

## レポート

# 日本の研究者によるデータ公開と 研究データ管理 (RDM) に関する実態調査 2022 — RDM 支援体制の構築と人材育成に向けて —

データ解析政策研究室 客員研究官 池内 有為、室長 林 和弘

## 概要

科学技術・学術政策研究所は、2022年にデータ公開や研究データ管理 (RDM) に関するウェブ質問紙調査を実施した。対象は科学技術専門家ネットワークであり、大学や企業に所属する多分野の研究者 1,237 名から回答を得た (回答率 73.9%)。

現在研究活動を行っている 1,159 名のうち、公開データの入手経験は 71.0% が、研究データの公開経験は 50.1% が、データマネジメントプラン (DMP) の作成経験は 28.6% が有しており、データの主な入手先や公開先は、論文の補足資料や学術機関のリポジトリであった。主な未公開理由は、論文を投稿した雑誌のポリシーではないこと (37.0%) や公開できない情報が含まれること (34.0%)、業績にならないこと (30.7%) であり、これらの理由が解消された場合に公開意思をもつ回答者は 26.3% であった。

RDMのための資源は、人材、時間、資金が不足していると認識している回答者の比率が高く、この傾向は本調査を開始した 2016 年からほとんど変化していない。RDM を図書館員やデータキュレーターに依頼したいと考える回答者は 48.4% であり、適切なデータ形式への変換 (82.2%) や適切なリポジトリの選択 (78.6%) の希望が高かった。RDMのための人材育成は、喫緊の課題であると言えよう。

キーワード：オープンサイエンス，研究データ公開，研究データ管理 (RDM)，  
データマネジメントプラン (DMP)，人材育成

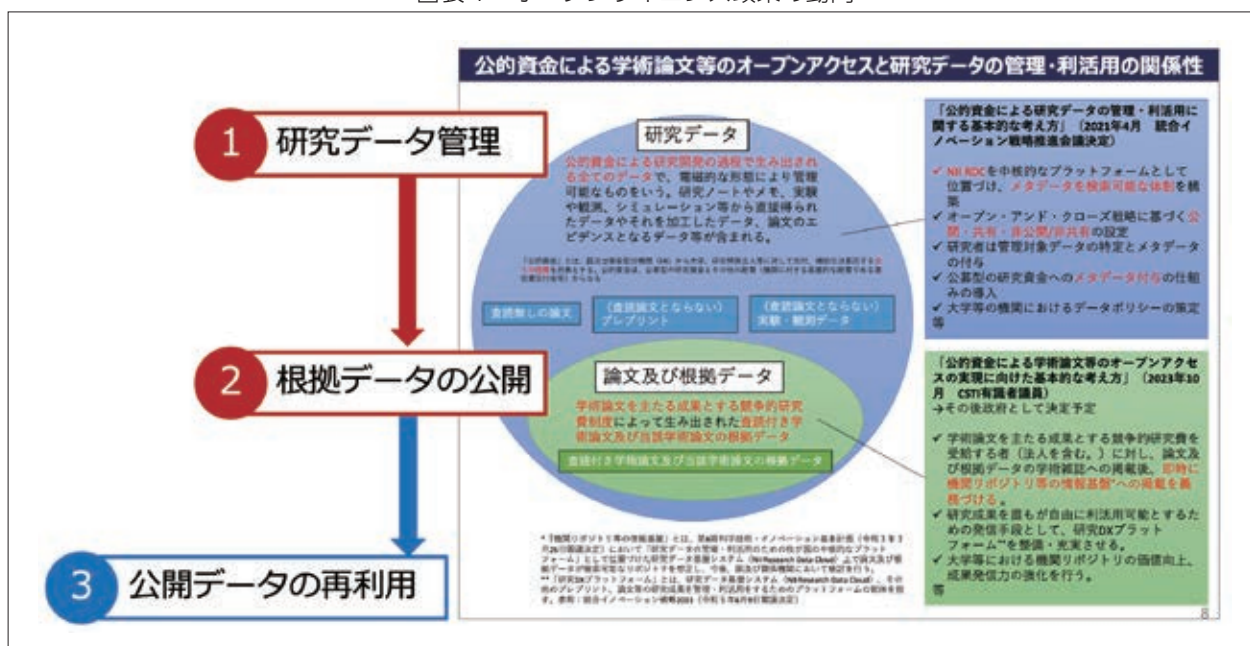
## 1. オープンサイエンス政策の動向と調査の概要

公的資金による学術研究の成果である論文やデータを広くインターネットで公開して利活用に供すること、及びデータを適切に管理することを求めるオープンサイエンス政策は、2023年のG7科学技術大臣会合<sup>1)</sup>等を経て、更に加速している。2021年の「公的資金による研究データの管理・利活用に関する基本的な考え方」<sup>2)</sup>では、公的資金による研究開発の過程で生み出される全てのデータについて、メタデータを付与し、管理することが求められている (図表 1 ①)。また、2023年の「公的資金による学術論文等のオープンアクセスの実現に向けた基本的な考え方」<sup>3)</sup>では、学術論文を主たる成果とする競争的研究費制度によって生み出された査読付き学術論文及び当該学術論文の根拠データについて、学術雑

誌への掲載後、即時に機関リポジトリ等の情報基盤に掲載することを義務づけるべきであると論じられている (図表 1 ②)。こうした取組によって、公開されたデータが再利用され、研究者のみならず市民やAIなどが新たな成果を生み出し、様々な課題を解決することが期待されている (図表 1 ③)。

それでは、①研究データ管理 (以下、「RDM」)、②根拠データの公開、③公開データの再利用はどの程度実践され、何が課題となっているのであろうか。科学技術・学術政策研究所 (以下、「NISTEP」) は、2022年10月から11月にかけて日本の研究者を対象としたオープンサイエンスに関するウェブ質問紙調査を実施した。これは2016年から隔年で実施してきた研究データ公開/RDMと論文のオープンアクセス (以下、「OA」) に関する実態調査<sup>4,5,6)</sup> (The State of Open Science in Japan) の一部である。図表 2 に示すように、2020年にはプレプリントに関する調

図表 1 オープンサイエンス政策の動向



出典：赤池伸一. 日本のオープンアクセス政策. SPARC Japan セミナー 2023 発表資料に基づき作成

査を実施した。そして2022年は論文のOAとプレプリントの調査を併せて実施し<sup>7)</sup>、研究データ公開/RDMの調査は別途実施した<sup>8)</sup>。本稿では、一連の調査結果を参照しながら、2022年の研究データ公開/RDM調査について報告する。

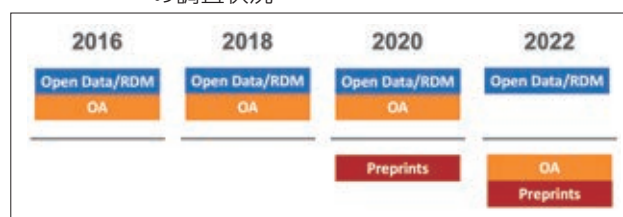
The State of Open Science in Japanの調査対象は、NISTEPが運営する科学技術専門家ネットワーク<sup>9)</sup>に所属する大学、公的機関・団体、企業の研究者・技術者等、約2,000名である。科学技術専門家ネットワークの構成員は推薦によって選出され、毎年一部入替えがある。すなわち、パネル調査ではない点に御留意いただきたい。

本稿では、1,349名(回答率70.5%)による有効回答のうち、現在研究活動を行っている1,268名の回答を分析した結果から、「データ公開の状況と障壁」(2章)、データ公開と強い関連をもつ「公開データの入手状況と課題」(3章)及び「データマネジメントプラン(DMP)<sup>注1</sup>の作成状況」(4章)、そして「RDMの状況と依頼意思」(5章)について報告する。調査の詳細については、報告書<sup>8)</sup>を御参照いただきたい。

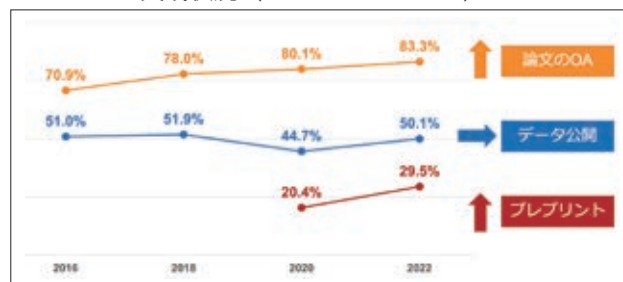
## 2. データ公開の状況と障壁

研究データ<sup>注2</sup>の公開経験を尋ねた結果、回答者の

図表 2 The State of Open Science in Japan の調査状況



図表 3 日本の研究者によるオープンサイエンスの実践状況 (2016/18/20/22)



50.1%が公開経験を有していた。図表3に示すように、論文のOAやプレプリント経験をもち研究者は経年的に増加している一方で、データ公開はそれほど増加していないことがわかった。

データの公開方法を複数選択方式で尋ねた結果、最も選択率が高かった項目は「論文の補足資料

注1 研究のために収集・作成するデータをどのように管理するか、取扱いや整備・保存・公開に関する計画を記した書類を指す。

注2 本調査では、研究データの定義を「研究のために収集・作成・観測したデジタルデータを指す。研究の成果である論文やスライドの根拠となるデータ、及び研究成果そのものであるデータの両方を含む。テキスト、画像、音声、動画など、形式は限定しない。また、ゲノムデータ、地理情報、ソフトウェアコード、インタビューの録音と書き起こしなど、内容も限定しない」とした。

(53.4%)」であり、次いで「学術機関のリポジトリ等(以下、「IR」) (38.9%)」、「個人や研究室のウェブサイト(31.5%)」の順であった。図表 4 に、2016/18/20 年調査の結果と併せて示す。

2016 年調査では個人や研究室のウェブサイトの選択率が最も高かったが徐々に順位を下げ、2018 年調査以降は論文の補足資料が最も良く利用されている。学術系 SNS やコード共有サービス、データ共有サービスは、相対的な順位は低いものの、いずれも選択率が上がっていた。

研究にデジタルデータを用いるが、データを公開した経験がない回答者を対象として、未公開理由を複数選択方式で尋ねた。図表 5 に、2016/18/20 年調査の結果と併せて示す。

最も選択率が高かったのは、「論文を投稿した雑誌のポリシーではないから (37.0%)」、2 位は「公開できない情報が含まれているから (34.0%)」、3 位は「業績にならないから (30.7%)」、4 位は「盗用やスクーピングの可能性があるから (28.8%)」であった。1 位 (雑誌のポリシーではない) と 3 位 (業績にならない) は、4 回の調査において毎回選択率が上昇していた。

これらの理由が解消された場合にデータ公開意思をもつ回答者は 26.3% にとどまり、公開しないと回答者が 25.8%、わからないとした回答者が 47.9% であった。公開データを再利用してデータ公開の利点を認知することは、データ公開の動機となる可能性があると考えられる。それでは、公開データの入手状況はどのように変化しているのだろうか。

### 3. 公開データの入手状況と課題

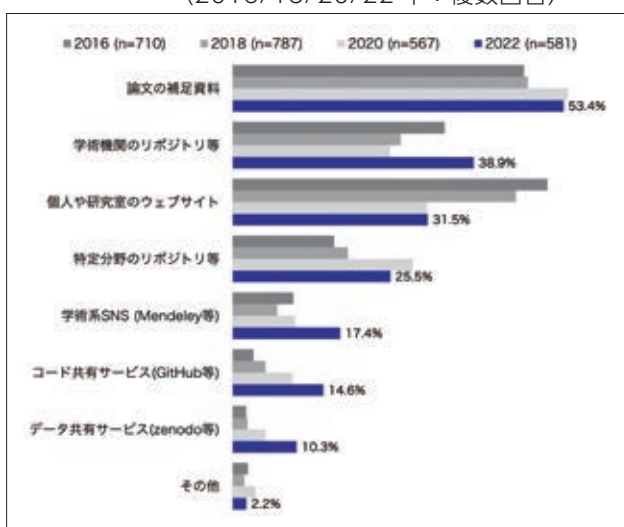
公開データ<sup>注3</sup>の入手経験を尋ねた結果、回答者の 71.0% が入手経験を有していた。次章で述べるように、DMP の作成経験をもつ研究者は経年的に増加している一方で、データの入手経験は公開と同様にそれほど増加していないことがわかった (図表 6)。なお、データの入手先は、公開先と同様に「論文の補足資料 (64.2%)」、「IR (60.1%)」、「個人や研究室のウェブサイト (52.6%)」の順であった。

図表 7 に示すように、分野別のデータ入手経験をもつ回答者の比率とデータ公開経験をもつ回答者の比率には、正の相関がみられた (Pearson の相関係数  $r=0.649$ ,  $p<0.05$ )。これは、2016 年調査から一貫した傾向である。

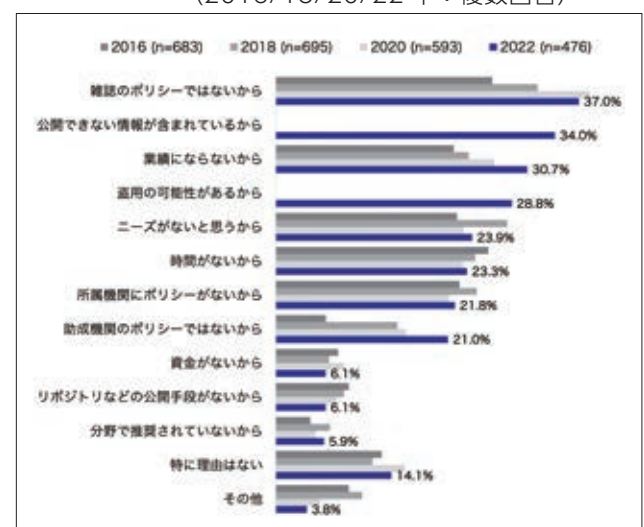
データの入手経験をもつ回答者のうち 62.9% は、データ入手において何らかの問題があると認識しており、具体的な内容を複数選択方式で尋ねた。図表 8 に 2016/18/20 年調査の結果と併せて示す。

1 位は「データごとにフォーマットが異なる (44.8%)」、2 位は「利用条件 (営利用が可能かどうかなど) がよくわからない (42.1%)」、3 位は「著作者情報がよくわからない (38.8%)」であった。いずれもデータの再利用の障壁となりうるため、データ公開に当たってはこれらの情報をメタデータとして記述し、わかりやすく提示する必要があると考えられる。次章では、データに関する情報を記述する DMP の作成状況について述べる。

図表 4 データの公開方法  
(2016/18/20/22 年：複数回答)



図表 5 データの未公開理由  
(2016/18/20/22 年：複数回答)

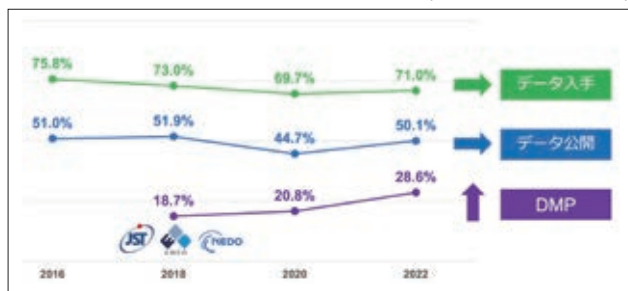


注 3 本調査では、公開データの定義を「ウェブサイトやリポジトリ、論文の補足資料などに掲載され、インターネットでアクセスして利用できるデータを指す。利用料金や利用者登録が必要な場合も含む」とした。

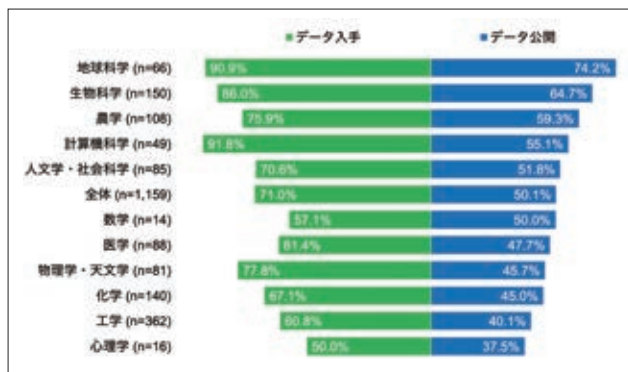
#### 4. データマネジメントプラン (DMP) の作成状況

2018年から国立研究開発法人科学技術振興機構 (JST)、国立研究開発法人日本医療研究開発機構 (AMED)、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) が DMP の作成を求めようになった<sup>10)</sup>。前章の図表 6 に示したように、2022年調査において DMP の作成経験をもつ研究者は 28.6% であり、2018年調査から経年的に増加している。

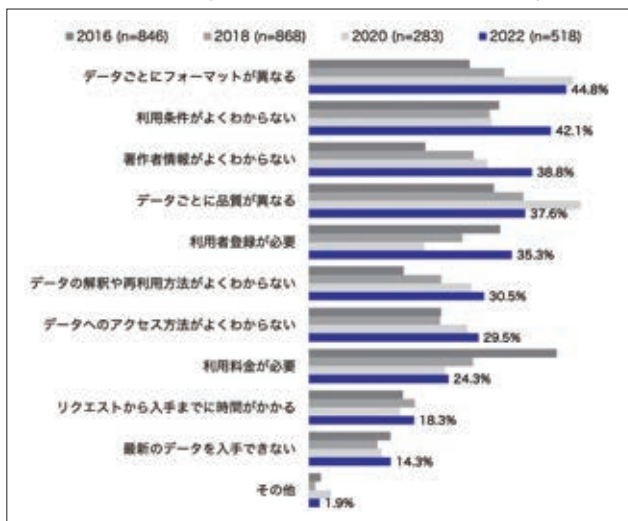
図表 6 データ公開に関する実践状況 (2016/18/20/22)



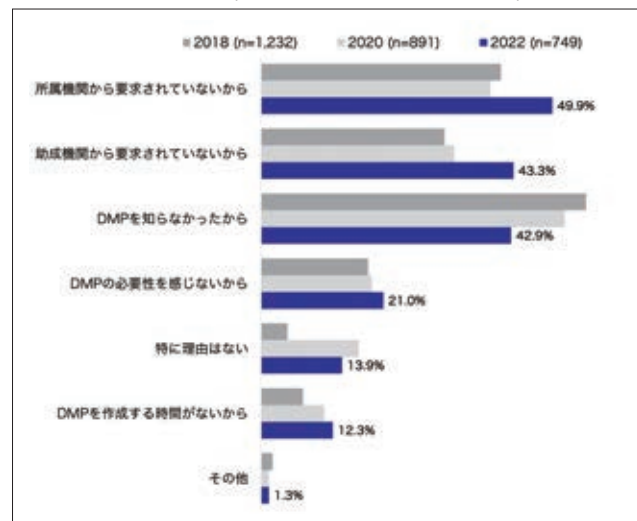
図表 7 分野別公開データの入手経験とデータ公開経験



図表 8 データ入手における問題 (2016/18/20/22: 複数回答)



図表 9 DMP の未作成理由 (2018/20/22: 複数回答)



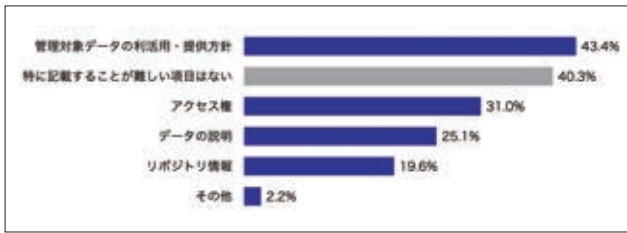
DMP の作成経験がない回答者を対象として、未作成理由を複数選択方式で尋ねた。図表 9 に 2018/20年調査の結果と併せて示す。

2018年調査及び2020年調査においては「DMPを知らなかったから」の選択率が最も高かったが、本調査では2018年から12.8ポイント減少して42.9% (3位) であった。国際調査においては、助成機関によるDMPの要求はデータ公開の主要な動機となっており<sup>11)</sup>、今後、日本においても助成機関によるDMPの要求がデータ公開を推進する要因の一つとなるのではないだろうか。

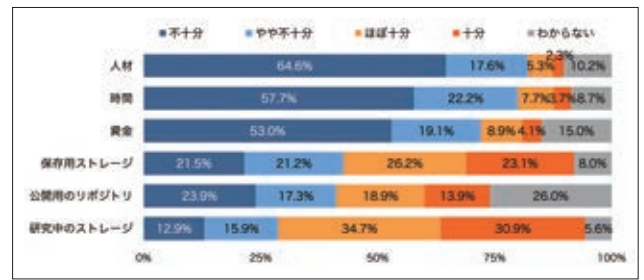
しかし、多くの研究者にとって、DMPの作成は新たな取組である。そこで「公的資金による研究データの管理・利活用に関する基本的な考え方<sup>2)</sup>」で示された、DMPに掲載する「メタデータの共通項目」(16項目)のうち4項目を挙げ、記述することが困難である項目を複数選択方式で尋ねた。図表10に集計結果を示す。

最も選択率が高かったのは、「管理対象データの利活用・提供方針(無償/有償、ライセンス情報、その他条件(引用の仕方等)等を記載)(43.4%)」であった。2位は「特に記載することが難しい項目はない」(40.3%)であり、メタデータ作成に困難を感じない回答者も一定数存在することが確認された。ただし、「その他」には「作成は可能だが労力がかかる」という趣旨の回答がみられた。例えば、「時間と精神的労力さえかければ、上記に難しいことはありません。しかし、上記を作成・記載するのに非常に時間・精神的労力が必要で、その意味では全て「難しい」です」、「記載すること自体は可能と思われるが、記載した経験がなく、ノウハウがないため、いずれについても規準がわからない。そういう意味で、全て難しく感じる」、「項目が多すぎるので、本当に必要であれば、自動生

図表 10 記述が困難な「メタデータ共通項目」  
(n=1,046：複数回答)



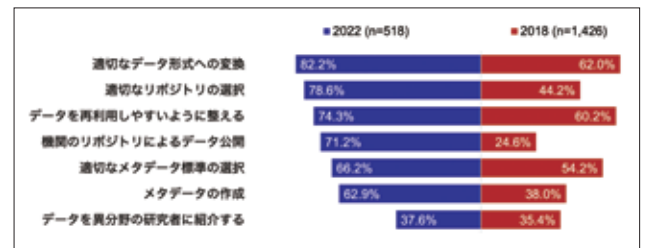
図表 11 データ公開・整備に必要な資源の充足度  
(n=1,084)



成できる枠組みが欲しい”といった指摘があった。

前章で示したように、公開データの情報が十分でないという再利用率に支障を来す恐れがあることからメタデータ作成は重要である一方で、作成には研究者の時間や労力を要する。適切な支援やガイドラインの整備、評価の仕組み、及びメタデータの必要性を丁寧に説明することが求められる。

図表 12 RDM の依頼希望項目と専門性を必要とする項目（複数回答）



## 5. 研究データ管理 (RDM) の状況と依頼意思

1章で述べたように、公的資金による研究開発の過程で生み出されるデータの管理や論文の根拠データの公開が求められつつある。それでは、RDMのための資源は十分であるのか。研究にデータを用いる回答者を対象として、“論文などの成果を発表済みの、最近の主要な研究1件のために収集・作成・観測したデータ（以下、「カレントデータ」）”を管理し、公開することを想定していただいた上で<sup>注4</sup>、資源の充足度を「不十分」から「十分」までの4件法で尋ねた。図表11に集計結果を示す。

2016年調査から引き続き、人材、時間、資金が「不十分」と認識している回答者が過半数であった。また、「わからない」の選択率も高く、公開用のリポジトリは26.0%であった。保存用ストレージ、公開用のリポジトリ、研究中のストレージは相対的に充足度が高いものの、2020年調査からほとんど変化していなかった。

カレントデータの整備・公開・保存プロセスを、自身や共同研究者にかかわって図書館員やデータキュレーターに依頼したいと思うかどうかを尋ねた結果、依頼したいと思う回答者は48.2%、依頼したいとは思わない回答者は29.5%、「わからない」を選択した回答者は21.9%、無回答は0.5%であった。2020年調査と比較すると、依頼意思をもつ回答者は7.1ポイント増加していた。

RDMの依頼意思をもつ回答者を対象として、具体

的な項目を複数選択方式で尋ねた。また、2016/18年調査では、RDMを第三者に依頼する場合に専門性を必要とする項目を複数選択方式で尋ねた。図表12に、本調査と2018年調査の結果を示す。

依頼を希望する項目は、2020年調査と順位に変化はなく、選択率もほとんど同じであった。1位の「適切なデータ形式への変換(82.2%)」や3位の「データを再利用しやすいように整える(60.2%)」は、2018年調査において研究者の多くが“専門性が必要である”と認識していた項目であった。RDMのための専門人材の育成は、喫緊の課題であると言えよう。

## 6. 終わりに

日本の研究者は、公的資金によって生み出されたデータの適切な管理や論文の根拠データの公開が要求されつつある。調査を開始した2016年以降、データの公開や公開データの入手経験をもつ研究者の比率は、論文のOAと比較してそれほど増加していないため、その要因の分析と対応が引き続き求められる。また、オープンサイエンス政策が目指す公開データの再利用を実現するためには、適切なデータ形式でデータを公開し、併せてデータに関する情報をメタデータとして提供する必要があると考えられるが、RDMに関する資源の充足度は依然として低い。外部資金を獲得した研究者がより多くの研究時間を確保

注4 研究によって扱うデータの量や種類が異なる場合があると予想されるため、カレントデータを公開することを想定して回答していただいた。

---

できるようにするためにもデータ公開や RDM を支援するための人材育成や体制の構築が進むことを期待している。

## 謝辞

調査及びプレテストに御協力を賜りました科学技術専門家ネットワーク専門調査員をはじめとする皆様に心より御礼申し上げます。

## 参考文献・資料

---

- 1) G7 科学技術大臣コミュニケ（仮訳）. 内閣府 . 2023, 9p.  
[https://www8.cao.go.jp/cstp/kokusaiteki/g7\\_2023/230513\\_g7\\_kariyaku.pdf](https://www8.cao.go.jp/cstp/kokusaiteki/g7_2023/230513_g7_kariyaku.pdf)
- 2) 統合イノベーション戦略推進会議. 公的資金による研究データの管理・利活用に関する基本的な考え方. 2021, 18p.  
<https://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/sanko1.pdf>
- 3) 総合科学技術・イノベーション会議有識者議員. 公的資金による学術論文等のオープンアクセスの実現に向けた基本的な考え方. 2023, 7p. [https://www8.cao.go.jp/cstp/231031\\_oa.pdf](https://www8.cao.go.jp/cstp/231031_oa.pdf)
- 4) 池内有為, 林和弘, 赤池伸一. 研究データ公開と論文のオープンアクセスに関する実態調査. 文部科学省科学技術・学術政策研究所, 2017, NISTEP RESEARCH MATERIAL No.268, 108p. <https://doi.org/10.15108/rm268>
- 5) 池内有為, 林和弘. 研究データ公開と論文のオープンアクセスに関する実態調査 2018. 文部科学省科学技術・学術政策研究所, 2020, NISTEP RESEARCH MATERIAL No.289, 96p. <https://doi.org/10.15108/rm289>
- 6) 池内有為, 林和弘. 研究データ公開と論文のオープンアクセスに関する実態調査 2020. 文部科学省科学技術・学術政策研究所, 2021, NISTEP RESEARCH MATERIAL No. 316, 124p. <https://doi.org/10.15108/rm316>
- 7) 池内有為, 林和弘. 論文のオープンアクセスとプレプリントに関する実態調査 2022：オープンサイエンスにおける日本の現状. 文部科学省科学技術・学術政策研究所, 2023, NISTEP RESEARCH MATERIAL No. 327, 124p.  
<https://doi.org/10.15108/rm327>
- 8) 池内有為, 林和弘. 研究データ公開と研究データ管理に関する実態調査 2022：日本におけるオープンサイエンスの現状. 文部科学省科学技術・学術政策研究所, 2023, NISTEP RESEARCH MATERIAL No. 335, 132p.  
<https://doi.org/10.15108/rm335>
- 9) 科学技術専門家ネットワーク. <https://www.nistep.go.jp/activities/st-experts-network>
- 10) 池内有為. データマネジメントプラン（DMP）—FAIR 原則の実現に向けた新たな展開. 情報の科学と技術. 2018, vol. 68, no. 12, p. 613-615. [https://doi.org/10.18919/jkg.68.12\\_613](https://doi.org/10.18919/jkg.68.12_613)
- 11) Digital Science, et al. The State of Open Data 2022. <https://doi.org/10.6084/m9.figshare.21276984.v5>