

令和 5 年 6 月 22 日現在

機関番号：82108

研究種目：若手研究

研究期間：2021～2022

課題番号：21K18024

研究課題名（和文）データ駆動型研究のための研究データパッケージング手法の開発

研究課題名（英文）Development of research data packaging method for data-driven research

研究代表者

田辺 浩介（TANABE, Kosuke）

国立研究開発法人物質・材料研究機構・統合型材料開発・情報基盤部門・主任エンジニア

研究者番号：70409788

交付決定額（研究期間全体）：（直接経費） 1,400,000円

研究成果の概要（和文）：本研究では、データ駆動型研究における研究データ管理を円滑に進めることを目的として「研究データパッケージング」の手法を検証し、材料科学における研究データの研究パッケージの作成とデータリポジトリへの登録・公開を行った。研究データパッケージの作成には、汎用データパッケージフォーマットであるRO-Crateを使用した。また、パッケージ内で研究データの作成者や使用装置・計測環境などのメタデータを記述するために、簡易メタデータスキーマであるMDR Schemaを開発した。これらを用いて、実際の材料科学の実験データに対して、本研究の手法を用いた研究データ管理を行えることを示した。

研究成果の学術的意義や社会的意義

本研究で示した手法では、研究データの作成者やメタデータ記述者が独自にメタデータの記述項目を定義可能であるため、本研究での例として示した材料科学分野以外でも応用することができる。本研究データで用いたRO-crateは、zipファイルやJSONファイルなどの一般的なデータフォーマットを用いており、さまざまな方法で作成が行える。データ駆動型研究においては、分野を問わず、データの生成・保存・再利用のサイクルが高い頻度で繰り返されることが想定されるが、本研究で示した手法はそれらの研究データ管理に広く活用できると考えている。

研究成果の概要（英文）：In this study, I examined the "research data packaging" method to facilitate research data management in in data-driven research. In addition, I tested a creation of research package files for experimental data in materials science and an automatic registration to a data repository.

RO-Crate, a generic data package format, was used to create the data package. MDR Schema, a simple metadata schema was developed to describe metadata such as the creator of the research data and the equipment and measurement environment. Finally, I showed that it is possible to manage research data using these methods for actual research data in materials science.

研究分野：図書館・情報学

キーワード：研究データ管理 データパッケージング 機関リポジトリ RO-Crate オープンサイエンス

1. 研究開始当初の背景

近年の学術研究では、研究手法の発展や学術政策の観点から「オープンアクセス」や「オープンサイエンス」の取り組みが強く求められるようになってきている。このため、研究データの共有や保存を目的として、研究機関や出版社によって研究データリポジトリが多く設置され、多くの研究データがそれらの研究データリポジトリにおいて公開されるようになってきている。しかし、しかし、研究に用いられる実験・計測データは単一ファイルとして出力されるとは限らず、測定日や測定条件など、コンテキストを共有する複数のファイル群として出力される場合がある。それらの研究データファイル群を管理するにあたっては、ファイルの散逸や他の環境下で作成されたファイルの混入が発生するといった事故が発生する可能性があり、ファイル群で一貫性を保った取り扱いを行うには注意を要する。

本研究では、実験データや計測データの実体を構成する複数のファイル群と、そのデータの性質を記述したメタデータファイルを単一のファイルにアーカイブする「研究データパッケージング」の手法を用いて、データ駆動に適した研究データ管理の手法の提案を行った。

2. 研究の目的

本研究では、研究データリポジトリへの研究データの登録において、複数の研究分野、特に材料科学を対象とした研究データの流通・再利用に必要な研究データ管理を可能とする「研究データパッケージング」の手法を検証・確立し、データ駆動型研究を円滑に進めるための枠組みを構築することを目的とした。

材料科学では物理的な試料とその計測、計算科学的アプローチをはじめ、データの性質により異なる観点による記述が必要とされるため、研究データの作成者やメタデータの記述者が、複数のメタデータスキーマを用いてメタデータを記述できる必要がある。このため、本研究で扱うデータパッケージファイルは、以下の特性を備えるものとした。

- 材料科学の研究データの内容を柔軟に記述できること
- 研究データリポジトリへのデータパッケージファイルの一括・大量の自動登録が行えること

3. 研究の方法

本研究では、複数の分野でのデータパッケージの作成と、データパッケージの研究データリポジトリへの登録・検索についてそれぞれ検証を行い、あわせてそれらの検証に必要なアプリケーションプログラムの開発を行った。

本研究ではまず、研究代表者の所属組織のデータリポジトリである Materials Data Repository (MDR, <https://mdr.nims.go.jp>)上で公開されている実験データのファイルを対象として、研究データパッケージの作成を行った。研究データパッケージングのフォーマットには RO-Crate (Research Object Crate, <https://www.researchobject.org/ro-crate/>)を用いた。RO-Crate は、汎用的な研究データパッケージングのフォーマットとして開発されているフォーマットであり、2019年にバージョン 1.0、2023年にバージョン 1.1 が公開されている。RO-Crate は、パッケージ自身のメタデータの記述に JSON-LD (JSON for Linked Data) を用いるようになってきているが、これとは別に研究分野固有のメタデータの記述を行うファイルを含めることもできる仕様となっているため、求める要件を満たすと判断した。

作成した研究データパッケージには、パッケージの作成対象である実験データのファイルに加えて、データ登録者やデータ名称のようなデータ流通に関するメタデータ、ならびに実験の測定値などの材料関係のメタデータを含んでいる。また、同データリポジトリに対して、RO-Crate パッケージファイルのインポート機能の追加開発を行い、リポジトリの試験環境において前述のパッケージファイルの登録を行えることを確認した。

一方、材料科学分野における研究データのメタデータ記述にあたっては、すでに MDR に材料研究データを登録している研究者に対してヒアリングを行い、簡易メタデータスキーマである MDR Schema (<https://doi.org/10.48505/nims.3239>)を新たに作成した。このスキーマに準拠した YAML ファイルを RO-Crate のパッケージ内に含めることで、パッケージの作成時に必要な材料データの記述を容易に行えるようにした。本研究での研究データパッケージの構成図を図 1 に示す。

これらの検証をもとに、RO-Crate と MDR Schema を用いて、材料研究データのメタデータの記述と研究データファイルの登録を行うことのできる Web アプリケーションを開発した。同アプリケーションから出力される研究データパッケージファイルを用いて、MDR に対して新たに約 1 万件のデータ登録を行い、研究者によるメタデータの記述、ならびに材料メタデータと研究データファイルの登録と検索が問題なく行えることを確認した。

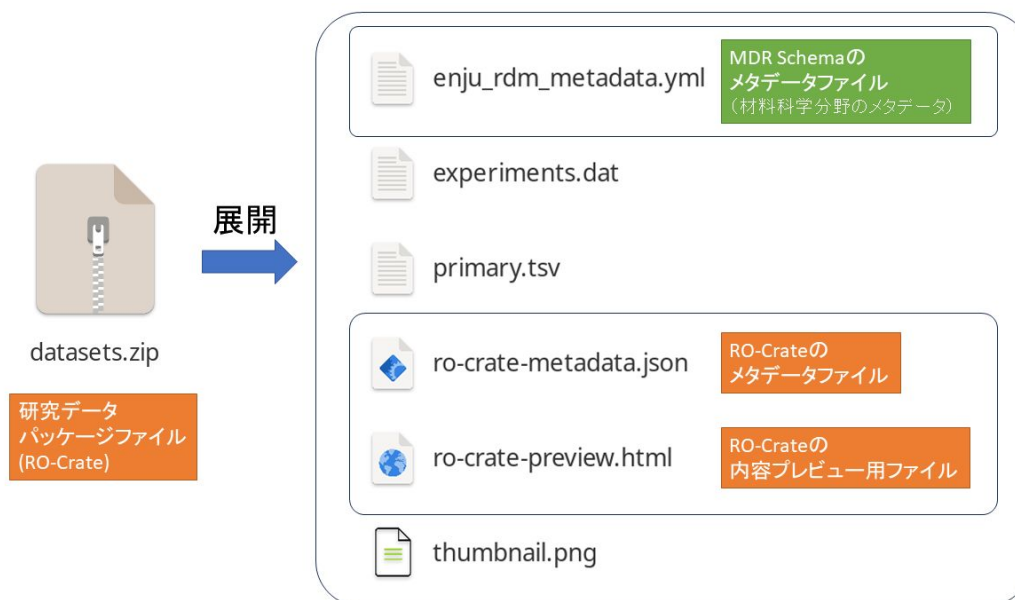


図 1 本研究での研究データパッケージの構成

4. 研究成果

本研究によって開発されたアプリケーションを用いた材料科学分野の研究データの登録事例として、「MDR XAFS DB」(<https://doi.org/10.48505/nims.1447>)が挙げられる。これは、放射光の測定データを収録しているデータベースであり、2023年6月現在およそ2700件のデータセットが収録されている。

本研究では、研究データパッケージの作成からデータリポジトリへの登録・公開の自動化を最終的な目標としているが、研究パッケージデータの作成や公開においては、メタデータの内容のチェックや公開の承認など、研究データの流通に関わるワークフローの詳細化が必要となる。今後はこれらのワークフローに関するメタデータの記述形式を確立し、そのメタデータを研究データパッケージ内に含めることを想定している。また、今回の事例を通して、研究データパッケージの手法の材料科学分野以外での応用について提案していきたい。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 田辺 浩介, 松田 朝彦	4. 巻 31
2. 論文標題 RO-Crateを用いた材料研究データのデータリポジトリへの登録	5. 発行年 2021年
3. 雑誌名 情報知識学会誌	6. 最初と最後の頁 497 ~ 502
掲載論文のDOI (デジタルオブジェクト識別子) 10.2964/jsik_2021_067	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Ishii Masashi, Tanabe Kosuke, Matsuda Asahiko, Ofuchi Hironori, Matsumoto Takahiro, Yaji Toyonari, Inada Yasuhiro, Nitani Hiroaki, Kimura Masao, Asakura Kiyotaka	4. 巻 3
2. 論文標題 Integration of X-ray absorption fine structure databases for data-driven materials science	5. 発行年 2023年
3. 雑誌名 Science and Technology of Advanced Materials: Methods	6. 最初と最後の頁 2197518
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/27660400.2023.2197518	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 田辺浩介, 松田朝彦
2. 発表標題 RO-Crateを用いた材料研究データのデータリポジトリへの登録
3. 学会等名 情報知識学会 第26回情報知識学フォーラム
4. 発表年 2021年

1. 発表者名 田辺 浩介
2. 発表標題 MDR Schemaの紹介と、その作成過程で得た教訓
3. 学会等名 RDUF総会・公開シンポジウム
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

MDR Schema
<https://doi.org/10.48505/nims.3239>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------