

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

池内 有為*

キーワード：データ管理計画 (DMP), FAIR 原則, アクティブ DMP, 研究データ公開, オープンサイエンス

1. はじめに

今回は、データマネジメントプラン (Data Management Plan, DMP) を取り上げる。DMP とは、研究のために収集・作成するデータをどのように管理するか、取り扱いや整備・保存・公開についての計画である。オープンサイエンス政策における“研究データ公開の義務化”とは、助成機関による DMP (文書) の提出要求を指すことが多い。

DMP は、2003 年に米国衛生研究所 (NIH) が義務化したのを先駆けとして、英国、欧州、オーストラリア、カナダ、南アフリカ…と各国の助成機関に広がり、現在は次世代 DMP の議論が高まっている。日本では、2017 年から科学技術振興機構 (JST) が、2018 年からは新エネルギー・産業技術総合開発機構 (NEDO) と日本医療研究開発機構 (AMED) が DMP の提出を求めている。日本学術振興会の科学研究費助成事業 (科研費) が DMP を要求するようになれば、いよいよ多くの研究者や機関が取り組むことになるだろう。本稿は、助成機関による DMP のねらいと動向をふまえて、実際に DMP を作成した所感を述べたい。

2. DMP とは何か

助成機関が DMP を要求する主な目的は、投資した研究成果の公開や再利用によって価値を最大化すること、データを効率的かつ適切に管理させることなどである。DMP の記述内容は助成機関によって異なるが、(1)データの詳細 (データの種類、形式、データ量、メタデータ)、(2)倫理と知的財産権、(3)アクセスと共有、再利用 (想定利用者、公開方法)、(4)短期保存とデータ管理 (バックアップ、データの管理者)、(5)長期保存 (保存期間、長期保存すべきデータ)、(6)必要なリソース (ハードウェア、ソフトウェア、技術的支援) などが含まれる¹⁾。

研究者は、図 1 に示すように、研究中に扱うアクティブデータを整備して、必要なデータを適切に公開・保存することをイメージしながら DMP を作成する²⁾。DMP は助成金を獲得するための書類にとどまらず、データセットの

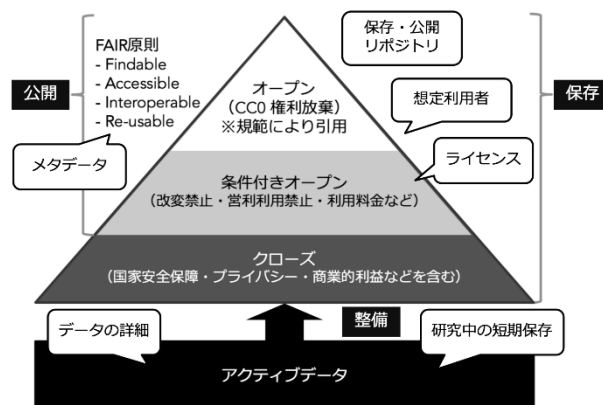


図 1 研究データ管理と DMP の記述内容の例

長期保存と有用性を確保するために重要であるとも指摘されている²⁾。

3. FAIR 原則と欧州委員会の DMP テンプレート

さて、研究データにとって公開はゴールではなく、利用されてこそ価値が生まれる。そこで国際イニシアティブの FORCE11 は、データが長期にわたって再利用できるように、Findable (見つけられる)、Accessible (アクセスできる)、Interoperable (相互運用できる)、Re-usable (再利用できる) 状態で公開しようという「FAIR 原則 (FAIR Data Principles)」を提唱した。「FAIR」はオープンサイエンス界隈のキーワードとなっており、日本ではバイオサイエンスデータベースセンター (NBDC) による解説が公開されている³⁾。

DMP の議論にも、しばしば FAIR 原則が登場する。欧州委員会 (EC) の「Horizon 2020 におけるデータマネジメントのガイドライン」は、第 3 版 (2016 年) からタイトルに「FAIR」を追加しており⁴⁾、付録の DMP テンプレートの冒頭には FAIR 原則を掲載している。DMP テンプレートは、[1]データの概要、[2]FAIR データ、[3]資源配分、[4]データセキュリティ、[5]倫理的側面、[6]その他の 6 章で構成されている。項目を数えると、本文は約 40 項目、本文をまとめた表でも約 30 項目あり、なかなかのボリュームである。特に FAIR に関する指示は細かい。

なるほどこの DMP テンプレートを完璧に作成し、実践することができれば、データが長期にわたって再利用でき

*いけうち 有為 筑波大学大学院図書館情報メディア研究科
〒305-8550 茨城県つくば市春日 1-2
E-mail: ikeuchi.ui@gmail.com
orcid.org/0000-0002-5680-1881 (原稿受領 2018.10.28)

そうだ — と思うものの、知識がないと書けない箇所（適切なリポジトリ、メタデータ標準）や、望ましい回答がわからない箇所も多い。OpenAIRE と EC の FAIR データ専門家グループは、2017 年に DMP の作成者と支援スタッフを対象としてテンプレートに関する調査を行い、289 名から回答を得ている⁵⁾。DMP テンプレートの作成・支援過程をポジティブに受け止めた回答者は 60%（意外と多い）、ネガティブな回答者は 16% であった。ネガティブな回答者は、作成のための労力が大きいことや馴染みのない用語が多すぎることを指摘している（まったくもって賛成である）。

4. 日本版 DMP を書いてみた

日本の DMP はどうだろうか。NEDO, JST, AMED の DMP も作成してみた（AMED は医学分野に特化されているため、適宜読み替えた）。項目は厳選されており、NEDO は 17 項目、JST は 9 項目、AMED は 12 項目である（表 1）。全体的に書きやすいよう工夫されていて、「データの概要」や「データ公開のレベル」などは箇条書きや選択式になっている。また、NEDO と JST はデータごとに、AMED は制限共有・制限公開・非制限公開データごとに記入欄が設けられている。

表 1 DMP の概要

項目	NEDO	JST	AMED
データ名称	○	○	
データの説明	○	○	◎
管理者/担当者	○		○
分類	○	○	
公開レベル	○	○	
DMP 対応項目		○	
秘匿理由	○		
秘匿期間	○		○
取得者	○		
取得方法	○		
その他	○	○	○
公開データ			
想定利活用用途	○	○	
利活用・提供方針	○	○	○
円滑な提供に向けた取り組み	○	○	
リポジトリ/データベース	○		◎
データフォーマット			○
想定データ量	○		
加工方針	○		
その他	○		○

項目名は NEDO の DMP に準じて作成し、適宜読み替えた。
◎は制限共有・制限公開・非制限公開データごとに記載する。

書くのが難しかったのは、NEDO と JST の「想定利活用用途」、「利活用・提供方針」、「円滑な提供に向けた取り組み」である。データ公開が習慣となっている分野の研究者には難しくないのかもしれないが — 筆者などは、例文に当てはまらない場合があること、データの種類ごとに記

載する必要があることから、かなりの時間を要した。それでいて完璧な記述ができたとも思えない。

そもそも、DMP のねらいは研究を適切に推進することだろう。研究者が DMP の作成に時間を取られてしまい、研究のための可処分時間が減るならば本末転倒ではないだろうか。実際に DMP を書いてみて、俄然、DMP ツール（作成補助ツール）への期待が高まった。

5. 日本版 DMP ツールへの期待

国外の DMP の多くは文章形式で記述しなければならない、負担が大きい。そこで助成機関の目的を達成しつつ負担を減らすためにオープンソースの DMP ツールが開発されてきた。研究者は無料でアカウントを作成して、助成機関名を選択すると、DMP の項目が質問の形で表示される。質問に答えるようにデータに関する情報を入力していくと文章形式の DMP が完成し、提出用のフォーマットで出力できるというものである。米国はカリフォルニア大学キュレーションセンター (UC3) が DMPTool (<https://dmp.cdlib.org>) を、英国はデジタルキュレーションセンター (DCC) が DMPonline (<https://dmponline.dcc.ac.uk>) を提供しており、他の国でも流用されてきた。現在は両者を統合した DMPRoadmap が、後述する Active DMP として開発されている。

DMP ツールには、回答例や回答に役立つガイドを掲載することができる。EC の DMP テンプレートに対する調査⁶⁾では、「自分の分野やデータタイプにとって適切な標準のサジェスト」、「記載例や推奨される回答の例を増やす」、「自分の分野での良い実践に基づくドロップダウンオプションの提示」の優先順位が高いことがわかった（こちらも大いに賛成）。模範的な回答を DMP ツールで選択できれば、回答者の負担が減り、データ管理の啓発にもなるだろう。

研究者が提出した DMP には、助成機関が想定しなかった「データ公開方法」や「データを公開しないほうが良い理由」なども含まれると考えられる。こうした記述をレビューして、常に DMP ツールの回答例やガイドに反映させられれば、よりよいデータ公開の方法や注意すべき事柄が研究者・助成機関・DMP ツールの開発者の間で共有できると考えられる。DMP から得られた知見は、研究データマネジメント教育などに活かすこともできるだろう。

6. Active DMP, あるいは機械で実行可能な DMP

次世代 DMP の主な論点は、FAIR 原則の実現、DMP の標準化、そしてアクティブな DMP の開発であり、FORCE11 の FAIR DMP Working Group (WG) や研究データ同盟 (RDA) の DMP Common Standards WG, Active DMPs Interest Group (IG) などで検討されている。#activeDMPs (<https://activedmps.org>) によれば、機械で実行可能な (machine-actionable) DMPs (maDMPs) は、従来の自由記述による DMP を構造化し、公開し、既存システムと連携することによって、多くのステークホルダーにとって価値を生むとのことである。「maDMPs の 10 原則」⁷⁾では、具体的な要件とメリットを挙げている。たとえ

ば、人が理解でき、かつ機械可読にすること、DOI や ORCID などの永続的識別子 (PID) および統制語彙を用いることによって、多様な分野の用語や作法に則って書かれていた DMP の検索や理解、研究と研究者情報の追跡が容易になる。すると評価者はもちろんのこと、研究者は幅広い分野の DMP やデータを、倫理委員会や法務担当者はベストプラクティスを、リポジトリ担当者はコストやライセンスの情報を参照して、それぞれ役立てることができるようになる。また、DMP を更新可能でバージョン情報をもつ文書にすること (データが公開されたら自動的に DMP が更新され、タイムスタンプが押される) なども挙げられている。つまり maDMPs は、データと同様に DMP の相互運用性を高めつつ、公開することによって、研究をより効率化することを目指しているといえるだろう。

7. おわりに

実際に DMP を作成してみて、データの管理や公開の参考になる点が多々あった。DMP ツールによって効率的な作成や、他の研究者の優れた実践・課題の共有が可能になれば、より有用なものになるだろう。とはいえ、DMP の作成とデータ公開には、相応のコストがかかる。DMP を要求するならば、米国科学財団 (NSF) のように^{注2)}公開データを業績とさせていただければ、と思う次第である。

註・参考文献

注1) 詳しくは、オープンアクセスリポジトリ推進協会 (JPCOAR) による教材「研究データ管理サービスの設計

と実践」や国立情報学研究所 (NII) によるオンライン講座「オープンサイエンス時代の研究データ管理」(<https://www.nii.ac.jp/service/jmooc/rdm/>) を参照。

注2) 2013年1月、NSFは業績記入欄の名称を「Publication」から「Products」に変更して、申請者が公開したデータやプログラムコードの記載を可能にした。
https://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_sigchanges.jsp, [accessed 2018-10-28]

- 1) Jones, Sarah. How to Develop a Data Management and Sharing Plan. Digital Curation Centre. 2011, 8p.
<http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf>, [accessed 2018-10-28]
- 2) Strasser, Carly. 研究データ管理. 機関リポジトリ推進委員会訳. 2016, p.5.
<http://id.nii.ac.jp/1280/00000195/>, (参照 2018-10-28)
- 3) NBDC 研究チーム. “データ共有の基準としての FAIR 原則”, 2018. <http://doi.org/10.18908/a.2018041901>, (参照 2018-10-28)
- 4) European Commission. H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020 (ver. 3.0, 26 July 2016). 2016, 12p.
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, [accessed 2018-10-28]
- 5) Grootveld, Marjan et al. OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans (Version 1.0.0) [Data set]. Zenodo, 2018. <http://doi.org/10.5281/zenodo.1120245>, [accessed 2018-10-28]
- 6) Miksa, Tomasz et al. Ten principles for machine-actionable data management plans [preprint]. zenodo, 2018. <http://doi.org/10.5281/zenodo.1434938>, [accessed 2018-10-28]

Series: Current trend of open science: Data Management Plan (DMP) – New dimension to implement FAIR Data Principles. Ui IKEUCHI (Graduate School of Library, Information and Media Studies, University of Tsukuba, 1-2 Kasuga, Tsukuba-shi, Ibaraki 305-8550)

Keywords: Data Management Plan(DMP) / FAIR Data Principles / Active DMP / Open Research Data / Open Science