

マテリアル先端リサーチインフラ（ARIM）における

データ共有のためのデータカタログの設計と運用

松波 成行¹⁾

1) 物質・材料研究機構 技術開発・共用部門

MATSUNAMI.Shigyuki@nims.go.jp

Design and Operation of Data Catalog for Promoting Data Sharing in Advanced Research Infrastructure for Materials and Nanotechnology in Japan (ARIM)

Shigeyuki Matsunami¹⁾

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

概要

本発表では、文部科学省が推進するマテリアル先端リサーチインフラ（ARIM）事業における実験データ共有の促進を目的としたデータカタログの設計・運用方針および Digital Object Identifier（DOI）付与に関する取り組みについて報告する。ARIM は、全国 26 の研究機関に設置された約 1,200 台の共用研究機器から生成される多様な実験データを対象としており、効率的かつ再利用可能なデータ流通を実現するため、データカタログの書誌情報（メタデータ）の標準化と品質保証を重視している。本発表では、ARIM のルールに基づく DOI 付与の要件を満たすための一貫した基準策定と、データセット全体の品質向上を目的としたカタログ設計の具体的手法について述べる。これらの取り組みを通じて、FAIR 原則に準拠した研究データ管理の実現と、長期的なデータ利活用基盤の強化を目指す。

1 ARIM 事業の概要

我が国では、2021 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 [1]。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・蓄積を進めている。

2025 年 7 月からは、半導体分野の研究基盤連携強化を目的とした「ARIM 半導体基盤プラットフォーム」のネットワーク構築により、多様

な半導体研究開発ユーザーからのアクセシビリティ向上を図り、我が国の半導体分野における研究開発力強化および人材育成基盤の拡充を戦略的に推進している。

さらに、2025 年 9 月からは、機器利用年度の翌年度から起算して 2 年間のデフォルトエンバゴ期間を経たデータセットについて、ARIM データポータル[2]における有償ライセンス方式によるデータ共有サービスを開始した[3]。

2 データ共有の概要

2.1 FAIR 原則に基づくデータ管理フレームワーク

効果的なデータ共有システムの実装には、データの発見可能性（Findability）、アクセス可能性（Accessibility）、相互運用性（Interoperability）、および再利用性（Reusability）を保証する FAIR 原則 [4] への



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



図 1 : ARIM データポータルとデータセット一覧

準拠が不可欠である。ARIM におけるデータ管理システムでは、この原則を基盤としたデータカタログの体系的作成・管理プロセスと、永続的識別子である Digital Object Identifier (DOI) の付与を運用の中で実装している。

本発表では、ARIM におけるデータカタログ作成の目的、書誌情報項目（書誌情報メタデータ）の標準化、記載要件の明確化、およびデータセット要件の体系化について述べる。加えて、データ管理の透明性向上とデータ共有促進を目的とした DOI 付与に関する ARIM 内部でのルール（内規）について述べ、ARIM が構築する研究データ管理エコシステムの全体的な設計思想と技術的実装について包括的に発表する。

2.2 データ共有に至るワークフロー設計

ARIM プラットフォームでは、参画 26 機関が保有する約 1,200 台の共用機器から生成される大規模実験データの効率的管理を実現している。これらの実験データは、物質・材料研究機構 (NIMS) がデータ基盤として整備・運用するデータ構造化システム（Research Data Express: RDE）[5] をホストプラットフォームとして活用し、各機関の保有機器に特化した AI Ready 対応データ構造化スキーマによる体系的蓄積を実現している。

RDE システムの特徴として、データの初期登録フェーズから装置固有のデータ構造化、そして装置メタデータ付与に至る統合的なタスクを一貫性のあるワークフローで実施できる点を挙げる事ができる。

データ共有は、機器利用年度の翌年度から起算して 2 年間のデフォルトエンバゴ期間終了後、ARIM データポータルより受け付けられる（図 1 参照）。データ利用申請は、機器利用申請と同様の「課題申請」システムを通じて処理され、ARIM データ共有事務局（物質・材料研究機構センタハブ）による手続き的確認を経て、申請者（または申請グループ）に対するダウンロードアクセス権の付与および有償ライセンスの形で実現される。

運用管理の観点から、データ利用期間は年度単位での設定とし、年度終了後には利用者に対してデータ利用に関する「利用報告書」の提出を義務化している。この利用報告書は、利用者に ARIM のデータ利用にかかる結果の記述を求め、本事業におけるデータ利用状況を把握し、将来の国の施策に反映させるとともに、利用を検討している将来の利用者の参考となるよう公開することを目的としている。



図 2：データカタログの構成

3 データカタログ作成

3.1 目的と基本原則

ARIM におけるデータカタログ（図 2）は、研究データの一定の品質管理と効率的なデータ共有を実現することを目指し、以下の二つの目的の達成を指向している。

- ① データの適切な管理およびデータ共有プロセスを通じた、データセット全体の品質向上および標準化の実現。
- ② DOI 付与に関する ARIM のルールを定め、一貫性のある評価基準の確立により、研究データの永続的識別と学術的な引用可能性の保証を実現。

データカタログの書誌情報項目（書誌情報メタデータ）の選定と設計においては、統合イノベーション戦略推進会議により策定された「公的資金による研究データの管理・利活用に関する基本的な考え方」（令和 3 年 4 月 27 日決定）[6]における「別紙メタデータの共通項目」への準拠することを心掛けた。この指針との整合性を確保することにより、研究データ管理・利活用に関するデータ連携にかかる適合性を維持しつ

つ、ARIM 固有の運用要件に最適化されたデータカタログ標準の確立を実現している。

3.2 データカタログの記載要件

データカタログには、次の項目が記載されていることが求められる。主要な記載要件は以下の通りである。

データセット名：データセット名は日本語で 50 文字程度以内（英語の場合は 100 文字程度以内）とし、データセットの内容を的確に表す適切なタイトルであること。なお、機器利用の利用課題名と同一である必要はなく、データセットの公開後にデータセット名を変更することも認めている。これにより、ポータルサイトからのデータの検索性と理解度を向上させる。

利用課題名：機器利用の利用課題申請時の課題名と一致していること。これにより、データセットと元の研究課題との紐付けを明確にする。

データセット登録者：機器利用の利用課題申請者であり、DICE アカウント（物質・材料研究機構がサービスを提供するデータ基盤へのアクセスするために発行されるアカウント）において

氏名及び組織名が正確に記載されていること。記載に不備がある場合は、各機関から当該利用者に通知し、修正を依頼する。これは、データ作成者を明確にし、問い合わせ先を確保するために重要である。

課題番号: JPMXP12 から始まる ARIM の課題番号 [7] とすること。これにより、ARIM 事業全体の中でのデータセットの位置付けを特定する。

データの説明 (要約): データセットの内容が理解できる日本語の場合は 200 文字以上 (英語の場合は 400 文字以上) の説明文及び代表図が添えられていること。当該内容は、既に提出されている機器利用における「利用報告書」の要約でも差し支えないものとする。これにより、データの概要を迅速に把握し、利用の可否を判断できる情報を提供する。

これらの記載要件の確認、未記入項目の補記、またはデータ登録者への追記依頼は、当該データセットのエンバーゴ解除後に各機関のスタッフが行う運用としている。また、各機関はデータ登録者に対し、記載要件を満たすための適切な案内もガイドブックなどを配布して行っている。

3.3 RDE におけるデータセット品質要件

データカタログの作成対象となる RDE におけるデータセットには、データカタログの品質を担保し、実際に共用されるデータ本体の有用性を高めるために以下の要件を定めている。

試料同定情報の完全性: 試料名に加え、組成式、化学式、分子式、化合物名等による材料同定可能情報の記載について確認し、データ再利用性の確保を実現する。

ファイル適合性: 利用課題の目的に整合した適

切なデータファイルが RDE に登録されていることを確認し、研究データとしての適合性を図る。

これらの要件を満たすことにより、ARIM のデータ基盤に蓄積されるデータについて一定の品質保証が実現され、構造化データとしての利活用が可能となる。

4 ARIM における DOI 付与の考え方

4.1 DOI 付与対象の定義と範囲

ARIM におけるデータ管理の透明性と信頼性を向上させ、データの共用と普及を促進するために、DOI 付与に関する ARIM 内でのルール (内規) を定めている。これは、データ登録者から預かった研究データについて、一定の基準を満たした場合には、永続的な識別子を付与することで、学術的な引用可能性を保証することを目的としている。

DOI の付与対象は、ARIM データポータルサイトに公開されているデータカタログとしている。これにより、データセット全体を一つのテーマ性のある研究データとして明確に識別できる。一方、データウェアハウスとして元データを管理している RDE のデータセットそのものは付与対象外とし、あくまでデータカタログに対して DOI が付与される点に留意が必要である。

さらに、上述の 3.2 節に規定するデータカタログの記載要件を満たしていない場合、もしくは 3.3 節の RDE のデータセットの要件を満たしていないデータカタログは DOI の付与対象外としている。また、匿名化が処置されているデータカタログは、DOI 付与の対象外としている。

4.2 DOI の付与申請と発行

具体的な運用として、週末までに申請された

依頼について、翌週において ARIM データ共用事務局（物質・材料研究機構センターハブ）による記載要件にかかる適合性チェックが行われ、要件を満たしているものは登録処理を実施する。これにより、効率的な DOI 発行プロセスを維持しつつ、一定のサイクルでデータの公開を支援する体制を構築している。

5 まとめ

本報告では、データの共用と普及を促進するために策定されたデータカタログ作成にかかる規定と DOI 付与に関する ARIM のルールについて説明した。データカタログ作成は、データセットの品質向上と DOI 付与に関する一貫した基準を満たすことを目的とし、政府指針のメタデータ共通項目に準拠した記載要件を定めている。これにより、ARIM データポータルを通じて公開されるデータカタログの品質と一貫性を保証し、利用者が信頼できる情報にアクセスできるようになる。

DOI 付与に関するルールは、データ管理の透明性と信頼性を高め、公開データの学術的な引用可能性を確保することを目的としている。DOI が付与されたデータカタログは、研究成果としての永続的な識別子を持ち、学術コミュニティにおける引用と評価の対象となることが期待される。

謝 辞

本取り組みは、文部科学省マテリアル先端リサーチインフラ事業 (ARIM) の支援のもとで行われた。ARIM の 26 機関のデータ業務に従事するスタッフには貴重な助言や指摘をいただいたことに深い感謝を申し上げたい。

参考文献

[1] 内閣府 “マテリアル革新力強化戦略”, 令和 3 年 4 月 27 日

[2] ARIM データポータル.

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

[3] 文部科学省プレスリリース ”大規模マテリアルデータ基盤を構築・共用開始 ～国内 26 機関連携により、科学と産業を支える知のインフラを整備～”, 令和 7 年 8 月 26 日

[4] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* “The FAIR Guiding Principles for scientific data management and stewardship”. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

[5] RDE (Research Data Express)

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

[6] 内閣府統合イノベーション戦略推進会議 “公的資金による研究データの管理・利活用に関する基本的な考え方”, 令和 3 年 4 月 27 日

[7] 内閣府 ”論文謝辞等における研究費に係る体系的番号の記載について”, 令和 2 年 1 月 14 日