



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

2021年12月18日 第26回情報知識学フォーラム

田辺浩介  <https://orcid.org/0000-0002-9986-7223>

松田朝彦  <https://orcid.org/0000-0001-5989-027X>

物質・材料研究機構 統合型材料開発・情報基盤部門

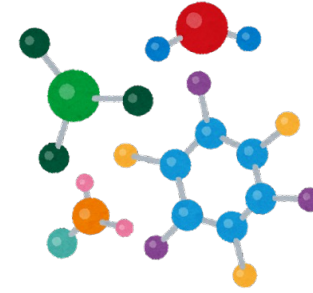
材料データプラットフォームセンター





研究の背景

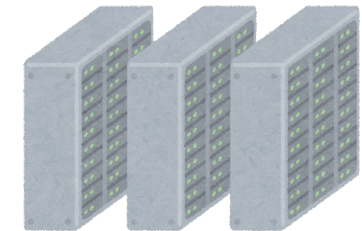
- 材料研究は複数の分野にまたがっており、データの形式や内容が多種多様
- データに付与するメタデータスキーマの設計や記述が困難
- 複数のデータファイルを、データ生成時のディレクトリ構造を保ったまま保存・共有する必要がある
- データ駆動型研究を実現するために、登録・公開処理の自動化が必要



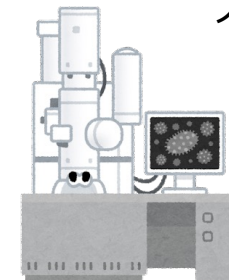
化学



物理



大規模データ



実験



計算





- これらの課題を解決する方法として、**RO-Crate** に着目
- 研究データパッケージの規格のひとつ
- Research Object communityによる開発
- 2019年11月にバージョン1.0がリリース
 - 最新バージョンは2021年2月リリースの1.1



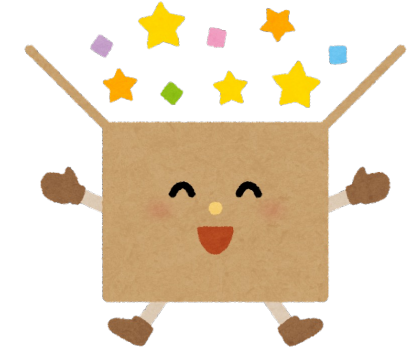
<https://www.researchobject.org/ro-crate/>





研究データパッケージとは

- なんらかの文脈を共有する研究データを、ひとかたまり（パッケージ）として扱う手法
- 類似の手法は、分野や用途によって古くから存在する
 - 科学計算分野の[HDF5](#)や、データ交換用の[BagIt](#)など
 - Officeファイル (Office Open XML Document) や EPUB もパッケージの一種
- RO-Crateは軽量・汎用の研究データパッケージングフォーマットとして開発





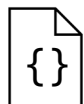
(ZIPファイルでもディレクトリでも可)



root directory: /

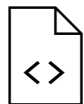


RO-Crate



ro-crate-metadata.json

パッケージのメタデータ。JSON-LD。必須。
データセット自体とペイロードについて記述。



ro-crate-preview.html

人が参照するためのHTMLファイル。任意。



ro-crate-preview_files/

上記HTML用画像等。任意。



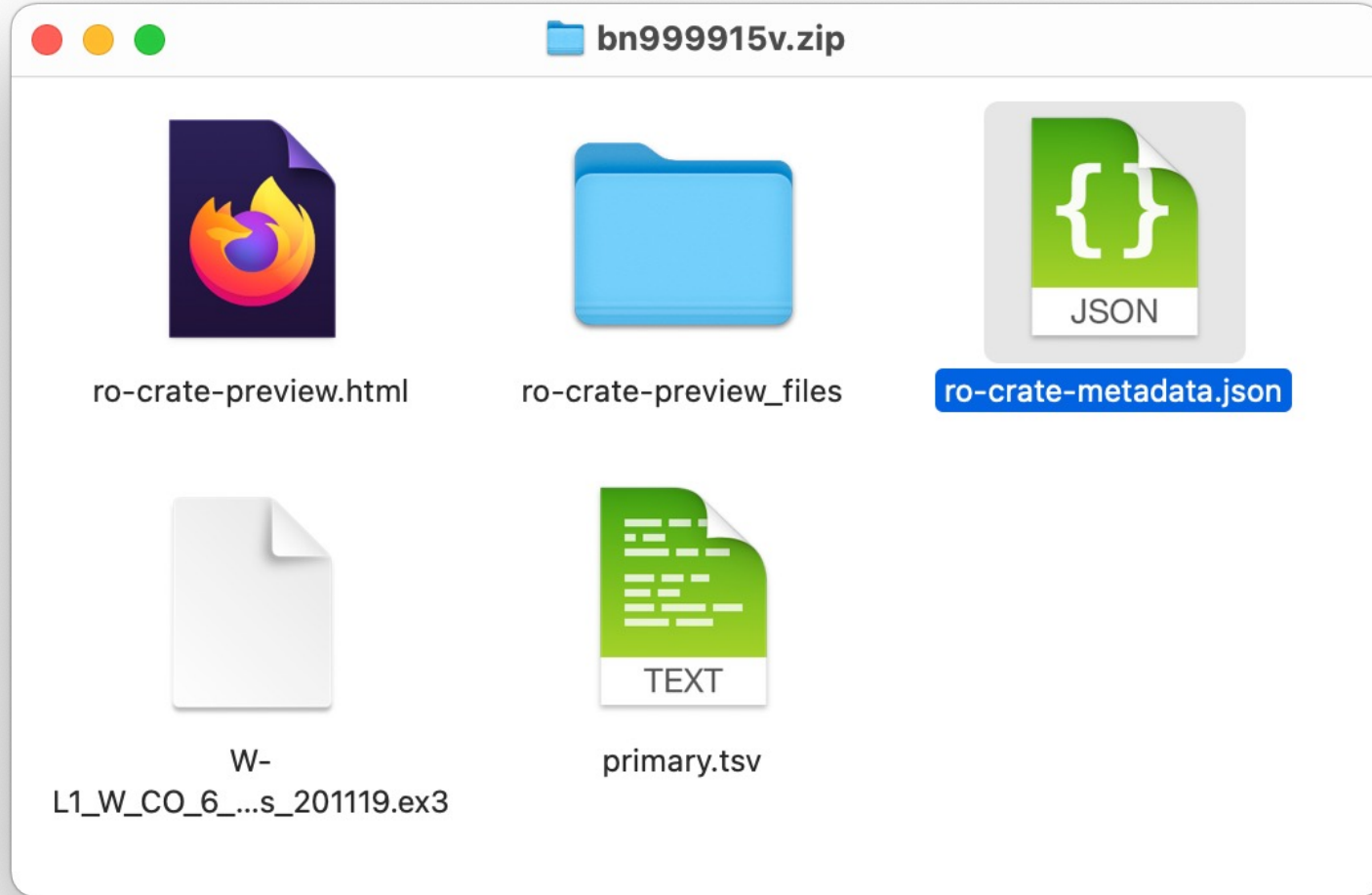
ペイロード

研究データ。複数可。





RO-Crateパッケージのイメージ





データ登録の実施



- 研究プロジェクトのひとつから約1500件の材料データを用意
 - 石井ほか、[第24回XAFS討論会](#)（日本XAFS研究会）
- RO-Crateのデータ登録用スクリプトを作成し、上記のデータを [Materials Data Repository](#) の試験環境に登録
- ro-crate-metadata.jsonに、データセット自体のメタデータと材料メタデータを記述
 - 材料メタデータの記述語彙として、[Bioschemas](#)と[Materials Schemas](#)を使用





ro-crate-metadata.jsonの各パートの説明（抜粋）



```
{
  "@context": [ "https://w3id.org/ro/crate/1.1/context", { "bio": "http://schema.org" } ],
  "@graph": [
    {
      "@type": "CreativeWork",
      "@id": "ro-crate-metadata.json",
      "conformsTo": { "@id": "https://w3id.org/ro/crate/1.1" },
      "about": { "@id": "." }
    },
    {
      "@id": ".",
      "@type": "Dataset",
      "url": "https://mdr.nims.go.jp/concern/datasets/bn99991",
      "identifier": "https://doi.org/10.48505/nims.2563",
      "name": "XAFS spectrum of Tungsten carbonyl",
      "datePublished": "2021",
      "description": "This dataset consists of X-ray absorption fine structure (XAFS) spectra at W L21-edge of Tungsten carbonyl measured at SPring-8 BL14B2",
      "hasPart": [
        { "@id": "W-L1_W(CO)6_Si311_50ms_201119.ex3" },
        { "@id": "primary.tsv" }
      ],
      "material": [ { "@id": "#material-0" } ]
    }
  ],
}
```

パッケージの情報

データセット
全体の情報

材料科学固有の情報

```
{
  "@id": "#material-0",
  "@type": "Material",
  "name": "Tungsten carbonyl",
  "description": "Standard Sample",
  "bio:ChemicalSubstance": {
    "bio:chemicalComposition": "W(CO)6",
    "identifier": "https://matvoc.nims.go.jp/wiki/Item:Q1765"
  }
},
{
  "@id": "W-L1_W(CO)6_Si311_50ms_201119.ex3",
  "@type": "File",
  "encodingFormat": "text/plain",
  "url": "https://mdr.nims.go.jp/downloads/73666690b",
  "mainEntityOfPage": { "@id": "https://mdr.nims.go.jp/concern/parent/bn999915v/file\_sets/73666690b" }
},
{
  "@id": "primary.tsv",
  "@type": "File",
  "encodingFormat": "text/tab-separated-values",
  "url": "https://mdr.nims.go.jp/downloads/k930c032p",
  "mainEntityOfPage": { "@id": "https://mdr.nims.go.jp/concern/parent/bn999915v/file\_sets/k930c032p" }
}
]
```

試料・対象物質

実験データ

独自メタデータ





- 研究データ作成時の文脈（ディレクトリ構造など）をそのまま保存可能
- RDFによる柔軟なメタデータ記述
- HTMLプレビューファイルで人間向けの可読性も確保
- **メタデータのファイル名が“ro-crate-metadata.json”固定**なので、アプリケーションでのメタデータ読み取りの実装が容易
- 研究コミュニティによるオープンな規格、かつ静的ファイルなのでアーカイブ化・長期保存にも有利





- 他の研究プロジェクトのデータセットの登録
- 材料メタデータの記述の拡充
- 研究データ公開ワークフローでの適用

