

CA1983

動向レビュー

Machine-actionable DMPs (maDMPs) の動向

つねかわ まお
常川真央*

1. はじめに

2011年以降、オープンサイエンスや研究公正の観点から、世界中の研究助成機関が研究者に対して研究データの管理・公開に向けた計画書であるデータマネジメントプラン(DMP)の提出を求めている⁽¹⁾。さらに、2016年より Machine-actionable DMPs (maDMPs) というコンセプトが登場し、DMPを研究データの管理・公開を担うシステムで活用するという取り組みが進められている。リポジトリや研究データ管理システムの運用を担うライブラリアンにとって、DMPは研究データの円滑な管理・公開を行う重要な情報源となり得る。そこで本稿では、maDMPsの概要を説明すると共に、その意義や動向について紹介する。

2. DMPが求められる背景

研究データが適切に公開されるためには、研究データの生成から公開に至るまでの「データライフサイクル」を適切に把握することが重要である。例えばライブラリアンが機関リポジトリに研究データを集約しようとするならば、自機関の中で公開すべきデータがあるかどうかを把握する必要がある。そして、研究データを登録・公開する場合、そのデータを公開してもよいか、あるいは個人情報などのセンシティブな情報が入っていないかを確認する必要がある。このように、一口に研究データを公開するといっても様々なタスクが発生し、その解決にはデータライフサイクル全体を見なければならぬ。

そこで、DMPがデータライフサイクルを支える存在として注目されている。DMPとは、「研究中ならびに研究終了後の両方でいかに研究データが取り扱われるかを概説した公的なドキュメント」である⁽²⁾。DMPには、研究の過程で生み出されるデータのうち、何を公開して何を秘匿するか、データにセンシティブな情報が含まれるか否か、もし含まれるなら、どのような対策をとるのか、公開に向けてどのような取り組みを行うのかといった事項を一般的には明記する。このような情報は、研究助成機関のマネジメントや学術組織のコンプライアンスのためだけでなく、ライブラリアンが円滑に研究データを公開する上で発生す

る様々なタスクを解決するための情報源となり得る。

DMPは、米国では2003年に米国国立衛生研究所(NIH)が研究データ共有の計画として求めたのを契機として広まり、2011年に全米科学技術財団(NSF)が提出を義務化するようになり浸透した⁽³⁾。欧州では2017年に、オープンサイエンスを志向した科学技術政策である Horizon2020において、研究助成を受けた研究プロジェクトの一部は FAIR原則に沿ったデータ公開の取り組みに関する DMPの提出が義務付けられるようになった⁽⁴⁾。

現在では国際的に、研究者が助成申請する際に DMPを要求されることが浸透しつつある。国内においても、科学技術振興機構(JST)をはじめとして主要な研究助成機関が採択プロジェクトの一部に対して DMPの提出を義務付けている⁽⁵⁾。さらに、2020年度からは日本学術振興会(JSPS)が科学研究費助成事業の一部において DMPの提出を応募の要件とする予定である⁽⁶⁾。

DMPを研究者が記述し、学術組織がこれを支援することは研究推進の観点からもコンプライアンス対応としても重要である。そこで DMPtool⁽⁷⁾ や DMPonline⁽⁸⁾、両サービスのコードベースを統合したオープンソースソフトウェア開発プロジェクトの DMProadmap⁽⁹⁾ など、DMP作成を支援するツール開発の取り組みが進められてきた。

しかし、データライフサイクルの様々なワークフローにおいて DMPを情報源として活用するという方向性での取り組みは、豪州などで実用化の例は見られるものの国際的な標準規格や技術が確立されておらず、検討が進められている。その一つとして現在、研究データ同盟(RDA:CA1875参照)を中心として活発に議論され、実現に向けて取り組みが行われているコンセプトに maDMPsがある。

3. maDMPsの定義と原則

maDMPsとは、研究データ管理を支援する機械またはソフトウェアが計画に沿って適切にプログラムおよび動作するように記述された DMPである。2016年9月に開催された RDA 第8回年次大会の Active Data Management Plans IGのセッションにて、ウィーン工科大学(オーストリア)のミクサ(Tomasz Miksa)氏らによって提唱された。その後、翌年にホワイトペーパーが発行され、そのコンセプトが明文化された⁽¹⁰⁾。

maDMPsにおける“machine-actionable”とは、機械可読性(Machine-readable)と類似した概念ではあるが、求められる要件はより広く、高度である。社会科学、行動科学、経済科学のデータを記述するための国際規格の作成に取り組むデータドキュメンテーション

*国立情報学研究所オープンサイエンス基盤研究センター

ン・イニシアティブ(DDI)のウェブサイトによれば、machine-actionableは、「機械またはコンピューターがその構造に従ってプログラミングされることができるように一貫性のある方法で構造化された情報」と定義されている⁽¹¹⁾。つまり、機械が情報を読み取れるだけでなく、その意味と構造を解釈して自動的に動作することを求めている。maDMPsは単なるメタデータフォーマットの範疇を超えたコンセプトである。

それでは、maDMPsを実現するには具体的にどのようなシステムの開発や導入が必要なのだろうか。ミクサ氏らが2019年に出版した「maDMPsのための10原則」によれば、maDMPsを実現する上で重要な原則として次の10項目を挙げている⁽¹²⁾。

- (1) 研究データのエコシステムに関わるすべての利害関係者のワークフローにDMPを組み入れる
- (2) 自動化されたシステムが利害関係者に代わって機能することを可能にする
- (3) 人々のためだけでなく、機械のために(も)ポリシーを作る
- (4) (機械と人間の両方に向けて)データ管理のエコシステムを構成する要素を記述する
- (5) 永続的識別子(PID)と統制語彙を使用する
- (6) maDMPsのための共通データモデルに従う
- (7) DMPを人と機械の両方で使用できるようにする
- (8) データ管理の評価とモニタリングを支援する
- (9) DMPを更新可能で、生きた、バージョン管理された文書とする
- (10) DMPを公開可能にする

このように、maDMPsであるということは、研究データを管理・公開する環境そのものを規定することなのである。

ミクサ氏らは、maDMPsを実現するには、以上の10原則を必ずしも満たす必要はないとしている。しかし、それぞれの原則同士は密接に関連している。研究データを管理する際には、多様な研究ツールやシステムが複雑に絡み合っていることが多い。そのような環境下でmachine-actionableであるには、DMPの構造に従って動作するだけでなく、他のシステムと協調して、研究者が定めたデータポリシーやプランと実態との間に矛盾が起きないように動作することが求められる。協調するシステム間が密に結合している状況では、一部のシステムやツールが更新または入れ替えられる度に環境の大幅な見直しが必要となり、DMPとの整合性を保つための多大なメンテナンスコストが発生する。そのため、DMPに基づく研究データ管理環境を作り出すには、相互運用性(Interoperability)の確

保が重要である。一般的には、相互運用性の向上には、システム間で共通に利用されるAPIが策定されることが重要である。maDMPsは、単なるメタデータの問題に留まらず、相互運用性を実現するような永続的識別子やAPIなどをいかに取り決め、広く共用されるようにするかという問題まで及ぶコンセプトなのである。

4. maDMPsの実現によるアウトカム

maDMPsの実現は、データライフサイクルを支える様々な関係者のタスクをシステムによって解決することを可能にする。それでは、ライブラリアンなどリポジトリの運用を担当する立場からは、どのようなアウトカムが期待できるだろうか。maDMPsのホワイトペーパーでは、リポジトリに関係するアウトカムとして、「リポジトリの推薦機能」や「研究データのアーカイブまたは保全のプロセスの開始」を例示している⁽¹³⁾。

リポジトリの推薦機能とは、DMPを研究者が作成する過程で、研究データの公開先として適したリポジトリをDMP作成支援システムが自動的に推薦するような機能である。研究データの公開先の決定にあたっては、研究者の専門分野や、データの種類、または研究助成機関からの要求を考慮する必要がある。これまでは、研究者に対する啓発活動やコンサルティング、または研究者コミュニティの慣習に依存しており、効果は限定的であった。また、研究データの特性に応じた適切な公開先を選択するには、リポジトリ担当者によるDMPの理解やヒアリングが必要になり、多大な負荷がかかる。maDMPsに対応したDMP作成支援システムは、作成の段階から研究データの特性を機械的に把握できるため、入力に応じた研究データの公開先のサジェストが可能になる。このように、要求に応じて研究者が利用可能なリポジトリをフィルタリングし、適切なリポジトリの利用へ容易に誘導することができる。これは、研究者とリポジトリ担当者双方にとっての負担軽減と公開データの増加につながるであろう。

こうして作成されたDMPは、研究終了後にデータをいかに公開・保全するかのポリシーを確認する情報源となる。「研究データのアーカイブまたは保全のプロセスの開始」のユースケースでは、こうした情報が機械可読になっていることで、リポジトリが自動的にDMPの中から登録すべきデータの存在を読み取り、ライブラリアンに対してデータの保全・公開が必要であることを通知することができる。また、データにセンシティブな情報が含まれているかも、DMPを確認することで研究者に改めて質問する必要がなくなる。このように、maDMPsによって多大なコミュニケーションコストがかかるようなタスクが自動化されるこ

とで、各自はDMPの作成やデータ登録業務などに集中することができるようになる。その結果として、研究データの適切かつ網羅的な公開が促されるようになるだろう。

5. DMP Common Standard

以上、maDMPsのコンセプトやアウトカムについて概説してきたが、ではこれを実現するために現在どのような取り組みが行われているだろうか。上述の通り、データストレージやリポジトリ、研究管理システムなど研究データ管理環境を構成する様々なシステムは、それ単体ではmaDMPsを実現することはできず、同じDMPを共有し、互いに関係する情報を交換して相互に動作できなければ実現しているとはいえない。

そのためには、maDMPsの10原則における第6原則「maDMPsのための共通データモデル」と、それに従ったAPIの策定にグローバルに取り組むことが必要である。そこでRDAのワーキンググループとしてDMP Common Standards WG⁽¹⁴⁾が発足し、maDMPsのための共通データモデルの策定に取り組んだ。同グループは、まずDMPのユースケースについて、様々

な利害関係者の視点から分析した⁽¹⁵⁾。その成果として同グループは2019年12月に“RDA DMP Common Standard for Machine-actionable Data Management Plans”を勧告した⁽¹⁶⁾。策定されたDMP Common Standardのデータモデルの概略を下図に示す⁽¹⁷⁾。同規格では、研究データに関する様々な情報を束ねる実体としてDMPが定義されており、研究プロジェクトや研究助成事業を記述する部分、機器調達や人材などのコストを記述する部分、研究データを記述する部分などに分かれている。

同データモデルは、JSON形式で記述され、JSON Schemaによるオントロジーも定義されている。つまり、実体の構造や関係性が明確に定義されており、machine-actionable なデータモデルとなっている。また、個々の実体の概念や属性は、DataCite(CA1849参照)や、データカタログの標準規格であるW3C-DATなど、グローバルな規格との互換性を有しており、他の研究データに関する様々なメタデータ規格と連携していけるように設計されている。

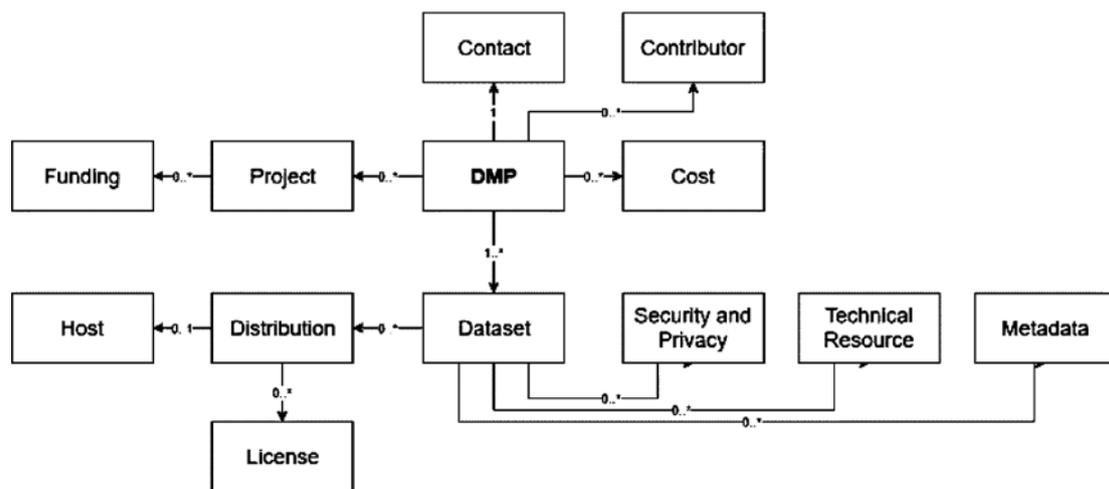


図 DMP Common Standardのデータモデル

6. 実装事例

DMP Common Standard が策定された後、現在進行形でプロトタイプシステムの開発や、既存のDMP作成支援ツールの同規格への準拠などが進められている。例えば、ウィーン工科大学のオブラスサー(Simon Oblasser)氏らは、DMP Common Standardに準拠したDMP作成支援ツールとしてDMapを開発した⁽¹⁸⁾。RDAでも、maDMPsの普及を活発化させるために、maDMPsの実装に関するコンペティションである「maDMP ハッカソン」を開催しており、多

くの人々が参加した⁽¹⁹⁾。さらに、DMP作成支援ツールのオープンソースソフトウェア開発プロジェクトであるDMProadmapでは、リード開発機関である米国のカリフォルニア電子図書館(CDL)を中心に、maDMPsに準拠したAPIなどの機能を実装中である。

他方で、これまでに述べた活動とは異なる文脈で、maDMPsを部分的に実現している事例が既に存在する。豪州の大学では、DMPを中心とした研究データ管理システムの開発・導入が進められており、特にクイーンズランド大学やシドニー工科大学などでは

DMPの内容に従って研究データ環境を整える仕組みを実現している⁽²⁰⁾。こうした独自にmaDMPsの実装を行っている事例も、maDMPsの実運用の観点から注目に値する。

7. まとめ

本稿では、maDMPsの定義、要件および関連技術の動向を、背景であるDMPの動向も踏まえて概説した。2020年7月執筆時点では、maDMPsは具体的な共通規格やプロトタイプの開発など、コンセプトの具体化にとどまっているが、今後は研究データ管理の現場と協調しながら、拡張とアップデートが重ねられていくだろう。ミクサ氏は、ホワイトペーパーの中でmaDMPsはボトムアップ的アプローチが適切であると述べている⁽²¹⁾。DMPがデータライフサイクルの情報源として真に活用され、オープンサイエンスが推進されていくためには、ライブラリアンなど学術組織の現場で業務にあたる立場からいかにmachine-actionableのコンセプトを理解し、自機関のシステムに導入していくかが重要である。国内でもmachine-actionableというコンセプトをいかに実現していくかの議論が学術組織やライブラリアンなどによるコンソーシアムの中で活発化し、世界でも類を見ないような事例が生まれることを期待したい。

- (1) DMP自体の動向については、以下が詳しい。
池内有為. データマネジメントプラン(DMP)— FAIR原則の実現に向けた新たな展開. 情報の科学と技術, 2018, vol.68, no.12, p. 613-615.
https://doi.org/10.18919/jkg.68.12_613, (参照 2020-07-21).
- (2) European Commission. Turning Fair Into reality: Final report and action plan from the European commission expert group on FAIR data. 2018, p. 76.
<https://doi.org/10.2777/1524>, (accessed 2020-07-21).
- (3) National Science Foundation. NSF Grant Proposal Guide, Chapter 11.C.2.j.
https://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_index.jsp, (accessed 2020-07-21).
- (4) European Commission. Guidelines on Fair Data Management in Horizon 2020. 2016, p. 6.
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, (accessed 2020-07-21).
DMPの提出義務の対象となるのは、Open Research Data Pilot (ORD pilot)と呼ばれる研究助成枠である。また、同事業は途中で脱退することができる。脱退後のDMP提出は必ずしも必要ではないが、提出及び計画に記されたデータのオープン化などは依然として推奨される。
- (5) 国立研究開発法人科学技術振興機構. 戦略的創造研究推進事業におけるデータマネジメント実施方針. 2016.
https://www.jst.go.jp/kisoken/crest/manual/data_houshin.pdf, (参照 2020-07-13).
- (6) 文部科学省. 令和2(2020)年度科学研究費助成事業科研費公募要領(学術変革領域研究(A・B)). 2020, p. 19.
https://www.mext.go.jp/content/20200106-mxt_gakjokik-000003634_01.pdf, (参照 2020-07-13).
- (7) DMPTool.
<https://dmptool.org/>, (accessed 2020-07-13).
- (8) Digital Curation Center. "DMPonline".
<https://dmponline.dcc.ac.uk/>, (accessed 2020-07-13).
- (9) The DMP Roadmap project. "DMP Roadmap". GitHub.
<https://github.com/DMPRoadmap/roadmap>, (accessed 2020-07-13).
- (10) Simms, Stephanie; Jones, Sarah; Mietchen, Daniel; Miksa,

Tomasz. Machine-actionable data management plans (maDMPs). Research Ideas and Outcomes. 2017, vol.3, e13086.

<https://doi.org/10.3897/rio.3.e13086>, (accessed 2020-07-21).

ただし、maDMPsという用語が登場する以前から、情報システムに親和的なDMPというコンセプト自体は“Active DMPs”という用語で欧米を中心に広く議論されていた。

- (11) Data Documentation Initiative. "Machine-actionable".
<https://ddialliance.org/taxonomy/term/198>, (accessed 2020-07-13).
- (12) Miksa, Tomasz; Simms, Stephanie; Mietchen, Daniel; Jones, Sarah. Ten principles for machine-actionable data management plans. PLoS computational biology. 2019, vol.15, no.3, e1006750.
<https://doi.org/10.1371/journal.pcbi.1006750>, (accessed 2020-07-21).
- (13) Simms, Stephanie; Jones, Sarah; Mietchen, Daniel; Miksa, Tomasz. op. cit., p. 9.
- (14) "DMP Common Standards WG". Research Data Alliance.
<https://www.rd-alliance.org/groups/dmp-common-standards-wg>, (accessed 2020-07-21).
- (15) Miksa, Tomasz; Neish, Peter; Walk, Paul; Rauber, Andreas; Park, Office; Way, Lynstock. Defining requirements for machine-actionable Data Management Plans. OSF.
<https://doi.org/10.17605/OSF.IO/CGP86>, (accessed 2020-08-04).
- (16) RDA DMP Common Standards WG. "RDA DMP Common Standard for machine-actionable Data Management Plans". GitHub.
<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>, (accessed 2020-07-21).
Miksa, Tomasz; Walk, Paul; Neish, Peter. RDA DMP Common Standard for Machine-actionable Data Management Plans. 2019-12-02.
<https://doi.org/10.15497/rda00039>, (accessed 2020-07-21).
- (17) 図の出典は以下のとおり。
RDA DMP Common Standards WG. "RDA DMP Common Standard for machine-actionable Data Management Plans". GitHub.
<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/diagrams/maDMP-diagram.png>, (accessed 2020-07-21).
- (18) Oblasser, Simon. "Machine-actionable DMP application (DMap)". Zenodo, 2019-10-29.
<https://doi.org/10.5281/zenodo.3522247>, (accessed 2020-07-21).
- (19) RDA hackathon on maDMPs 2020.
<https://rda-dmp-common.github.io/hackathon-2020/>, (accessed 2020-07-13).
- (20) 常川真央, 尾城孝一, 込山悠介, 藤原一毅, 山地一禎. データマネジメントプラン活用に関する研究データ基盤の要件定義を目的とした国際事例研究. 研究報告インターネットと運用技術(IOT). 2020, vol.2020-IOT-4, no.15, p. 1-6.
<http://id.nii.ac.jp/1001/00203411/>, (参照 2020-07-21).
- (21) Miksa, Tomasz; Simms, Stephanie; Mietchen, Daniel; Jones, Sarah. op. cit., p. 4.

[受理: 2020-08-20]

Tsunekawa Mao

Machine-actionable DMPs(maDMPs): A Review



本著作(CA1983)はクリエイティブ・コモンズ 表示 4.0 国際パブリック・ライセンスの下に提供されています。ライセンスの内容を知りたい方は

<https://creativecommons.org/licenses/by/4.0/legalcode.ja>でご確認ください。