

RDMトレーニングツールの開発

西園 由依

鹿児島大学学術情報部

オープンアクセスリポジトリ推進協会作業部会員（研究データTF）

概要

2016年度研究データTF

■活動内容

RDMトレーニングツールの開発

ケーススタディによる研究データ管理ノウハウの蓄積

■メンバー

尾城 孝一

東京大学附属図書館 (主査)

山地 一禎

国立情報学研究所 (副主査)

前田 翔太

北海道大学附属図書館

三角 太郎

千葉大学附属図書館

天野 絵里子

京都大学学術研究支援室

大園 隼彦

岡山大学附属図書館

西園 由依

鹿児島大学学術情報部

南山 泰之

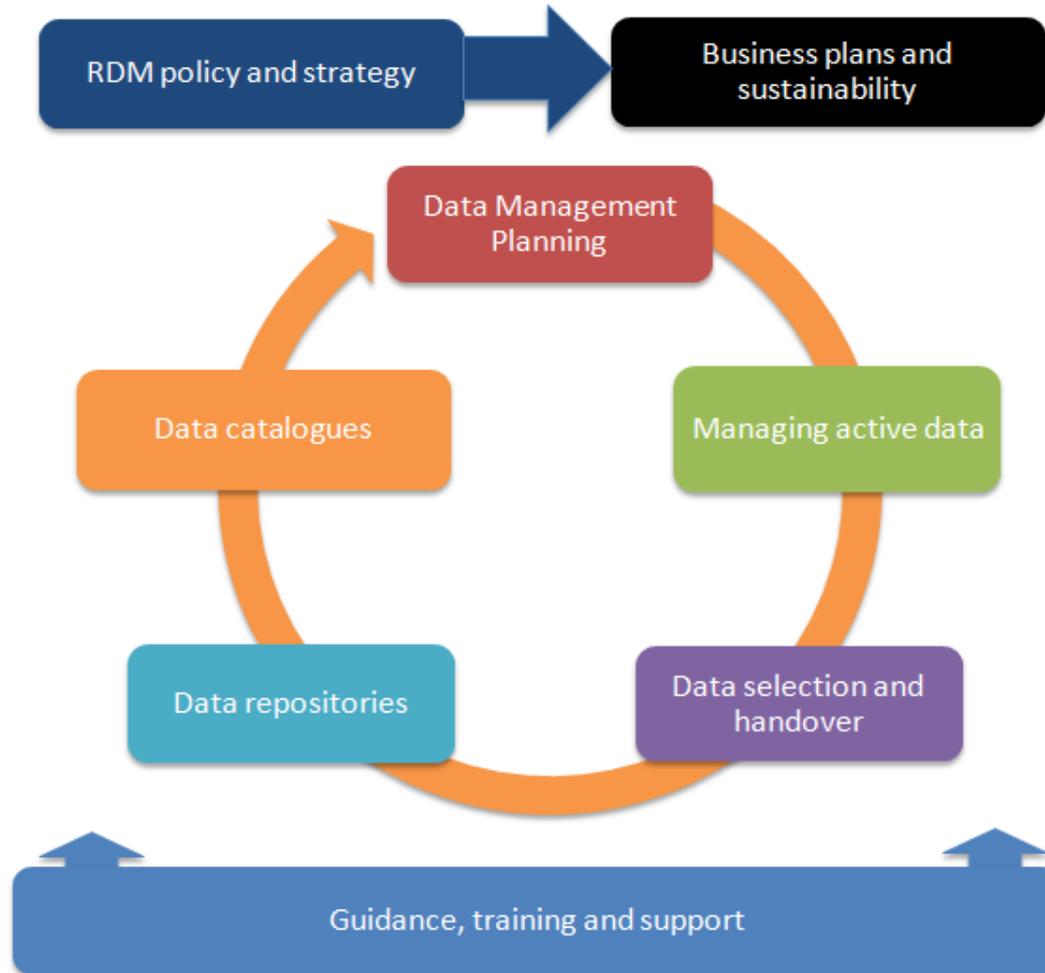
国立極地研究所情報図書室

RDMトレーニングツールの開発

- 目的
 - 学習者が研究データ管理（RDM: Research Data Management）に関する基礎的な知識を習得する
 - 各機関におけるRDMサービス構築の足掛かりを得る
- 内容
 - 全7章から構成されるeラーニング教材（MOOCにて提供）
 - 動画講義、理解度確認テスト、ワーク
- 受講対象者
 - 図書館、IT部門、研究支援部門の職員（URA）等
 - オープンサイエンスや研究データ管理に関心のある教職員

機関による研究データ管理サービス

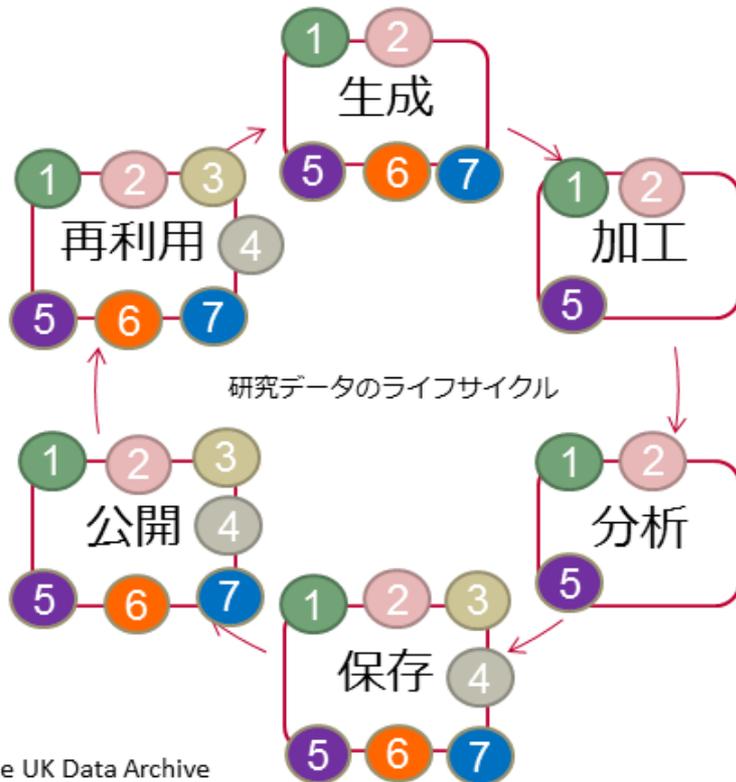
研究データの一貫した計画的な管理をサポート



<http://www.dcc.ac.uk/resources/developing-rdm-services>

内容紹介

データのライフサイクルと教材の構成



- The UK Data Archive
(<http://www.data-archive.ac.uk/create-manage/life-cycle>)
- RDM Support - basic training course for information specialists
(<http://dx.doi.org/10.6084/m9.figshare.1285313>) より

- 1章 導入
- 2章 データ管理計画
- 3章 保存と共有
- 4章 組織化、文書化、メタデータ作成
- 5章 法・倫理的問題
- 6章 ポリシー
- 7章 研究データ管理サービスの設計

第1章 導入

- 研究データ管理の重要性が増している背景や、研究データ、研究データ管理の定義について

1.2.1 研究データとは

- 研究データ収集・生成されたり、結果が正し
- 電子的なデータの情報も
- 研究データり、記録さ
多岐にわた

1.2.3 研究データの具体例

- 観測データ (Observations)
 - 自然現象等を観測し取得さ
 - 例：天体観測、気象観測、
- 実験データ (Experimental)
 - 実験によって得られたデー
 - 例：DNAシークエンス、ク
- シミュレーション (Simulation)
 - モデルやコードを用いて生
 - モデルやコード、パラメ
 - 例：気象モデル、経済モ
- 派生データ・編集されたデータ
 - 取得したデータを加工、お
 - 例：データマイニング、マ
- 参照/標準データ (Reference)
 - 研究の過程で参照されるデータ
 - クション。整理・共有され
 - 例：遺伝子配列データバ

1.3.2 研究データ管理のメリット

- 研究効率
 - 研究効率の上昇
 - 長期的に時間や資源を節約
 - データセキュリティの強化とデータ消失リスクの最小化
 - データを他の研究者が再利用することにより、労力の重複を避ける
 - ある時点のデータを他の時点と比較できるデータを提供することにより、データの変化の分析を容易にする
- ポリシー
 - 資金提供元の要求を満たす
 - 産業界や商業界の慣行に従う
- 研究倫理
 - 研究の再現性の確立
 - 研究データや研究記録の正確性、完全性、真正性、信頼性の保障

第2章 データ管理計画

- 効果的なデータ管理に欠かせないデータ管理計画に関し、作成義務化の動向や、その構成要素について

2.1.1 データ管理計画 (DMP) とは

- 「研究...
であ...
有の...
れる...
委員...
め」
- 「研究...
がど...
この...
切に...」
- 「研究...
の前...」

2.1.2 データ管理計画策定の意義

- 計画を策定することは適切なデータ管理を行うための重要な要素である。
- 研究の公正性と再現性を確保する。
- 自分の研究データやソフトウェアのセキュリティを確保する。
- データのセキュリティを確保する。
- 研究の初期段階から、適切なデータ管理に関与することで、データ管理に関する課題を事前に発見し、解決策を講ずることができる。
- これらは研究の質や効率性を向上させる。
- 助成機関等の意向 (2017年度から) に不可欠なものになる。

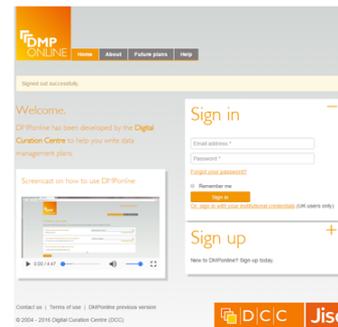
2.3.1 何をどう計画するのか

- チェックリスト (Digital Curation Centre (DCC))
<http://www.dcc.ac.uk/resources/data-management-plans/checklist>

ツール例

- DMPTool (カリフォルニア大学)
<https://dmptool.org/>
- DMPOnline (DCC)
<https://dmponline.dcc.ac.uk/>

誰でも無料で
アカウント作成
可能



と
に
関
す
る
こ
と

J P C O A R
オープンアクセスリポジトリ推進協会

第3章 保存と共有

- 研究データの、研究期間中の保管や長期保存に関する留意点について
- 研究データの共有に関し、その意義や検討すべき点、共有方法について

3.1.1 保存場所

- ▶ 研究資料（研究データ）を失うことは研究の進展に大きく影響
- ▶ ファイルの意図しない削除

3.1.2 バックアップ

研究を円滑に進めるためにも、バックアップを適切に実施することは不可欠。不用意な害にも対応可能。

3.1.3 セキュリティ

- ▶ 物理的なセキュリティ対策
 - ▶ 離席する際のPCをロック
 - ▶ 部屋の施錠
 - ▶ モバイルPCを放置しない
 - ▶ USBメモリ等を紛失しないための対策
- ▶ PC上でのセキュリティ対策
 - ▶ アンチウイルスソフトウェア/ファ
 - ▶ ソフトウェアやOSのアップデート
 - ▶ 適切な強度のパスワードや認証方法
 - ▶ ファイルやハードディスク、USBメモ
 - ▶ 例 WindowsのBitLockerやMacのFi
 - ▶ など、機関のセキュリティポリシー
- ▶ 個人情報・プライバシー保護
 - ▶ 個人情報やプライバシー情報などの
 - ▶ する取り扱い方法を明確にした上で
 - ▶ 例 高等教育機関の情報セキュリティ
 - ▶ (C2103 情報格付け基準) <http://w>

3.2.2 フォーマット

える場合、適切なファイルフォーマットを選択す

(.pdf, .jpg)は、ファイルフォーマットに関する情報を提

フォーマット
独立したフォーマット
るフォーマット
ット

とにより、ファイル内部に記録されたメタデータや画
情報を失う可能性大。

依存せず開くことができ人間可読性高。
ッチな情報を包埋していることが多いが人間可読性低。

3.3.1 オープン/クローズ

- ▶ 研究データを公開することの意義
 - 1 研究成果の透明性や公正性の確保
 - 2 研究資金助成機関からの要求への対応
 - 3 研究成果への直接的な引用や再利用によるインパクト
 - 4 新たな研究コミュニティの創出や産学連携への発展
 - 5 派生的な成果や新たな解釈への発展によるイノベーションの創出
 - 6 適切に管理し保存することによる自身のための再利用性の向上
 - 7 若手研究者や学生に対するデータの収集や解析に関する教育的効果
 - 8 市民や納税者からの研究に対する理解の向上
- ▶ 研究データを非公開とする必要性

- 1 知的財産的な価値がある研究成果に対する戦略的な占有
- 2 複数の関係者が知的財産権を保持したり秘匿義務がある場合
- 3 個人情報やプライバシー情報などのセンシティブな情報を含む場合

単にオープン化を推進するのではなく、機関、研究者自身、研究コミュニ
ティなどでオープン・クローズ戦略を十分に検討することが重要

第4章 組織化、文書化、メタデータ作成

- 研究データを長期的に管理・活用するために欠かせない、一定のルールに則ったデータの組織化や、データについて説明する文書やメタデータの作成について

4.1.2 ファイルを組織化する際の3つの判断基準

1. ファイル名の構成

4.1.3 ファイル命名規則の例

2. ファイルの構成

- 特殊記号は用いない。&*\$%\$#@はOSで特別な意味を持つ。

3. 文書化

4.2.1 データについて文書で記録する意味

- 1. 自分自身の覚えのため
研究を実施しているが、時間の経過により忘れやすくなる
- 2. 他の研究者の再利用
研究者が既存のデータ
例えば、先行研究の結果を再現する、同

4.3.2 様々なメタデータの標準

- 一般的なもの : Dublin Core、DataCite
- 分野特有のもの : Data Documentation Initiative (DDI, 社会科学)
DCCのwebサイトで検索可能

選択の基準

- 誰が収集しているデータはどのようなタイプか？
- 研究グループ内の慣習は何か？
- メタデータはどこに登録予定か？

4.4.3 データ引用の方法

- 研究で参照したデータセットは、他の利用者がアクセスできるように適切な情報を含まなければならない。
- DataCite は、以下の最小限の引用要素を推奨
Creator (PublicationYear): Title. Publisher. Identifier
- Identifier は、DOI や URL 等の一意に識別するもの。
- 可能ならば Version と ResourceType も追加する。
Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier
- UK Data Service は、主題を表すタイトルに加えて、データがカバーする地理的情報と時間的範囲を示すことを推奨。

は分野特有のメタデータより一般的なメタデータ

J P C O A R
オープンアクセスリポジトリ推進協会

第5章 法・倫理的問題

- 研究データをめぐる著作権や、再利用を促進するためのライセンスの仕組みについて
- センシティブデータを取り扱う上での留意点や研究倫理について

5.1.2 データの著作物性

著作物の定義：

「思想又は感情を創
学術、美術又は音楽の
法第2条第1項第1号

- ▶ 単なる客観的事実や
ない
- ▶ データを得るために
たとしても、保護対

もつとも、
一定の考え方のもとに
「創作性」が認められ
例：データベースの著

5.2.1 ライセンシング

- ▶ ライセンシング（利用許諾）とは
 - ▶ コンテンツの利用許諾の条件を明示すること。
- ▶ 趣旨
 - ▶ 利用許諾に関する曖昧さの回避
 - ▶ 使ってよい範囲を明確にしておくことで、再利用の促進
 - ▶ 各国で異なる著作権ルールを標準化
- ▶ 種類
 - ▶ Creative Commons License
 - ▶ <https://creativecommons.jp/licenses/>
 - ▶ Open Data Commons
 - ▶ <http://opendatacommons.org/licenses/>
 - ▶ Open/Non-Commercial Government Licence
 - ▶ <http://www.nationalarchives.gov.uk/doc/non-commercial-government-licence.htm>
 - ▶ Public Domain

5.4.1 センシティブデータとは

▶ センシティブデータとは？

1. 個人情報
 - ▶ 種族・民族に関するもの
 - ▶ 政治的意見・宗教に関するもの
 - ▶ 労働組合での活動など
 - ▶ (肉体的・精神的な)健康状態に関するもの、性に関するもの
 - ▶ 守秘義務に係るもの(例：被弁護人に係る情報)
2. 国家安全保障に関するもの
3. 知的財産権
4. その他特に配慮を必要とするもの
例：絶滅危惧種の生息地情報

J P C O A R
オープンアクセスリポジトリ推進協会

J P C O A R
オープンアクセスリポジトリ推進協会

第6章 ポリシー

- 国や助成団体、機関、出版者等による、研究データの保存や共有を求めるポリシー策定の動向や要件について

6.1.2 研究者が遵守を求められる三つのポリシー

- ▶ 1. 研究機関のポリシー
 - ▶ 大学、研究所等
- ▶ 2. 配分機関のポリシー
 - ▶ 研究機関に対し
- ▶ 3. 学会・出版者等

6.1.3 研究機関における研究データ管理の二つの流れ

- ▶ 1. オープンサイエンスの流れ

「我が国におけるオープンサイエンス」
「オープンイノベーションに資する研究データの利活用」

- ▶ 2. 研究公正の流れ

「研究活動における不正行為への対応」
研究データは検証が目的
→研究データをオープンにしているところがあるであろうが、多

1と2では、オープン化の研究公正の流れのポリシーを策定するのであ

6.1.6 想定される課題

- ▶ 各機関・各部署で論文公表の実態のモニタリングはできているのか？
- ▶ 研究データの管理はシステム化されているのか
 - ▶ データの検索は？（大規模大学では数千本の論文が公表されるが、その根拠データをどのように管理？）
 - ▶ 公表論文とデータの紐づけは？

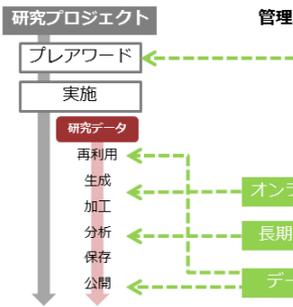
【研究支援担当者側で検討できそうなこと】

- 研究データ管理ポリシーの具体化をはかるためには、研究成果のモニタリングおよびデータベース化が必要になってくるであろうし、そこでオープンサイエンスとの課題の共有ができるのでは？
- 研究データは分野により性質があまりに異なるため、部局レベルの運用ガイドラインで、論文および根拠データ（実験ノートまで含めるのであれば根拠資料）は、機関リポジトリで管理、とすることは可能か？
 - ← 現在の機関リポジトリシステムでは、データ保全機能（凍結・非改変・データのバージョン管理・タイムスタンプ等）が不十分か？

第7章 研究データ管理サービスの設計

- 前章までの内容を踏まえ、学習者が自機関での研究データ管理サービスを構築していくためのステップについて

7.1.1 研究データ管理サービスとは？



- 研究機関が研究者に対してサービスから成り立っている
- 自機関の状況に合わせてサービスを設計する必要がある。

7.2.4 サービスの設計：調査(1)

- サービス設計の前にはまず調査を実施し、の実態やニーズを把握する

目的

- 研究者が取り扱っているデータの実状を定量的に把握する
- 研究者の研究データ管理に対する予備知識や意識を計画に活かす。

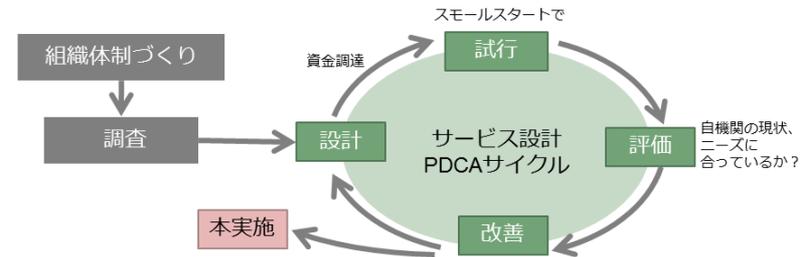
推奨される方法

- アンケート調査
- インタビュー調査

参考

- DAF (Data Asset Framework) = 研究データを資産として実施例がある。
<http://www.data-audit.eu/>
http://www.data-audit.eu/docs/DAF_Implementation.html
- RDM Rose 2015 > 3.1 DAF surveys / 2.6 Interview
http://rdmrose.group.shef.ac.uk/?page_id=1061

7.2.5 サービスの設計：試行、評価



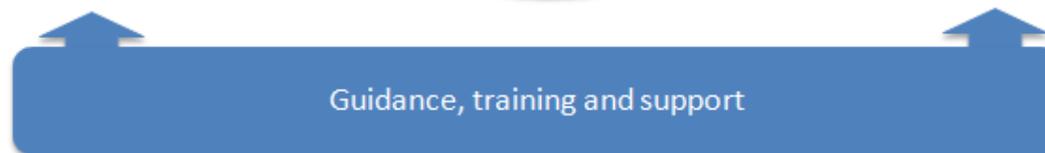
- 研究データ管理サービスの定着は長期的計画で
- オンラインストレージなど管理基盤の提供（ハード）、相談窓口やトレーニングなどの人的支援（ソフト）ともに、試行と評価、改善をくりかえしながら、自分の機関に合ったサービスを構築し、組織内の意識を高めていくことが重要

機関による研究データ管理サービス

研究データの一貫した計画的な管理をサポート



各機関がその実情に即して
自分の機関に合わせた独自のサービス設計を行うことが重要
↓
本トレーニングツールがその一助となることを期待



<http://www.dcc.ac.uk/resources/developing-rdm-services>

今後の展開と課題

- RDMトレーニングツールの公開
 - JMOC/gaccoでの開講へ（2017.11予定）
 - 教材素材の公開（JPCOAR Webサイトにて6/6公開）
<http://id.nii.ac.jp/1458/00000023/>
 - スライド、スクリプト
 - 再利用可能な形での提供 
- RDMトレーニングツールの活用促進
- 持続可能な教材開発体制の確立
 - 内容のアップデート
 - RDMトレーニングツールの段階的拡充