

ラボラトリでの 研究データ管理ポータル の役割

京都大学 青木学聡

Gakunin RDM 対応状況

- IdP の接続 – 済
- ストレージの準備 – 進行中
 - キャンパスICTラボが試行する Nextcloud – 利用可・試験中
 - 教職員向け Google Drive – API利用が解放されない?
 - 組織向けo365より利用する OneDrive – osf.io でも利用できない?
- 利用者の拡大 – どうする?
 - G-RDM (及びRCOSのオープンサイエンス基盤全体)が目指すものが、研究者・機関の需要とどのようにマッチするか、

様々な研究支援ツール

Bosman and Kramer @ Utrecht Univ. Lib. <https://doi.org/10.6084/m9.figshare.1286826.v1>

101 Innovative tools and sites in 6 research workflow phases
(< 2000 - 2015)



Most important developments in 6 research workflow phases

	Discovery	Analysis	Writing	Publication	Outreach	Assessment
Trends	social discovery tools	data-driven & crowdsourced science	collaborative online writing	Open Access & data publication	scholarly social media	article level (alt)metrics
Expectations	growing importance of data discovery	more online analysis tools	more integration with publication & assessment tools	more use of "publish first, judge later"	use of altmetrics for monitoring outreach	more open and post-publication peer review
Uncertainties	support for full-text search and text mining	willingness to share in analysis phase	acceptance of collaborative online writing	effect of journal/publisher status	requirements of funders & institutions	who pays for costly qualitative assessment?
Opportunities	discovery based on aggregated OA full text	open lebanotes	semantic tagging while writing/editing	reader-side paper formatting	using repositories for institutional visibility	using author-, publication- and affiliation-IDs
Challenges	real semantic search (concepts & relations)	reproducibility	safety/privacy of online writing	globalization of publishing/access standards	making outreach a two-way discussion	quality of measuring tools
Most important long-term development	multidisciplinary + citation-enhanced databases	collaboration + data-driven	online writing platforms	Open Access	more & better connected researcher profiles	importance of societal relevance + non-publication contributions
Potentially most disruptive development	semantic/concept search + contextual/social recommendations	open science	collaborative writing + integration with publishing	circumventing traditional publishers	public access to research findings, also for agenda setting	moving away from simple quantitative indicators

Typical workflow examples



LIMSやELN

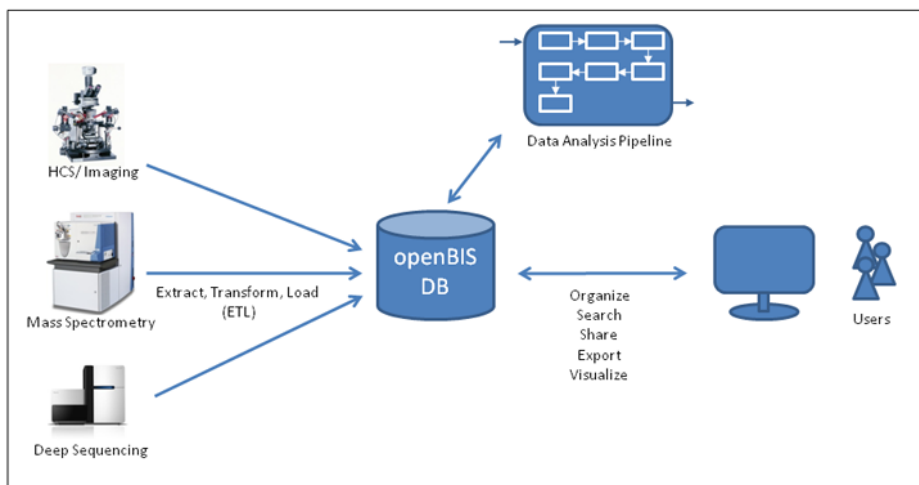
ある特定の業界(医薬、生命)では、市場が構築されている

- Laboratory Information Management System

- 計測器からの出力から、解析、レポート、アーカイブまで総体的に扱い、データの整理と自動化を支援する

- Electronic Laboratory Notebook

- 実験の計画と記録、ディスカッションなどを電子的に記録、共有する



OpenBIS のwebページより

<https://wiki-bsse.ethz.ch/display/bis/Home>

研究者のツール選択は保守的？

- 自身の理解、スキルレベルに見合ったものしか利用できない
 - 身についた方法はなかなか変わらない
 - 今のやり方で最適化されている
 - 新しいツール、方法論に対する抵抗感
 - 個人的な感覚では「石の上にも3年」
- イニシエーションが重要
 - 全く新しい概念を受け入れさせる
 - 全く新しい人に受け入れさせる

ツールによりどのようなRDMが実現されているか

- RDMルーブリックを用いた自身のRDMの評価
- California Digital Library より提案
John A Borghi, et al., "Support Your Data: A Research Data Management Guide for Researchers",
<https://doi.org/10.3897/rio.4.e26439>
- 研究データ管理の段階
計画、生成/収集、保存、解析の準備、
解析の実施、共有/公開
それぞれにおいて、
その都度、一度限り、
有効かつ有用、再利用のため最適化
された手続きが取られているかを確認

RDM Rubric

John A Borghi, et al. "Support Your Data: A Research Data Management Guide for Researchers",
<https://doi.org/10.3897/rio.4.e26439>

	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ルーブリック(和訳)

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

	その都度	1度のみ	有効かつ有用	再利用のために最適化
データマネジメント計画	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はいない。	他人が自分でデータを検索、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。	優れたデータマネジメントを実践し、自身のデータをいつでもだれとも効率的に共有できる。

「有効かつ有用」な状態

Active and Informative

- ‘Active and Informative’ refers to circumstances in which data management is a regular part of the research process. Practices and procedures are standardized, well documented, and well integrated with those implemented at other phases.
- 「有効かつ有用」データマネジメントが研究プロセスの標準的な内容となっている状態を指す。方法と手順は標準化、文書化され、さらに他の段階において実装された方法、手順と連携できている。

研究データ管理ポータル必要性

- 研究活動の「これまで」と「これから」を俯瞰できる
 - 利用するツールが拡大、多様化する中で「アソコのアレ」だけではコミュニケーションできない。「どこにナニがある」の共通認識が得られる場所
 - データに限らず、実験や解析手順の記述、これまでなされてきたディスカッションも含むだろう
- 効果を発揮するシーン
 - 新人、新任教育
 - 衝突する価値観の調停役(複数の共同研究など)
 - 「アリバイ作りため」ではなく、より自発的な「明確、継続性のある研究プロジェクト進行」

最後に

- 研究者が「マネジメント」すべきものは大量
 - データ
 - 資金
 - 資産(装置、薬品、・・・)
 - 人(最近はメンタルも)
 - 時間
- これらに対し
 - 「迷わせない」ソリューション
 - 専門性の充実(人、教育の充実)といったバックエンドのサポートはさらに重要