

ループブリックによる RDMへの共通理解

青木 学聡

研究データマネジメント

Research Data Management

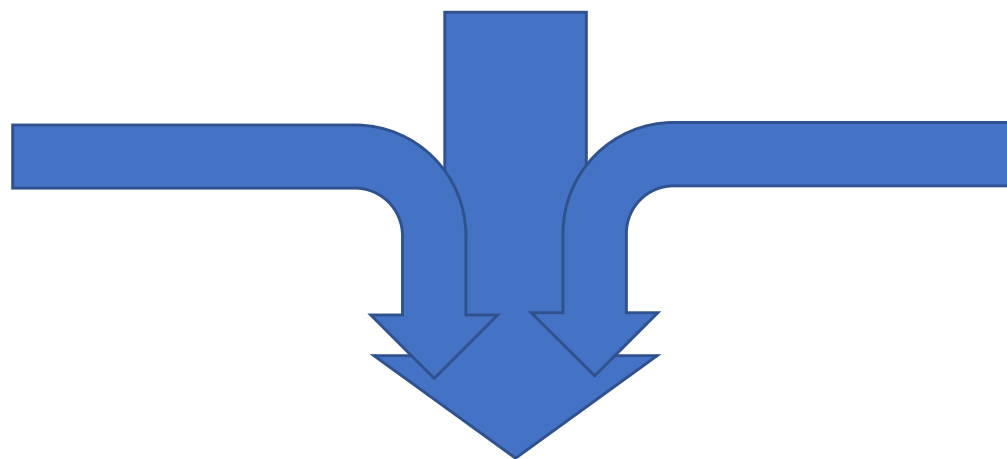
- 研究を進めるときに、どのように
 - データを生成 or 収集するか？
 - 安全に保管するか？
 - 共同研究者とデータを共有するか？
 - 有効な解析手法を適用するか？
 - 研究成果として広く公開するか？
- 既に皆が「なんとなく」行っている行為
 - 研究公正維持
 - オープンサイエンス推進といった現代の文脈で再定義する必要がある

現場(=教員)から見た研究データ管理

研究データの信頼性が損なわれた場合、最大の被害者は研究者自身
研究データの価値を最大限生かすことができるのも、研究者自身

自己流(研究室伝統)の方法

大学当局他の
ガバナンス



自身の研究業界
の動向

- 研究者個人で解決できない問題に
明確なルールに加え, 豊富な事例が必要
- 信頼できる(免責されうる) 方法とは?
 - コストが許容できる方法とは?

今風の「研究データ管理」に 対する意識調査

- 倉田ら「日本の大学・研究機関における研究データの管理，保管，公開：質問紙調査に基づく現状報告」情報管理, 60(2), 119–127, (2017)
- いわゆる「研究データ10年保存」の号令がかかった直後の、機関(事務、図書館)、研究者の意識調査
- 「データ保存のルールが作られたのは知っている」
- (オープンサイエンスを含む)包括的なRDMについてはよくわからない
- 未知の概念、新しい制度に対する心理的障壁？

京都大学研究データマネ ジメントワークショップ KU RDM workshop

- 2018/10/6(土)に開催
- California Digital Library (CDL)より Stephanie Simms 氏を招待
- RDMループリックによる成熟度の自己評価
 - CDL による提案
 - 研究ライフサイクルの各段階で、FAIR原則を満たすべくどのような取り組みを実施しているかを4段階で自己評価。
 - 葛ユニットで参考和訳を作成



RDM Rubric

<https://uc3.cdlib.org/2018/01/11/support-your-data/>

	Ad Hoc	One-Time	Active and Informative	Optimized for Re-Use
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 at the start of a project, 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 have created plans for managing my data that are designed to streamline its 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 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 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 document 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 understood and put into action by others.
Sharing and publishing your data	I share the results of my research, but generally I do not share the underlying data.	I share 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.

RDMルールブック(和訳)

https://github.com/kyoto-u/rdm_rubric

	その都度	1度のみ	常に見直す	再利用のために最適化
データマネージメント計画	データを入手した段階で、自分なりに考える。標準化や文書化はしていない。	最初に計画を作成するが、通常は、途中で見直すことはしない。	詳細な計画を作成し、プロジェクト期間を通じ参照と見直しを行う。	自分あるいは他人が将来スムーズにデータを利用できるように計画している。
データの体系的整理	一貫した手法に従っていない。そのため、しばしば探し出すのに時間を要する。	手法はあるが、それを適用するのは、プロジェクト終了時のみ。	先を見越した手法を導入しているが、必ずしも標準化されていない。	他人が自分でデータを探索、内容を理解し、利用できる。
データの保存とバックアップ	作業中にどのデータが重要であるかを決め、通常はこれを特定の1か所に保存する。	どのデータを保存すべきかを理解している。損失のリスクを緩和するため、作業後にバックアップをとる。	作業中でも定期的にデータを保存するシステムを用いている。複数のバックアップを保持している。	再利用する機会を最大限生かせるよう考慮した方法と場所にデータを保存している。
データを解析可能な状態にする	標準化もしくは十分に文書化された手続きはない。	解析に適した形式を検討し整備するが、事例ごとに異なる方法をとっている。	データ処理方法は標準化され、かつ十分に文書化されている。	将来も自分自身及び第3者が利用しやすい方法で、データ処理を行っている。
データ解析と成果の取扱い	解析や実験がどのような手順やパラメータにより実施されたか確認するために、これらをやり直すことが多々ある。	解析作業終了後、パラメータの詳細や解析手順、プロトコルを文書化している。	データ解析の際、決められた方法で解析のワークフローや判断のプロセスの詳細をいつも記録している。	自身が行った解析のワークフローや判断プロセスの詳細が他人によっても実行できることが確認できている。
データの共有と公開	研究成果を共有しているが、その根拠となるデータについては共有できていない。	リクエストがあった場合にのみ研究データの共有を行っている。	自身の研究成果や根拠となるデータは、第三者が利用できる形式で共有できる。	優れたデータマネージメントを実践し、自身のデータをいつでもだれとでも効率的に共有できる。

グループワークによるルーブリック作成

Group working on RDM rubric

- 参加者10名
- Google Docs にあらかじめ質問事項を記載した文書を配置、各人が同時に回答を書きこむ
 - 他人の回答から、「どのように回答すればよいか、記述精度はどの程度か」がリアルタイムで認識できるので、「手が止まる」ということがなかった
- 回答時間は約30分、そのあと結果を見直しディスカッション



ループリック以前の問題

Problem before Rubric

- 「どのような立場で」「何を対象として」回答しようとしているのかを明確にしないと議論がかみ合わない
- 参加者の立場
 - (Solo) ほぼ一人で実施している研究
 - (PI) ラボのリーダー、プロジェクトマネージャ
 - (Arc) アーキビスト、コンテンツ作成者
 - (Repo) データリポジトリ運営者

回答の傾向

Summary of Comments

- Planning Data
 - ほぼ全員が「Adhoc(その場しのぎ)」と評価
 - Repo は「受け入れルールを決め、利用者に通知している」ため、ある程度形式だった方法が適用されている
- Organizing Data
 - Solo の多くは「自分で決めたルールでフォルダ整理」
 - Pl, Arc, Repo は組織で動くことが前提なので、ローカルルールを定めている場合が多い
- Saving and Backing up Data
 - いずれの立場でも何らかの方法でもってバックアップを取っている。ただしそれぞれが確立した手法を模索している

回答の傾向

Summary of Comments

- Getting your data ready for analysis
 - Solo は「自分がわかればいい」ことを基準にして、解析手順などを管理
 - PI, Arc は共同作業や第3者提供を前提としたルールを決めている
 - Repo の場合、データ提供者の技量によるところが大きい
- Analyzing your data and handling the outputs
 - いずれのクラスでも、再利用や再現性を考慮している
 - ただし、手順は記録されていても、「なぜそれを行ったか」に関する部分は残されないことが多い
- Sharing and publishing your data
 - 論文についてはオープンアクセスが普及し始めたが、データについては未整備。したがってデータの公開についてはできていない部分が多い
 - データのクレンジング、質保証の問題
 - データの公開について、提供者との合意形成が必要。プロジェクト開始時に瑕疵があると、公開に支障が出る(アンケート等)

その他の話題

Miscellaneous Topics

- 権利関係は研究プロジェクト開始時に考慮しなければならない。将来共同研究におけるMOU内にDMPを記載するようになるかもしれない。
- 「データは資産である」という考えが浸透すると、逆に「正しくデータを捨てる」技術が必要になる。
- ファイル、ディレクトリの命名規則は個人の中では完結することが多いが、グループでの作業になると、それらが衝突することが多い。

組織での研究データ管理体制構築の必要性

- 大学ICT推進協議会(AXIES)において「学術機関における研究データ管理に関する提言」を取りまとめ中

(目的・効果)

- I. 研究データ管理における学術機関の役割
- II. 学術機関における研究データ管理の導入目的の実際
- III. 学術機関が管理・提供する研究データと利活用の場面

(方法・機能・維持管理)

- IV. 学術機関における研究データ管理を成り立たせる条件
- V. 学術機関における研究データ管理のための仕組み
- VI. 研究データ管理のためのデジタルプラットフォームの機能要件
- VII. 研究データ管理のための人材育成

- ステークホルダー間の調整、意識合わせが重要
(マルチステークホルダーアプローチ)

組織的なRDM体制構築のための ツールキット

- どの様なデータや方法論でもって、議論を進めるか。良い事例集をツールキットとしてまとめる？
 - 研究者に対するヒヤリング
 - アンケート
 - 対面調査
 - 研究者、ラボ向けRDMルーブリック
 - 教育プログラム(ex. JPCOARのRDMトレーニングツール)
 - 機関データリポジトリ活用ガイドライン(現状)
- 機関としてのRDM導入のガイドライン、ルーブリック整備へ

組織のためのRDMルーブリック

- Syracuse University (MLIS, MS in Library and Information Science)
<http://rdm.ischool.syr.edu/xwiki/bin/view/Blog/A+Rubric+for+the+CMM4RDM?language=>
- Australian National Data Service (ANDS)
<https://www.ands.org.au/guides/creating-a-data-management-framework>
- RDA “fair data maturity model wg”
<https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>
- Angus Whyte, Magdalena Getler (DCC), “Building the capabilities for a professional RDM support service”, Danish National Forum for Research Data Management,
<https://slideplayer.com/slide/14755415/>