The University of Osaka
Institutional Knowledge Archive

Title	大学でのオープンサイエンスと研究データ管理
Author(s)	青木, 学聡
Citation	
Version Type	
URL	https://hdl.handle.net/11094/71306
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

大学でのオープンサイエンスと研究データ管理

京都大学
情報環境機構 / 学術情報メディアセンター /
アカデミックデータ・イノベーションユニット

青木 学聡
aoki.takaaki.6v@kyoto-u-.ac.jp

最新版は↓より入手してください(2019/2/28 まで)
https://nextcloud.rd.iimc.kyoto-u.ac.jp/index.php/s/CzSiEXHXW2Q36oy

簡単な自己紹介

- (2000.3~) 各種研究員/産学官連携助手等
  - ナノ加工・製造プロセスの計算機シミュレーション

- (2007.2~) 京都大学工学研究科附属情報センター
  兼 電子工学専攻 講師
  - 研究科レベルの情報インフラ整備
    - 情報セキュリティポリシー実装
    - 論文データベース
    - 入試・学部成績データ分析基盤
    - 研究データ保存システムの検討, プロトタイプ(2015.12~)

- (2016.3~) 情報環境機構
  兼 学術情報メディアセンター 准教授
  - 研究支援部門として, 全学的なITCの設計・導入・運用
    - スーパーコンピュータシステム (2016.10 更新, 稼働)
    - 汎用コンピュータシステム(2016.12 更新, 稼働)
    - データセンター, ハウジングサービス
    - 全学レベルの研究データマネジメント環境の整備
    - 大学における研究者情報環境の整備

京都大学情報環境機構
研究支援部門の主な現行業務

- コンピューティングサービス
  (メディアセンタースーパーコンピュータシステム運用)
  - 全国共同利用・共同研究拠点(ネットワーク型) を含む各種利用手続き対応
  - 障害/QA対応
  - 講習会、プログラミング高度化事業等の企画・実施
- データセンターサービス
  - ハウジング(ハードウェア預かり)
  - VMホスティング(仮想コンピュータ資源の提供)
  - Webホスティング(web公開機能の提供)
  - データ保存サービス(長期間保存ストレージの提供)
- 研究情報、研究者情報サービス
  - 教育研究活動データベース運用、開発
  - ORCIDの導入、活用の検討

トピック

- オープンサイエンスと研究データ管理
- 大学における組織的な研究データ管理の位置づけ
- ボトムアップからの大学での研究データマネジメント環境の理解
- その他の話題

基本的立ち位置

- おそらく一番の興味は
  - いつ義務化されるの?
  - その時どんな仕事があるの?
  - 執行部の意向を反映していて「自発的に動きにくい」

- 図書館, URA, 情報基盤センター
  - 専門知識・技術を持っている
  - 研究者のマインドを持っている
  - 「トップダウン」と「ボトムアップ」をすり合わせることができる、はず

オープンサイエンスと研究データ管理

内閣府「統合イノベーション戦略」(2018.6)

これまで大量印刷と物流による知的基盤が支えてきた科学は、ＩＣＴの発展により、サイバー空間が支える科学へと大きく変容し、学界、産業界、市民等あらゆる者がサイバー空間にある研究データを利活用し、協働によって知の創出をするというオープンサイエンスが進展してきている。このような社会の変化に応じて、新たな制度を整備しつつ、研究データの取扱いについての対応方針や運用を再定義することが求められる。

「オープンサイエンス」への誤解と反発

「何でもかんでも公開」というものではない

- 統合イノベーション戦略の中では「オープン・アンド・クローズ戦略」を冒頭に示している。
- 最近は「Open」よりも「Sharing」という言葉が好まれている(らしい)

「何を」「誰に」オープンにするか、正しくコントロールされなければいけない。

→ 時代に即した「研究データ管理」の実践
→ 機関は何をサポートすることが求められるか?

「研究データマネジメント」とは?

Research data management concerns the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information.
(From, Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre.
http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm)

「研究データマネジメント」とは?

ある研究プロジェクトにおいて使用された、あるいは生成された情報を、どのように組織化、構造化、保管、管理していくのか、を指す言葉です。

JPCOAR,「RDMトレーニングツール 第1章」より
http://id.nii.ac.jp/1458/00000023/

研究データライフサイクルと研究データ管理計画(DMP)

https://library.sydney.edu.au/research/data-management/research-data-management.html

各ステージで適切な
データの取り扱い方針
(=policy)
と、それを実現するための方法
(=procedure)
をDMP(Data Management Plan)として記述。

これに従い研究データを管理

FAIR原則

- 「データをオープンにする」ための要件
  NDBCの記事(DOI:10.18908/a.2018041901 が簡潔)
  - Findable (データを見つけられる)
  - Accessible (データにアクセスできる)
  - Interoperable (相互利用できる)
  - Re-usable (再利用できる)

- 「誰に対して」を考えると、応用範囲が広がる
  - 「明日の自分」に対してFAIRか?
  - 「身近な共同研究者」に対してFAIRか?
  - 「学術コミュニティ」に対してFAIRか?
  - 「公共、公衆」に対してFAIRか?

大学での組織的な研究データマネジメントの位置づけ

研究スタイルによる研究データのとらえ方の違い

- 学説、手法の正当性を主張するための手段
  - 「新規性を主張する」データを持つことに意義がある
  - データ取得の正しさそのものが議論になる

  - 知財管理、研究公正の観点から研究データマネジメントを進めることが必要

- データを取得・収集することが目的
  - 「他では取れない1点もの」のデータを持つことに意義がある
  - 公開、共有のためのノウハウの蓄積

  - オープンサイエンスの先進的事例として研究データマネジメントが実践されていることが多い

「研究データ管理」の多義性

- 「研究公正とコンプライアンス」のため

- 「学術領域の発展と社会貢献」のため

義務化
形式化
：

自発的
多様性
：

「研究データ管理」の多義性

- 「研究公正とコンプライアンス」のため

義務化
形式化
：

- 「学術領域の発展と社会貢献」のため

自発的
多様性
：

研究公正のためのデータ保存

京都大学での対応
http://www.kyoto-u.ac.jp/ja/research/ethic/research_guide

- 2014(H26)年度
  - 8月: 文部科学省「研究活動における不正行為への対応等に関するガイドライン」
  - 2月:「京都大学における公正な研究活動の推進等に関する規程(※)」
    - 教職員, 部局, 全学の役割, 責任の確認. 研究データ一定保存義務
  - 3月: 日本学術会議「科学研究における健全性の向上について」

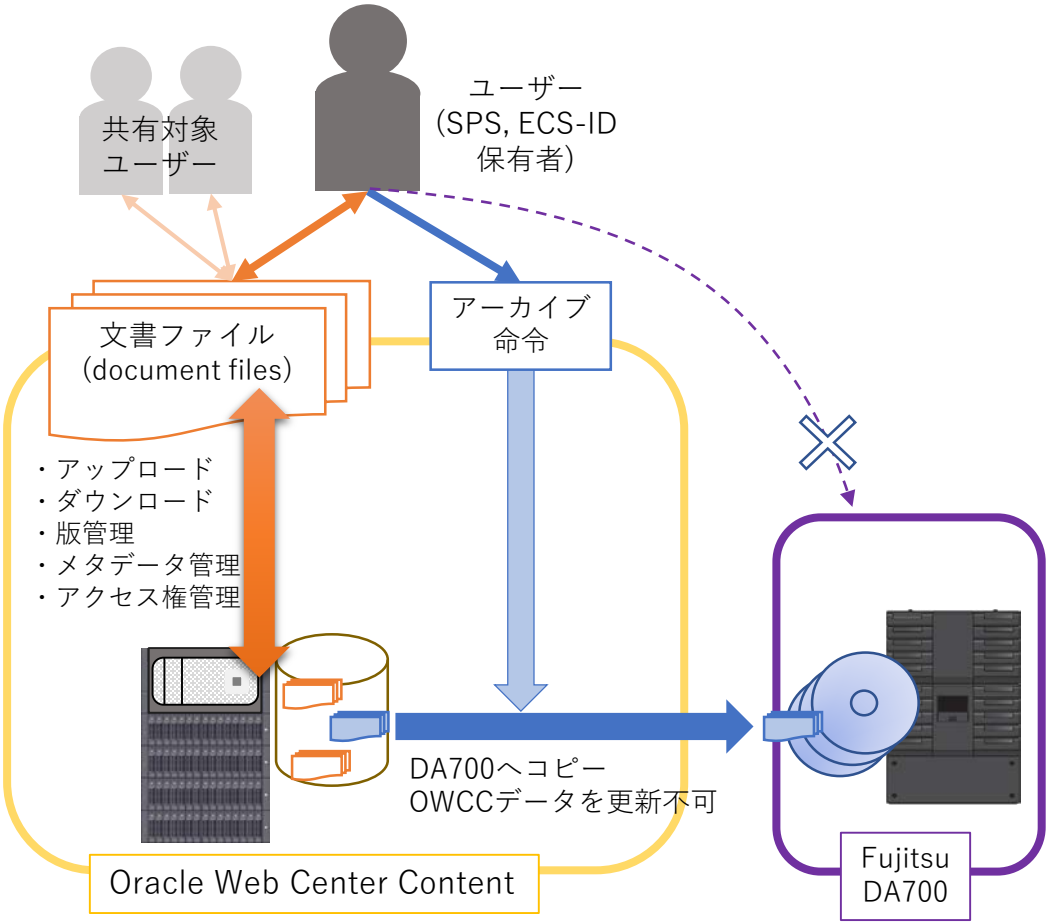
- 2015年度(H27)年度
  - 7月: 「(※)の研究データの保存、開示等について定める件」年度内に管理方針, データ保存計画を, 部局毎に制定

- 2016年度(H28)年度より
  - 部局ごとに決定した管理方針, データ保存計画を元に行動中

ダークアーカイブ型データ保存システム

- エンタープライズコンテンツマネジメントシステム(ECM)と大容量光ディスクの組み合わせ
  - 前者をライトアーカイブ(ユーザーが随時利用できる)
  - 後者をダークアーカイブ(ユーザーは直接利用できない)

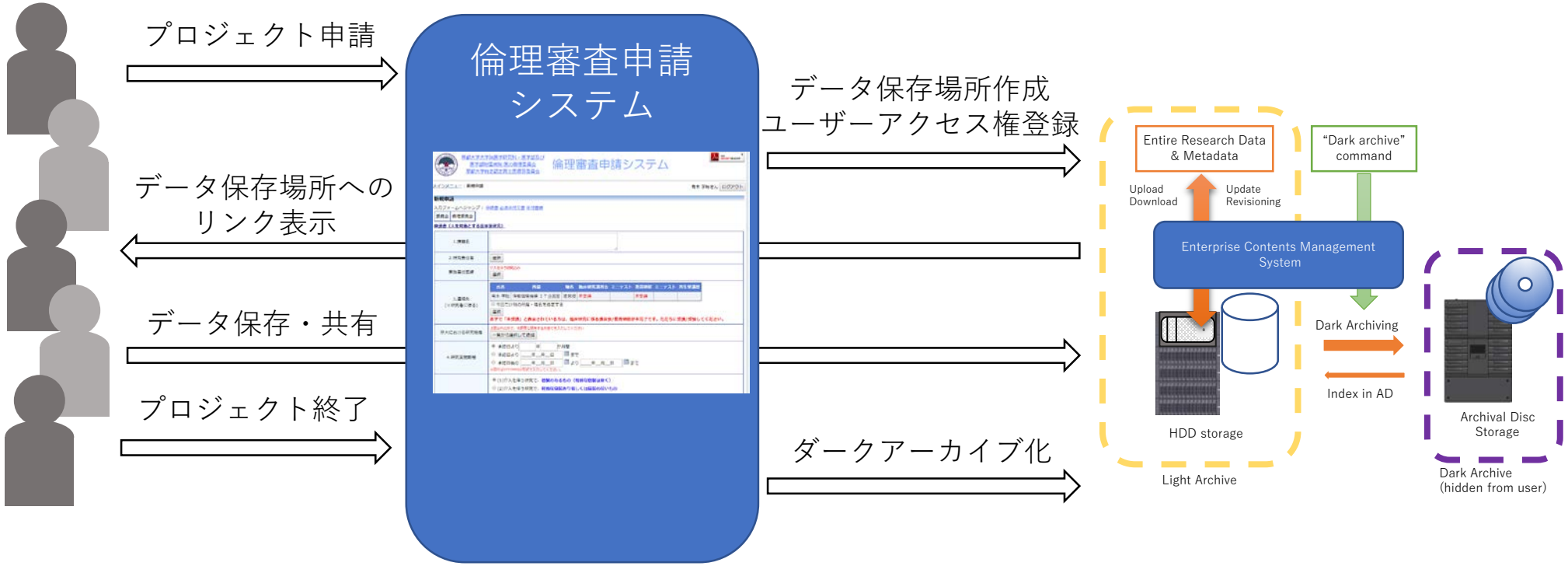
  として異なる保存ポリシーを適用



ライトアーカイブ　　ダークアーカイブ

部局の研究プロジェクト管理システムとの連携

京大病院、医学研究科の臨床研究データ保存(2019.1 本格稼働開始)



「研究公正とコンプライアンス」のための研究データ管理の視点

- データの長期保存ルール(policy)が先行, 現実的な実施手順(procedure)の策定と実施(deployment)が後追い

- 長期保存に対する客観的な質保証は研究者個人の能力をはるかに超える。組織的な対応が後回し

- 残念ながら「ほとんどが死蔵されるデータ」
  - 研究者個人, 限られた研究グループ内で保持するバックアップ最終手段
  - 最悪の場合(研究公正にかかわるインシデント)への備え、というリスク管理の問題

「研究データ管理」の多義性

（「研究公正とコンプライアンス」の視点）

- 「研究公正とコンプライアンス」のため
  - ポリシー先行で、具体的な組織的対応が追い付いていない
  - 機関にとっては「リスク管理」の課題

- 「学術領域の発展と社会貢献」のため

義務化
形式化
：

自発的
多様性
：

「研究データ管理」の多義性

（「学術領域の発展と社会貢献」の視点）

- 「研究公正とコンプライアンス」のため
  - ポリシー先行で、具体的な組織的対応が追い付いていない
  - 機関にとっては「リスク管理」の課題

- 「学術領域の発展と社会貢献」のため（オープンサイエンス的取り組み）

義務化
形式化
：

自発的
多様性
：

京都大学の特色ある学術データベース(1)

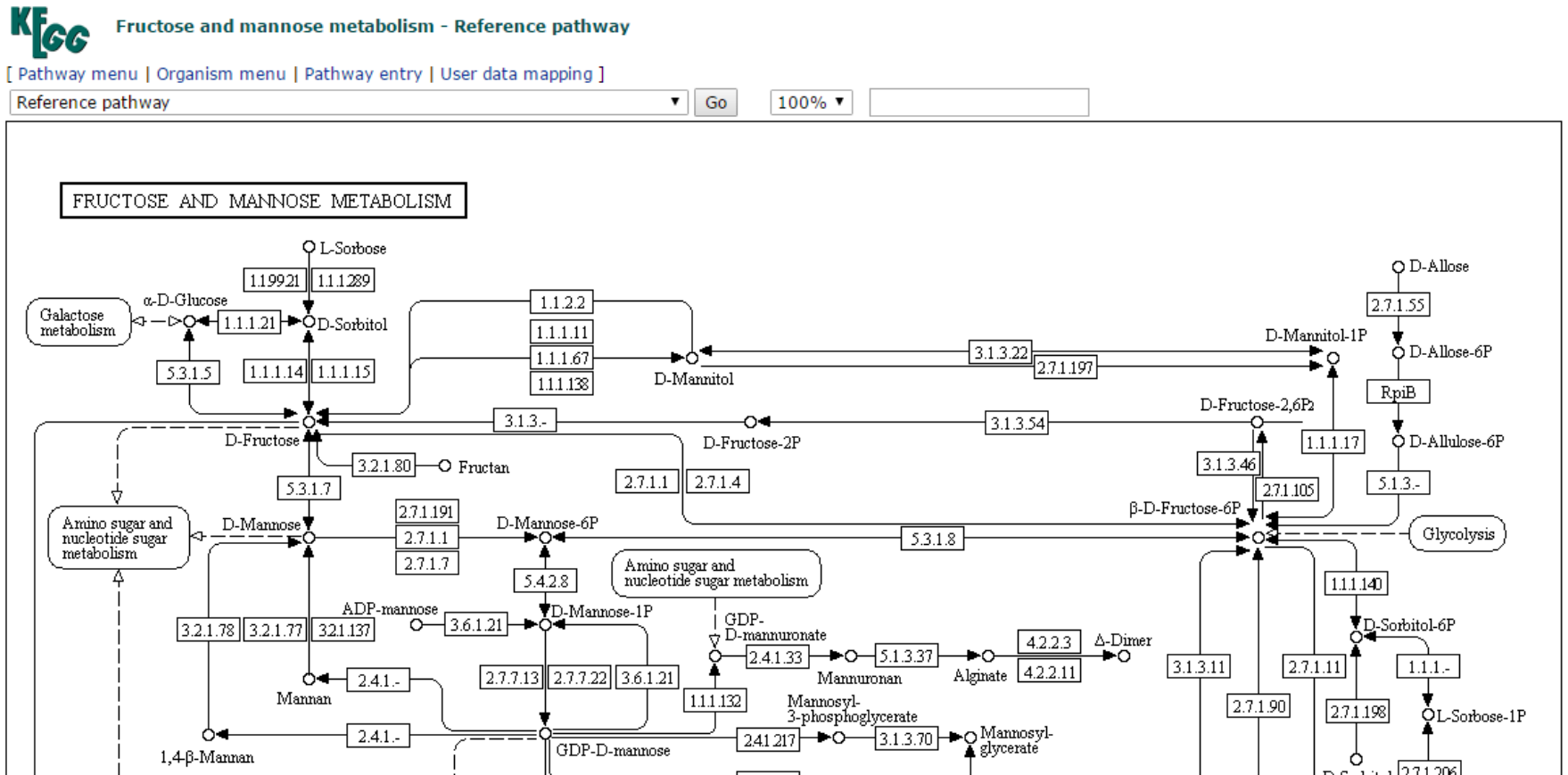
- 貴重資料画像(図書館機構)
  - http://edb.kulib.kyoto-u.ac.jp/exhibit/
- 研究資源アーカイブ(博物館等)
  - http://www.rra.museum.kyoto-u.ac.jp/
- KEGG: Kyoto Encyclopedia of Genes and Genomes (化学研究所)
  - www.genome.jp/kegg
- World Data Center for Geomagnetism, Kyoto (理学研究科)
  - http://wdc.kugi.kyoto-u.ac.jp/index.html
- 地域研究統合DB
  - http://www.cias.kyoto-u.ac.jp/database/
- Etc. etc…

京都大学の特色ある学術データベース(2)

- 貴重資料画像(図書館機構)
  http://edb.kulib.kyoto-u.ac.jp/exhibit/

京都大学の特色ある学術データベース(3)

- KEGG: Kyoto Encyclopedia of Genes and Genomes
  - www.genome.jp/kegg
  - バイオインフォマティクス情報のキュレーションとDB化
  - DBと共にHPCリソースも提供

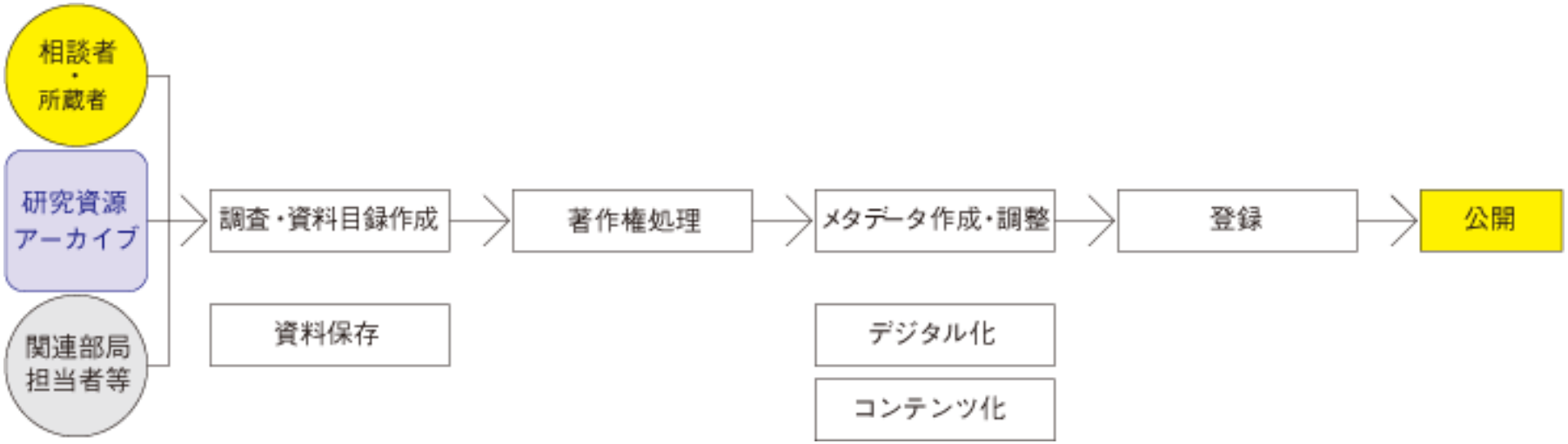


京都大学の特色ある学術データベース(4)

- 京都大学研究資源アーカイブ
  peek.rra.museum.kyoto-u.ac.jp

研究資源アーカイブによる
デジタルデータ取得

- http://www.rra.museum.kyoto-u.ac.jp/activity/
- 博物館スタッフが支援し、デジタイズ、メタデータ付与などを実施



「学術領域の発展と社会貢献」のための研究データ管理の視点

- 学術コミュニティ主導で多大な労力, 時間をかけて醸成
  - 自発的な取り組みでもって, 高い専門性とデータ利活用を実現

- 持続的な発展が大きな課題
  - 機関からの継続的なサポート
  - 余計な手出しは無用

  と相反する思い

- 機関側から見れば、全容を把握できていない
  - 全容がわからないのでサポートできない
  - 大学の「コンテンツ発信力」に興味が持たれない

「研究データ管理」の多義性

（「学術領域の発展と社会貢献」の視点）

- 「研究公正とコンプライアンス」のため
  - ポリシー先行で、具体的な組織的対応が追い付いていない
  - 機関にとっては「リスク管理」の課題

- 「学術領域の発展と社会貢献」のため（オープンサイエンス的取り組み）
  - 主にコミュニティレベルで多大な労力，時間をかけて醸成
  - 持続的な発展の問題
  - 機関にとっては「研究コンテンツ戦略」の課題

義務化
形式化
：

自発的
多様性
：

ロングテールサポートの必要性

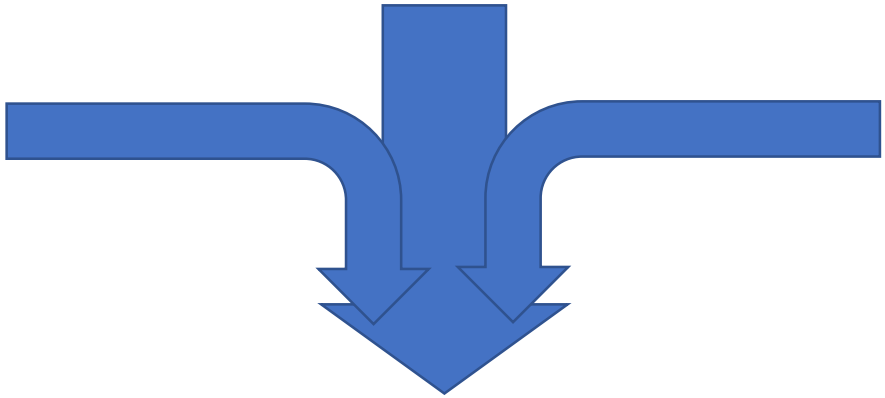
- 研究データの(一部)公開義務化が進むと・・・
- どの様な研究分野であっても、研究データを何らかの方法でAccessibleにしないといけない
  - コミュニティのレポジトリを準備する
    - ある程度確立されたデータ・メタデータ形式、利活用シーン
  - 出版社が用意するデータジャーナル
    - 出版社が学術分野を規定する? コミュニティよりもさらに保守的?
  - ↑から外れた研究データをどのようにサポートするか、 が大学の役割?
    - 機関がデータリポジトリを持つか、サードパーティと契約するか?
- データ保存と同様、低コスト、省力化が大きな課題

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?

「現場任せ」から脱却できるか? する必要があるか?
機関がこれらの課題に対しどのようなソリューションが提供できるか?

ボトムアップからの大学での研究データマネジメント環境の理解

大学の学術データ資産

- どの大学も、何らかの学術データを保有している
- 学術データの利活用は、多くの場合研究者や研究センター任せ
- 大学全体として
  - 誰が
  - 何を
  - どの様に

  学術データを保有しているか、ほとんど理解されていない
- <u>ボトムアップでの連携の必要性</u>

ボトムアップからの研究データマネジメント環境の理解

2016年度から有志による「アカデミックデータマネジメント」検討会を開始

2017年11月に学際融合創造センターの「アカデミックデータ・イノベーションユニット(葛ユニット)」として承認

http://www.cpier.kyoto-u.ac.jp/about/acd/

主な参加者:

- 情報環境機構 ■■■
- 研究推進部研究推進課 ■
- 学術研究支援室 ■■
- 附属図書館 ■■
- 総合博物館 ■
- 学術データレポジトリ運用者 ■
- 他研究者有志■

「コンプライアンス」

来るべき
「オープンデータ化」

「学術領域の発展と社会貢献」

ミッション:

大学における学術研究のライフサイクルに沿った研究データの蓄積・共有・公開および長期保管に必要な研究データマネジメント環境に求められる要求要件のとりまとめ

葛ユニットの活動

- RDM、オープンデータに関する学内、国内外動向調査
- RDMルーブリックの検討、実践
- 学内アンケート
- 学内ヒヤリング
- ワークショップの開催

京都大学研究データマネジメントワークショップ

- 2018/10/6(土)に開催
- California Digital Library (CDL)より Stephanie Simms 氏を招待
- RDMルーブ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ルーブリック(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ス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。

- 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

データマネジメントに関するアンケート

- 内容はごく簡単
  - 研究データ（データセット、データベースもしくはデータベース化したいデータ）持っていますか?
  - (あるならば)
    データセットの名称
  - 分野、目的
  - データ形式、容量
  - キーワード
  - 管理主体
  - データの公開状況(今後の公開の意図も含む)

データマネジメントに関するインタビュー

- アンケート回答者から選出

- 研究分野により、データのとらえ方、課題は様々
  - 論文投稿時にデータの公開を義務付けられたが、大学としてサポートがあるか?
  - センシティブな内容(政治、思想、犯罪)に関するインタビュー調査はどのように管理するのが安全か? ガイドラインもないが、大学を信頼してデータを預けてよいのか。

第2回京都大学研究データマネジメントワークショップ

2019年02月28日木曜日午前10時～午後5時 (終了後，懇親会を予定)
京都大学理学研究科セミナーハウス
（http://www.kyoto-u.ac.jp/ja/access/campus/yoshida/map6r_n.html の10番）

10:00 - 12:00 報告「京都大学におけるデータマネジネントの現状」（仮）
アンケート調査・ヒアリング調査報告
京大データマップ・ルーブリック初版説明
ワークショップ (会場からの意見を収集)

12:00 - 13:30 ランチ／Lunch

13:30 - 14:30 招待講演「研究データマネジメントの現状と課題」
村山泰啓（国立研究開発法人情報通信研究機構・研究統括）

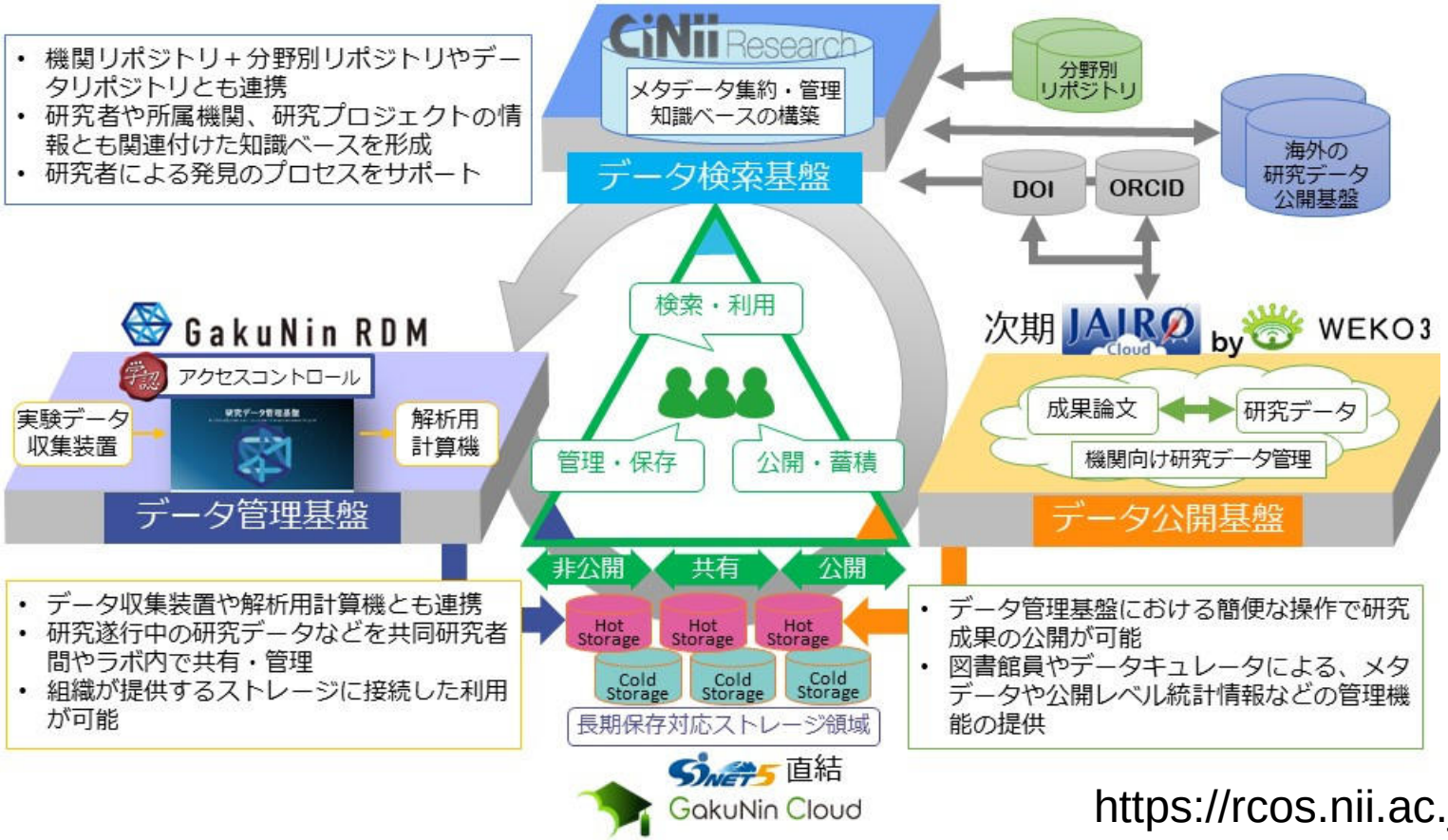
14:50 - 16:50 パネルディスカッション
「京都大学における研究データマネジメントのあり方」

16:50 - 17:00 まとめ「京都大学におけるアカデミックデータマネジメント環境」
に関する 提言策定に向けて

その他の話題

NIIでのRDM基盤開発

RCOSによるサービスの概要



https://rcos.nii.ac.jp/service/

JPCOAR(オープンアクセスリポジトリ推進協会)によるRDMトレーニングツール

- 2017年度: RDMに関する包括な教材作成
  - http://id.nii.ac.jp/1458/00000023/
    「RDMトレーニングツール」で検索
- 2018年度:
  オンライン講座「研究データ管理サービスの設計と実践」の開発中
  - 研究者を支援するスタッフ向けに、研究者の研究プロセスに沿ってRDMサービスを設計し、それを実践するノウハウを学ぶための教材

その他国内コミュニティの動き

- 大学ICT推進協議会(AXIES)
  - 研究データマネジメント部会を設立(2017)
  - 「学術機関における研究データ管理に関する提言」を作成中(11/19-21 の年次大会にてアルファ版公開)
- 図書館関係(オープンアクセスレポジトリ推進協議会)
  - 「論文のオープンアクセス」から「データのオープンアクセス」を検討する段階に
  - “統合イノベーション戦略についての見解” https://jpcoar.repo.nii.ac.jp/?page_id=74#_href_243
- 研究データ利活用協議会にて(https://japanlinkcenter.org/rduf/)
  - データマネジメント計画
  - 研究データライセンス検討プロジェクト
  - 等
- RA協議会
  - 研究支援の在り方としての、データ保存、オープンサイエンス対応を課題にしつつある?

まとめ

- オープンサイエンスと研究データ管理
  - 「何を公開するか」を含めて「研究データ管理」を実施することの必要性
- 大学における組織的な研究データ管理の位置づけ
  - 「ガバナンス、コンプライアンス」の視点から、「研究成果を生かすデータマネジメント」へ
- ボトムアップからの大学での研究データマネジメント環境の理解
  - トップダウンとボトムアップの調整役が今後重要に
- その他の話題

付録: RDM ルーブリック ワークショップのまとめ

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ルーブリック(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ス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。

ルーブリック以前の問題
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、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」
  - PI, 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)