



JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE

日本学術振興会

2021年度統計関連学会連合大会
2021年9月7日13:00-15:00 B会場

人文学・社会科学分野における
データインフラストラクチャー構築推進事業の展開

[2BPM1-02]

人文学・社会科学におけるデータ共有のための手引きについて

独立行政法人日本学術振興会

人文学・社会科学データインフラストラクチャー構築推進センター研究員(PO)

前田幸男(東京大学大学院情報学環・教授)

1. 背景
2. 海外機関における手引き
3. 日本学術振興会における作成過程
4. 想定される読者とコミュニティ形成
5. 今後の課題

*本報告では、前田の個人的見解を含めて、お話をさせていただきます。

1. 背景

- 政策的背景

- OECD. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*.
- OECD. 2015. *Making Open Science a Reality*.

- Open data の定義

- “In a nutshell, open data are data that can be used by anyone without technical or legal restrictions. The use encompasses both access and reuse.”
(OECD 2015)

FAIR Principle

- FAIR Principle (Wilkinson et al. 2016).
 - Findable (見つけられる)
 - Accessible (アクセスできる)
 - Interoperable (相互運用できる)
 - Reusable (再利用できる)
- 研究者がデータをFAIRにするための『手引き』
 - 後述のCESSDAのガイドには以下の説明
“This guide is designed by European experts to help social science researchers make their research data Findable, Accessible, Interoperable and Reusable (FAIR).”

2. 海外機関における先例

- ICPSR
 - *Guide to Social Science Data Preparation and Archiving* (1997, 2000, 2005, 2009, 2012, 2021).
- UK Data Archive
 - *Managing and Sharing Data* (2009, 2011).
 - *Managing and Sharing Research Data: A Guide to Good Practice*, Sage (2014, 2019).
- Finnish Social Science Data Archive (FSD)
 - *Data Management Guidelines* (no date)
- Consortium of European Social Science Data Archives
 - *CESSDA Data Management Expert Guide* (2017, 2019, 2020)

日本における翻訳

社会科学データの準備と保存のための手引き

データのライフ・サイクルを通じた最良の実践



ICPSR | SSJDA

ICPSRのGuide 3rd (2005)の和訳(2007)
ICPSRの許可した公開期限が切れたため現在は閲覧不可



研究者向け最良事例

2011年5月(英語版)

SSJDA 社会調査・データアーカイブ共同利用・共同研究拠点
東京大学 社会科学研究所
附属社会調査・データアーカイブ研究センター

2013年2月(日本語版)

「近年我が国でも官庁統計を含む社会調査データの保存と公開についての関心が高まっています。しかし、具体的にデータの保存と公開に必要な手続きや準備すべき資料について解説している文書は極めて少ないのが現状です。これは英語でも日本語でも事情は変わりません。ICPSR がGuideを編集したのは、社会科学における調査データの保存と公開のために必要な技術に関する情報をデータ寄託者に提供するためです。」

3. 日本学術振興会における作成過程

日本学術振興会『人文学・社会科学におけるデータ共有のための手引き』の作成と公開

- 事業の一環として『手引き』を作成
- データの共有・利活用をより一層進めることによって、日本における人文学・社会科学分野の研究が国内的にも国際的にもさらに進展することを目的とする。

具体的な進め方

- 運営委員会の下に策定に係る作業部会を設置
 - 事前準備：海外の複数の機関のガイドラインを参考に『手引き』の構成を検討
 - 2019年10月から社会科学分野を策定開始
 - 2020年4月から人文学分野をコラム（一部章立て）として策定開始
 - 2020年11月～12月にかけて、関連学協会等へ意見照会を実施
 - 意見照会に基づく修正を経て、2021年6月に完成
 - まもなく電子的に公開予定

作成過程に関与した専門家

(作業部会委員8人)

- 社会学・社会調査
- 統計学・教育心理学
- 政治学・投票行動論
- 図書館情報学・情報工学
- 法学・行政法
- 人文情報学
- 図書館情報学
- 医療社会学・研究倫理

(JSPSセンター)

- 統計学・経済学
- 経済統計学
- 図書館情報学
- 政治学

各分野の研究者だけではなく、図書館情報学の専門家にも協力を求めた。

海外ガイドの目次構成比較

| | |
|----|-------------|
| 1 | メタデータ |
| 2 | ファイルフォーマット |
| 3 | バックアップの作成 |
| 4 | データのセキュリティ |
| 5 | データの匿名化 |
| 6 | 著作権 |
| 7 | データ管理計画 |
| 8 | 役割と責任 |
| 9 | データのフォーマット |
| 10 | 品質保証 |
| 11 | バージョン管理と信頼性 |
| 12 | 法的小よび倫理的問題 |
| 13 | アクセス管理 |

| | |
|----|--------------|
| 14 | データ管理の費用計算 |
| 15 | データの文書化 |
| 16 | 参考資料 |
| 17 | データアーカイブ |
| 18 | データライフサイクル |
| 19 | インフォームドコンセント |
| 20 | 定量データ |
| 21 | 定性データ |
| 22 | データストレージ |

| | |
|--|--------------|
| | ・・・ 4 機関項目あり |
| | ・・・ 3 機関項目あり |
| | ・・・ 2 機関項目あり |

当初の目次構成(案)

UKDAに近い案

直感的な
配列

- 1 序文
- 2 データを保存・共有する意義
- 3 データ管理計画
- 4 メタデータ
- 5 データのフォーマット
- 6 データの保管
- 7 研究倫理
- 8 匿名化
- 9 データに関する著作権
- 10 データ・アーカイブの役割
- 11 文献リスト

こちらを採択

FSDに近い案

データのライフサイ
クルに基づく配列

- 1 序文
- 2 データを保存・共有する意義
- 3 データ管理計画
- 4 データに関する著作権
- 5 研究倫理
- 6 データのフォーマット
- 7 データの保管
- 8 匿名化
- 9 メタデータ
- 10 データ・アーカイブの役割
- 11 文献リスト

特徴(1)DMPの導入

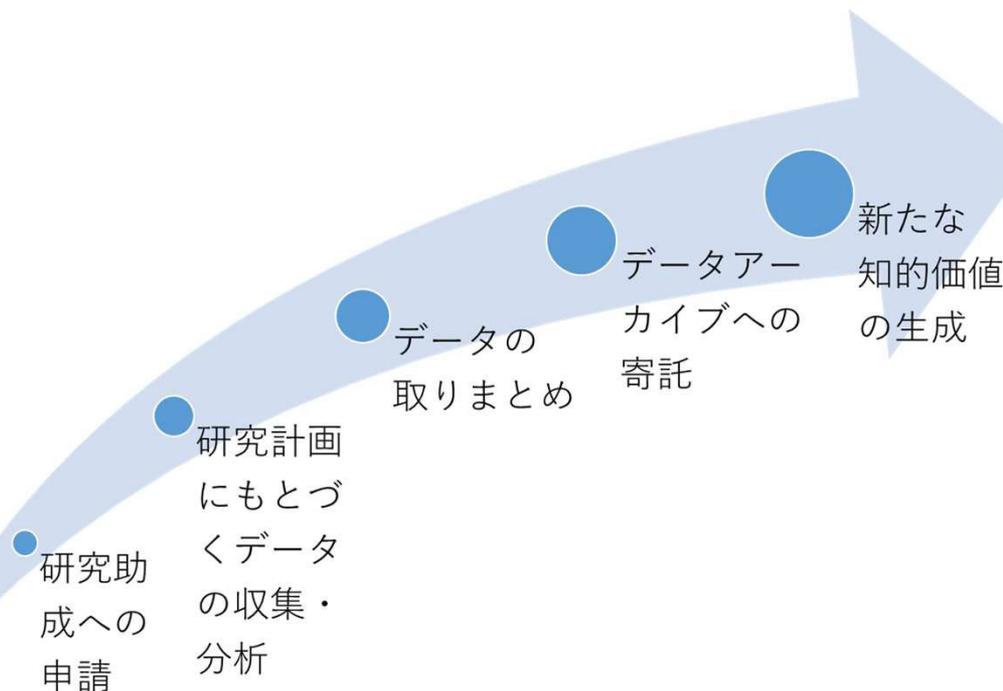
- データ管理計画 (Data Management Plan : DMP) の考え方を導入
- 研究プロジェクトにおいてデータをいかに管理するかを**事前に**定めた計画のこと、具体的には以下の事項等を定める
 - データの種類
 - フォーマット
 - アクセス及び共有方針
 - 研究成果やデータの所在と保管責任者
 - 長期保存と二次利用に必要なとなる文書の特定
- 当然研究の進捗に合わせて調整することはある。

* 国内ではAMEDやJST等が、データ管理計画の提出を要求している。JSPSも2024年度科研費以降、DMPを求める予定。

特徴(2) データのライフサイクルの考えを導入

- 研究費助成への申請から、第三者によるデータの二次利用による新たな知的価値の生成までをデータのライフサイクルとよぶ。

- データ管理計画の導入と表裏一体
- 研究の各段階において実施すべき作業を明確にする。
- 目的は事前準備・計画による作業の効率化



特徴(3)従来は触れられることが少なかった論点

- メタデータについての解説
 - 社会科学者向けの解説は稀
- データの保管
 - ソフトウェア、ハードウェアの両面から解説
- データ共有の文脈における研究倫理
 - 事前の研究倫理審査やインフォームドコンセントにおけるデータ共有の扱いを議論

特徴(4)日本の制度を前提にした説明

- 個人情報と匿名化
 - 個人情報保護法等について解説
- データの著作権
 - 日本における法制度・著作権の考え方を前提に解説
- 公的統計についても必要に応じて補足

4. 想定される読者とコミュニティ形成

- 読者の世代は若い研究者や大学院生を想定
 - 授業の副教材としての利用も想定
 - 研究データの収集を開始する前に読んで欲しい
- 読者の分野は以下を想定
 - 人文学(主に歴史的文献資料を扱う分野)
 - 社会科学(主に社会調査・統計調査を扱う分野)
 - 図書館情報学分野
 - 研究データの利活用に興味がある人

データ利活用のコミュニティ

- データの保存と利活用に関連したコミュニティは伝統的な学問分野の垣根を越えたもの
 - 社会科学、人文学、図書館情報学等の垣根が急激に下がっている。
- データ利活用基盤を支える専門家のコミュニティ
 1. 研究者と協力してデータの受け入れ・整理を行うドメインのデータ・アーキビスト
 2. データの長期保存、提供に従事するデータ・ライブラリアン
 3. データ提供の技術的基盤を支えるIT専門家

5. 今後の課題

- 執筆者について
 - 今回は各研究分野の専門家が執筆
 - 海外機関のものは全てデータアーカイブの専従スタッフが執筆
- 社会科学分野と人文学分野のバランス
 - 人文学の扱いが部分的
 - 人文学で扱うデータの幅は社会科学よりも広い
 - むしろ人文学と社会科学を分冊にするべきか

今後の課題(2)

- 長期的な維持と改訂
 - ICPSRのGuide はおおよそ4年毎に改訂
- 未対応の分野
 - 質的データ
 - テキストデータ、フィールドノート、インタビューのスク립ト etc.
 - 録音・動画
 - ソーシャルメディアデータ

参考文献

- CESSDA Training Team (2017 - 2019). *CESSDA Data Management Expert Guide*. Bergen, Norway: CESSDA ERIC. DOI: 10.5281/zenodo.3820473
- Corti, Eynden, Bishop, and Woollard. 2019. *Managing and Sharing Research Data, 2nd edition*. Sage.
- Eynden, Corti, Woollard, Bishop, and Horton. 2011. *Managing and Sharing Data*. <http://data-archive.ac.uk/media/2894/managingsharing.pdf> (日本語訳、SSJDA『データの管理と共有』 <http://csrda.iss.u-tokyo.ac.jp/UKDAguide.pdf>)
- FSD. n.d. Data Management Guidelines. <https://www.fsd.tuni.fi/en/services/data-management-guidelines/>
- ICPSR. 2021. *Guide to Social Science Data Preparation and Archiving, 6th edition*. <https://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- OECD. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. (日本語訳、日本学術振興会2020『公的資金による研究データへのアクセスに関する原則およびガイドライン』 <https://www.jsps.go.jp/j-di/data/jigyo/OECDGuidelines.pdf>)
- . 2015. *Making Open Science a Reality*.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

ご静聴ありがとうございました。

質問・感想等は前田までご送付下さい。

ymaeda@iss.u-tokyo.ac.jp