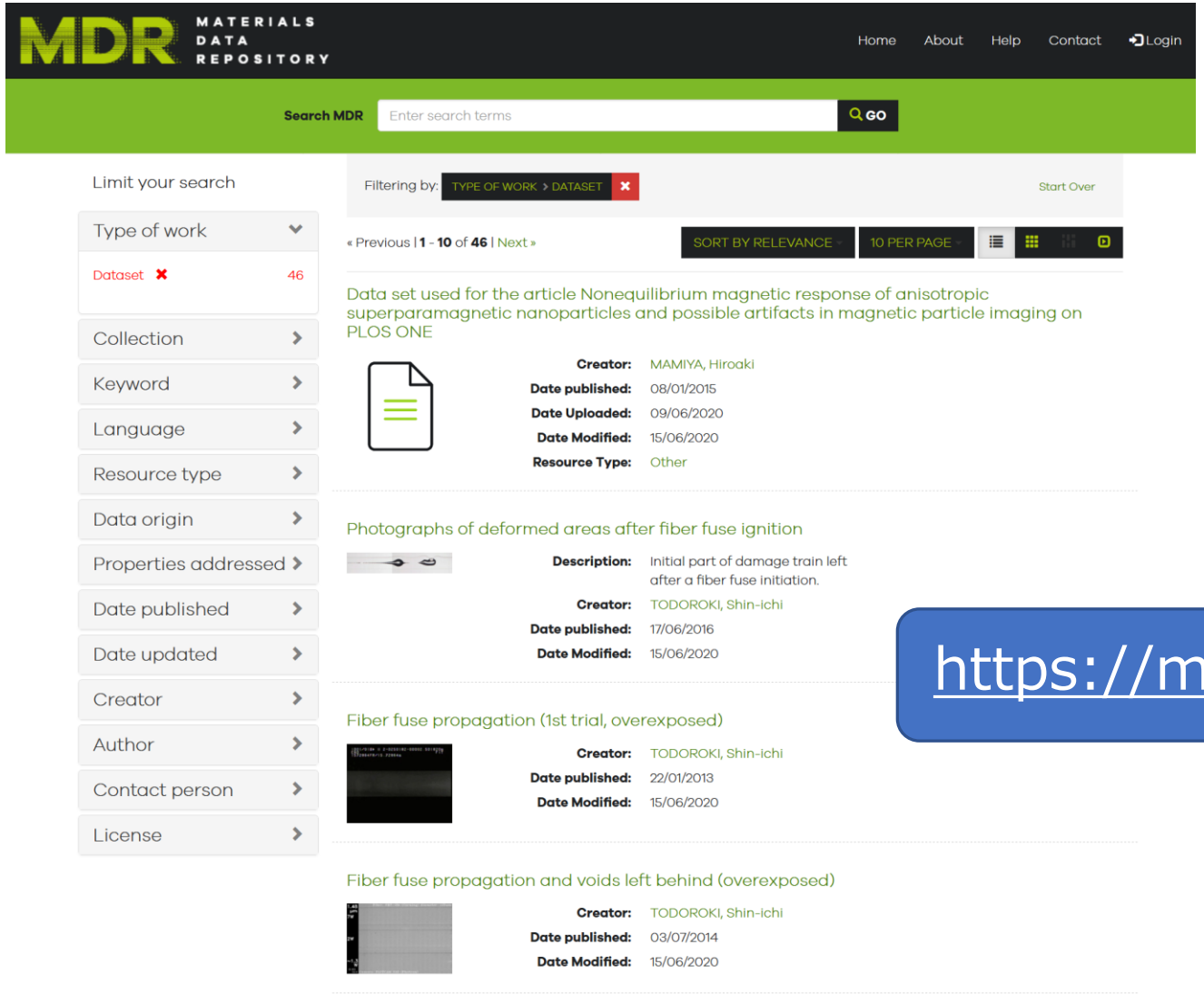
CC BY

1. Materials Data Repository (MDR)とは

Materials Data Repository (MDR)

- **物質・材料研究機構(NIMS)が開発・運用するデータリポジトリ**
  - **https://mdr.nims.go.jp**
- NIMSの研究成果である論文ファイルや、材料科学に関するデータ（実験・計測データ）を収録
- 2020年6月15日より運用開始
- 2023年6月20時点で、**約1500件のPublication（出版物）**と**約12,700件のDataset（研究データ）**を収録
- NIMSの研究者総覧データベース「SAMURAI」とも連携して動作
  - https://samurai.nims.go.jp

Materials Data Repository (MDR)



https://mdr.nims.go.jp

MDRに収録するデータ

- **大きく分けてDatasetとPublicationの2種類を扱う**
- Dataset: 材料データ
  - 実験で使用した数値や画像のデータなどを想定
  - **材料関係のメタデータを記述可能**
- Publication: 論文のデータ
  - 論文ファイル（の著者版）を登録することを想定
  - 一般的な機関リポジトリと同様の運用

BROWSE ALL PUBLICATIONS

BROWSE ALL DATASETS

MDR XAFS DB

User Collection MDR Open

MDR XAFS DB

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

MDR XAFS DB
国内6組織によるX線吸収微細構造の計測データ集
https://doi.org/10.48505/nims.1447

This collection of data has been compiled at the National Institute fo
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

Total items	2264
Size	unknown
Resource type	Dataset
Creator	ISHII, Masashi
Contributor	MATSUDA, Asahiko TANABE, Kosuke NAGAO, Hiroko YOSHIKAWA, Hideki
Keyword	spectroscopy x-ray absorption local structure
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	https://user.spring8.or.jp/sp8info/?p=2751 https://support.spring8.or.jp/BL/bl14b2/xafs

Dataset

MDR XAFS DB Readme

MDR Open | Deposited

ANALYTICS

DOWNLOAD THIS WORK (ZIP)

CITATIONS | EXPORT:

In Collection:

MDR XAFS DB

This MDR XAFS DB is an integrated database of XAFS spectral data provided by the participating institutions in this project. The metadata ("local metadata") and data formats used by each institution have many points in common, but there are also differences. This readme summarizes the status of data provided by each institution (2. Statistics), basic

https://mdr.nims.go.jp/concern/datasets/vh53wz94c?locale=en

Instruments

Instrument

Title	BL12_XAFS
Description	SAGA-LS Beamline XAFS setup
Instrument function	
Description	x-ray absorption spectroscopy
Manufacturer	
Organization	SAGA-LS

実験・計測データのメタデータ
装置・試料の情報が記録されている

Specimen details

Specimen type

Title	Nickel(II) oxide
Description	on carbon tape

Chemical composition

Description	NiO
Category	http://matvoc.nims.go.jp/entity/Q1631

Structural feature

Description	local structure
Category	http://matvoc.nims.go.jp/entity/Q686

https://doi.org/10.48505/nims.4192

Preview: primary.tsv

Show 10 entries

Search:

key	value
@data_info@data_depositor@organization	Kyushu Synchrotron Light Research Center
@data_info@data_depositor@affiliation	Beamline Group
@data_info@data_depositor@name	Kyushu Synchrotron Light Research Center
@data_info@data_depositor@role	Beamline staff
@data_info@contact_name@organization	Kyushu Synchrotron Light Research Center
@data_info@contact_name@affiliation	Beamline Group
@data_info@contact_name@name	Kyushu Synchrotron Light Research Center
@data_info@contact_name@role	Beamline staff
@data_info@access_rights	open
@data_info@license	CC-BY-NC-SA

Showing 1 to 10 of 47 entries

Previous 1 2 3 4 5 Next

This dataset consists of X-ray absorption fine structure (XAFS) spectra at Ni K-edge of HAVAR mea
BL14B2, and is a part of XAFS database (MDR XAFS DB, https://doi.org/10.48505/nims.1447) as a c

Description

Alternative title

spring8.2f057cc2-8d9d-4ebd-83dc-e0d054e7cecb

DOI

https://doi.org/10.48505/nims.2930

MDRで付与したDataCite DOI

Creator

Name	Industrial Application and Partnership Division
Role	contact person

Keyword

Alloy
BL14B2
HAVAR

Items

Thumbnail	Title	Date Uploaded	Size	Visibility	Actions
	_最終稿版_機械学習による光電子収量分光PYSスペクトルの自動閾値予測20 0629.pdf	30/06/2020	1.04 MB	MDR Open	SELECT AN ACTION
	Readme.md	30/06/2020	3.36 KB	MDR Open	SELECT AN ACTION
	requirements.txt	03/07/2020	636 Bytes	MDR Open	SELECT AN ACTION
	pys_mdr.zip	03/07/2020	82.8 MB	MDR Open	SELECT AN ACTION

論文の著者最終稿ファイル

論文の付随データファイル

https://doi.org/10.48505/nims.1440

MDRでのDataCite DOIの付与

- MDRに登録する論文・データに対して、**DataCite DOI**を付与
- 2023年6月20日時点で、**2872件のDataCite DOI**を付与している
  - 日本の研究機関別付与数では、JAMSTEC（3659件）の次に多い
  - JAXA1059件、DIAS153件、極地研147件、九大115件、環境研64件、京大4件
- DOIの付与対象は研究データ全般と論文著者版
  - Gold OAになっている論文も収録しているが、MDRではDOIは付与しない
  - その他、すでに他の場所でDOIが付与されているデータには付与しない
- NIMSはJaLC経由ではなく、DataCiteの会員になってDataCite DOIを付与している
  - **JaLC経由でDataCite DOIを付与する場合、本来DataCiteがサポートしているメタデータ項目や機能が使用できないことがある（後述）**

MDRでのDataCite DOIの文献種別ごとの付与状況

resourceType（文献種別）	付与数
Dataset	2677
Journal Article	141
Text	16
Other	14
Collection	10
Software	5
Preprint	4
Conference Paper	2
Book Chapter	1
Book	1
Conference Proceeding	1
Report	1

Materials data platform DICE

2. なぜDataCite DOIを付与しているのか

MDRでDataCite DOIを付与している理由

- 研究データのDOIとして広く用いられているから
- **外部データベースへの収録に優位性があるから**
  - OpenAIRE Explore, Data Citation Index, Mendeley Data, Dimensions, Ex Libris Central Discovery Index
    https://support.datacite.org/docs/where-can-i-search-doi-metadata
- 豊富で使いやすいAPIを備えているから
  - NIMSはJAIRO Cloudを使用していないため、DOIの付与の方法を別途用意する必要があった
- **研究データだけでなく、文献に対するDOIも付与できるから**
  - 論文（著者版）にもDOIを付与したいという要望があった
  - figshareやZenodo, arXivで付与されているDOIはDataCite DOI

DataCiteのresourceType（資料種別、2023年6月時点）

- Audiovisual
- Book
- BookChapter
- Collection
- ComputationalNotebook
- ConferencePaper
- ConferenceProceeding
- DataPaper
- Dataset
- Dissertation
- Event
- Image
- InteractiveResource
- Journal
- JournalArticle
- Model
- OutputManagementPlan
- PeerReview
- PhysicalObject
- Preprint
- Report
- Service
- Software
- Sound
- Standard
- Text
- Workflow
- Other
- **（ただし現在のJaLCでは、文献に対するDataCite DOIの付与をサポートしていない）**

論文や書籍に加えて、DMPやプレプリント、ソースコードや査読歴など、新しい種別の研究情報にも対応している

DataCite API

- REST・OAI-PMH・GraphQLの各インターフェースが用意されており、いろいろなアプリケーションににかんたんに組み込める
- JaLC APIと異なり、JSONでDOI登録用のメタデータが書ける
  - 以下のようなJSONをPOSTで送信するだけで、DOIが付与できる
  - 実際にはリゾルブ先のURLやタイトル・著者名などが必要

```
{
  "data": {
    "type": "dois",
    "attributes": {
      "doi": "10.5438/0012"
    }
  }
}
```

https://support.datacite.org/docs/api-create-dois

- **（ただし現在のJaLCでは、DataCite APIによるDOIの付与をサポートしていない）**

DOI付与のためのメタデータ

- DataCite DOIが研究データ用のメタデータだからといって、**個別の研究分野の詳細なメタデータが記述できるわけではない**
  - 地理情報（緯度・経度）や「データ収集日」といった項目はある
  - むしろ文献以外にもDOIが付与できるよう、メタデータの記述が簡略化されているとも言える
  - MDRでは、材料科学のメタデータは別途CSV・JSON・YAMLファイルで記述

- DataCite Schema 4.4で、文献のメタデータの項目が大幅に増加した
  - DataCite Schema 4.3までは、文献のメタデータを記述するための項目が少なく、ジャーナル名など論文のメタデータを記述するのに苦労した

DataCite DOI付与用のメタデータ（抜粋）


```
{
  "data": {
    "type": "dois",
    "attributes": {
      "event": "publish",
      "url": "https://mdr.nims.go.jp/pid/c7c8a33f-e754-4b87-8c4b-49b463e72b54",
      "doi": "10.48505/nims.4192",
      "types": {
        "resourceTypeGeneral": "Dataset"
      },
      "titles": [
        {
          "title": "XAFS spectrum of Nickel(II) oxide"
        }
      ],
      "creators": [
        {
          "name": "SAGA-LS",
          "nameIdentifiers": [

          ],
```


タイトルやデータ作成者名など。
研究データそのもののメタデータはほとんど含まれない

MDRでの材料メタデータの記述例

MDR独自のメタデータ。**現在のDataCiteでは、これらの研究分野個別のメタデータは扱えないので、** CSVやJSON、YAMLなどで別途メタデータを記述する必要がある

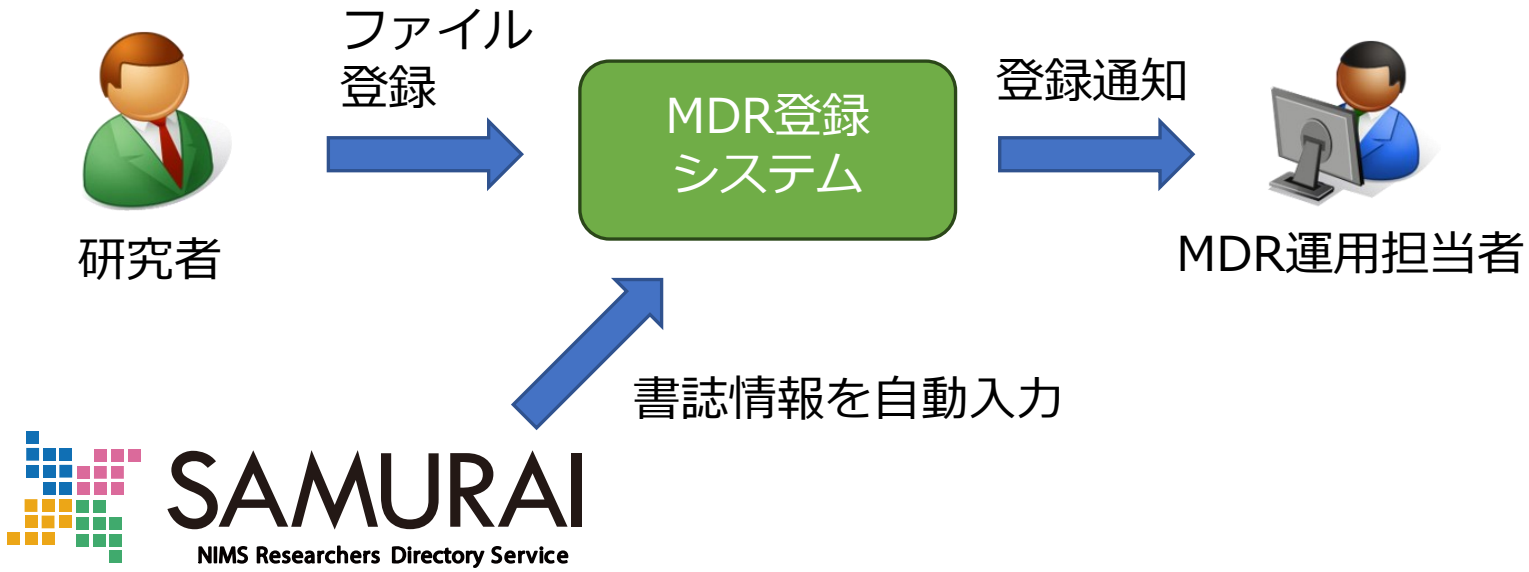
```yaml
# 試料（複数記述可）
specimens:
- name: A002 #試料の名前
  description: NiO polycrystal film サンプル8 #試料の説明
  identifier: DCStagid-47910155 #試料のID

chemical_compositions:
- identifier: 'PubChem: 23954' #試料の化学組成に関するID
  description: TiCoooo #試料の化学組成の説明・記述

crystallographic_structures:
- category_vocabulary: https://matvoc.nims.go.jp/wiki/Item:Q556 #試料の結晶構造に関するID
  description: （試料の結晶構造の説明・記述）
```

Materials data platform DICE

3. MDRでのDataCite DOIのワークフロー

1. 研究データファイルの登録

- 研究者は「**MDR登録システム**」（データ登録用の内部アプリケーション）にログインし、MDRに登録するファイルをアップロードする
- 関連URLに論文DOIを入力すると、著者名やタイトルが自動入力される
  - NIMSの研究者総覧「SAMURAI」などの所内データベースから取得
- 研究者がデータを登録すると、MDR担当者に通知が届く



MDR登録システム

Resource type
資料種別

Required / 入力必須

資料種別: poster

論文ファイルを登録する場合は"article"や"proceeding"、口頭発表のスライドは"presentation"や"poster"、論文の付随データファイルは"dataset"を選択してください。

Creator
著者

Required / 入力必須

氏名: 松田 朝彦

論文を登録する場合、論文での著者表記とあわせてください。

ORCID: 0000-0001-5989-027X

0000-1111-2222-3333 の形式で入力してください。

所属組織: 物質・材料研究機構

著者の所属する組織の名称を入力してください。論文を登録する場合、論文での著者表記とあわせてください。

所属部署: 統合型材料開発・情報基盤部門/材料データプラットフォームセンター/データシス

著者の所属する部署の名称を入力してください。論文を登録する場合、論文での著者表記とあわせてください。

ROR: 026v1ze26

役割: author

著者順序: 1

行を削除

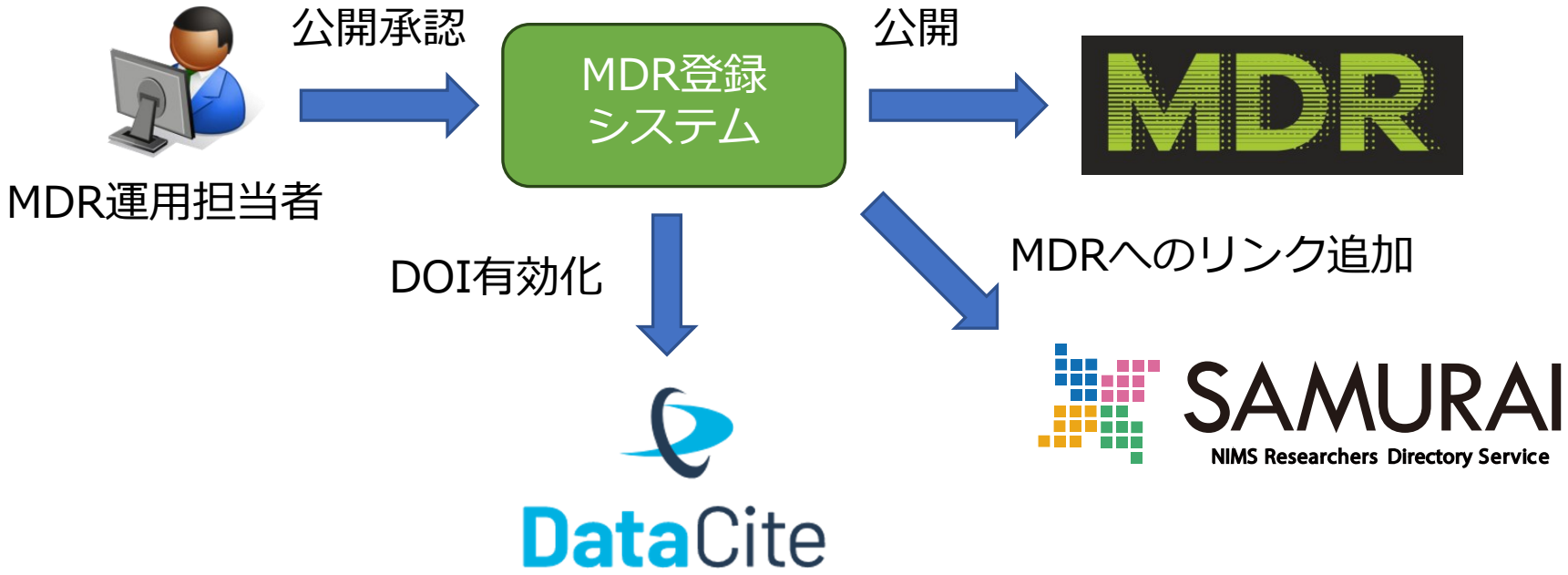
タイトルや著者名は、NIMS内のデータベースから自動的に取得して入力される

2. 研究データの公開準備作業とDataCite DOIの付与

- MDR担当者は登録された書誌事項とファイルをチェックし、公開可否、ならびにDataCite DOIの登録可否を判断する
- DataCite DOIの登録が行える場合、MDR登録画面からDataCite DOIを予約する
  - MDR登録システムがDataCite APIを呼び出す
  - 論文付随データの場合、論文DOIとデータDOIの関連も自動的に作成される
  - 必要に応じて、予約したDOIを研究者に連絡し、執筆中の論文にデータのDOIを記入できるようにする

3. 研究データの公開

- MDR担当者は公開を承認し、DataCite DOIを有効化する
- 公開が承認されたデータは、自動的にMDRに送付されて公開される。
  あわせて、SAMURAIにもMDRへのリンクが自動的に追加される



NIMSの研究者総覧SAMURAIとのリンク

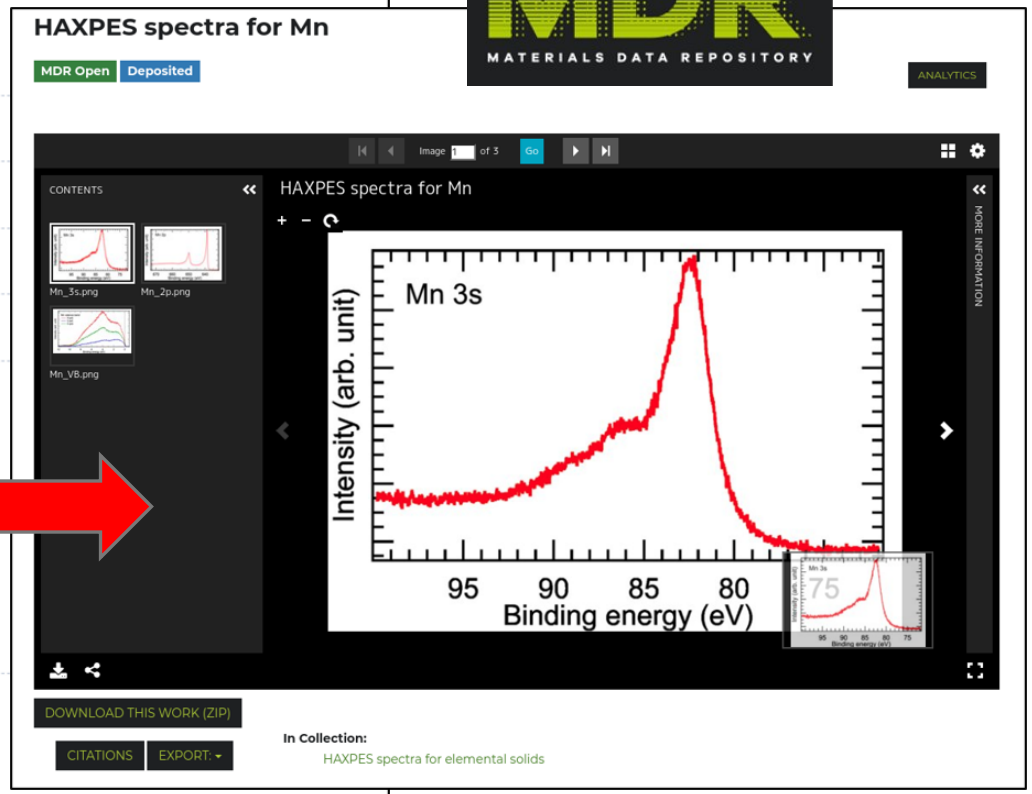
SAMURAIの論文詳細ページに、MDRに登録した論文付随データのリンクが追加される

SAMURAI
NIMS Researchers Directory Service

Polarization Dependent Bulk-sensitive Valence Band Photoemission Spectroscopy and Density Functional Theory Calculations: Part I. 3d Transition Metals

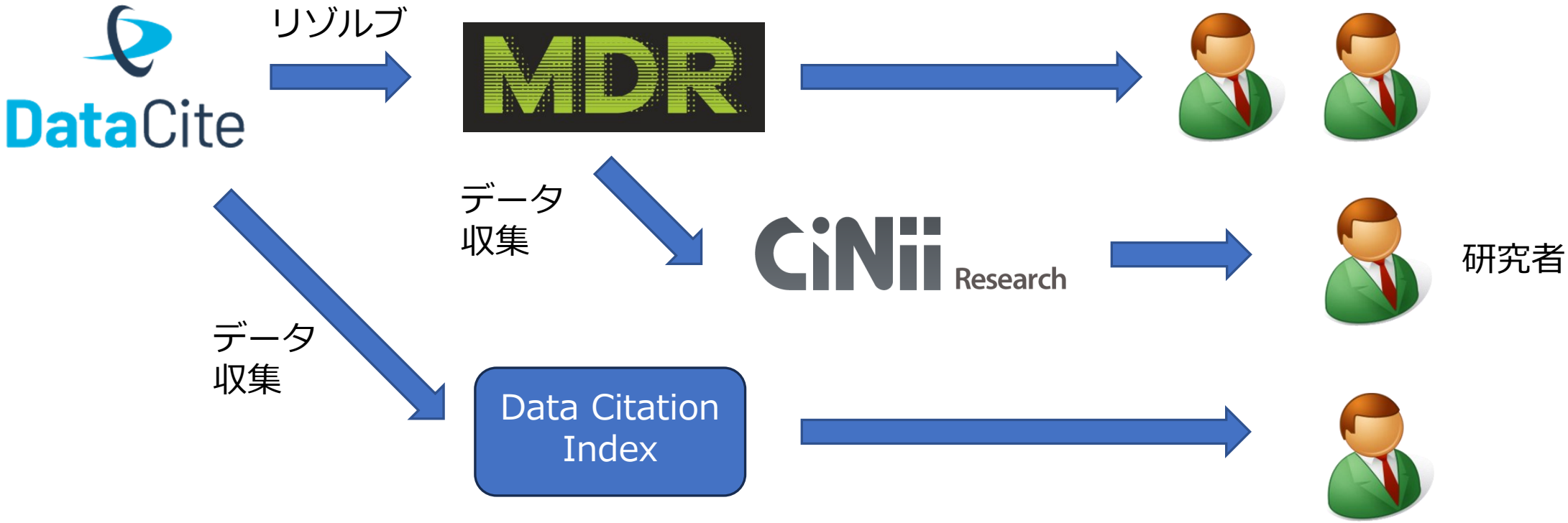
著者	Shigenori Ueda, Ikutaro Hamada.
掲載誌名	Journal of the Physical Society of Japan 86 [12] 124706 ISSN: 13474073, 00319015 ESIでのカテゴリ: PHYSICS
出版社	Physical Society of Japan
発表年	2017
言語	English
DOI	https://doi.org/10.7566/jpsj.86.124706
本文ファイル・データセット	MDR available HAXPES spectra for Fe MDR available HAXPES spectra for Co MDR available HAXPES spectra for Zn MDR available HAXPES spectra for Ti MDR available HAXPES spectra for Ni MDR available HAXPES spectra for Cr MDR available HAXPES spectra for V MDR available HAXPES spectra for Mn MDR available HAXPES spectra for Cu
この文献をMendeleyにインポート	MENDELEY

MDR
MATERIALS DATA REPOSITORY

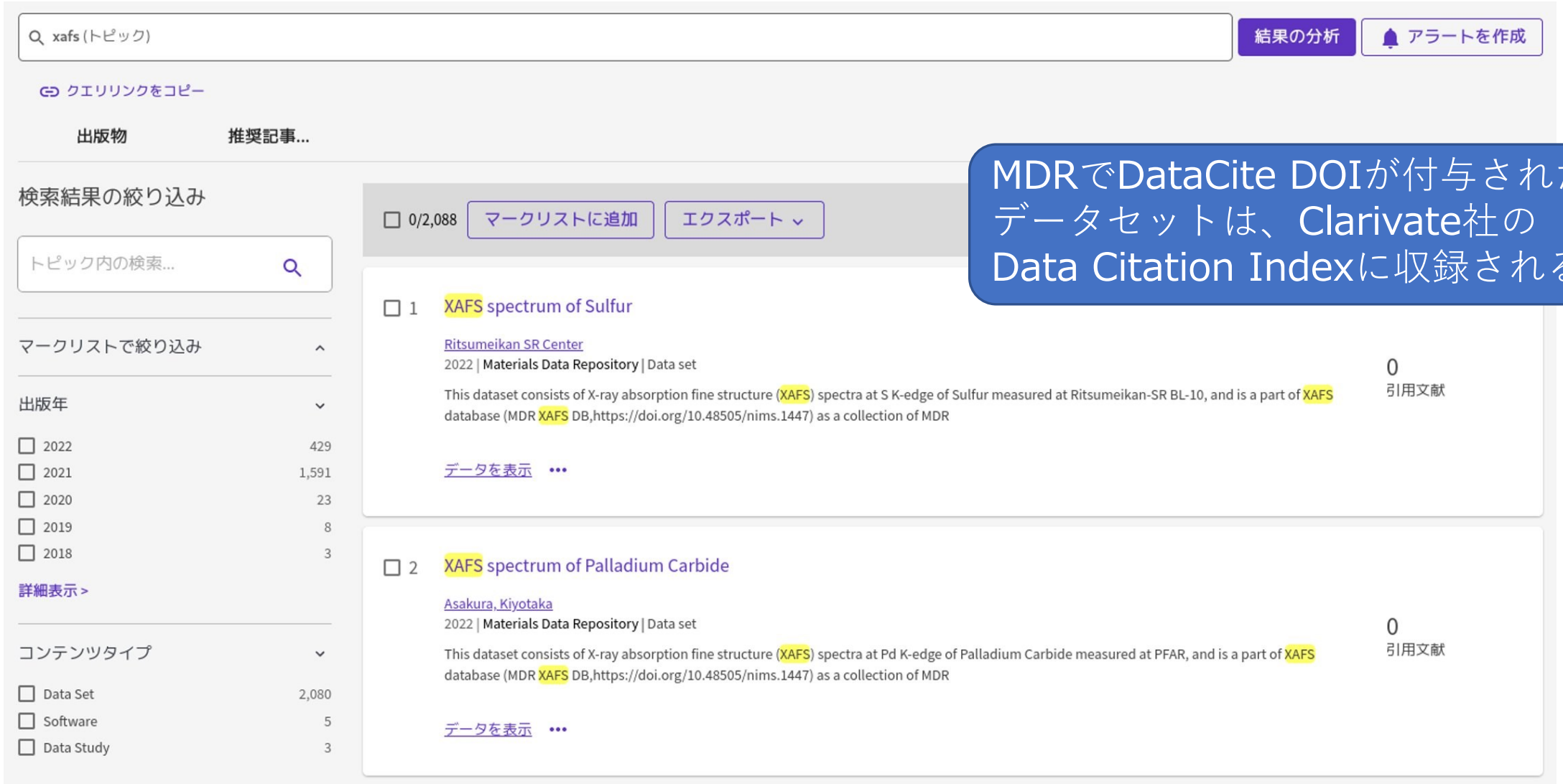


4. 研究データの流通

- DataCite DOIはMDRのURLにリゾルブされる
- DataCiteに収録された研究データのメタデータは、DataCite APIを通して、DataCite自身の検索サービスや外部検索データベースに収録され、研究者が検索できるようになる



外部データベースへのデータ提供



MDRでDataCite DOIが付与されたデータセットは、Clarivate社のData Citation Indexに収録される

4. MDRでのDataCite DOIの今後の活用予定

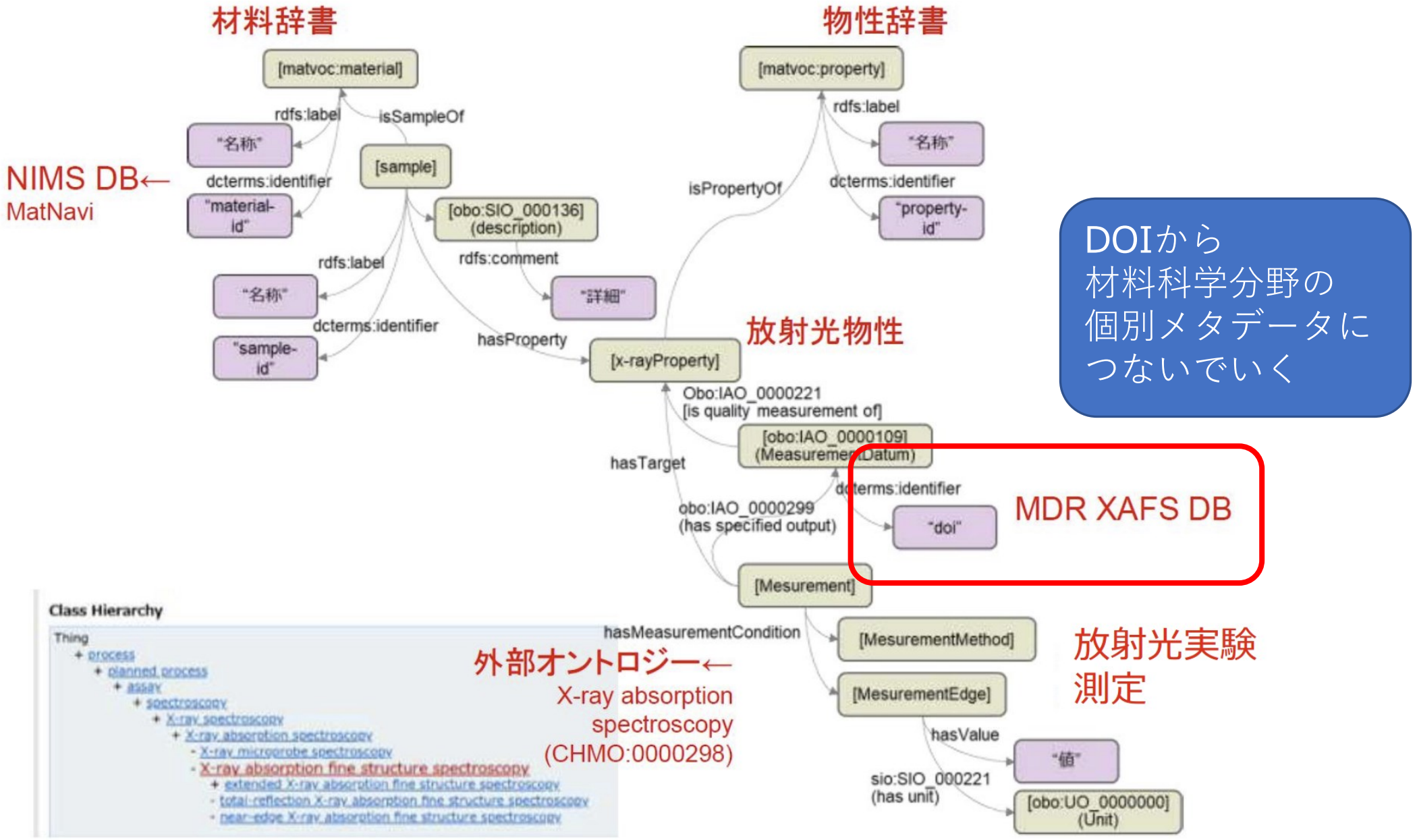
DOIの付与対象の拡大

- 実験装置・試料へのDOI付与
  - 研究データの検索や追跡のために、「どの実験でどの試料を用いたか」「どの実験でどの装置を用いたか」といった情報を管理する必要が生まれてきている
  - **現在策定中のDataCite Schema 4.5で、resourceTypeにInstrumentが追加**
  - 実験装置のPID付与については、RDAのPersistent Identification of Instruments WGで議論
  - RDUFの「研究資料・実験機器へのPID付与検討小委員会」で、ホワイトペーパーの和訳を作成
    https://pindinst-white-paper.readthedocs.io/ja/latest/
  - DataCiteでは試料に対するIDとして、**IGSN**(International Generic Sample Number)をサポートしている

研究データの利用統計

- データ共有・公開を促すためにも、MDRで提供している**データの利用状況**を研究者に示すことは極めて重要
- DataCiteは**COUNTER準拠のデータの利用統計**を提供しており、MDRでもこれを採用したい
- 以前は集計用のログを自前で収集しなければならなかったが、2023年にDataCiteがJavaScriptによるトラッカーのβ版をリリース
  - Google Analyticsのように、HTMLにタグを埋め込むだけで利用統計が取れる
  - https://support.datacite.org/docs/datacite-usage-tracker
- ダウンロード数や引用数の表示も、HTMLにタグを埋め込むことで可能
  - 以下が表示例

https://support.datacite.org/docs/displaying-usage-and-citations-in-your-repository



松本崇博. 試料PID付与に関する動向. 第25回XAFS討論会 XAFSデータベース セッション. 2022
https://www.jxafs.org/w/wp-content/uploads/2022/08/2022_Aug_3_Sample_PID_Matsumoto.pdf

まとめ

- **PIDは、複雑なデータの結びつきを解きほぐすための鍵**
- 研究活動も、それに付随するデータ管理も、ほんとうに多様化している
- 論文だけでなく、研究データや実験装置など、さまざまなものにPIDが付与されるようになっている

- DOI、ならびにDOIと連携して動作する他のPIDを通して、複雑で多様化する研究活動が、研究者にもデータ管理者にももっとFAIRなものになるはず

Materials data platform
DICE

TANABE.Kosuke@nims.go.jp

NIMS