

Title	データ管理計画の国際標準モデルと各分野における研究データライフサイクルとの比較分析
Author(s)	南山, 泰之; 林, 和弘
Citation	年次学術大会講演要旨集, 39: 735-739
Issue Date	2024-10-26
Type	Conference Paper
Text version	publisher
URL	<a href="http://hdl.handle.net/10119/19613">http://hdl.handle.net/10119/19613</a>
Rights	本著作物は研究・イノベーション学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Research Policy and Innovation Management.
Description	一般講演要旨

## データ管理計画の国際標準モデルと各分野における研究データライフサイクルとの比較分析

○南山泰之, 林和弘 (科学技術・学術政策研究所)

yasuyuki.minamiyama@gmail.com

### 1. はじめに

近年、諸外国のオープンサイエンス政策において、データ管理計画 (DMP) の義務化が進められている[1]。DMP は、研究データをそのライフサイクル全体にわたってどのように取り扱うかを定めた文書、と定義される[2]。その性質上、DMP は研究の開始前に作成されるのみならず、研究活動の進展に応じて随時更新されることが推奨されており、研究者は DMP の管理を通じて円滑な研究データ管理を実現することが可能になるとされる。

このような理解のもと、国内外の助成機関では DMP のテンプレートを作成し、研究者による DMP の作成を支援している[3]。また、この動きをさらに押し進めるものとして、DMP に記述された情報を機械的に交換し、利用することを可能にする machine actionable DMP (maDMP) [4] の開発が進められている。maDMP は、国際的な研究データ関連コミュニティである研究データ同盟 (RDA) を中心として議論が展開され、共通データモデルの策定やプロトタイプシステムの開発などが進められている[5]。

一方で、現在の maDMP に関する議論は主に助成機関の視点からのみなされており、各研究分野の現場で取り扱われる情報との対応関係は明らかになっていない。そこで本研究では、RDA が策定を進める maDMP のデータモデルを念頭に、いくつかの分野において定められている研究データのライフサイクルとの違いを分析する。さらに、「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」におけるメタデータの共通項目との比較を通じて、maDMP を国内における研究の現場へ実装していく際の課題について考察する。

### 2. maDMP の概要

maDMP とは、DMP を機械可読形式で記述し、利害関係者間でデータを効率的に交換・共有できるようにしたものである。maDMP の目的は、従来の DMP を単なる「人が読んで利用する文書」から「機械も活用できる動的な文書」へと変える

ことにあり、これによってデータの管理・監視プロセスの自動化を企図している。初期のコンセプトは RDA の Active Data Management Plans IG の有志によって提唱され、その後同メンバーらによって「maDMPs のための 10 原則」の形で整理されている[6]。同原則では、研究データエコシステムのすべての利害関係者のワークフローに DMP を組み入れることを始めとし、機械による利用可能性を高める個々の取り組みについて詳細に記述している。

maDMP をアプリケーションのレベルに落とし込む取り組みとしては、RDA のワーキンググループとして 2019 年に発足した DMP Common Standard WG の活動がある。同 WG では、maDMP の具体的なユースケース分析、及び同ユースケースを実現するためのデータモデル策定を実施している[7]。図 1 に同 WG が分析した maDMP のユースケース例を示す。

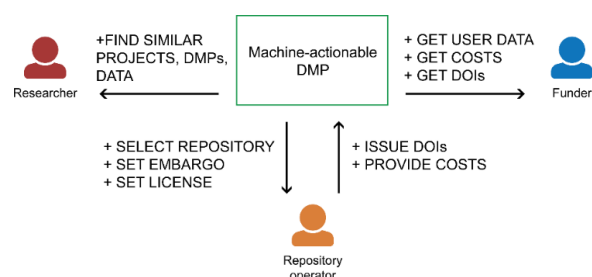


図 1 RDA DMP Common Standard WG による maDMP のユースケース例

想定するステークホルダーは研究者、支援者、リポジトリ管理者、助成機関であり、プロジェクトやデータの発見、リポジトリの選択、DOI の発行、コスト情報の共有など幅広いユースケースをカバーすることを志向している。さらに、このユースケースを実現するためのデータモデルとして、RDA DMP common standard が策定されている (図 2)。

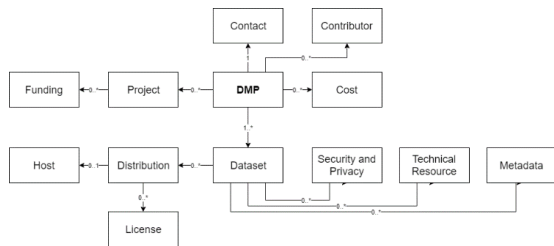


図2 RDA DMP common standard WG のデータモデル

RDA DMP common standard は、情報システム間の基本的な相互運用性を提供するためのメタデータアプリケーションプロファイルとして策定され、DMP に関する一般的な情報を提供する DMP クラスを中心に、データセット、プロジェクト、コスト、貢献者、連絡先などの情報がリンクされている。アプリケーション開発者は、

RDA DMP common standard に互換性のあるデータモデルを採用することにより、複数のアプリケーション間における情報交換のコストを下げることができる。RDA DMP common standard は2021年時点で10のサービスが実装済み、あるいは検討中とのことである。

### 3. 分析

本章では、2章で紹介した RDA DMP common standard を念頭に、いくつかの分野における研究データライフサイクルとの違いを分析する。DMP の構想が登場する以前から、各研究分野では研究データのライフサイクルを定め、円滑な研究データ管理を実現してきている。表1に代表的な研究コミュニティが採用している研究データのライフサイクル[8]を示す。

表1 代表的な研究コミュニティが採用している研究データのライフサイクル

Fields	Earth Sciences		Life sciences	Social sciences		Humanities/Linguistics	Multiple		
Institutions	USGS	DataONE	EMBL	UK Data Archive	ICPSR	CLARIN-NL	DPCVocab	Data Curation Network	DCC
Steps	Plan	Plan	Collecting	Transfer of data	Proposal development and data management plan	Identification and assessment	Ingest	Ingest	Conceptualise
	Acquire	Collect	integrating	Assigning processing standard	Project start-up	Development of a curation plan	Representation	Appraise/ Accept	Create or receive
	Process	Assure	processing	Data processing	Data collection and file creation	Curation	Provenance management	Curate	Appraise & select
	Analyze	Describe	analyzing	Documentation processing	Data analysis	Validation	Systems management	Access	Dispose
	Preserve	Preserve	storing	Metadata creation	Preparing data for sharing	Archiving	Data storage	Preserve	Ingest
	Publish/Share	Discover	sharing	Additional user information	Depositing data		Policies		Preservation action
		Integrate Analyze	publishing finding	Publishing data	Delivering data		Preservation Public access provision		Store
				Preserving data					Access, use & reuse
									Transform

“Field”の行には対応する研究分野名, “Institution”の行には団体名, “Steps”の行には各団体が採用している研究データ管理のステップを示している。いくつかのステップには RDA DMP common standard と一部共通する用語も散見されるが, 各ステップは研究の進展に沿って実施すべき手順や戦略を含んだ形で設計されており [9], 研究データ管理上の行動そのものも記録の対象となる。この点において, ある時点におけるデータ管理計画書をモデル化している RDA DMP common standard とは構造上のギャップがある。また, 各ステップの設計は同じ分野内であっても異なっており, 分野に共通するデータ管理手法は確立されていないことが示唆されている。

#### 4. 考察

本章では, maDMP のコンセプトを国内における研究の現場へ実装していくための課題について考察する。既に見たように, 各研究分野のライフサイクルと国際的な DMP 項目とは構造上のギ

ャップが存在しており, DMP によって研究データのライフサイクル全体を記述するための標準的なプロトコルは管見の限り存在しない。また, 各研究分野のライフサイクルには多様性があり, 分野に共通するデータ管理手法は確立されていないことが示唆されている。政策的な観点からは, 当面の課題として国際標準モデルへの準拠を見据えつつ, 中長期的には研究データ管理活動のモデル化に関する試みを繰り返す必要がある。

上記のような理解のもと, ここでは RDA DMP common standard と「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」におけるメタデータの共通項目 (以下「政府共通メタデータ項目」という。) [10] との比較を通じて, 当面の課題である国際標準モデルへの準拠について考察する。表 2 に両者の項目別マッピングを示す。なお, 政府共通メタデータ項目には構文規則 (Syntax Encoding) が設定されていないため, 項目を構成する要素レベルでのマッピングは行っていない。

表 2 RDA DMP Common Standard と政府共通メタデータ項目のマッピング

名称	RDA DMP Common Standard		政府共通メタデータ項目における対応箇所
項目名	DMP	DMP のタイトル, 更新日など, DMP に関する俯瞰的な情報	資金配分機関情報
	Contact	DMP に関する情報を提供できる者の情報。DMP の作成者である必要はなく, 個人でも組織でもよい	データ管理者の連絡先
	Cost	データ管理に関する費用の一覧	--
	Project	DMP に関連するプロジェクトの情報。資金提供プロジェクトだけでなく, 内部プロジェクト, 博士論文などを含む	体系的番号/プロジェクト名
	Funding	資金提供されたプロジェクトの詳細情報	体系的番号におけるプログラム情報コード/プログラム名
	Contributor	本 DMP に記載されているデータマネジメントのプロセスに関わるすべての関係者, および DMP 自体の作成・管理に関わる関係者の情報	データ作成者/データ作成者の e-Rad 研究者番号
			データ管理機関/データ管理機関コード
			データ管理者/データ管理者の e-Rad 研究者番号
Dataset	データに関する俯瞰的な情報。データセットの品質, 配布, 保存, 性質, 生成過程など多様な情報を含む	データ No./データの名称/データの説明/データの分野/データ種別/掲載日・掲載更新日	
Technical Resource	顕微鏡など, データを作成または処理するために必要/使用される機器の情報	--	

Metadata	データを記述するために使われるメタデータ標準に関する情報。データセットに関する実際のメタデータは含まれない	--
Distribution	データセットの配布に関する情報。データサイズ、アクセス権、公開日、ファイルフォーマット等に関する情報を含む	アクセス権／公開予定日／概略データ量
License	配布されるデータセットごとに設定されるライセンス情報。エンバーゴ期間を含む	管理対象データの利活用・提供方針／アクセス権
Host	データが保存されているシステムに関する情報。バックアップの頻度、認証情報、採用している技術（PID システムなど）などの情報を含む	リポジトリ情報／リポジトリ URL・DOI リンク
Security and Privacy	特定のデータセットのセキュリティとプライバシーに関する情報	--

※RDA DMP common standard の項目説明は、ウェブサイトの記述を元に著者が作成。

政府共通メタデータ項目で定義された項目は、現在のところ全て RDA DMP common standard でカバーされており、概ね互換性を保った形で準拠することが可能である。RDA DMP common standard 特有の項目としては、Cost, Technical Resource, Metadata, Security and Privacy の4つがあり、それぞれデータ管理にかかるコスト情報、データの匿名化ステータス、メタデータ標準の特定、実験機器に関する情報の共有といったユースケースに対応可能である。また、Distribution にはデータのファイルフォーマット情報や品質に関する情報を記述する項目がある、Host にはデータのバックアップ体制や採用している PID システム情報、取得した認証情報を記述する項目が用意されているなど、既存のユースケースを詳細化した要求に応えることが可能である。

## 5. 結論

本研究では、maDMP の国際標準である RDA DMP common standard を念頭に、いくつかの分野における研究データライフサイクルとの違いを分析した。分析の結果、各研究分野のライフサイクルと RDA DMP common standard の間に存在する構造上のギャップ、及び各研究分野のライフサイクルの多様性が明らかになった。さらに、RDA DMP common standard と政府共通メタデータ項目との比較を通じて、maDMP のコンセプトを国内における研究の現場へ実装していくための課題を考察した。同考察からは、RDA DMP common standard への準拠にあたり互換性の問題は存在せず、より多くのユースケースに対応できる可能性が示唆された。

残された課題としては、中長期的な視点から研究データ管理活動のモデル化に関する試みを繰り返し、研究データライフサイクルと maDMP の

モデルを統合していくことが挙げられる。研究データ管理活動のモデル化に該当する取り組みとしては、欧州委員会 (European Commission) がデータポータルのために導入した DCAT-AP[11]のほか、欧州 OpenAIRE のガイドラインにおける記述例[12]、国立情報学研究所が研究データ基盤間のコミュニケーションハブとして開発した RDM オントロジー[13]などの取り組みがある。特に、RDM オントロジーは DMP における研究データ管理上の行動に関する情報を機械実行可能な形で表現する試みも進められており[14]、注目に値するだろう。引き続き、DMP の作成や更新作業の機械化、さらには自動化に向けた研究の動向を注視していきたい。

## 謝辞

本研究の遂行に当たっては、平木俊幸氏、横山重俊氏 (国立情報学研究所) から有益な助言をいただいた。ここに記して感謝の意を表したい。本研究は JSPS 科研費 JP24K03229 の助成を受けたものである。

## 参考文献

- [1] 内閣府, バイオデータ連携・利活用に向けたガイドブック, 47-49 (2023), [https://www8.cao.go.jp/cstp/bio/data\\_renkei\\_2.pdf](https://www8.cao.go.jp/cstp/bio/data_renkei_2.pdf)
- [2] European Commission, **Turning Fair Into reality: Final report and action plan from the European commission expert group on FAIR data**, 76 (2018), <https://doi.org/10.2777/1524>
- [3] Gajbe, S. B., et al., Evaluation and analysis of Data Management Plan tools: A parametric approach, **Information Processing & Management**, 58(3), 102480 (2021),

<https://doi.org/10.1016/j.ipm.2020.102480>  
[4] Simms, S., et al., Machine-actionable data management plans (maDMPs), **Research Ideas and Outcomes**, **3**, e13086 (2017),  
<https://doi.org/10.3897/rio.3.e13086>  
[5] Walk, P., Miksa, T., & Neish, P., **RDA DMP Common Standard for Machine-actionable Data Management Plans**, Research Data Alliance, (2019),  
<http://doi.org/10.15497/rda00039>  
[6] Miksa, T., et al., Ten principles for machine-actionable data management plans, **PLOS Computational Biology**, **15**(3), e1006750 (2019),  
<https://doi.org/10.1371/journal.pcbi.1006750>  
[7] Miksa, T., et al., Application Profile for Machine-Actionable Data Management Plans, **Data Science Journal**, **20**(1), 1-32 (2021),  
<https://doi.org/10.5334/dsj-2021-032>  
[8] Minamiyama, Y., et al., A study on formalizing the knowledge of data curation activities across different fields, **PLOS ONE**, **19**(4), e0301772 (2024),  
<https://doi.org/10.1371/journal.pone.0301772>  
[9] Gupta, S., & Müller-Birn, C., A study of e-Research and its relation with research data life cycle: a literature perspective, **Benchmarking: An International Journal**, **25**(6), 1656-1680 (2018),  
<https://doi.org/10.1108/BIJ-02-2017-0030>  
[10] 内閣府, 「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」におけるメタデータの共通項目, (2024),  
[https://www8.cao.go.jp/cstp/common\\_metadata\\_elements.pdf](https://www8.cao.go.jp/cstp/common_metadata_elements.pdf)  
[11] European Commission, **DCAT Application Profile for data portals in Europe**, (2024),  
<https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe>  
[12] OpenAIRE, **OpenAIRE Guidelines for Data Archives. Application Profile Overview**, (2022),  
[https://guidelines.openaire.eu/en/latest/data/application\\_profile.html](https://guidelines.openaire.eu/en/latest/data/application_profile.html)  
[13] Research Center for Open Science and data platform (RCOS), **The Research Data Management ontology**, (2024),  
<https://purl.org/rdm/ontology/>  
[14] 平木俊幸ほか, データ管理計画における研究データ管理上の行動に関する部分の機械行動可能な形式での表現可能性の検討, **研究報告インターネットと運用技術 (IOT)**, **2024-IOT-64**(78), 1-6 (2024),

<http://id.nii.ac.jp/1001/00232984/>