

マテリアル先端リサーチインフラ事業における共用機器 PID 付与の実験

田辺 浩介¹⁾, 松波 成行¹⁾

1) 物質・材料研究機構 技術開発・共用部門

TANABE.Kosuke@nims.go.jp

An experiment on assigning PIDs to shared research instruments in Advanced Research Infrastructure for Materials and Nanotechnology in Japan (ARIM)

Kosuke Tanabe¹⁾, Shigeyuki Matsunami¹⁾

1) Research Network and Facility Services Division, National Institute for Materials Science

概要

材料科学データ駆動型研究において、共用機器とその機器を用いて作成された研究データセットの情報を紐付けて管理することは、研究成果の把握という点において重要である。本報告では共用機器と研究データセットの両方に PID (永続的識別子) を付与することで、共用機器とその成果となる研究データの紐付けを行い、研究データの検索性の向上、ならびに共用機器の情報の永続的な利用への応用の可能性を示した。

1 ARIM 事業の概要

1.1 背景

我が国では、令和 3 年 4 月に内閣府によって決定された「マテリアル革新力強化戦略」の下、「マテリアル DX プラットフォーム」の構築が進められている。その中で、文部科学省「マテリアル先端リサーチインフラ事業(ARIM : エイリム)」では、全国的な最先端共用設備体制と高度な技術支援提供体制に加え、リモート化・自動化・ハイスループット化された先端設備を導入し、設備共用支援を継続する。加えて、設備共用に伴って創出されるマテリアルデータは、第三者が利活用しやすく、かつ機械可読性の高い構造化された形で、収集・蓄積およびデータ共用を推進している。

1.2 プロジェクトの概要と特長

ARIM では、約 1100 台の共用機器が利用可能である。これにより、研究者は高額な機器を自ら購入することなく、最新技術を活用したマ

テリアル分野に関する研究開発を進めることができる。また、共用機器は精密なデータ取得や高度な加工技術が求められる分野において、高品質なデータの創出が期待されている。

さらに、ユーザーの承諾を得た上で、共用機器のデータは、物質・材料研究機構がデータ基盤として整備するデータ構造化システム（サービス名：RDE）を通じて AI Ready な機器ごとに固有のデータ構造化様式で蓄積されている。これらのデータは、一定のエンバゴ期間（デフォルトでは、機器利用年度の翌年度から起算して 2 年間）を経た後、ARIM データポータル [1] において、機器利用の利用課題ごとにデータカタログが掲載される。日本国内に在住し、かつ法人に属する者であれば、ARIM データポータルを通じてデータ利用申請を行うことで、当該データのライセンスを受けることができる。この仕組みにより ARIM で収集・蓄積されたデータは広く利活用されることが期待されており、研究者や企業が新たな解析手法の開発、先端機器開発のための強力なデータ基盤として寄与することが見込まれている。

1.3 本検討の目的

大学や公的機関が保有する機器に関する現状と課題として、特にハードウェアの保管体制だけでなく、マニュアルや関連ドキュメントの保存体制が不十分である点が指摘されている。また、永続的識別子 (PID) の持続性に関しては、付与された ID がリゾルブ可能（転送先があること）であることが重要であり、機器が撤去された場合や管理対象から外れた場合でも、何らかの情報にアクセスできる状態が求められる。本検討では、ARIM における機器情報の永続性を確保するための方法について議論し、特に DOI や Handle といった PID のリゾルブ可能性を担保する手法について検討する。また、これらの PID が、ARIM に登録された機器に対してどのように適用可能か、その実施可能性を議論する。

2 共用機器への PID 付与の重要性

2.1 データ登録フロー

ARIM のデータ共用システムを効果的に活用するためには、機器利用プロセスとデータ登録フローの正確な理解が不可欠である。本節では、これらのプロセスについて詳述する。

まず、利用者は各機関のウェブサイトを通じ

て機器の利用を申し込む。申請内容は該当機関で審査され、承認されると「利用課題番号」が発行される。この利用課題番号は、データ登録およびデータ共用の中心的な要素であり、データの識別や管理において極めて重要な役割を果たす。

次に、RDE システム内に該当の利用課題に対応するデータセットが作成される。特に注目すべき点として、RDE では機器ごとにデータセットが個別に管理されることが挙げられる。したがって、複数の機器を利用した場合、それぞれの機器に対して個別のデータセットが作成されることになる。

続いて、データ登録の具体的な手順を説明する。利用者が機器を用いて測定を完了した後、機器に接続された PC から測定データを RDE にアップロードする。このデータは、まず RDE の前受サーバーにてウイルスチェックを受け、その後、ARIM のデータ構造化仕様に基づく Python スクリプトが実行される。このスクリプトにより、データは自動的に構造化され、対応するデータセットに保存される。各データセットには、利用課題番号、機器情報、利用者情報が関連付けられており、これによってデータの追跡や管理が容易に行える。

データの公開・非公開に関しては、原則と

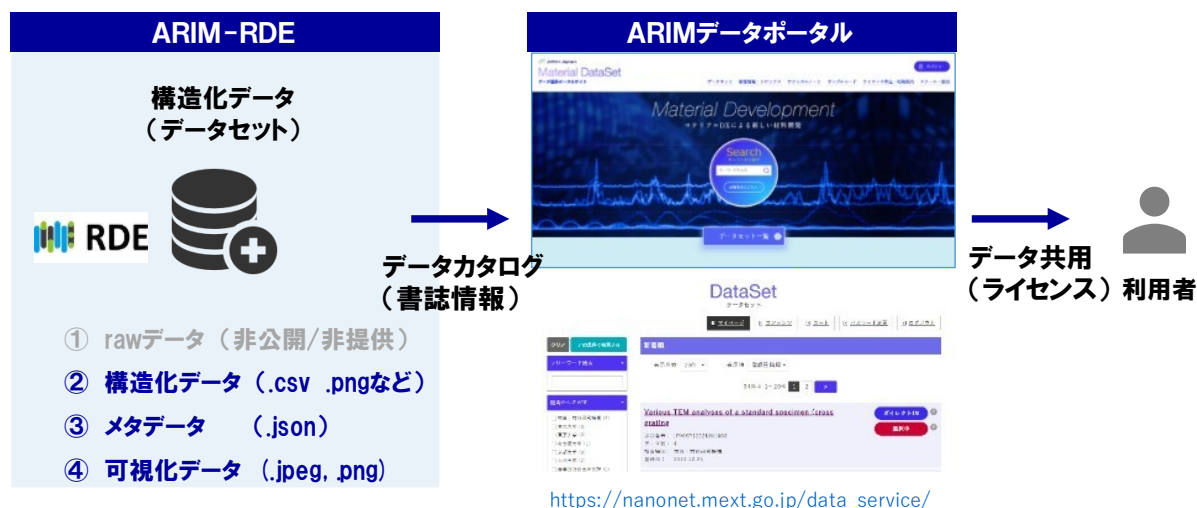


図1：ARIM 事業におけるデータ構造化システム（RDE）とデータ共用のための ARIM データポータルとの関係

して機器利用年度の翌年度から 2 年間のエンバゴ期間が設定されている。エンバゴが解除されると、第三者へのデータ共有が可能になる。ARIM プロジェクトでは、外為法などの規制を考慮し、第三者へのデータ共有は日本国内に在住し、かつ法人に所属する者に限定されている。このような仕組みにより、ARIM のデータ共有システムは、データの透明性と信頼性を維持しつつ、利用者にとって柔軟で効率的なデータ管理環境を提供している。

2.2 設計概念

システム工学の観点から見ると、RDE の前受サーバーはデータレイクとして機能し、構造化スクリプトが生成する構造化ファイル群を集積する部分はデータウェアハウス (DWH) に相当する。RDE はデータ共有機能も備えているが、ARIM では、さらに広範な第三者へのデータ共有を促進するために ARIM データポータル [1] という専用サイトを構築し、データマート形式でデータ配信を行っている (図 1)。

この ARIM データポータルでは、書誌情報に基づくデータカタログ形式のランディングページを提供しており、ユーザーはここからデータセットの概要を閲覧できる。必要なデータセットはこのカタログを通じて検索し、個別のデータセット単位で利用申し込みを行うことが可能である。

データカタログには、利用課題番号、データセット名、データセットの要約が記載される欄が設けられており、登録者の希望に応じてデータセットに DOI を付与する予定である。また、ARIM の共用機器に関する機器情報は、事業全体の全機器が登録されている「共用設備検索」 [2] で機器ごとの情報が明記されているランディングページとリンクしている。

なお、ARIM では、共用機器の登録および抹消が厳格に管理され、各機器には一意の機器 ID が付与されている。この機器 ID はデータ登

録時に必須の情報であり、利用課題に関連する機器の特性を識別する基本情報となるだけでなく、データセットの管理を効率化し、データの整合性と追跡性を向上させる重要な要素である。今後、機器 ID に加えて機器固有の PID を導入することで、データセットに関する情報の一貫性が保たれ、データの信頼性がさらに向上することが期待されている。

2.3 先行事例

計測機器に対する PID の付与は、研究機関やデータ共有プラットフォームにおける透明性とトレーサビリティの確保に寄与し、機器利用履歴の追跡や、データとその取得元となる機器とのリンクを明確にするための重要な要素となっている。特に、研究データが再利用される際、そのデータがどの機器で、どのような条件下で取得されたかを特定することは、再現性や科学的信頼性の観点から極めて重要である。このような背景から、計測機器に PID を付与する取り組みは、国際的に進展している。

この分野で特筆すべき動きの一つが、Research Data Alliance (RDA) による Persistent Identification of Instruments Working Group (PIDINST WG) の活動である。同グループでは、共用機器への PID 付与に関する具体的なユースケースが議論されており、とりまとめられているホワイトペーパー [3] では、DataCite DOI や ePIC [4] を用いた実例が示されている。また、PID に関連するメタデータの標準化も進められており、同グループは共用機器のためのメタデータスキーマを開発し、その仕様を公開している [5]。

DataCite においては、Metadata Schema 4.5 で共用機器への DOI 付与がサポートされており、PIDINST WG のメタデータスキーマとマッピングが行われている [6]。このような取り組みにより、2024 年 10 月現在、DataCite DOI が付与された共用機器は 138 件に達している [7]。

欧州においては、EUDAT が研究データに加え、B2INST[8]というデータベースで共用機器情報を収集しており、2024 年 10 月時点で 306 件の機器に Handle が付与されている。また、オーストラリアでは、Australian Research Data Commons (ARDC) が主導し、PID の普及促進を目的としたウェビナーの開催など、機器への PID 付与に関する積極的な取り組みが進められている[9]。

3 共用機器への PID 付与の実装

3.1 使用する PID の種別

共用機器に付与可能な永続的識別子 (PID) には、DOI、ePIC、RRID[10]など複数の種類が存在するが、今回の実験では Handle を採用した。その理由として、同PIDを付与するために必要なソフトウェアである Handle Server[11]が物質・材料研究機構ですでに運用されており、新たな費用や手続きが発生しないことが挙げられる。なお、DOIの利用も検討しているが、これについては第4章で詳述する。

一方、共用機器によって生成された研究データに対しては、DataCite のテスト環境で発行された DOI を使用した。DataCite DOI は、Web 管理画面や REST API を通じて容易に発行が可能である。加えて物質・材料研究機構が運用している研究データリポジトリである Materials Data Repository [12] でも、研究データに対して DataCite DOI を付与しているため、本実験でも同様の方式を採用した。

3.2 PID の設計

Handle や DataCite DOI の付与には、リゾルブ先 (転送先) の URL が必要である。つまり、PID の付与対象の情報 (機器や研究データ) は、URL によって一意に識別できるようになっている必要がある。このため、ARIM の共用機器、ならびに研究データにおいて、それらが

一意の URL を保持しているかを調査した。

ARIM の共用機器の情報は、上述した「共用設備検索」で公開されており、機器ごとに個別の URL が割り当てられている。たとえば、物質・材料研究機構で管理している共用機器の一つである「単原子分析電子顕微鏡」の情報は当該装置に一義的な URL [13]で恒常的にアクセスできるようになっている。このため、現時点においてはPIDの付与に問題はないと判断した。

Handle・DataCite DOI とも、組織ごとに割り当てられるプレフィックスと、機器やデータごとに割り当てられるサフィックスによって構成される。例えば、

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

という DOI の場合、10.48505 が DataCite によって物質・材料研究機構に割り当てられたプレフィックス、nims.1447 が研究データに割り当てられるサフィックスである。プレフィックスは PID の管理元 (Handle であれば CNRI Corporation for National Research Initiatives [14])、DataCite DOI であれば DataCite) から割り当てられるため、PID の付与の際に研究機関側で変更することはできない。一方、サフィックスは研究機関で任意の値を設定することができる。

本報告では、共用機器・研究データとも、基本的に連番を用いてサフィックスを生成している。共用機器に対しては、ARIM で用いている機器の管理番号を用いることも検討したが、将来の管理体制の変更によって管理番号の採番方針に影響が出る可能性を鑑み、単純な連番を採用した。また、研究データに対しては文字列と連番の組み合わせとした。Handle の付与は、バッチ付与用のテキストファイルを作成後、コマンドラインで実施した。DataCite DOI の付与は、DataCite の提供する Web ユーザーインターフェースで実施した。

```

1 {
2   "doi": "10.83134/CD2D-GY95",
3   "url": "https://nanonet.next.go.jp/data_service/arim_data.php?
mode=detail&code=38&key=Rbhw5YSzxFMhEYoykTdXKoavE2IGBDWS&inst_array[]=
1&display_order=0&display_result=0&page=1",
4   "types": {
5     "resourceType": "json",
6     "resourceTypeGeneral": "Dataset"
7   },
8   "creators": [
9     {
10      "name": "KIMOTO, Koji",
11      "nameType": "Personal",
12      "givenName": "Koji",
13      "familyName": "KIMOTO",
14      "affiliation": [
15        {
16          "name": "National Institute for Materials
Science",
17          "schemeUri": "https://ror.org",
18          "affiliationIdentifier":
"https://ror.org/026v1ze26",
19          "affiliationIdentifierScheme": "ROR"
20        }
21      ],
22      "nameIdentifiers": []
23    }
24  ],
25  "titles": [
26    {
27      "lang": "en",
28      "title": "Various TEM analyses of a standard specimen
(cross grating)",
29      "titleType": "AlternativeTitle"
30    }
31  ],
32  "publisher": {
33    "name": "ARIM"
34  },
35  "publicationYear": 2023,
36  "descriptions": [
37    {
38      "lang": null,
39      "description": "TEM and STEM images, a diffraction
pattern and an EEL spectrum of a cross grating replica (Agar S106).",
40      "descriptionType": "Abstract"
41    }
42  ],
43  "relatedIdentifiers": [
44    {
45      "schemeUri": null,
46      "schemeType": null,
47      "relationType": "IsCollectedBy",
48      "relatedIdentifier": "20.500.11932/1034",
49      "resourceTypeGeneral": "Instrument",
50      "relatedIdentifierType": "Handle",
51      "relatedMetadataScheme": null
52    }
53  ],
54  "state": "draft"
55 }

```

研究データの DOI

研究データの作成者の情報

研究データの名称・概要の情報

実験装置の情報

図 2 共用機器の PID を含めた研究データセットの DataCite DOI のメタデータ

3.3 研究データのメタデータ作成

研究データには DataCite DOI を付与するため、DataCite のメタデータスキーマに沿ってメタデータを作成した。作成例を図 2 に示す。共用機器の情報として、機器に付与した Handle である 20.500.11932/1034 を、検索対象となるメタデータ項目である relatedIdentifier に含めている。DOI のリゾル部分は、ARIM の共用データの公開 URL としている。DataCite の REST API では

クエリの例

<https://api.test.datacite.org/does?query=relatedIdentifiers.relatedIdentifier:20.500.11932/1034%20AND%20relatedIdentifiers.relationType:IsCollectedBy>

図 3 DataCite API を通じて、共用機器 20.500.11932/1034 によって収集された (IsCollectedBy) データセットの DOI の一覧を取得するクエリ

relatedIdentifier を検索対象に含められるようになっており、図3のように REST API を呼び出すことで、当該共用機器の relatedIdentifier に紐づく研究データを検索することができるようになる。

4 考察

4.1 共用機器への PID 付与の粒度

共用機器では、キャリブレーションや機器の改造などによって、機器自体の構成が大きく変わることがある。構成の変更ごとに異なる PID を付与するにしても、機器の構成が取りうる変更は無数に存在する。たとえば、透過電子顕微鏡 (TEM) には二次電子検出器や元素分析機器などが付属しており、詳細な部品やアタッチメント (付帯機器) が取り得る構成に対して PID を付与することは、PID 自体の管理を難しくする可能性がある。このため、本報告では機器全体に対して一つの PID を付与することとしている。また、付与に対する費用が発生する PID (DataCite DOI の場合、2023 年 10 月時点で 1 件あたり 0.8 ユーロ[15]) を使用する場合は、PID の付与の粒度によって付与件数が大きく変わってくるため、PID 付与の費用に影響が及ぶ可能性がある。本報告においても、共用機器で使用する PID には DataCite DOI の付与も検討しているが、上記の議論が進んでいないため、現時点では Handle を用いることとしている。

4.2 共用機器の永続性

論文や研究データは電子的なデータとして作成されることが多く、その長期保存においては、少なくとも物理的な制約は大きくない。特に論文は出版物であり、永続的なアクセスのための仕組みが、組織や技術などさまざまな点から整備されている。一方で、共用機器は一般的に利用後の長期保存を前提とされておらず、保存のための物理的な制約も非常に大きい。

このため、PID によって提供されるメタデータやランディングページからコンテンツ自体が

利用できる論文や研究データと異なり、共用機器の情報は機器自身の情報の記述に限定される。本報告時点では、ランディングページの内容を保存したファイルを Materials Data Repository に登録する方法と、国立国会図書館インターネット資料収集保存事業(WARP)[16]や Internet Archive[17]などの Web アーカイブに保存する方法を検討している。

4.3 実環境での運用の課題

本報告時点では、Handle の付与はコマンドラインによるバッチ、DataCite DOI の付与は Web ユーザインターフェースを用いて、それぞれ手動で行っている。しかし、対象の共用機器が増えるにつれて、その機器から出力されるデータセットも増えることから、今後実環境で動作させるためには、PID 付与作業の自動化が求められる。Materials Data Repository で DataCite DOI の付与は REST API で行っており[18]、Handle Server も REST API を備えているため、技術的には DataCite DOI と Handle の付与を一元的に管理するアプリケーションも作成可能であるが、本報告時点では未実装となっている。

5 おわりに

本報告では、Handle と DataCite DOI を用いて、ARIM の共用機器と研究データへの PID の付与、ならびに ARIM の共用機器と研究データの紐付けを行った。これによって、共用機器の情報を研究データの DOI のメタデータを通じた公開・流通が行えることを実証するとともに、研究データの検索対象に共用機器の情報を含めることによる検索性の向上、ならびに共用機器の情報の永続的な利用への応用の可能性を示した。

参考文献

[1] ARIM データポータル.

https://nanonet.mext.go.jp/data_service/

(accessed on 2024-10-18)

[2] ARIM の共用設備検索.

<https://nanonet.mext.go.jp/facility.php>

(accessed on 2024-10-18)

[3] PIDINST White Paper.

<https://docs.pidinst.org/en/latest/white-paper/index.html> (accessed on 2024-10-18)

[4] ePIC. <https://www.pidconsortium.net/>

(accessed on 2024-10-18)

[5] Krah, R., Darroch, L., Huber, R., Devaraju, A., Klump, J., Habermann, T., Stocker, M., & RDA PIDINST WG Members. Metadata Schema for the Persistent Identification of Instruments (1.0). Zenodo. 2022.

<https://doi.org/10.15497/RDA00070> (accessed on 2024-10-18)

[6] PIDINST Schema 1 Mapping.

<https://datacite-metadata-schema.readthedocs.io/en/4.5/mappings/pidinst/> (accessed on 2024-10-18)

[7] DataCite Commons.

<https://commons.datacite.org/doi.org?query=types.resourceTypeGeneral%3AInstrument> (accessed on 2024-10-18)

[8] B2INST. <https://b2inst.gwdg.de/> (accessed on 2024-10-18)

[9] Identifiers for Instruments Webinar.

<https://ardc.edu.au/resource/identifiers-for-instruments-webinar/> (accessed on 2024-10-18)

[10] Research Resource Identification.

<https://www.rrids.org/> (accessed on 2024-10-18)

[11] Handle.Net® Software.

https://www.handle.net/download_hnr.html (accessed on 2024-10-18)

[12] Materials Data Repository.

<https://mdr.nims.go.jp/> (accessed on 2024-10-18)

[13] 単原分析電子顕微鏡.

<https://nanonet.mext.go.jp/facility.php?mode=detail&code=34> (accessed on 2024-10-18)

[14] Corporation for National Research Initiative. <https://www.cnri.reston.va.us/> (accessed on 2024-10-18)

[15] DataCite Fee Model.

<https://datacite.org/fee-model/> (accessed on 2024-10-18)

[16] 国立国会図書館インターネット資料収集保存事業. <https://warp.ndl.go.jp/> (accessed on 2024-10-18)

[17] Internet Archive.

<https://www.archive.org/> (accessed on 2024-10-18)

[18] 田辺浩介. DataCite を選んだ理由 ～ Materials Data Repository における DataCite DOI の活用～. Japan Open Science Summit 2023, オンライン, 2023.

<https://doi.org/10.48505/nims.4198> (accessed on 2024-10-18)