

研究データリポジトリの事例紹介 「Materials Data Repository」

田辺 浩介

物質・材料研究機構 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

TANABE.Kosuke@nims.go.jp <https://www.nims.go.jp>

2020年11月2日 JPCOAR Monday

1. はじめに



データリポジトリってなんだろう



- 論文だけではなく、実験データや計測データなど、研究に関わるさまざまなデータを保存・公開するためのリポジトリ
 - 特定のソフトウェアやサービスと結びついているわけではない
- 世界のデータリポジトリの一覧はre3dataなどを参照
 - <https://www.re3data.org/>
 - 研究分野ごとに多種多様なリポジトリがある
 - 特定の組織のみからデータ登録を受け付けるものも、誰でもデータを登録できるものもある



Materials Data Repository (MDR)



MDR MATERIALS DATA REPOSITORY

Home About Help Contact Login

Search MDR **GO**

Limit your search

Type of work ▾

Dataset ✕ 46

Collection >

Keyword >

Language >

Resource type >

Data origin >

Properties addressed >

Date published >

Date updated >

Creator >

Author >

Contact person >

License >

Filtering by: TYPE OF WORK > DATASET ✕

Start Over

« Previous | 1 - 10 of 46 | Next »

SORT BY RELEVANCE

10 PER PAGE



Data set used for the article Nonequilibrium magnetic response of anisotropic superparamagnetic nanoparticles and possible artifacts in magnetic particle imaging on PLOS ONE



Creator: MAMIYA, Hiroaki

Date published: 08/01/2015

Date Uploaded: 09/06/2020

Date Modified: 15/06/2020

Resource Type: Other

Photographs of deformed areas after fiber fuse ignition



Description: Initial part of damage train left after a fiber fuse initiation.

Creator: TODOROKI, Shin-ichi

Date published: 17/06/2013

Date Modified: 15/06/2020

Fiber fuse propagation (1st trial, overexposed)



Creator: TODOROKI, Shin-ichi

Date published: 22/01/2013

Date Modified: 15/06/2020

Fiber fuse propagation and voids left behind (overexposed)



Creator: TODOROKI, Shin-ichi

Date published: 03/07/2014

Date Modified: 15/06/2020

<https://mdr.nims.go.jp>

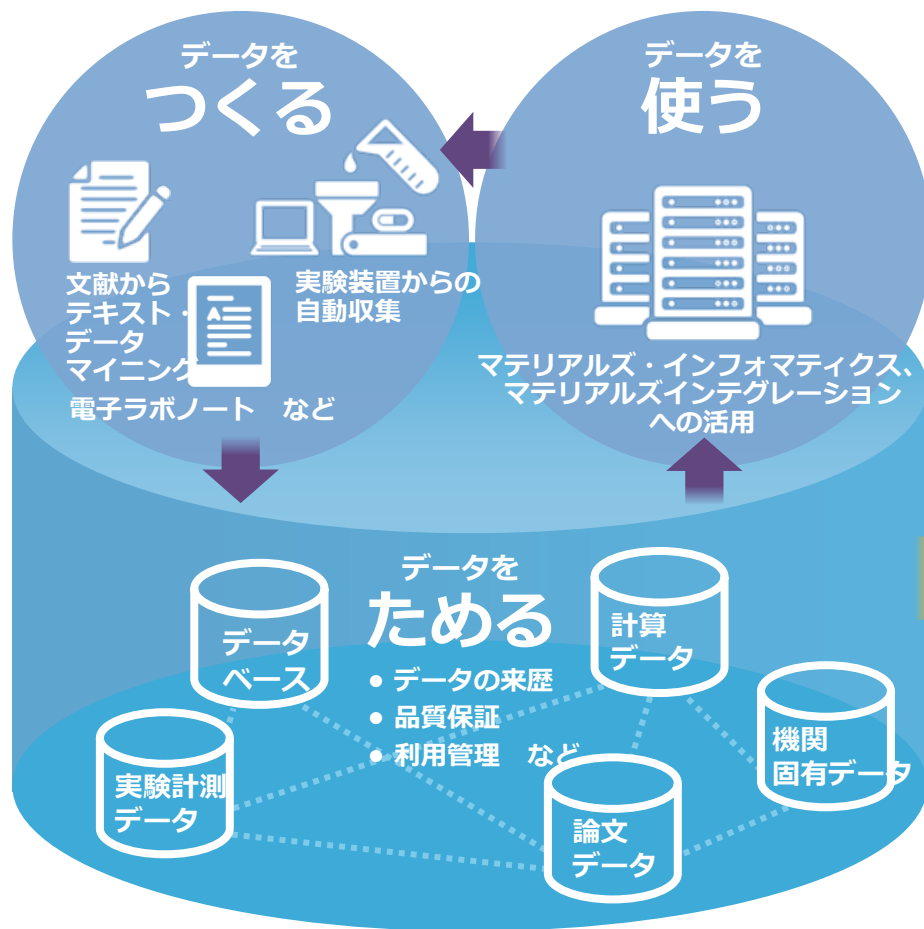


- 物質・材料研究機構(NIMS)が開発・運用するデータリポジトリ
 - 2020年6月15日より運用開始
- 機構で生成された、材料科学に関するデータ（実験・計測データ、論文データなど）を収録
 - 2020年10月末時点で、**約1200件の論文データと約50件のデータセット**を収録
- NIMSが構築中の材料データプラットフォーム「DICE」を構成するアプリケーションのひとつ
 - 情報学を応用した材料の研究開発の手法「**マテリアルズ・インフォマティクス**」を実現するためのプラットフォーム



<https://dice.nims.go.jp>

材料データの収集から
公開までを一貫して
行うシステム



データを公開する

- データリポジトリ
- データ出版
- データベース
- 他機関リポジトリとのリンクなど外部データプラットフォームとの相互運用



DICEを構成するアプリケーション群



データ収集

IoT

IoT file transfer system

M-IDac
Materials Data Conversion Tools

データ分析

FigResourceMiner
powered by TDM

TDM

mi
MATERIALS
INTEGRATION

データ公開

MatNavi

MDR MATERIALS
DATA
REPOSITORY

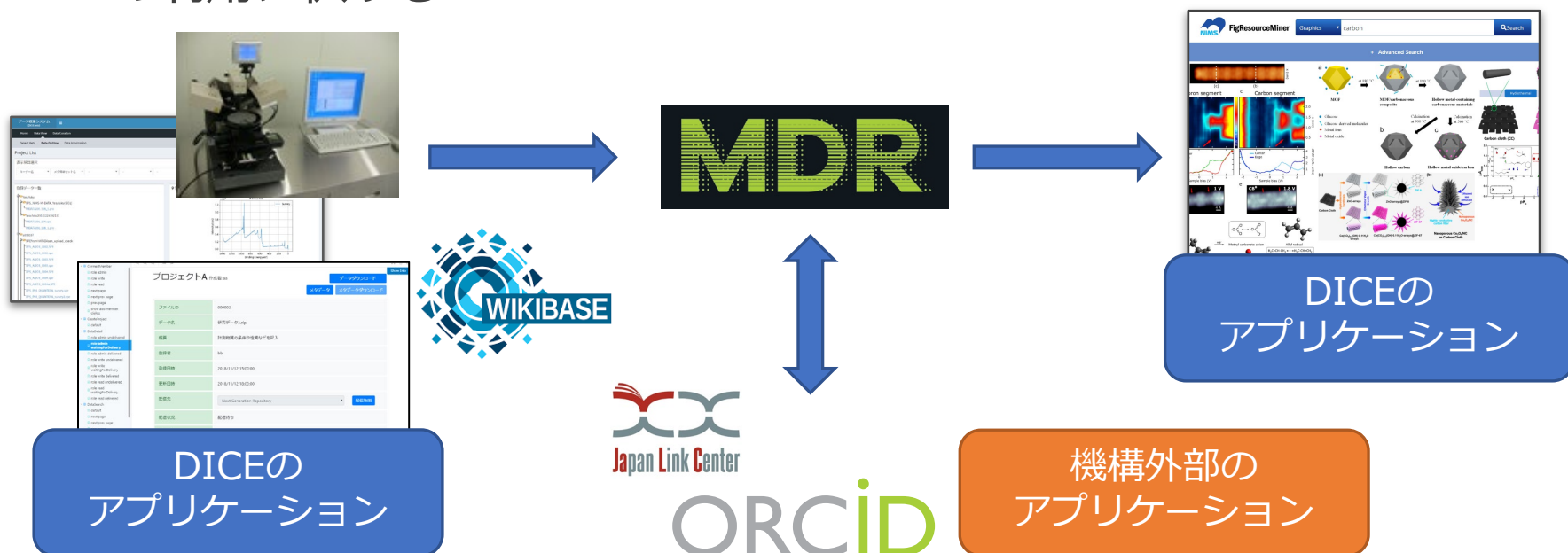
MDRは「DICE」のアプリケーションのひとつ



DICEの中でのMDRの位置づけ



- NIMS内部サービスと外部の研究機関・サービスの結節点
- DICEの各アプリケーションで作成されたデータは、最終的にMDRで公開される
 - ただし、MDR単体でもデータ登録を行うことができる
- MDRで公開されるデータは、グローバルな流通のための識別子（DOI, ORCIDなど）を付与して、研究者や他の検索サービスへの利用に供する



2. MDRの構成



- 大きく分けてDatasetとPublicationの2種類を扱う
- Dataset: 材料データ
 - 実験で使っ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



- **DICE全体のメタデータスキーマ**として、NIMS独自のメタデータスキーマである「**共通メッセージ形式**」を定義
- 材料メタデータを「**共通メタデータ**」と「**分野別メタデータ**」に分けて記述
- MDRではこのスキーマのサブセットを実装

共通メタデータ（必須） 識別子・記入者・物質・装置・データ起源…				
計測 メタデータ 計測法 測定環境	試料 メタデータ 物質タイプ 構造的特徴	特性 メタデータ 特徴的性質	合成・プロセス メタデータ 処理日 処理温度	計算 メタデータ 計算機 ソフトウェア
計測 主要パラメータ	試料 主要パラメータ	特性 主要パラメータ	合成・プロセス 主要パラメータ	計算 主要パラメータ
任意データ	任意データ	任意データ	任意データ	任意データ



DICE共通メッセージ形式スキーマ

DICEの各システム間で研究データを流通・交換・保管する際、共通のメタデータスキーマを定め、す。このメタデータはXML Schemaで定義の上、JSON Schema形式に変換して利用されます。記述その研究データの特徴を記述するための項目群も含んで構成され、複雑な構造を内包します。

JSON形式

2020-07-22 depositUploadReq_20191105.json

XML形式

2020-07-22 depositUploadReq_20191105.xsd

材料メタデータ
スキーマの標準と
なることを目指す

応用例

材料データリポジトリ (Materials Data Repository, MDR) では、DICE共通メッセージ形式をもとにす。

<https://dice.nims.go.jp/about.html>



MDRでの書誌メタデータ項目



Field	Required	想定用途	記入方式
First published at URL	System	公知であることのエビデンス申告	登録者記入
Supervisor approval	System	上長承認済申告	登録者記入
Title	System	題名	登録者記入
Alternative title	No	別の題名（日本語等）	登録者記入
Data origin	System	データの性質	登録者選択
Abstract or Summary	System	アブストラクト	登録者記入
Keyword	System	キーワード	登録者記入
Specimen	Policy	試料・対象物質	登録者記入
Surname	Policy	関係者姓（英字）	登録者記入
Given name	Policy	関係者名（英字）	登録者記入
Name	Policy	関係者 姓, 名（英字）	登録者記入
Role	Policy	関係者が本Workにどう携わったか	登録者選択
Orcid	Policy	ORCID	登録者記入
Organization	Policy	著者所属組織	登録者記入
Sub organization	No	著者所属組織部門	登録者記入
choose type	No	著者に関する別IDがあれば	登録者選択
Identifier	No	著者に関する別IDがあれば	登録者記入
choose type	No	本Workに関する別IDがあれば	登録者選択
Identifier	No	本Workに関する別IDがあれば	登録者記入
choose type	Policy	日付の種類	登録者選択
Date	Policy	日付（作成日等）	登録者記入
Rights	Policy	ライセンス	登録者選択
Date	No	当該ライセンスのもと公開した日	登録者記入
Version	No	バージョン	登録者記入
Date	No	当該バージョンリリース日	登録者記入
Title	No	その他関係するリンク先のタイトル	登録者記入
Url	No	その他関係するリンク先のURL	登録者記入
Relationship	No	上記リンク先が本Workにどう関係するか	登録者選択
Label	No	カスタムメタデータの項目名	登録者記入
Description	No	カスタムメタデータの値	登録者記入
(NIMS DOI)	No	フォームには表示されず、運用側で付与	運用者記入

出典: <https://dice.nims.go.jp/services/MDR/MDR-metadata-20200717-ext.xlsx>



MDRでの試料メタデータ項目



Field	Required	想定用途	記入方式
Title	No	試料／対象物質（以下「試料」）の名前	登録者記入
choose type	No	試料の化学組成に関するIDの種類	登録者選択
Identifier	No	試料の化学組成に関するID	登録者記入
Description	No	試料の化学組成の説明・記述	登録者記入
choose type	No	試料の結晶構造に関するIDの種類	登録者選択
Identifier	No	試料の結晶構造に関するID	登録者記入
Description	No	試料の結晶構造の説明・記述	登録者記入
Description	No	試料の説明	登録者記入
choose type	No	試料のIDの種類	登録者選択
Identifier	No	試料のID	登録者記入
Material type	No	物質の分類	登録者記入
Material sub type	No	物質の副分類	登録者記入
Description	No	物質の分類の説明・記述	登録者記入
choose type	No	物質の分類のIDの種類	登録者選択
Identifier	No	物質の分類のID	登録者記入
Title	No	試料購入記録のタイトル	登録者記入
Date	No	試料購入記録に関する日付	登録者記入
choose type	No	試料購入記録のIDの種類	登録者選択
Identifier	No	試料購入記録のID	登録者記入
Organization	No	試料供給元組織	登録者記入
Sub organization	No	試料供給元組織部門	登録者記入
Organization	No	試料製造元組織	登録者記入
Sub organization	No	試料製造元組織部門	登録者記入
Purchase record item	No	試料購入記録データ	登録者記入
choose type	No	試料形状のIDの種類	登録者選択
Identifier	No	試料形状のID	登録者記入
Description	No	試料形状の説明・記述	登録者記入
choose type	No	試料の態のIDの種類	登録者選択
Identifier	No	試料の態のID	登録者記入
Description	No	試料の態の説明・記述	登録者記入
Category	No	試料の構造的特徴の分類	登録者記入
Sub category	No	試料の構造的特徴の副分類	登録者記入
Description	No	試料の構造的特徴の説明・記述	登録者記入
choose type	No	試料の構造的特徴のIDの種類	登録者選択
Identifier	No	試料の構造的特徴のID	登録者記入

出典: <https://dice.nims.go.jp/services/MDR/MDR-metadata-20200717-ext.xlsx>



- Hyraxを採用
 - Ruby on Rails, Apache Solr, Fedoraで構成
 - アメリカの大学を中心に、機関リポジトリとして多数の採用例が存在
- RDFによるメタデータ定義とフォームの自動作成機能を持つ



```
class ComplexInstrument < ActiveTriples::Resource
  include CommonMethods

  configure type: ::RDF::Vocab::NimsRdp['Instrument']

  property :alternative_title, predicate: ::RDF::Vocab::NimsRdp['alternative_title']
  property :complex_date, predicate: ::RDF::Vocab::NimsRdp['complex_date'],
    class_name:"ComplexDate"
  accepts_nested_attributes_for :complex_date

  property :description, predicate: ::RDF::Vocab::DC11['description']

  property :complex_identifier, predicate: ::RDF::Vocab::NimsRdp['complex_identifier'],
    class_name:"ComplexIdentifier"
  accepts_nested_attributes_for :complex_identifier
```



Instrument

Title

Alternative title

Date

Processed

Description

Identifier

choose typ

Instrument function

Column number

Category

Sub category

Description



MDRの入力フォームの全体図



Descriptions Method Instruments Specimen details Files

Supervisor approval required

To create a separate work for each of the files, go to Batch upload

For the names of your supervisor and the dates of approval, if you have already got approval on the MDR System of Process Management System (品質管理システム), please use it in the specification number (仕様番号). The current format of Publication number will be published to the public with yours and your supervisor's consent. See the user's Manual for details. eg. Materials Year 2020-10-20 20200100000

Materials Year 2020-10-20 20200100000

[Add another supervisor approval](#)

First published at required

The DOI of the article related to this dataset on the DOI, where this dataset was first made public. DOI example: <https://doi.org/10.26434/chemrxiv-2020-01000>

[Add another first published at](#)

Tag required

A keyword used in identifying a work. English is preferred.

Sample data notebook

[Add another tag](#)

Alternative title

[Add another alternative title](#)

Data origin required

Where the data was obtained

[Add another data origin](#)

Abstract or Summary required

Please brief notes about the work. Examples include abstracts of a paper or citation information for a journal article.

Sample abstract

[Add another abstract or summary](#)

Keyword required

Keywords or phrases you added to describe what the work is about. These are used to search for content.

sample

[Add another keyword](#)

Specimen

hello world

Creator

Surname Butuzov

Given name Taro

Name BUTUZOV Taro

Role author (著者)

Orcid <https://orcid.org/0000-0000-0000-0000>

Organization National Institute for Materials Science

Sub organization Division of Functional Psychoceramics

[Remove](#)

[Add another person](#)

Identifier

choose type [Remove](#)

[Add another identifier](#)

Date

Created 10/10/2020 [Remove](#)

[Add another date](#)

Rights

Rights Creative Commons BY-NC-ND 4.0 International

Date 2020-10-20 [Remove](#)

[Add another rights](#)

Version

Version 1.0

Date 10/10/2020 [Remove](#)

[Add another version](#)

Related item

Title Sample file

URI <http://example.com/sample-dataset>

Relationship New related material [Remove](#)

[Add another relationship](#)

Additional metadata

Label Custom metadata 1

Description value 1 [Remove](#)

[Add another metadata](#)

Descriptions Method Instruments Specimen details Files

Characterization methods

microscopy (顕微鏡法)

[Add another Characterization methods](#)

Computational methods

[Add another Computational methods](#)

Properties addressed

structural (構造的特性)

[Add another Properties addressed](#)

Synthesis and processing

[Add another Synthesis and processing](#)

Descriptions Method Instruments Specimen details Files

Instrument

Title Sample Microscope

Alternative title

Date

choose type

Description Optical microscope of psychoacromics group

Identifier

choose type

Instrument function

Category Q37

Sub category Q340

Description optical microscopy

Manufacturer

Organization Microscope Ltd.

Sub organization Optical Division

Model number MS-100

Operator

Surname Butuzov

Given name Taro

Name BUTUZOV Taro

Role operation (オペレータ/測定者・計測者)

Orcid <https://orcid.org/0000-0000-0000-0000>

Organization National Institute for Materials Science

Sub organization Division of Functional Psychoceramics

Managing organization

Organization National Institute for Materials Science

Sub organization Division of Functional Psychoceramics

[Remove](#)

[Add another instrument](#)

Descriptions Method Instruments Specimen details Files

Specimen type

Title Hello world specimen

Chemical composition

Identifier

choose type

Description MCQ

Crystallographic structure

Identifier

choose type

Description polycrystalline

Description A sample p-MCQ specimen

Identifier

choose type

Material type

Material type Q36

Material sub type Q380

Description samples

Identifier

choose type

Purchase record

Title

Date 01/04/2020

Identifier

choose type

Supplier

Organization World Specimen Inc.

Sub organization

Manufacturer

Organization

Sub organization

Purchase record item

Shape

Identifier

choose type

Description

State of matter

Identifier

choose type

Description solid

Structural feature

Category Q338

Sub category Q345

Description surfaces

Identifier

choose type

[Remove](#)

[Add another specimen type](#)

3. MDRの特長



- 材料データを表現するための柔軟なメタデータ定義を行える
- メタデータはRDFストア上に保存され、RDF/TurtleとJSONによって出力することができる
- **データリポジトリ全体を複製するためのAPIである ResourceSyncに対応している**
 - Sitemap ProtocolをベースにしたAPI
 - ユーザはリポジトリ上の複数のメタデータを一度に取得することが可能
 - 大量のデータを利用するマテリアルズ・インフォマティクスに適する





- resync (<https://github.com/resync/resync>) を使用した例

```
kosuke@mdr-dev:~$ resync https://mdr.nims.go.jp --noauth
Using URI mapping: https://mdr.nims.go.jp -> mdr.nims.go.jp
Status:      NOT IN SYNC (same=0, to create=2968, to update=4, to delete=0)
Will GET 2972 resources
Status:      PART SYNCED (same=0, created=2968, updated=4, to delete=0)
```

メタデータ
一括取得の
コマンド実行

```
kosuke@mdr-dev:~$ ls -l mdr.nims.go.jp/
total 8
drwxr-xr-x 2 kosuke docker 4096 Sep  4 02:52 collections
drwxr-xr-x 5 kosuke docker 4096 Aug 28 2019 concern
kosuke@mdr-dev:~$ ls -l mdr.nims.go.jp/concern/
total 104
drwxr-xr-x 2 kosuke docker 4096 Sep  4 02:52 datasets
drwxr-xr-x 2 kosuke docker 57344 Sep  4 02:54 file_sets
drwxr-xr-x 2 kosuke docker 36864 Sep  4 02:57 publications
```

```
kosuke@mdr-dev:~$ ls -l mdr.nims.go.jp/concern/datasets/
total 1352
-rw-r--r-- 1 kosuke docker 23501 Jun 16 03:57 05741s554
-rw-r--r-- 1 kosuke docker 15364 Aug 19 2019 08612n52b
-rw-r--r-- 1 kosuke docker 39775 Jun 15 16:26 0c483k10j
-rw-r--r-- 1 kosuke docker 20854 Jun 16 03:57 1n79h516t
-rw-r--r-- 1 kosuke docker 31299 Jun 16 03:57 1r66j1947
-rw-r--r-- 1 kosuke docker 23229 Jun 16 03:54 2227mq55d
-rw-r--r-- 1 kosuke docker 31177 Jun 16 03:57 2z10wr05h
-rw-r--r-- 1 kosuke docker 31769 Jun 16 03:57 37720d711
-rw-r--r-- 1 kosuke docker 25458 Jun  1 23:32 3j333309j
-rw-r--r-- 1 kosuke docker 26935 Jun 15 16:26 3t945r48t
```

取得した
メタデータ一覧



MDRから出力されるRDF (一部)



```
<https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source> <http://purl.org/dc/terms/hasPart>
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g47298744140380>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953696467160>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953662368900>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953632935260>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953659112760>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953710325320>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953705788080>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953707573640>,
  <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953699353840>;
<http://www.iana.org/assignments/relation/first> <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g47298744140380>
<http://www.iana.org/assignments/relation/last> <https://mdr.nims.go.jp/concern/datasets/x059c824t/list_source#g46953699353840>
<info:fedora/fedora-system:def/model#hasModel> "ActiveFedora::Aggregation::ListSource" .

<https://mdr.nims.go.jp/concern/datasets/x059c824t#dateg47260583842680> a <http://www.w3.org/2006/vcard#hasName>
  <http://bibframe.org/vocab/classification> "Published";
  <http://bibframe.org/vocab/eventDate> "17/06/2016" .

<https://mdr.nims.go.jp/concern/datasets/x059c824t#persong47447410232820> a <http://xmlns.com/foaf/0.1/Person>,
  <http://purl.org/spar/datacite/hasIdentifier> "https://orcid.org/0000-0003-3986-1900";
  <http://www.w3.org/2006/vcard/ns#hasName> "TODOROKI, Shin-ichi";
  <http://www.w3.org/ns/org#organization> "National Institute for Materials Science";
  <http://xmlns.com/foaf/0.1/familyName> "TODOROKI";
  <http://xmlns.com/foaf/0.1/givenName> "Shin-ichi" .
```

4. MDRの運用



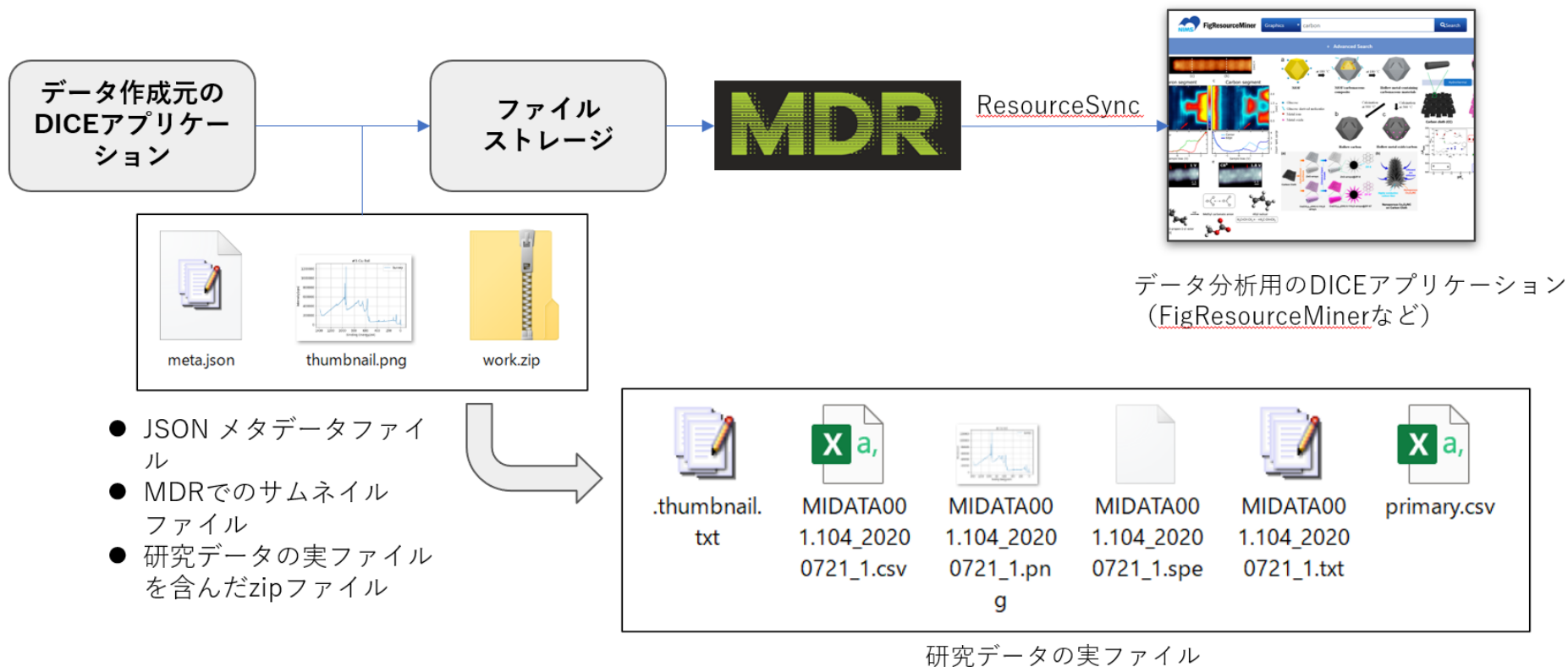
- データ登録は現時点では機構の職員のみが行える
 - 将来的に共同研究機関からの登録をサポート予定
- 一般ユーザ（研究者自身）のWebフォーム経由による登録と、管理者によるバッチ登録をサポート
- ファイルサイズの上限は、フォームの場合は100MB、バッチの場合は特に制限なし
- GUI上で入力を必須とするメタデータ項目は以下のとおり
 - **Title, Data origin, Abstract, Keyword**
 - **First published URL**（そのデータが公知であることを示すURL。DOIを推奨）
 - **Supervisor approval**（発表を承認した上長の氏名、承認日、機構内の発表決裁の番号。DOIがない場合のみ必要）



バッチ登録のフロー



- バッチ登録を行う場合、MDRがアクセスできる所定の共有ディレクトリに、所定の構造によるメタデータと実データを含んだzipファイルを作成する
- MDRは当該ディレクトリのファイルを定期的にインポートする





- MDRでは**公知となっているデータのみ**登録可能
 - たとえば、**出版済みの論文の作成時に使用したが、論文と同時に公開しなかった実験データ**
- ファイルサイズの上限は、フォームの場合は100MB、バッチの場合は特に制限なし



- 一般ユーザがデータを登録した時点では、データは外部に公開されない
 - ログインしたユーザのみが参照可能な状態となる
- データの公開依頼をMDR管理者にメールで送信すると、管理者が公開作業を実行する
 - 公開作業の実行後は、データの編集は不可
- 公開作業には以下の作業を含む
 - 著者へのORCID番号の紐付け（機構職員のみ）
 - 機構の研究職・エンジニア職の約90%がORCID番号を取得
 - 現時点ではORCID番号の紐付けは手作業
 - ライセンスの確認（論文ファイルのみ）
 - **DOIの付与**



- MDRでは国内外の他の検索サービスへのデータ提供を重視し、**JaLC DOIを付与せず、DataCite DOIを付与している**
 - Scientific Data（Natureのデータジャーナル）は、投稿時に**DataCite DOIを付与できるリポジトリにデータを登録することを**要求している
 - Data Citation Index（Clarivateのデータセット検索サービス）は、**DataCiteやCrossrefのWebAPI**を用いて、収録対象のメタデータを収集している

5. MDRの課題



- 「論文とその付随データのセット」は、論文とデータセットのどちらになるのか？
 - 別々のデータとして登録する場合、それらの間の関係を、メタデータとリポジトリの画面内でどのように表現するか？

MDRでは登録者と逐次相談の上で、どちらに登録するかを決めている



- そもそも「論文なし・データのみ」が登録されるケースがほとんどない
 - 公開できるデータは論文投稿時に、論文と同時に出版社に送っていて、論文と同時に出版社のサイトで公開されることが多く、研究者にとって新たに公開できる・公開したいデータが存在しないことが多い
 - 研究者「出せるデータは出版社のサイトでオープンアクセスになってるのに、なんでMDRにも登録する必要があるの？」
- データの登録のインセンティブが薄い

ダウンロード数や引用数など、データの価値を「見える化」するための仕組みが必要



- 現在のMDRでは、**論文投稿時のデータ登録先として利用するには機能が不十分**
 - 具体的には**査読者とのデータの共有方法**が問題
 - 限定公開URLの発行機能があるが、24時間しか有効にならず、しかも一度しかダウンロードできないため、査読者の数だけあらかじめURLを用意しなければならない
 - 現状ではDOI付与も管理者の作業が必要
 - 研究者「登録を待っている間に、研究の競争相手に論文投稿で先を越されるのでは本末転倒」
 - 外部の汎用リポジトリ（Zenodo, figshareなど）と比較した場合の優位点をどこで打ち出すか？

論文投稿時の際のデータリポジトリでの
データ共有のあり方は検討が必要



- **「共通メッセージ形式」によるメタデータの記述が困難**
 - 複数の研究者から「**自分の研究で使いたい項目がスキーマに定義されていない**」というフィードバックが寄せられた
 - 研究者「この項目では細かすぎて書けない」「荒すぎて書けない」
 - 材料科学は多くの研究分野で成り立っており、ユーザが増えていくと、同様の問題に次々に遭遇することが予想される
- **データ登録フォームが巨大で、ユーザフレンドリーとはいえない**
 - スキーマ追加の要望に単純に答えていくと、フォームがさらに複雑になる
- メタデータに相当する情報を**ディレクトリ（フォルダ）の構造**で表現している場合がある
 - MDRでは**ディレクトリ構造を保ったままでのデータ登録ができない**ため、ディレクトリをzipファイルにまとめる必要がある

データ登録のUIについて再検討中



MDRの入力フォームの全体図



Descriptions Method Instruments Specimen details Files

To create a separate work for each of the files, go to Batch upload

Supervisor approval method
Click the name of your supervisor and the date of approval. If you have already got approval on the first release of the specimen management system (国産特許番号: 2,742, 742, please also fill in the application number (国産特許) The current database is published here and will be published to the public with your and your supervisor's consent. See the user's manual for details. eg. Suzuki, Ken 2020-08-01 02:00:00 GMT+0900

[Add another Supervisor approval](#)

First published on method
This is the date of the article related to this dataset on the URL, where the dataset was first made public. For example: <https://doi.org/10.26434/chemrxiv-2020-08-01>

Title method
A name to aid in identifying a work. English is preferred.
Sample title:

[Add another Title](#)

Alternative title

Date origin method
acquisition 取得

[Add another Date origin](#)

Abstract or Summary method
Free text notes about the work. Examples include abstracts of a paper or citation information for a journal article.
Sample abstract:

[Add another Abstract or Summary](#)

Keywords method
Words or phrases you want to describe what the work is about. These are used to search for content.
sample:

[Add another Keyword](#)

Specimen
Hello world

Creator
Surname: Butsuzaki
Given name: Taro
Name: Butsuzaki Taro
Role: operator 研究者
Email: https://doi.org/10.26434/chemrxiv-2020-08-01
Organization: National Institute for Materials Science
Sub organization: Division of Functional Psychoacoustics
[Remove](#)
[Add another person](#)

Identifier
choose type: [Remove](#)
[Add another identifier](#)

Date
Created: 09/05/2020 [Remove](#)
[Add another date](#)

Rights
Creative Commons BY Attribution 4.0 International
Date: 09/05/2020 [Remove](#)
[Add another rights](#)

Version
1.0
Date: 09/05/2020 [Remove](#)
[Add another version](#)

Related item
Title: Sample title
URL: <http://example.org/sample-dataset>
Relationship: New related material [Remove](#)
[Add another relationship](#)

Additional metadata
Label: Custom metadata 1
Description: value 1 [Remove](#)
[Add another metadata](#)

Descriptions Method Instruments Specimen details Files

Characterization methods
 microscopy 顕微鏡法
[Add another Characterization methods](#)

Computational methods

[Add another Computational methods](#)

Properties addressed
 structural 構造の性質
[Add another Properties addressed](#)

Synthesis and processing

[Add another Synthesis and processing](#)

Descriptions Method Instruments Specimen details Files

Instrument
Title: Sample Microscope
Alternative title:

Date
choose type:

Description
 Optical microscope of psychoacoustics group

Identifier
choose type:

Instrument function
Category: Q37
Sub category: Q340
Description: optical microscopy

Manufacturer
Organization: Microscope Ltd
Sub organization: Optical Division

Model number
 MS-100

Operator
Surname: Butsuzaki
Given name: Taro
Name: Butsuzaki Taro
Role: operator オペレーター 計算機
Email: https://doi.org/10.26434/chemrxiv-2020-08-01
Organization: National Institute for Materials Science
Sub organization: Division of Functional Psychoacoustics

Managing organization
Organization: National Institute for Materials Science
Sub organization: Division of Functional Psychoacoustics
[Remove](#)
[Add another instrument](#)

Descriptions Method Instruments Specimen details Files

Specimen type
Title: Hello world specimen

Chemical composition
Identifier: choose type:
Description: MO2

Crystallographic structure
Identifier: choose type:
Description: polycrystalline

Description
 A sample p-MOS specimen
Identifier: choose type:

Material type
Material type: Q37
Material sub type: Q30
Description: oxide
Identifier: choose type:

Purchase record
Title:
Date: 09/04/2020
Identifier: choose type:
Supplier
Organization: World Specimen Inc.
Sub organization:
Manufacturer
Organization:
Sub organization:
Purchase record
Item:

Shape
Identifier: choose type:
Description:

State of matter
Identifier: choose type:
Description: solid

Structural feature
Category: Q38
Sub category: Q345
Description: surface
Identifier: choose type:
[Remove](#)
[Add another specimen type](#)

実際にはこの何倍もの
入力フォームが必要



ディレクトリ構造によるデータの表現の例

Full-text search

Experimental Data Transfer System

UID: 0000664 [LOGOUT](#)

Search under folder your proposal [DOWNLOAD CHECKED ITEMS](#)

Directory structure with explore style

Download file via zip

Metadata view for selected file

FILE NAME	PATH
☒ Zr-K_stabilized_ZrO2_YSZ_SG11_50ms_170419_fe.json	/SPring-8/BL14B2/Standard
☐ Zr-	/SPring-8/BL14B2/Standard
☐ K_stabilized_ZrO2_YSZ_SG11_50ms_170419_fe.system.json	/SPring-8/BL14B2/Standard
☐ Zr-K_stabilized_ZrO2_YSZ_SG11_50ms_170419_fe.user.json	/SPring-8/BL14B2/Standard
☐ README.csv	/SPring-8/BL14B2/Standard
☐ Zr-K_stabilized_ZrO2_YSZ_SG11_50ms_170419.dat	/SPring-8/BL14B2/Standard

Metadata	Value
@subject@correspondence@affiliation	JASRI
@subject@create_time	2019-03-27 21:20:17
@subject@disk_name	Standard
@subject@facility	SPring-8
@subject@pid	spring8.32caae
@subject@proposal_number	201450000
@subject@register_name	/SPring-8/BL14B2/Standard/Zr/K/311/s/K_stabilized_ZrO2_YSZ_SG11_50ms_170419_fe.json
@subject@representative@username	0007719

BENTEN SYSTEM: Copyright (C) 2019 JASRI All Rights Reserved.

SPring-8の所外実験データ転送システム「BENTEN」
<https://user.spring8.or.jp/sp8info/?p=37465>



- (現時点では) 登録できるのは「**公知データ**」のみ
- **データ公開時のリスクを避けるため**
 - データは論文と異なり、**どのような中身なのかが見た目ではすぐにわからないし、どのような利用のされ方をするのかもわからない**
 - 論文であれば、それは**公開されることが前提の情報である**ことが明らかであるが、**データはそうではない**
- このため「論文で使用した、公知になっているデータ」のみを登録できるようにしている
 - 登録時には公知になっている証拠として、論文のURL (DOI推奨) をメタデータに記入させるようになっている
- 一方で「**未公知のデータを共同研究機関と共有したい**」というユースケースも存在する
 - 限定公開のための機能を追加予定

第1回 SPARC Japan セミナー2020
「研究データ公開：フルオープンと制限公開の境界線」
での議論もぜひごらんください



- データリポジトリといっても、仕組み自体はこれまでの機関リポジトリと根本的に違うものではない
- メタデータの扱いはやはり難しい
 - 研究で作られるデータは多種多様で、**中身は基本的にはデータの作成者本人しかわからない**
 - 分野ごとのメタデータ項目の設計は、それ自体が研究と言ってもよいほどの大仕事
- **まず、図書館でつけられるメタデータをきちんとつけよう**
 - 具体的には識別子 (ID)
 - DOIの付与は当たり前。DOIのないデータは、**ISBNのない本と同じ**
 - **ORCIDの紐付けも当然**。学内の研究者のORCID番号を確認しよう



- データ公開の作法はまだ確立していない部分が多い
 - データの種別の問題
 - 「論文とデータのセット」は論文なのかデータなのか
 - 公開したデータの評価
 - ダウンロード数、引用数、それともほかの指標？
 - データの「セキュリティ」
 - 公開してよいデータか？ どこまで共有してよいデータか？
- ほかにも課題は多数ありますが…

ぜひいっしょに考えていきましょう！

TANABE.Kosuke@nims.go.jp
