

社会調査データのオープン化に向けたデータマネジメント ～イギリスの事例との比較から～

田中康裕

大学共同利用機関法人情報・システム研究機構

データサイエンス共同利用基盤施設

社会データ構造化センター

Email: y-tanaka@ism.ac.jp

要旨

近年、オープンサイエンスやオープンデータ化の潮流の中で、学術研究データの公開が進められている。我が国でも、2025年度から公的研究助成を受けた学術兼研究について、学術論文とその根拠データの公開が義務化される。これに伴い、研究データの管理計画（DMP）の作成が求められ、研究者はデータの適切な管理と公開を行う責務を負うようになる。本稿では、イギリスの事例を参考に、日本における社会調査データの公開の在り方を検討し、データライフサイクルマネジメントの観点から、研究データのマネジメントと公開の在り方を検討する。

キーワード

オープンデータ、データライフサイクルマネジメント、オープンサイエンス、データマネジメントプラン、データアーカイブ

1. はじめに

近年オープンサイエンスやオープンデータ化の世界的な潮流の中で、社会調査データに限らず、公的資金を受けた学術研究に関して、その成果である学術論文の公開だけではなく、学術データに関しても、公開を義務づける方針がEUやアメリカを中心に進められている。このデータ公開にあわせて、研究プロジェクトにおけるデータの管理や公開に関する計画をまとめたDMP（Data Management Plan）を作成し、研究助成機関へ提出することが研究助成の必須要件として求められている。また、DMPに基づき学術研究データを適切に管理し、研究データの公開を滞りなく遂行することが研究者の責務とされている。

我が国においては、今年の2月に内閣府科学技術・イノベーション推進事務局が公表した「我が国のオープンサイエンス政策について」[1]において、科研費など競争的資金による助成を受けた研究に関して、2025年度新規公募分から学術論文等の即時オープンアクセスの実現とその根拠データのリポジトリ登録の義務化する方針が示された。

この内閣府の方針を受けて日本学術振興会（JSPS）でも科研費に関する対応として、2024年度新規採択課題及び継続課題から原則全種目においての作成を義務づける方針が示され

た[2]。

しかしながら、我が国における学術研究データの公開やそれに向けた DMP 作成の義務化は、開始の途についたばかりで、まだ課題も多く、研究者も戸惑うことが多いものと想定される。

そこで、本稿では、我が国に先行して、公的研究助成制度の中で学術研究論文のオープンアクセス化や研究データの公開の義務化が進展するイギリス、特に社会科学系研究に関する Economic and Social Research Council (ESRC) の研究助成制度とデータ公開の義務化の取組から、我が国における社会調査データ公開の在り方を検討する。

2. 日本における学術研究データ公開への取組と DMP の位置づけ～JSPS 科研費を事例として～

我が国では、2025 年度新規公募課題からの競争的研究助成による学術論文の即時オープンアクセスと根拠データの公開に先駆けて、科研費での DMP 作成の義務化が 2024 年度新規採択課題及び継続課題から実施されることとなった。

科研費に関する DMP 作成の方針に関しては、2023 年時点では、交付申請時に提出が必要とされていたものが、提出不要となり、また、2024 年度新規採択分が対象としていたものが、継続課題についても対象とするなど、変更が行われている。今後の展開として、科研費の報告書の一部として、取得した研究データのメタデータを提出することが求められることとなる（図 1）。

DMP 作成やメタデータ提出に関する方針も、義務化やその開始時期など調整される可能性はあるが、JSPS は、公的研究助成の成果である論文のオープン化と、研究データのリポジトリ登録を通じた公開を進めていく方針を図 1 に示すように公開をしている。

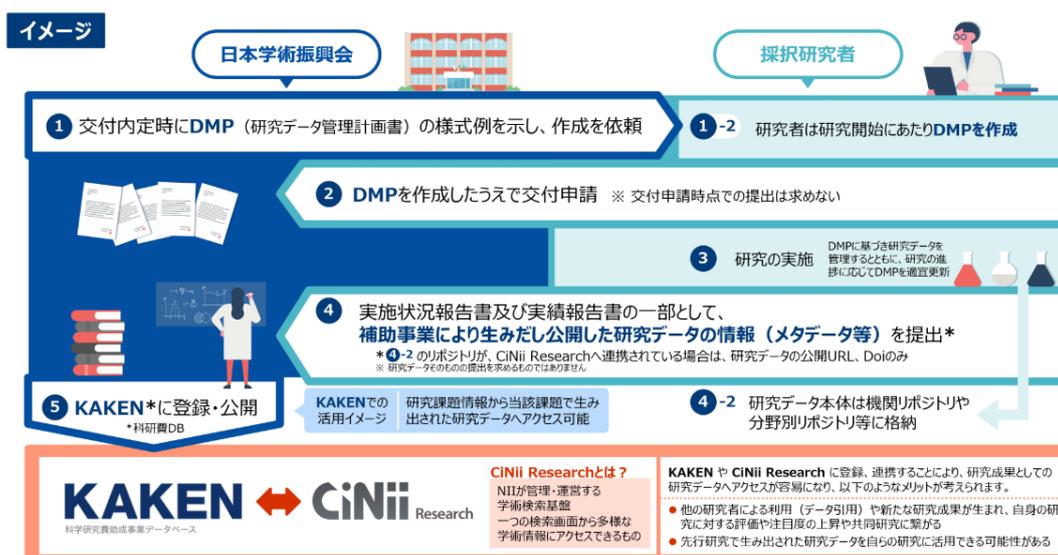


図 1 科研費での研究成果のオープンアクセス・オープンデータ化に向けた方針[3]

当初公開されていた方針やスケジュールから調整を行いつつ、概ね 2024 年度から DMP の作成を義務づける運用が開始されている。。2024 年度より作成が義務づけられた研究データの公開に向けた DMP 作成については、2024 年度以前に採択された継続課題も対象とする方針が明示されている。ただ一方で、DMP 作成に関する FAQ[3]の中で、継続課題に関して下記に示すように令和 6 年以降に実施する研究内容を対象として、過去に遡っての作成は不要ということが明示されている。

Q1. 令和 6 年度に実施の課題が DMP 作成の対象とのことですが、令和 6 年度以前に採択された継続課題も対象ですか。

A1. DMP の作成は、研究者による研究データの適切な管理や効率的な研究進捗の把握を目的としています。そのため、新規課題、継続課題問わず、DMP を活用し、研究データの適切な管理や利活用の促進に努めてください。なお、対象は令和 6 年度以降に実施する研究です。(過去に遡って作成する必要はありません)

JSPS, 「科研費における研究データの管理・利活用について」FAQ より抜粋

DMP 作成の義務化に関する方針において、「過去に遡っての作成は不要」という意味をどのように捉えるかというのは、我が国における DMP の位置づけを示すものとして、検討すべき課題であると考えられる。

この課題を検討していくために、まず、我が国、特に JSPS が定める科研費において DMP を作成する目的をどのようにとらえているのかを概観する。DMP に関して公開されている資料では、DMP について、『研究過程において、どのような「研究データ」を、どのように管理・利活用するか等について、整理した文書』と定義されている。この方針に基づき、科研費で作成が義務づけられる DMP の記載様式例では、研究データの管理体制と、公開方針等を記載することが求められている (図 2)。

図 2 は科研費の DMP 様式例にある記載項目をもとに階層図を作成したものである。主な記載項目は、研究データに関する情報として、名称や概要を記載し、管理体制として取得者や収集者、調査の責任者と全体の管理責任者、管理するデータに個人情報など機微情報が含まれる場合には、その対応を記載することが求められる。

このような研究データの管理情報に加えて、研究データの公開方針として、公開する対象となるデータと、非公開とするデータの区分、公開するデータについては、さらに公開方法や公開予定日等を記載することが求められている。

世界的に公的資金による助成を受けて遂行される学術研究に関し、研究データの公開と公開に向けた DMP 作成と提出を義務づける動きが顕著になっている。そのような世界的な潮流の中で、我が国においても、図 2 に示すようなデータ公開に向けたデータの管理者とデータの公開方針を記載することを主目的とする DMP の作成を義務づける制度が開始された。しかしながら、本来 DMP に求められるのは、データの公開だけではなく、研究プロ

セス全体を通して、研究データをどのようにマネジメントするかを記載することである。この考え方が、公的研究助成において DMP 作成を義務づける世界各国の制度・取組では主流となっている。研究プロセスの適正な管理が研究の公正性の担保だけでなく、(質の高い)データ公開の促進に繋がるという前提である。

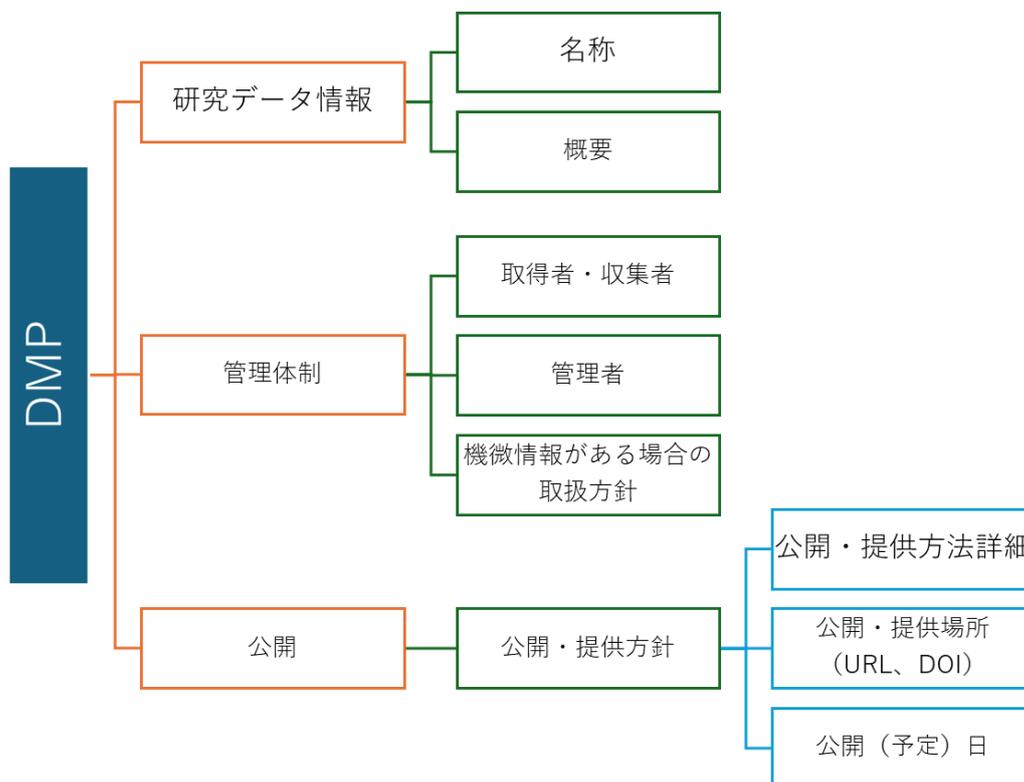


図2 DMP 記載様式

JSPS 科学研究費助成事業「DMP 様式例・記入例」[4]を元に作成

研究プロセスの中でのデータマネジメント観点でもう一度、科研費の DMP 作成に関する FAQ で言及されている継続研究については、『過去に遡って DMP 作成は不要』が示唆することを捉え直す必要がある。

学術振興会からは先に示した FAQ の文章のみが公開されているため、その意図を正確に読み取することは難しく、その意図の解釈は推測の域を出ないものではあるが、いくつかの解釈が考えられる。

解釈の1つとしては、令和6年度以前に取得した研究データを令和6年以降の継続課題で利活用しない場合、DMP 作成義務化の対象外とするが、令和6年に実施する継続課題の中で利用する場合は、DMP 作成を義務づける対象とするとする解釈が可能である。

ただ、この解釈の場合、一般的に過年度に取得したデータを、以降の継続研究において一切利用しないということは想定しづらく、過去に遡って作成不要と言及することが、混乱にも繋がるため、おそらくこのような解釈を意図したものではないと推測される。

JSPS が定める DMP の様式では、データ公開とそれに向けた管理方針を記載することが定められている。現時点では、2024 年度以前に収集した研究データは公開義務化の対象外となっており、公開方針等を定める DMP の作成は不要と捉えているのではないかと推測される。

3. DMP に基づくデータライフサイクルマネジメント

DMP とは、データの管理方法を定めた文書であるが、データの管理手法として、ビッグデータなどを取り扱うビジネス分野においては、データライフサイクルマネジメント (DLM) という概念に基づくデータマネジメント手法が取り入れられている。

ビッグデータや AI を活用したビジネスを展開する場合、利活用されるデータの容量がギガやテラを大きく超えてペタやセタ単位の容量に及ぶことがある。このため、ビジネスの中で利活用するデータを永続的に管理することが難しく、データの活用について、生成から、利用、廃棄までの一連の過程をデータライフサイクルと定義して、各段階を管理するための手法となる[5]。

この DLM の概念を学術研究では、データの廃棄をオープンデータ化の過程に置き換えて、研究・調査計画の立案、調査・実験実施によるデータ生成・収集、データ分析の為の前処理・加工、データ分析、データの保管、共有、二次利用サイクルとして捉え、データライフサイクル全般を管理するというアプローチとして捉えられている (図 3)。

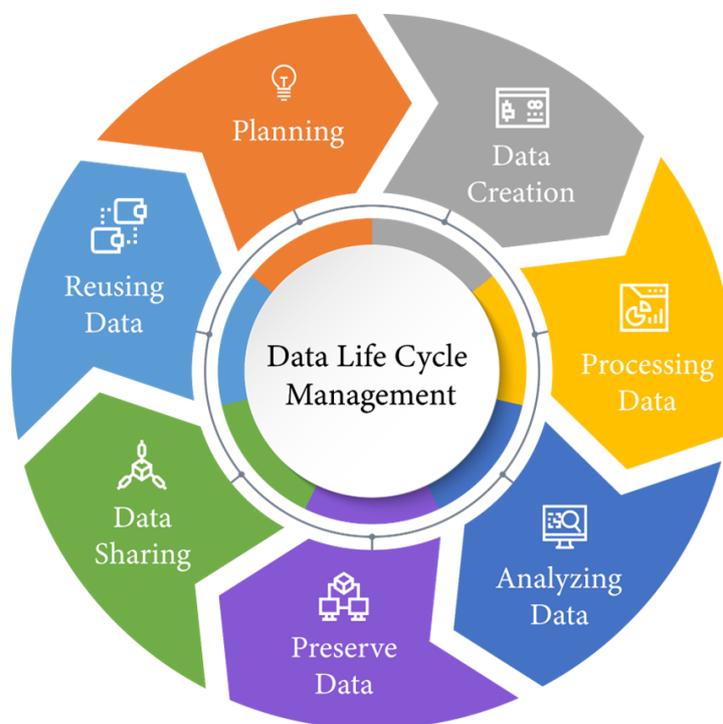


図 3 学術研究におけるデータライフサイクルマネジメントの概念図
University of Ottawa, Research data management [6] を元に作成

このデータライフサイクルマネジメントの考え方に基づけば、研究データは、調査計画から研究データの収集・取得やデータ分析や論文の執筆など研究プロセス全般のデータライフサイクルとして DMP に基づいて管理されるべきもので、研究データの公開や公開に向けた管理方法だけを切り取って管理すべきものではない、と考えられる。

この DMP に基づいて研究プロセス全体をマネジメントするとい考え方を前提とするならば、「過去に遡っての DMP の作成は不要」とする意味をもう一度問い直す必要があるのではないかと考える。

また、今後公的助成を受けた研究成果である学術論文とその論文の根拠データの即時公開（オープン化）の方針が示されている[4]が、根拠データの公開についても、公開すべき内容に内容について、今一度検討することが必要ではないかと考える。

論文と根拠データを公開することは、確かに、研究の公正性や再現性を担保するという観点では大きな意義はあるものである。一方で、それがそのまま、オープンデータの目的であるデータの二次利用など、オープンデータの利用促進や社会での有効な利活用に繋がるのかということを検討していかなければならないと考える。

4. 社会調査データ公開に向けたデータアーカイブの動向とイギリスにおける社会科学系学術データ公開義務化の取組

ここで、社会調査データを公開するデータアーカイブの動向から、学術データのオープン化に関する動向を検討してみる。社会調査データのアーカイブでは、例えば、我が国であれば、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（SSJDA）[7]のような分野別リポジトリとよばれるようなアーカイブ組織にデータを寄託して公開することが従来から行われてきた。これに対してオープンサイエンスやオープンデータ化による迅速なデータの公開や公開データの増大に対応するために、openICPSR[8]や UK Data Service[9]などのように研究者自身が研究データを公開する場を提供するセルフデポジット型のアーカイブ組織も登場してきている。

セルフデポジット型のデータアーカイブの運用例として、UK Data Service の取組を概観する。UK Data Service は後述するとおり、イギリスにおける社会科学系研究データの公開先として運用されているアーカイブである。

UK Data Service は UK Research and Innovation（UKRI）と Economic and Social Research Council（ESRC）から助成を受けてエセックス大学が構築・運用を行っている経済、人口、社会調査データの分野別リポジトリに該当する公的アーカイブである。UKRI は我が国における JSPS に類する公的研究助成を取りまとめる機関で、ESRC は社会科学系研究に関連する分科会に位置づけられ、社会科学系研究に関する研究助成などを担っている。

UK Data Service では、大規模プロジェクトの研究データの寄託を受けて、公開するデータアーカイブとして、データ公開にあたってアーカイブ組織の支援のもと研究データのアーカイブを行う従来の分野リポジトリ型の Curated Repository と、研究者や研究プロジェ

クト自身で、データを公開することが可能な、セルフデポジット型の ReShare と呼ばれるリポジトリを運用している。

UK Data Service にアーカイブされている学術データの公開レベルは主に下記の Open Data、Safeguarded data、Controlled data の3段階に分けて管理されている[10]。

Open data:

- ・ オープンデータとして自由な利用が可能なデータで、データの利用者は UK Data Service のアカウント登録がなくても利用可能。データの寄託・登録者は公開時に Open Government Licence (OGL) や Creative Commons Licence (CC) を付加することも可能。

Safeguarded data:

- ・ 研究データの利用に、UK Data Service のアカウント登録とエンドユーザライセンス契約への同意が必要。また、寄託・登録者による利用条件、利用可否の審査、個別ライセンス契約などを設定することが可能。

Controlled data:

- ・ 個人情報に分類される、または商業上の理由など機密性が高いデータ。利用者には、データの取扱いに関する講習の受講し、認定を取得することが義務づけられる。また、データ利用には、データ寄託者による利用目的等、申請書による審査のほか、別途条件が設定される。
- ・ 利用が認められた場合、Safe Room などを通じて、データが提供される。

UK Data Service の出資母体にもなっている ESRC は、イギリスにおける社会科学系研究に関する助成を行っている機関である。UKRI の傘下にある助成機関は ESRC を含めて、それぞれ要件の違いはあるが、研究助成の受領者に対して、論文のオープン化やデータ公開を義務づけている。

ESRC の場合は、助成を受けた研究に対して、学術論文などは即時オープン化、研究データに関しては、原則研究期間終了後、3ヶ月以内に研究データをリポジトリに登録することを義務づけている。このときデータのアーカイブ先として推奨されているのが UK Data Service で、UK Data Service 以外にデータを登録し、公開する場合には、DOI のような永続的識別子を付与することが可能なリポジトリに登録することを条件として定めている。

また、研究データの登録・公開を UK Data Service 以外で行う場合も、メタデータに関しては、UK Data Service に登録することが義務づけられている。これによりデータカタログの作成と管理は UK Data Service が集約して担うこととなっている[11]。

このように ESRC では、研究助成対象者に対して UK Data Service を通じた研究データとメタデータの公開を義務づけているが、研究データの適切な管理及びデータ公開と再利用のために、2011年4月から助成金の申請の段階で DMP の作成と提出を義務づけていて、

DMP も申請手続きの中でのレビュー対象として位置づけられている[12]。

では、ESRC では、DMP にどのような項目の記載を求めているのか、以下に示す[13]。

Information on new data：新規データの概要

- ・ 研究プロジェクトにより生成・収集されるデータの概要。
- ・ データ量やデータのタイプその他、メタデータやデータ収集方法の概要などを含む。

Quality assurance of data：データの品質保証

- ・ データ収集（調査・実験）における品質保証
- ・ サンプルング・データ収集方法などを具体的に記述。
- ・ 訪問調査やインタビュー調査などの場合、調査員による調査データの質のバラツキを抑制する工夫など。
- ・ 調査票からデジタル化する際のエラー抑制の手段。

Backup and security of data：データセキュリティ

- ・ 研究プロジェクトにおいてデータ及びメタデータを安全に管理・運用するためのセキュリティ対策。
- ・ 特に個人情報を含むデータの管理方法や調査協力者・被験者のプライバシー保護対策。

Management and curation of data：データ管理とキュレーション

- ・ データ公開・二次利用を前提として、他の研究者がデータの理解し、解釈することができるようなメタデータ・付帯文書等の準備・検討状況。
 - 調査目的や方法論、使用した変数やコードの意味などを明示的に説明する。
 - 匿名化処理やデータクリーニングの方法論。

Difficulties in data sharing and measures to overcome these：データ共有の課題とそれを克服する対処法

- ・ 個人情報を含む場合などデータ共有・公開の課題。
- ・ 課題を解決するための手段。
 - 匿名化処理による対応
 - インフォームドコンセントによる同意の取得

Consent, anonymisation and strategies to enable further re-use of data：同意取得・データの匿名化・データの二次利用促進策

- ・ 調査・実験における同意取得のプロセス・方法。
- ・ 将来的なデータの共有・公開・二次利用を前提とした同意。
 - 個人情報を含む研究データの公開方法の明示（匿名化処理・同意による個人情報を含むデータの公開）
 - 同意撤回時の対応方法など。

Copyright and intellectual property ownership：知的財産権の帰属

- ・ 研究プロジェクトにより生成・収集したデータの知的財産権に関する

Responsibilities：データマネジメントに関する責任の所在

- ・ データ管理・メタデータ作成・データの品質管理・公開データの作成に関して、管理体制及び責任の所在を明記。

Preparation of data for sharing and archiving：データ共有とアーカイブのための準備作業

- ・ ESRC が定めたデータ公開ポリシーに準拠したデータや文書化の計画。
- ・ 研究プロジェクト内で生成・収集するデータに関して、構造化されたメタデータや付帯文書を作成するプロセス。

ESRC が定める DMP では、データの公開方針を重視しながらも、公開方針だけではなく、研究プロセスの適正な管理が、データ公開の促進に繋がるという考え方のもと、研究プロセス全般にわたるデータライフサイクルについてのデータ管理計画の作成が求められている。

特に調査や研究の適正な実施に繋がるデータの品質保証やデータ公開に繋がるデータキュレーションを研究者の責任として求めていると考えられる。

ただ、一方で、ESRC は助成金申請時に、DMP の提出を求めて、評価を行うが、その評価は、あくまでも申請時の研究計画との整合性、例えば、調査を実施する場合には、被験者のプライバシー保護や収集した個人情報含むデータの保護対策が必要であるが、それに関する言及がない、といった、観点で評価される[9]。

ESRC は、研究が進展していく中で、常に研究計画は修正し、更新されるものであることを理解し、研究プロジェクトの進展状況により、適切に DMP を更新することを研究者に求めている。DMP の充実が研究プロセスや研究データの公正性の担保だけではなく、最終的にデータ公開に必要なメタデータや付帯文書の作成にも繋がっていると捉えられているとも考えられる。

5. まとめ

オープンデータなどデータ公開を推進するためにセルフデポジット型のアーカイブの運用が開始されている。これは、従来は分野別リポジトリなどアーカイブ組織が担ってきたデータ公開に関するコストを、今後研究データ公開が義務化された際には、アーカイブ組織で公開コストをまかないきれないだろうということが背景にあることが推測される

我が国においても研究データ公開を義務づけた場合のデータの公開先として、機関リポジトリの活用などが想定されているが、データ公開が義務化された際は、研究者自身がデータ公開を担うセルフアーカイブに近い運用になることが見込まれる。つまり、データ公開の義務化に伴うコストは今後研究者が担っていかなければならないことが想定される。DMP に基づいて適正に研究プロセスを管理していく、また、研究プロセスの進展にあわせて、DMP を適切に更新し、充実させていくことで、データ公開に係るコストを、大幅に削減することが可能になるのではないかと考えられる。

社会調査を含めていわゆるデータ駆動型研究においては、DLCMは言い換えれば、そのまま研究プロセスのマネジメントとなる。つまりDMPは研究計画書や調査計画書そのもの、あるいはこれにデータ公開の要素を加えたものと捉えることができるのではないかと考える。

学術論文や研究データのオープン化を推進するという観点では、我が国だけではなく、イギリスの例などでも。DMP作成の義務化の過程でデータ公開を強調している面はあるため、データ公開を前提としたDMPの作成が求められているということが。世界的なトレンドであると考えられる。

一方で、データ公開の文脈だけで、DMPを捉えてしまうと、作成そのものの面倒な作業で、メタデータの作成やデータ公開は余計な労力の係る作業としてイメージされるものになってしまう。しかしながら、本来のDMPは社会調査に携わる研究者が普段から当たり前で作成、管理する研究計画書や調査計画書そのものであり、普段から研究の進捗にあわせて、研究計画書を適切に管理することで、データ公開に関するコストも大幅に削減し、データの2次利用を含めたオープンデータの利活用促進に繋がるものと考えられる。

謝辞

本報告はJSPS科研費JP 23K17577の助成を受けたものです。

参考文献

- [1] 内閣府 科学技術・イノベーション推進事務局, “我が国のオープンサイエンス政策について”, 2024年2月28日, https://www.jsps.go.jp/file/storage/kaken_g_1609/shiryoku04.pdf (最終アクセス 2025年1月20日)
- [2] 日本学術振興会, “科研費における研究データの管理・利活用等について”, 2024年2月, https://www.jsps.go.jp/file/storage/kaken_g_1609/shiryoku05.pdf (最終アクセス 2024年12月16日)
- [3] 日本学術振興会, “科研費における研究データの管理・利活用について”, https://www.jsps.go.jp/j-grantsinaid/01_seido/10_datamanagement/index.html (最終アクセス 2025年1月16日)
- [4] 日本学術振興会科学研究費助成事業, “DMP 様式例・記入例”, https://www.jsps.go.jp/file/storage/kaken_01_10_datamanagement/dmp_yoshiki_rei.xlsx (最終アクセス 2025年1月16日)
- [5] 日本IBM, “データ・ライフサイクル管理とは”, <https://www.ibm.com/jp-ja/topics/data-lifecycle-management> (最終アクセス 2025年2月1日)
- [6] University of Ottawa, “Research data management”, <https://www.uottawa.ca/library/research-data-management> (最終アクセス 2025年2月11日)
- [7] 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, “SSJDA”, <http://www.ssjda.u-tokyo.ac.jp/>

- s://csrda.iss.u-tokyo.ac.jp/infrastructure/ssjda/ (最終アクセス 2025 年 2 月 1 日)
- [8] The Institute for Social Research at the University of Michigan, “openICPSR”, <https://www.openicpsr.org/openicpsr/about> (最終アクセス 2025 年 2 月 7 日)
- [9] UK Data Service, “UK Data Service”, <https://ukdataservice.ac.uk/> (最終アクセス 2025 年 2 月 7 日)
- [10] UK Data Service, “guide for depositing data in our curated collection”, https://ukdataservice.ac.uk/app/uploads/depositing_data_in_the_curated_collection.pdf (最終アクセス 2025 年 2 月 7 日)
- [11] ESRC, “ESRC research data policy”, <https://www.ukri.org/publications/esrc-research-data-policy/> (最終アクセス 2025 年 2 月 7 日)
- [12] ESRC, “ESRC research funding guide”, 2025.1.21, <https://www.ukri.org/publications/esrc-research-funding-guide/> (最終アクセス 2025 年 2 月 11 日)
- [13] ESRC, “ESRC data management plan and policy”, <https://ukdataservice.ac.uk/learning-hub/research-data-management/plan-to-share/esrc-data-management-plan-and-policy/> (最終アクセス 2025 年 2 月 10 日)