

研究データ管理（RDM）の目的地と現在地

池内 有為*

キーワード：研究データマネジメント（RDM）、大学・研究図書館、RDM 評価、研究データ公開、オープンサイエンス

1. はじめに

オープンサイエンスにおける研究データの公開について、推進に向けた課題や論点が整理されつつある。それでは、研究者、研究機関、図書館等のステークホルダーは、どのように研究データ公開を実践すればよいのだろうか。

本稿では、研究データ管理（Research Data Management, RDM）の評価ツールから検討してみたい。一般には、実践→評価の順に行われるが、既に RDM を進めている国外の評価ツールを参考に、バックキャストで実践内容や優先順位を考えてみようとする試みである。

2. 研究データ管理（RDM）とは

本稿で扱う RDM の範囲は、(1)研究者による RDM の実践、(2)研究機関による RDM の方針、(3)図書館等による RDM の支援としたい。RDM には 2018 年 12 月号で紹介した¹⁾データマネジメントプラン（Data Management Plan, DMP）^{注1)}、すなわち研究データの整備、公開、保存についての計画（文書）が含まれる。すなわち、(1)には DMP

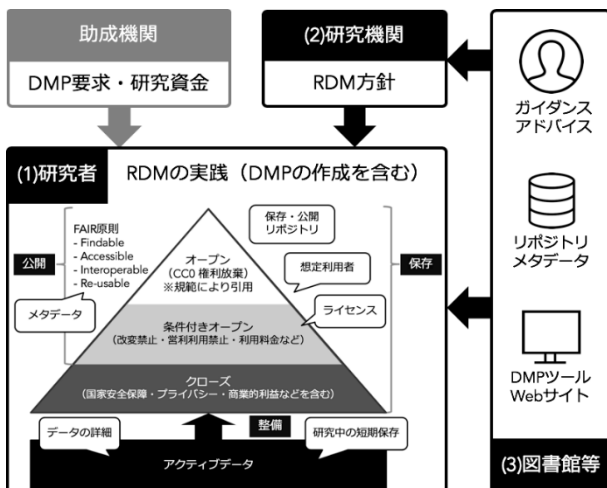


図1 研究データ管理（RDM）とステークホルダー

の作成や実践が、(2)には DMP または DMP に準ずる要求が、(3)には DMP の作成・実践の支援が含まれる（図1）。RDM と DMP は字面も内容も似ていてややこしいのだが、DMP < RDM と考えていただければと思う。

3. 研究者のための RDM 評価ツール

3.1 RDM ルーブリック

カリフォルニア大学キュレーションセンター（University of California Curation Center, UC3）の Borghi らは、RDM の実践状況について研究者が自己評価を行うためのルーブリック（評価基準の表）を考案した（図2）²⁾。本文の訳語は、京都大学オープンソースプロジェクトによる「研究データマネジメントルート」の抄訳³⁾に従って記す。

		低 ← 成熟度レベル → 高			
		Ad Hoc	One-Time	Active and Informative	Optimized for Re-Use
研究の段階	開始 Planning your project	When it comes to my data, I have a "way of doing things" but no standard or documented plans.	I create some formal plans about how I will manage my data, but I generally don't refer back to them.	I develop detailed plans about how I will manage my data that I actively revisit and revise over the course of a project.	I design my plans for managing data to streamline future use by myself or others.
	Organizing your data	I don't follow a consistent approach for keeping my data organized, so it often takes time to find things.	I have an approach for organizing my data, but I only put it into action after my project is complete.	I have an approach for organizing my data that I implement prospectively, but it is not necessarily standardized.	I organize my data so that others can navigate, understand, and use it without me being present.
	Saving and backing up your data	I decide what data is important while I am working on it and typically save it in a single location.	I know what data needs to be saved and I back it up after I'm done working on it to reduce the risk of loss.	I have a system for regularly saving important data while I am working on it. I have multiple backups.	I save my data in a manner and location designed to maximize opportunities for re-use by myself and others.
	Getting your data ready for analysis	I don't have a standardized or well documented process for preparing my data for analysis.	I have thought about how I will need to prepare my data, but I handle each case in a different manner.	My process for preparing data is standardized and well-documented.	I prepare my data in such a way as to facilitate use by both myself and others in the future.
	Analyzing your data and handling the outputs	I often have to redo my analyses or examine their products to determine what procedures or parameters were applied.	After I finish my analysis, I document the specific parameters, procedures, and protocols applied.	I regularly report the specifics of both my analysis workflow and decision making process while I am analyzing my data.	I have ensured that the specifics of my analysis workflow and decision making process can be put into action by others.
	終了 Sharing and publishing your data	I share the results of my research, but generally I do not share the underlying data.	I share my my data only when I'm required to do so or in response to direct requests from other researchers.	I regularly share the data that underlies my results and conclusions in a form that enables use by others.	Because of my excellent data management practices, I am able to efficiently share my data whenever I need to with whomever I need to.

図2 RDM ルーブリック⁴⁾

RDM ルーブリックは、成熟度レベル（列）と、研究の各段階における評価基準（行）からなる二次元の表として構成されている。成熟度レベルは「その場しのぎ的な対応」、「1回はできている」、「有効かつ有用」、「再利用のために最適化済み」の4段階に分けられている。評価基準は、「研究プロジェクトの計画段階」から「データの共有と公開」までの6段階である。たとえば、「研究プロジェクトの計画段階」で「研究データを入手した場合、『何らかの方法をとらなくてはならない』のだが、標準化あるいは文書化された計画を持っていない」ならば「その場しのぎ的な対応」である。一方、「自分自身あるいは他者が将来合理的にデータを利用できるように、データマネジメント計画を作成して

*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 2019.2.5)

いる」場合は「再利用のために最適化済み」となる。

3.2 もう少し簡単に — チェックリスト

eRSA^{注2)}による13項目のチェックリスト⁵⁾は、より簡単に利用できると考えられる(図3)。項目によっては補足が必要だと思われるが、DOIの付与やデータ引用に関する解説をつければ、啓発にもつながりそうである。

- DMP作成済み
- データとファイルは論理的に組織化されている
- 一貫した語彙とラベルが使用されている
- 対象と方法のメタデータが記録されている
- ファイルは永続性のある一般的なフォーマットで作成されている(例: XLSではなくCSV)
- データは安全な場所に保存されている
- バックアップが作成されている
- 保存期間が決められている
- データの所有者が特定されている
- 倫理ガイドラインに準拠している
- コレクションレベルのメタデータが適切なディスクバリーポータルに登録されている
- 適切なライセンスが付与されている
- データセットが引用できるようにDOIが付与されている

図3 RDM チェックリスト⁵⁾

4. 研究機関・図書館等のためのRDM評価ツール

4.1 RISE (Research Infrastructure Self Evaluation Framework)

英国のデジタルキュレーションセンター(Digital Curation Centre, DCC)は、研究機関によるRDMサービスを自己評価するための枠組みとして、RISEを公開している⁶⁾。図4に示す研究データサービスモデルの10項目

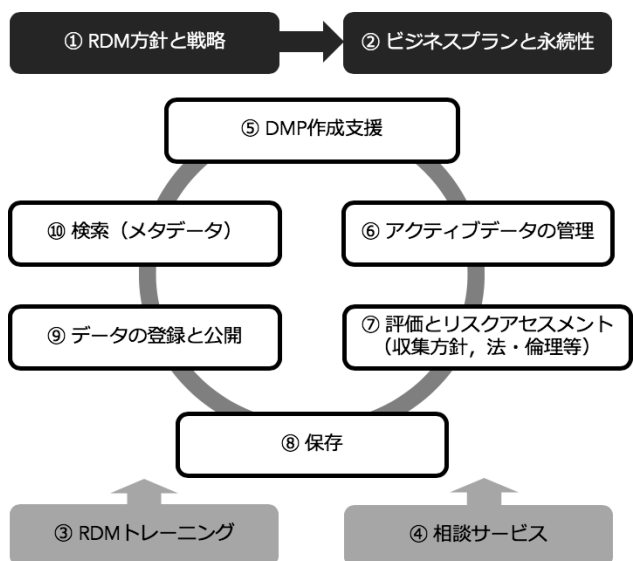


図4 DCC Research Data Service Model⁶⁾

について、それぞれ解説と「レベル1」から「レベル3」までの評価項目が設定されている。

たとえば、③RDMトレーニングには、「オンライン」と「対面」の2種類がある。オンラインの「レベル1」は(組織の)RDMページに外部のトレーニングページがリンクされている、「レベル2」は組織のニーズやサービスにあわせたものが追加されている、「レベル3」は組織の研究者やスタッフのニーズにあわせて多くのオンライントレーニングツールを作成しており再利用されている、である。

4.2 より具体的に — チェックリスト

SPARC Europeは、RISEに基づくチェックリスト式の評価ツールを公開しており⁷⁾、RISEの項目を「方針と戦略」、「ライセンス」などの11項目に整理した上で、それぞれのチェック項目を示している。たとえば「データリポジトリ」には、「DMPが含まれる」、「保存方針がある」、「ROAR^{注2)}にリストアップされている」、「データ登録を要求している」、「発見可能性(OAI-PMI, Googleカスタム検索エンジン等)」といったチェック項目がある。

RISEが目標を記述しているのに対して、このチェックリストには具体的な取り組みが示されている。結果はレーダーチャートで表示されるため、機関の強みと弱みが明白となる。

5. RDMの目的地と現在地の明確化

こうしたRDM評価ツールには、ある時点での理想や目標が反映されていると考えられる。その全てを実現しようと思うとハードルが高いが、研究者や研究機関にとって不十分な点を可視化し、優先順位を検討するためのツールとして有用であるとも考えられる。

また、研究者による評価結果は、RDMサービスを構築する上でニーズを把握するための良い手がかりになるのではないだろうか。評価ツールに研究機関の方針や図書館等が提供可能なサービスを盛り込み、研究者に提示すれば、方針やサービスの周知にも役立つと考えられる。

ただし、ここで紹介した評価ツールは国外のものであるため、まずは日本の状況にあわせて項目の妥当性を検討したり、項目を追加したりする必要があるだろう。たとえば、「国立研究開発法人におけるデータポリシー策定のためのガイドライン」とのすり合わせ、Google Dataset Searchや開発中のCiNii Researchへの対応を盛り込むことなどが考えられる。改定した評価ツールの達成度から、組織の現在地、および相対的な強みや弱みを把握することによって、今後の優先課題を具体的に検討することができるのではないだろうか。

6. おわりに

ここ数年、RDMやRDM支援サービスをどう進めればよいのかを考えてきた。その難しさは、研究分野や研究者によって困難な点やニーズが異なること、結局「要求にあわせてできる範囲で泥縄式」に落ち着くのではないかと

う諦観にある。しかし、今回はこうした懸念を一旦脇に置いて、一足先に RDM を進めている国外の評価ツール（目的地）から現在地を把握する方法や、研究者の個別ニーズを把握する方法を考えてみた。

さて、本リレー連載も早一年。読者や編集委員の皆様からの反響に助けられつつ書き進めることができた。また、佐藤翔さん（同志社大学）や林豊さん（九州大学）による時宜を得た記事からは、多くの知見や示唆を得ることができた。心よりお礼申し上げます。

註・参考文献

- 注 1) 本稿では management を「マネジメント」、「管理」と訳出している。「データマネジメントプラン」は科学技術振興機構（JST）などの方針に、「研究データ管理」は国立情報学研究所（NII）の MOOC (<https://www.nii.ac.jp/service/jmooc/rdm/>) に準じた。
- 注 2) 南オーストラリアの University of Adelaide, Flinders University, University of South Australia による研究支援ベンチャー。 <https://www.ersa.edu.au>
- 1) 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, (accessed 2019-02-05).
- 2) Borghi, John A. et al. Support your data: A research data management guide for researchers. Research Ideas and Outcomes, 2018, vol.4, e26439. <https://doi.org/10.3897/rio.4.e26439>, (accessed 2019-02-05).
- 3) takaakiaoki. 研究データマネジメントルーブリック (Research Data Management rubric). GitHub, 2018. https://github.com/kyoto-u/rdm_rubric/blob/master/rubric.md, (accessed 2019-02-05).
- 4) Borghi, John A. Support your Data. University of California Curation Center (UC3), 2018. <https://uc3.cdlib.org/2018/01/11/support-your-data/>, (accessed 2019-02-05).
- 5) Data management checklist. eRSA. <https://www.ersa.edu.au/data-management-checklist/>, (accessed 2019-02-05).
- 6) Rans, Jonathan; Whyte, Angus. Using RISE, the Research Infrastructure Self Evaluation Framework. 2017, DCC. <http://www.dcc.ac.uk/resources/how-guides/RISE>, (accessed 2019-02-05).
- 7) How Open is Your Research? A Checklist for Institutions. SPARC Europe. <https://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-research-checklist-institutions/>, (accessed 2019-02-05).

Series: Current trend of open science: Research Data Management (RDM) – Goals and strategy. Ui IKEUCHI (Graduate School of Library, Information and Media Studies, University of Tsukuba, 1-2 Kasuga, Tsukuba-shi, Ibaraki 305-8550)

Keywords: Research Data Management (RDM) / University/Research Libraries / RDM Evaluation / Open Research Data / Open Science