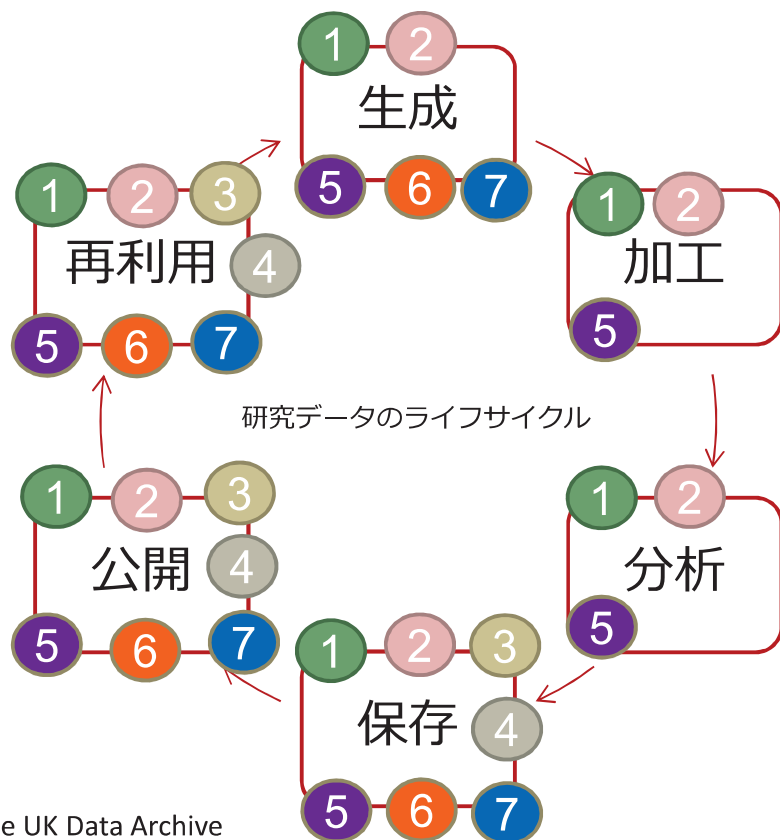


研究データのライフサイクルと 本コンテンツとの関わり



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

1. 導入

Introduction

1.1 背景

1.1.1 オープンサイエンスの潮流(1)

オープンサイエンスとは、**オープンアクセスと研究データのオープン化（オープンデータ）を含む概念である**。オープンアクセスが進むことにより、学界、産業界、市民等あらゆるユーザーが研究成果を広く利用可能となり、その結果、研究者の所属機関、専門分野、国境を越えた新たな協働による**知の創出を加速し、新たな価値を生み出していくことが可能**となる。また、オープンデータが進むことで、社会に対する**研究プロセスの透明化や研究成果の幅広い活用**が図られ、また、こうした協働に市民の参画や国際交流を促す効果も見込まれる。さらに、研究の基礎データを市民が提供する、観察者として研究プロジェクトに参画するなどの新たな研究方策としても関心が高まりつつあり、市民参画型のサイエンス（シチズンサイエンス）が拡大する兆しにある。近年、こうした**オープンサイエンス**の概念が世界的に急速な広がりを見せており、**オープンイノベーション**の重要な基盤としても注目されている。

<http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>
第5期科学技術基本計画 本文 より

1.1.1 オープンサイエンスの潮流(2)

こうした潮流を踏まえ、**国は、資金配分機関、大学等の研究機関、研究者等の関係者と連携し、オープンサイエンスの推進体制を構築する。**公的資金による研究成果については、その利活用を可能な限り拡大することを、我が国のオープンサイエンス推進の基本姿勢とする。その他の研究成果としての研究二次データについても、分野により研究データの保存と共有方法が異なることを念頭に置いた上で可能な範囲で公開する。

<http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>
第5期科学技術基本計画 本文 より

1.1.2 不正対策

- ▶ 『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』
(文部科学省)*₁
 - ▶ p8「このことから、研究機関において、**研究者に対して一定期間研究データを保存し、必要な場合に開示することを義務付ける**旨の規程を設け、その適切かつ実効的な運用を行うことが必要である。」

- ▶ 『科学研究における健全性の向上について』 (日本学術会議)*₂
 - ▶ p8「**資料（文書、数値データ、画像など）の保存期間は、原則として、当該論文等の発表後10年間とする。**電子化データについては、メタデータの整理・管理と適切なバックアップの作成により再利用可能な形で保存する。」

*1 http://www.mext.go.jp/b_menu/houdou/26/08/___icsFiles/afieldfile/2014/08/26/1351568_02_1.pdf より

*2 <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-k150306.pdf> より

1.1.3 助成機関のポリシー

研究資金の助成機関によっては、研究資金提供の条件として、研究データ管理を文書化したデータ管理計画 (Data Management Plan / DMP) の提出を義務付けているところもある。

e.g.) イギリスの助成機関における資金提供ポリシー一覧

助成機関	ポリシーの範囲		ポリシーの条項					支援体制			
	論文等	データ	時限	DMPの提出	共有	長期的整備・保存	監視	ガイダンス	リポジトリ	データセンター	諸費用の負担
AHRC	●	●	●	●	●	●	●	●	●	●	●
BBSRC	●	●	●	●	●	●	●	●	●	●	●
CRUK	●	●	●	●	●	●	●	●	●	●	●
EPSRC	●	●	●	●	●	●	●	●	●	●	●
ESRC	●	●	●	●	●	●	●	●	●	●	●
MRC	●	●	●	●	●	●	●	●	●	●	●
NERC	●	●	●	●	●	●	●	●	●	●	●
STFC	●	●	●	●	●	●	●	●	●	●	●
Wellcome Trust	●	●	●	●	●	●	●	●	●	●	●

●はポリシーでカバーしている、●は部分的にカバーしている、●はカバーしていないことを示す。
<http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies> より
 (2017/03/12現在)

1.2 研究データとは

1.2.1 研究データとは

- ▶ 研究データとは研究の過程で、あるいは研究の結果として収集・生成される情報である。仮説を検証するために使用されたり、結論を導くための根拠となる。また、後に研究結果が正しいかどうか確認するために使用されることもある。
- ▶ 電子的なデータのみ研究データと定義することも、紙ベースの情報も研究データに含める場合もある。
- ▶ 研究データのとり形式はテキストや3Dモデルなど様々であり、記録されるものも研究ノートや写真、音声テープなど多岐にわたる。

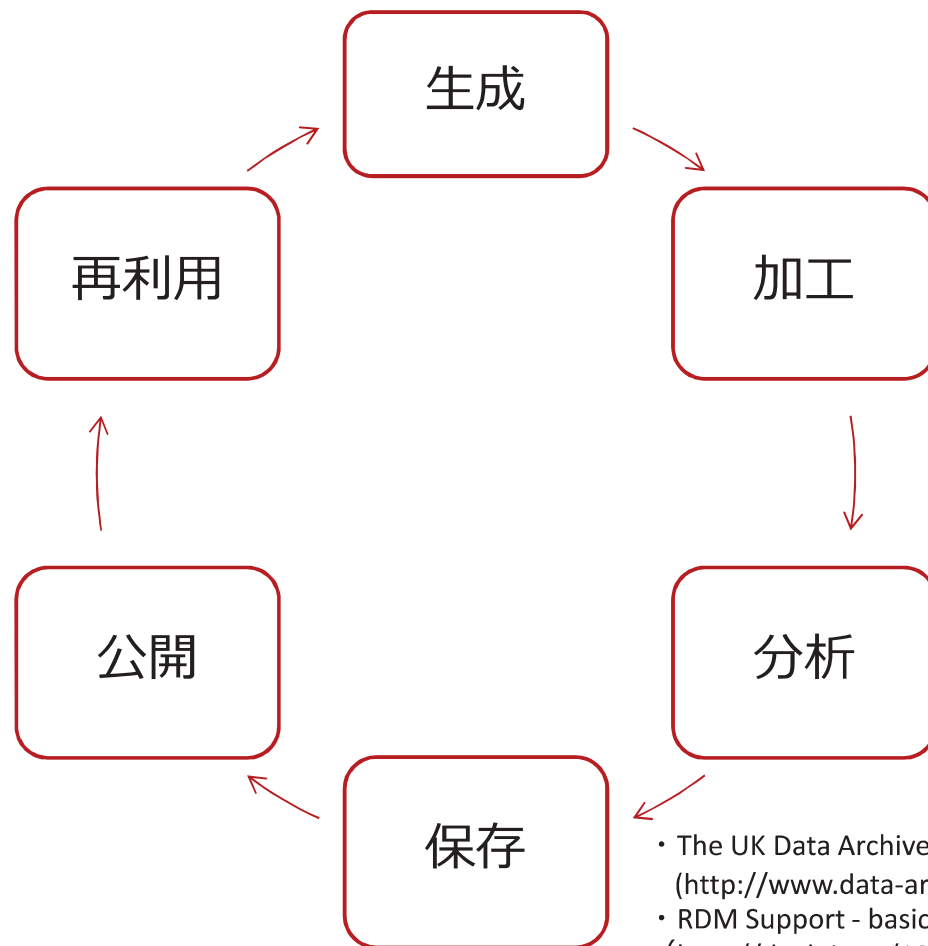
1.2.2 研究データを構成するもの

- ▶ 文書ファイル、スプレッドシート
- ▶ 研究ノート、フィールドノート（野外調査の記録）、日誌
- ▶ アンケート、転写物、コードブック
- ▶ 音声テープ、ビデオテープ
- ▶ 写真、フィルム
- ▶ 試験・検査への反応、応答
- ▶ スライド、人工物、標本、サンプル
- ▶ 研究の過程で獲得され、生成されたデジタル資料のコレクション
- ▶ 統計的なデータファイル
- ▶ データベースのコンテンツ（ビデオ、音声、テキスト、画像）
- ▶ モデル、アルゴリズム、スクリプト
- ▶ アプリケーションソフト
- ▶ 方法論、ワークフロー
- ▶ 標準的な操作手順

1.2.3 研究データの具体例

- ▶ 観測データ (Observational)
 - ▶ 自然現象等を観測し取得されたデータ。
例：天体観測、気象観測、社会調査
- ▶ 実験データ (Experimental)
 - ▶ 実験によって得られたデータ。高価になることもある。
 - ▶ 例：DNAシーケンス、クロマトグラム、微量検定法により得られたデータ
- ▶ シミュレーション (Simulation)
 - ▶ モデルやコードを用いて生成されたデータ。得られたデータに加え、使用されたモデルやコード、パラメータセットも重要。
 - ▶ 例：気象モデル、経済モデル
- ▶ 派生データ・編集されたデータ (derived/compiled)
 - ▶ 取得したデータを加工、あるいは結合することによって得られたデータ。
 - ▶ 例：データマイニング、テキストマイニング、3Dデータ
- ▶ 参照/標準データ (Reference or canonical)
 - ▶ 研究の過程で参照されるデータ、データセット。あるいは標準的なデータのコレクション。整理・共有されているものも多い。
 - ▶ 例：遺伝子配列データバンク、結晶データベース、歴史画像アーカイブ

1.2.4 研究データのライフサイクル



- 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

1.2.5 研究と研究データ

- ▶ 研究データはしばしば、その研究データが得られた研究プロジェクトよりも寿命が長い。研究者は資金援助が終わった後もデータに関する作業を続け、後続く研究ではそのデータが分析されたり、新たに追加され、さらには他の研究者によって再利用されることもありうる。
- ▶ 組織化、文書化、保存・共有がしっかりとされたデータは科学研究を推進することや、学習及びイノベーションの機会を増やすことにおいて計り知れない価値がある。

<https://www.ukdataservice.ac.uk/manage-data/lifecycle> より

1.3 研究データ管理とは

本ツールでは、電子的なデータのみを対象として、その扱い方について学んでいきます。

1.3.1 研究データ管理 (Research data management) とは

「研究データ管理」とは、一般的に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。

研究データ管理には次のようなことが含まれる：

- ▶ どのようにデータを取り扱っていくかの計画（Data Management Plan）策定 > 第2章も参照
- ▶ 研究プロジェクトが行われている間の、日々の情報の取り扱い方
- ▶ 長期的にデータをどのように取り扱うか（研究プロジェクトが終わった後はそのデータをどうするのか）

<http://researchdata.ox.ac.uk/home/introduction-to-rdm/> より

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

1.3.3 研究データ管理と大学図書館

- ▶ すでに海外の大学ではデータ管理計画作成等の研究データ管理業務に関する支援サービスを図書館や他の関連組織が共同で行っているところもある
 - ▶ 例 1 : バージニア大学
 - ▶ <http://data.library.virginia.edu/data-management/>
 - ▶ 例 2 : エディンバラ大学
 - ▶ http://www.ed.ac.uk/files/atoms/files/rds_booklet_may2016.pdf
 - ▶ 例 3 : ケンブリッジ大学
 - ▶ <http://www.data.cam.ac.uk/support>
 - ▶ 例 4 : アムステルダム大学
 - ▶ <http://rdm.uva.nl/en>

> 第7章も参照

2. データ管理計画

Data Management Plan (DMP)

2.1 データ管理計画とは

2.1.1 データ管理計画（DMP）とは

- ▶ 「研究プロジェクト等における研究データの取り扱いを定めるものであり、具体的にはデータの種類、フォーマット、アクセス及び共有のための方針、研究成果の保管に関する計画などについて記載されるもの」（文部科学省科学技術・学術審議会学術分科会学術情報委員会（2016）「学術情報のオープン化の推進について（審議まとめ）」用語解説より）
- ▶ 研究の実施段階から終了後に至るまでの期間において、研究データがどのように生成、管理、共有、保存される予定か、を文書化する。
- ▶ この計画に従った管理を行い、研究データが利活用可能な状態で適切に管理されるようにする。
- ▶ 研究データの適切な保管・管理は、研究データの公開を進めるための前提である。

2.1.2 データ管理計画策定の意義

- ▶ 計画を策定することにより、研究全体を通して、具体的で適切なデータ管理を行える。適切なデータ管理は、
 - ▶ 研究の公正性と再現性を保証する。
 - ▶ 自分の研究データや記録が、正確、完全、真正で、信頼できることを保証する。
 - ▶ データのセキュリティを高め、データ損失のリスクを最小化する。
- ▶ 研究の初期段階から、データ管理に関する事柄を検討することで、データ管理に関する入念な準備を行える。
- ▶ これらは研究の質や効率性を高めるのに役立つ。
- ▶ 助成機関等の意向（2.2.1参照）に関わらず、データライフサイクルに不可欠なものとしてデータ管理計画策定に取り組むべき。

2.2 データ管理計画をめぐる 国内外の動向

2.2.1 データ管理計画策定の義務化

- ▶ 国際的に、助成機関等が助成の要件として、データ管理計画の策定を求めるようになってきている。

例：

- ▶ 英国 <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>
- ▶ 欧州 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- ▶ 米国 <https://dx.doi.org/10.6084/m9.figshare.1367165.v7>
- ▶ 日本 JST CREST・さきがけ等（平成28年度～） http://senryaku.jst.go.jp/teian/koubo/data_houshin.pdf

＞ 第1章、第6章も参照

2.2.2 データ管理計画策定の支援

- ▶ 海外では、助成機関等からのデータ管理計画策定義務化方針を受け、機関による研究データ管理サービスの一環として、大学図書館がデータ管理計画策定の支援に取り組む例が見られる。
- ▶ 例 1 : バージニア大学
<http://data.library.virginia.edu/data-management/>
- ▶ 例 2 : エディンバラ大学
http://www.ed.ac.uk/files/atoms/files/rds_booklet_may2016.pdf
- ▶ 例 3 : ケンブリッジ大学
<http://www.data.cam.ac.uk/support>
- ▶ 例 4 : アムステルダム大学
<http://rdm.uva.nl/en>

2.3 データ管理計画の実際

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>
- ▶ 計画書に含める可能性のある情報：
 - ① 管理上のデータ
 - ② データ収集に関すること
 - ③ 文書化とメタデータに関すること
 - ④ 倫理・法律のコンプライアンスに関すること
 - ⑤ 保管とバックアップに関すること
 - ⑥ 選定と保存に関すること
 - ⑦ データ共有に関すること
 - ⑧ 責任とリソースに関すること

①管理上のデータ

- ▶ 助成機関や所属機関が定めた関連IDは？
- ▶ 助成機関・助成金番号は？
- ▶ プロジェクト名称・内容・連絡先は？
 - ▶ データ収集・生成の目的
- ▶ PIもしくは主要研究者名、その個人識別子は？
- ▶ データ管理計画書の作成日・最終更新日は？
- ▶ 関連するポリシーは？

考慮すべき問題：

- ▶ 何か既存の手続きに則っているか。
- ▶ 所属部局等はデータ管理ガイドラインを有しているか。 > 第5章、第6章も参照
- ▶ 所属機関はデータの保護やセキュリティに関するポリシーを有しているか。 > 第3章も参照
- ▶ 所属機関や助成機関は研究データ管理ポリシーを有しているか。 > 第5章、第6章も参照
- ▶ 何らかの公式規格を採用したか。 > 第4章も参照

②データ収集に関すること

- ▶ どのようなデータを収集・生成するのか > 第3章も参照

考慮すべき問題：

- ▶ データの種別・フォーマット・容量
- ▶ 選択したフォーマット・ソフトウェアはデータの共有・長期アクセスに適しているか。
- ▶ 再利用できる既存のデータはあるか。

- ▶ どのようにデータを収集・生成するのか > 第4章も参照

考慮すべき問題：

- ▶ どの標準・方法論を用いるのか。
- ▶ フォルダやファイルをどのように構造化するのか。
- ▶ バージョン管理をどのように行うのか。

③ 文書化とメタデータに関すること

- ▶ データをどのように文書化し、どのようなメタデータを付与するのか

> 第4章も参照

考慮すべき問題：

- ▶ データが将来にわたって可読性を有し、解釈されうるためには、データにどのような情報が必要とされるか。
- ▶ どのように文書化し、メタデータを作成するか。
- ▶ どのメタデータ標準を用いるか。また、その理由は何か。

④倫理・法律のコンプライアンスに関すること

▶ 倫理的問題にどのように対処するのか

> 第5章も参照

考慮すべき問題：

- ▶ データの保存や共有について同意を得たか。
- ▶ もし必要な場合、参加者の個人識別性をどのように保護するのか。
(例えば匿名化する)
- ▶ センシティブデータが確実に安全に保管・移動されるようにするため、どのような対処をするか。

▶ 著作権や知的財産権に関する問題にどのように対処するのか

> 第5章も参照

考慮すべき問題：

- ▶ データの所有者は誰か。
- ▶ 再利用のためどのようなライセンスを付与するか。
- ▶ 第三者による再利用について何らかの制限があるか。

⑤ 保管とバックアップに関すること

- ▶ 研究期間中、データをどのように保管・バックアップするのか考慮すべき問題：
 - 十分なストレージを有しているか。あるいは、付加サービス利用に必要な料金を含める必要があるか。
 - データをどのようにバックアップするか。
 - 障害発生時、データはどのようにリカバリされるか。
- ▶ アクセス制限やセキュリティ確保をどのように管理するのか

> 第3章も参照

⑥選定と保存に関すること

▶ どのデータを保持・共有・保存すべきか

考慮すべき問題：

- ▶ 契約上、もしくは法的、もしくは規制上の理由で、保持あるいは破棄しなければならないデータは何か。
- ▶ 保存するその他のデータをどのように決定するか。
- ▶ データはどの程度の期間、保持・保存するか。

▶ データセットの長期保存計画はどのような内容か

考慮すべき問題：

- ▶ どこに（リポジトリやアーカイブ等）データを保存するか。

> 第3章、第5章も参照

⑦データ共有に関すること

- ▶ どのように共有するか
- ▶ データ共有に何らかの制限が必要か

> 第3章も参照

⑧責任とリソースに関すること

- ▶ 責任とリソースに関すること
 - ▶ データ管理の責任者は誰か
 - ▶ 計画を実行するために、どのようなリソースが必要か

2.3.2 データ管理計画作成ツール

- ▶ 無料で使えるデータ管理計画作成ツールを利用すると、作成の負担が少ない。
- ▶ テンプレート等を用いて、各助成機関等の要件に則った計画書を容易に作成できる。
- ▶ 共有・公開機能により、他者の作成した計画書を参考にすることもできる。

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

The screenshot shows the DMPTool website homepage. The header includes the DMPTool logo and navigation links: Home, DMP Requirements, Public DMP's, News, Help, Contact Us, About, and Log In. The main content area features a large green graphic with the text "Data Management Planning Tool" and a sub-headline "Create, review, and share data management plans that meet institutional and funder requirements." A "Get Started" button is prominently displayed. Below this, there are three columns of content: "PUBLIC DMPs" with a list of sample plans, "DMPTOOL NEWS" with the latest information and a "More News" button, and "DMPTOOL HELP" with an overview of the tool and a "View All" button. The footer contains copyright information for the University of California and the Digital Library, along with social media icons.

The screenshot shows the DMPOnline website homepage. The header includes the DMP ONLINE logo and navigation links: Home, About, Future plans, and Help. A yellow banner at the top says "Signed out successfully". The main content area is divided into two sections: "Welcome" on the left, which states "DMPOnline has been developed by the Digital Curation Centre to help you write data management plans", and "Sign in" on the right, which includes a login form with fields for "Email address" and "Password", a "Remember me" checkbox, and a "Sign in" button. Below the login form is a "Sign up" section with a plus sign and the text "New to DMPOnline? Sign up today". A video player is visible in the middle section, titled "Screenshot on how to use DMPOnline". The footer includes contact information, terms of use, and logos for DCC, Jisc, and AR (Association of Research Librarians).

2.3.3 データ管理計画のサイクル

研究の初期段階

- ▶ 計画の作成

研究の途中段階

- ▶ 計画の実行
- ▶ 必要に応じて計画の修正

研究終了後

- ▶ 計画達成状況について検証