

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

2017 年 12 月

文部科学省 科学技術・学術政策研究所

科学技術予測センター

池内 有為, 林 和弘, 赤池 伸一

【調査研究体制】

| | |
|-------|---------------------|
| 池内 有為 | 科学技術予測センター 客員研究官 |
| 林 和弘 | 科学技術予測センター 上席研究官 |
| 赤池 伸一 | 科学技術予測センター センター長 |

【Authors】

| | |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Ui IKEUCHI | Affiliated Fellow Science and Technology Foresight Center, National Institute of Science and Technology Policy (NISTEP), MEXT |
| Kazuhiro HAYASHI | Senior Research Fellow Science and Technology Foresight Center, National Institute of Science and Technology Policy (NISTEP), MEXT |
| Shinichi AKAIKE | Director Science and Technology Foresight Center, National Institute of Science and Technology Policy (NISTEP), MEXT |

本報告書の引用を行う際には、以下を参考に出典を明記願います。

Please specify reference as the following example when citing this NISTEP RESEARCH MATERIAL.

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」, *NISTEP RESEARCH MATERIAL*, No.268, 文部科学省科学技術・学
術政策研究所.

DOI: <http://doi.org/10.15108/rm268>

Ui IKEUCHI, Kazuhiro HAYASHI, Shinichi AKAIKE “A Survey on Open
Research Data and Open Access,” *NISTEP RESEARCH MATERIAL*, No.268,
National Institute of Science and Technology Policy, Tokyo.

DOI: <http://doi.org/10.15108/rm268>

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

池内 有為, 林 和弘, 赤池 伸一

文部科学省科学技術・学術政策研究所 科学技術予測センター

要旨

公的資金による研究の成果である論文やデータを公開し、学術関係者のみならず、企業や市民による利活用を可能にするオープンサイエンス政策が世界的に推進されている。オープンサイエンスの実現によって、科学の発展やイノベーションの創出が期待されている一方で、データの公開については分野ごとの特性をふまえる必要があると指摘されている。

そこで科学技術・学術政策研究所(NISTEP)科学技術予測センターは、データ公開を中心とした日本のオープンサイエンスの実態と課題を明らかにするために、2016年11月から12月にかけて、科学技術専門家ネットワークを活用したアンケートシステムによる調査を実施した。その結果、回答者1,398名(回答率70.5%)のうち、51.0%がデータの、70.9%が論文の公開経験を有していた。データを公開しようとする場合、資源の不足感が強く、79.2%は人材が、74.5%は資金が、74.1%は時間が、それぞれ「不足」または「やや不足」としていると認識していた。また、データ公開に対する懸念も強く、87.8%が引用されずに利用される可能性を、84.6%が先に論文を出版される可能性を、「問題」または「やや問題」とであると認識していた。回答者の75.8%は公開されているデータを入手した経験を有しており、研究にも活用しているが、利用料金や利用者登録など、入手の際に問題があると考えていることも明らかになった。

A Survey on Open Research Data and Open Access

Science and Technology Foresight Center, National Institute of Science and Technology Policy (NISTEP), MEXT

Ui IKEUCHI, Kazuhiro HAYASHI, and Shinichi AKAIKE

ABSTRACT

For figuring out current status and challenges for Open Science, authors conducted a survey to Japanese researchers of Science and Technology Experts Network of NISTEP during Nov. to Dec. 2016 by asking the experience of sharing and use of their article and data, recognition of open research data, sufficiency of resources and items that support researchers.

Results by analyzing 1,398 responses (response rate 70.5%) shows that 51% of researchers opens their data and 70.9% of them has their Open Access journal article. When they try to publish the data, they are aware of strong shortage of resources, such as human resources (79.2%), funds (74.5%), and time (74.1%) are "insufficient" or "somewhat insufficient". In addition, they concern about data publishing, and possibility of using without being cited (87.8%), possibility of publishing a paper earlier (84.6%) are recognized as a "problem" or "somewhat problematic." It was also revealed that 75.8% of respondents have experience of obtaining published data and are also used for research, but it is thought that usage fee, user registration, etc. are problems.

目次

| | |
|----------------------------------|----|
| 概要 | i |
| (1) データ公開とオープンアクセス論文の現状 | ii |
| (2) データ公開の障壁 | iv |
| (3) 公開データの利用状況 | v |
| 1. 調査背景・目的 | 1 |
| 1.1 オープンサイエンスの隆盛と学術情報流通の変化 | 1 |
| 1.2 調査目的 | 3 |
| 1.3 調査の意義と有効性 | 3 |
| 2. 先行研究 | 4 |
| 2.1 データの公開状況 | 4 |
| 2.2 データ公開の方法 | 5 |
| 2.3 データの公開理由と障壁 | 5 |
| 3. 調査・分析方法 | 7 |
| 3.1 主要な概念の定義 | 7 |
| 3.2 調査項目 | 8 |
| 3.3 プレテスト | 11 |
| 3.4 調査対象 | 11 |
| 3.5 実施期間 | 11 |
| 3.6 分析方法 | 12 |
| 3.7 回答率と回答者の属性 | 12 |
| 4. 調査結果 | 14 |
| 4.1 データと論文の公開状況 | 14 |
| (1) データと論文の公開経験 | 14 |
| (2) データの公開方法 | 17 |
| (3) データと論文の公開理由 | 19 |
| (4) データと論文の非公開理由 | 22 |
| (5) データと論文の公開意思 | 23 |
| 4.2 データの提供・被提供状況 | 24 |
| 4.3 公開データの利用状況と課題 | 26 |
| (1) 公開データの入手経験 | 26 |
| (2) 公開データ入手の障壁 | 28 |

| | |
|-------------------------------|----|
| (3) 公開データの利用目的 | 29 |
| (4) データと論文の利用分野 | 31 |
| (5) データと論文の探索方法 | 33 |
| (6) 信頼性の判断基準 | 34 |
| 4.4 カレントデータのプロフィール | 35 |
| (1) データの保存期間規定 | 35 |
| (2) データ量 | 36 |
| (3) 所有権 | 37 |
| (4) 機密情報 | 39 |
| (5) 望ましい保存期間 | 41 |
| 4.5 カレントデータを公開する場合の障壁 | 42 |
| (1) データ公開に必要な資源の状況 | 42 |
| (2) データ公開に対する懸念 | 43 |
| (3) 他の研究者によるデータの理解 | 47 |
| 4.6 データ管理・公開に対する関心と専門性 | 48 |
| (1) メトリクスへの関心 | 48 |
| (2) データリテラシー教育への関心の有無 | 49 |
| (3) データ管理・公開の専門性 | 50 |
| 4.7 自由回答 | 52 |
| (1) データ公開のインセンティブ | 52 |
| (2) データの保存と公開のためのインフラ | 52 |
| (3) データ公開におけるその他の問題 | 53 |
| 4.8 質問項目とデータ公開経験・属性との関連 | 53 |
| 5. まとめ | 55 |
| 謝辞 | 56 |
| 参考文献 | 57 |
| 資料 | 65 |
| (1) 質問票 | 65 |
| (2) 単純集計結果 | 91 |

概要

概要

オープンサイエンスとは、“幅広い分野の公的資金による研究成果(論文や関連するデータセット等)に学術関係者だけでなく、民間企業や一般市民が、広く利用・アクセスできるようにする”¹取組である。論文のオープンアクセス(以下、「OA」)や研究データの公開と利活用によって、新たな科学の発展やイノベーションの創出、研究の透明性の向上などが期待されることから、G7 科学大臣会合をはじめとする国際組織や各国の政府機関がオープンサイエンスの推進を表明している。第5期科学技術基本計画(2016年度～2020年度)²では、“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”するとともに、“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”と述べられている。

しかし、データの公開については分野によってデータの種類や機密性、取扱の慣習などが異なるため、それぞれの特性をふまえた政策が必要であると指摘されている。2015年に公開された内閣府による報告書では、“各省庁等のステークホルダーは、オープンサイエンスを推進すべき領域、プロジェクトを選定し、研究活動上の利益・損失や研究途上の取扱及び機微の判断など各分野の専門家・研究者、技術者の意見を十分に取り入れ、その分野の活動・研究成果が最大化されることを旨として、オープンサイエンス実施方針を定める”³と述べられている。

そこで科学技術・学術政策研究所(NISTEP)科学技術予測センターは、政策立案や研究マネジメントに資することを目的として、2016年11月30日から12月14日にかけて、日本の研究者によるデータ公開を中心としたオープンサイエンスの実態や課題を把握するために、アンケートシステムによる調査を実施した。調査対象は、大学、企業、公的機関・団体に所属する研究者や専門家、技術者等によって構成される約2,000名の科学技術専門家ネットワークである。以下では、1,398名(回答率70.5%)の回答から、(1)データ公開とOA論文の現状、(2)データ公開の障壁、(3)公開データの利用状況について述べる。なお、各図のキャプションの「n」は、それぞれの回答者数を示す。

¹ G7 茨城・つくば科学技術大臣会合。つくばコミュニケ(共同声明)。内閣府, 2016, p. 9.
http://www8.cao.go.jp/cstp/kokusaiteki/g7_2016/2016communique.html

² 内閣府。第5期科学技術基本計画。2016, 53p.
<http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>

³ 国際的動向を踏まえたオープンサイエンスに関する検討会(2015)。我が国におけるオープンサイエンス推進のあり方について～サイエンスの新たな飛躍の時代の幕開け～。内閣府, 23p.
<http://www8.cao.go.jp/cstp/sonota/openscience/>

(1) データ公開とオープンアクセス論文の現状

研究のために収集・作成・観測したデジタルデータで、論文など研究成果の根拠となるもの（以下、「データ」）の公開経験がある回答者は 713 名（全体の 51.0%）、OA の論文がある回答者は 991 名（70.9%）であった（図 1）。企業の研究者についても、約 3 割はデータ公開の経験を有していた。

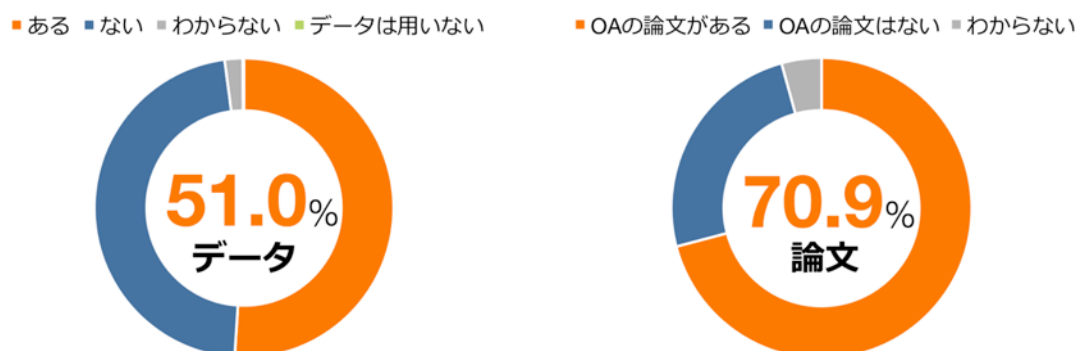


図 1 公開データと OA 論文の有無（いずれも n=1,398）

データの公開方法は「個人や研究室のサイト(50.8%)」、次いで「論文の補足資料(47.0%)」の順に選択率が高かった（図 2）。オープンサイエンス政策や学術雑誌のデータ共有ポリシーで推奨、あるいは想定されている、永続性のあるリポジトリによる公開は「所属機関のリポジトリ」が 34.2%、「特定分野のリポジトリ」が 16.4%にとどまっていた。

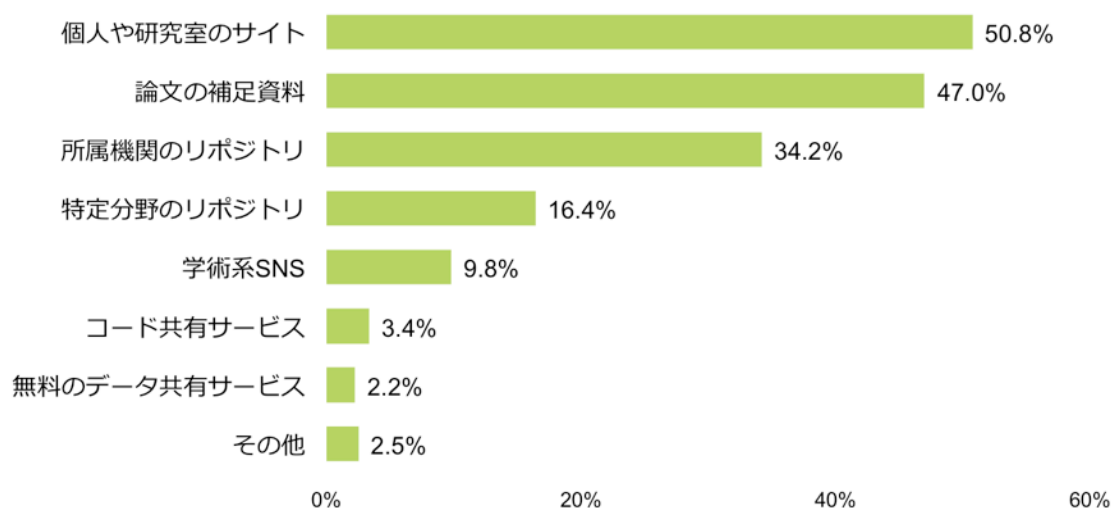


図 2 データの公開方法（n=713）

分野別にデータ公開経験の有無を確認すると、生物科学分野(66.7%)、農学(64.2%)、計算機科学(63.4%)の順に多く、分野による差がみられた。

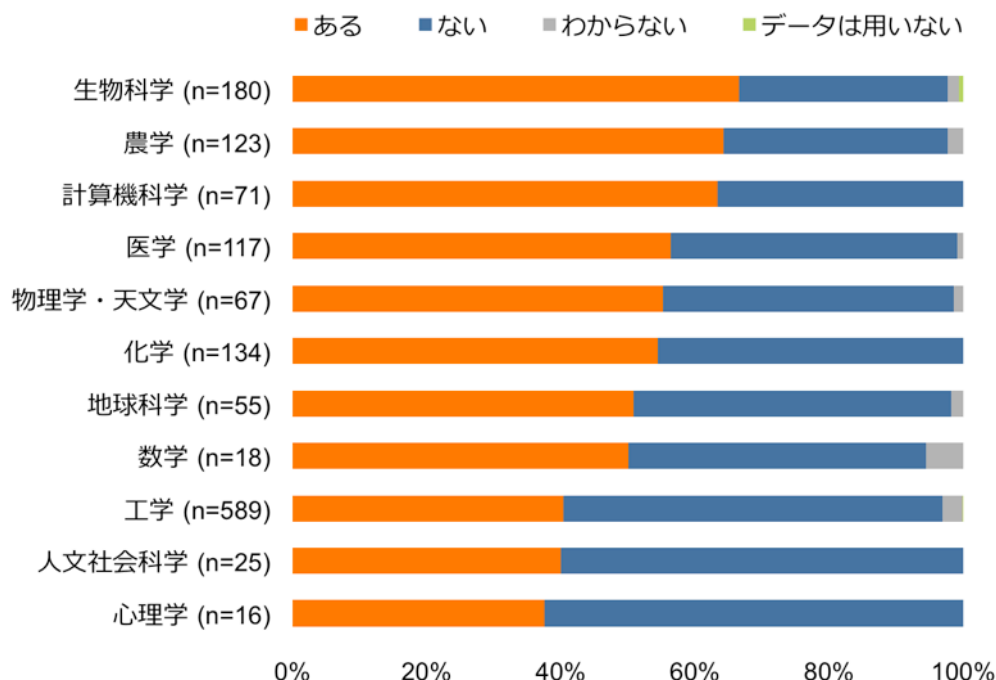


図 3 分野別データ公開経験 (n=1,398)

データを公開した理由は「研究成果を認知してもらいたいから(58.5%)」が最も多く、論文の OA は「投稿した雑誌のポリシーだったから(81.7%)」であった。データを公開しない理由については、「雑誌のポリシーではないから(26.4%)」や「公開のための時間が必要だから(25.9%)」などが比較的に選ばれているものの突出した理由はみられなかった。一方、論文を OA にしない理由は「雑誌のポリシーではないから(60.4%)」と「資金が必要であるから(39.6%)」に集中していた。非公開理由が解決された場合の公開意思を比較すると、データは論文よりも「いいえ」や「わからない」の選択率が高く、慎重な姿勢がうかがえた(図 4)。

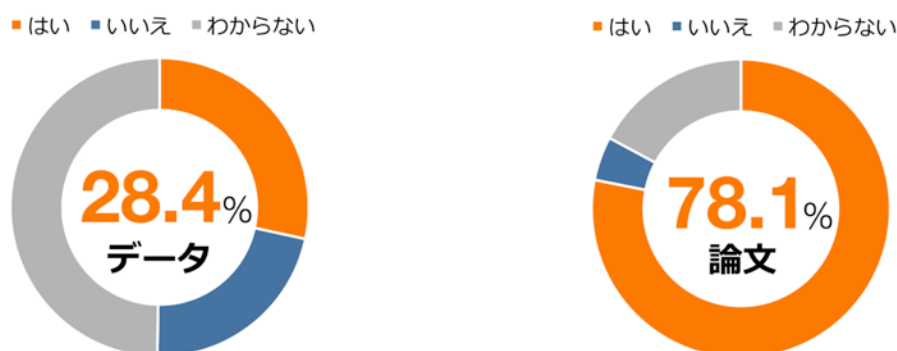


図 4 非公開理由が解決した場合のデータと論文の公開意思(データ n=595, 論文 n=379)

(2) データ公開の障壁

データの公開の障壁を明らかにするために、データ公開経験の有無にかかわらず、研究にデータを用いる回答者全員を対象として、資源の充足度や懸念の強さを尋ねた。その結果、データを整備・公開するために必要な資源は全体的に不足しており、特に人材や時間、資金が不足していると認識されていることがわかった(図 5)。また、データ公開用のリポジトリについては「わからない」とする回答が 27.1%にのぼった。分野リポジトリや機関リポジトリの整備が行われているものの、認知度が低い、あるいは十分ではないということが示唆された。

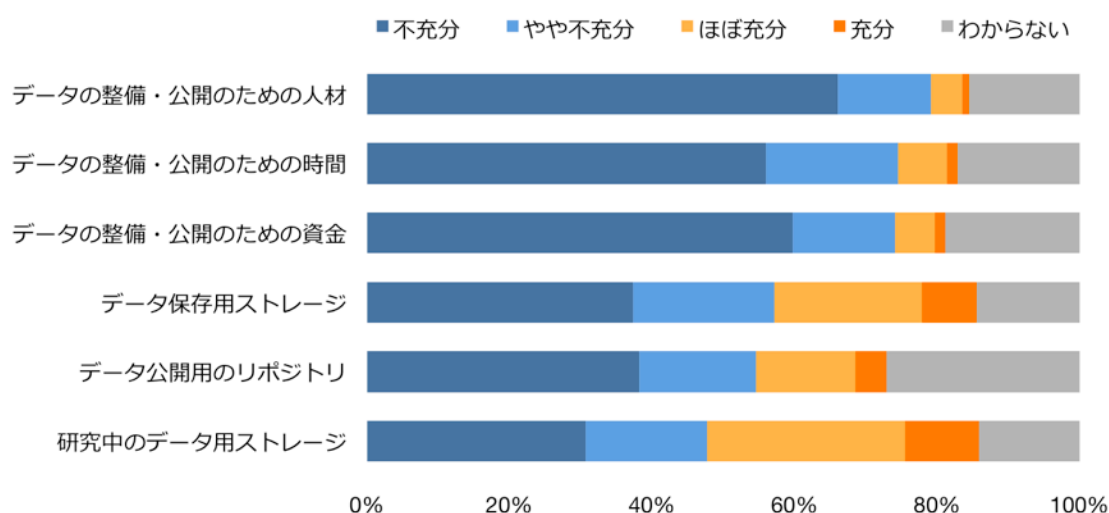


図 5 データ公開に関する資源の充足度 (n=1,396)

データを公開する場合の懸念については、「引用せずに利用される可能性」と公開データを使って「先に論文を出版される可能性」を問題視する回答者が多かった(図 6)。データ公開に関する議論では、論文の出版前にデータを公開することが前提となっている場合が多いが、「先に論文を出版される可能性」に対する懸念は特に若年層で強く、この傾向は先行研究である Tenopir⁴や Schmidt⁵による調査においても同様であった。公開を求めるタイミングは、慎重に検討する必要があると考えられる。

⁴ Tenopir, Carol et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PLOS ONE. 2015, vol. 10, no. 8, e0134826. <http://doi.org/10.1371/journal.pone.0134826>

⁵ Schmidt, Birgit et al. Open data in global environmental research: The Belmont Forum's open data survey. PLOS ONE. 2016, vol. 11, no. 1, e0146695. <http://doi.org/10.1371/journal.pone.0146695>

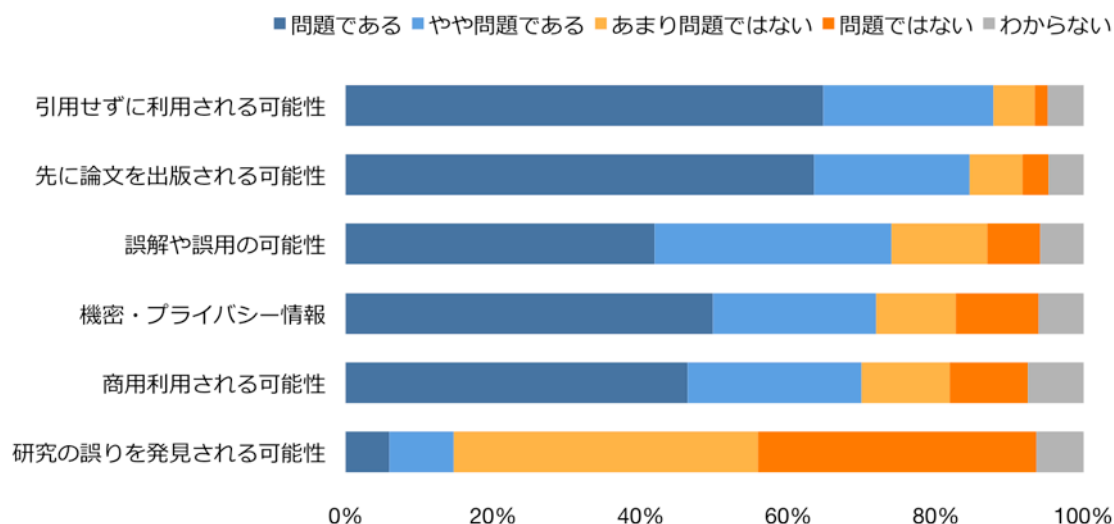


図 6 データを公開する場合の懸念の強さ (n=1,396)

(3) 公開データの利用状況

公開データの入手経験を尋ねたところ、回答者の 75.8%が何らかの方法でデータを入手しており(図 7)、うち、91.2%が研究の参考に、55.3%が再利用・再分析を、46.1%が再現・追試を行っていることがわかった。冒頭で述べた通り、オープンサイエンス政策は公開データを企業や市民が活用することを期待しているが、企業の研究者の 70.7%が入手経験を有しており、特に研究の再現や追試を比較的好く行っていることが明らかになった(54.6%)。

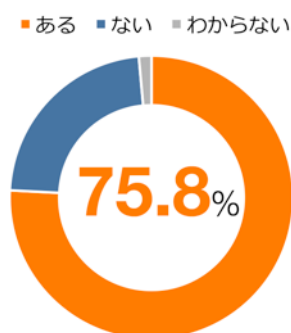


図 7 公開データの入手経験の有無 (n=1,398)

データの入手に最も利用されているのは、公開と同様に「個人や研究室のサイト(64.8%)」であり、次いで「論文の補足資料(53.1%)」であった。また、データの利用においては 79.8%が何らかの問題を感じており、「利用料金が必要(43.1%)」、「利用者登録が必要(33.3%)」、「利用条件がよくわからない(33.1%)」の順に選択率が高かった(図 8)。

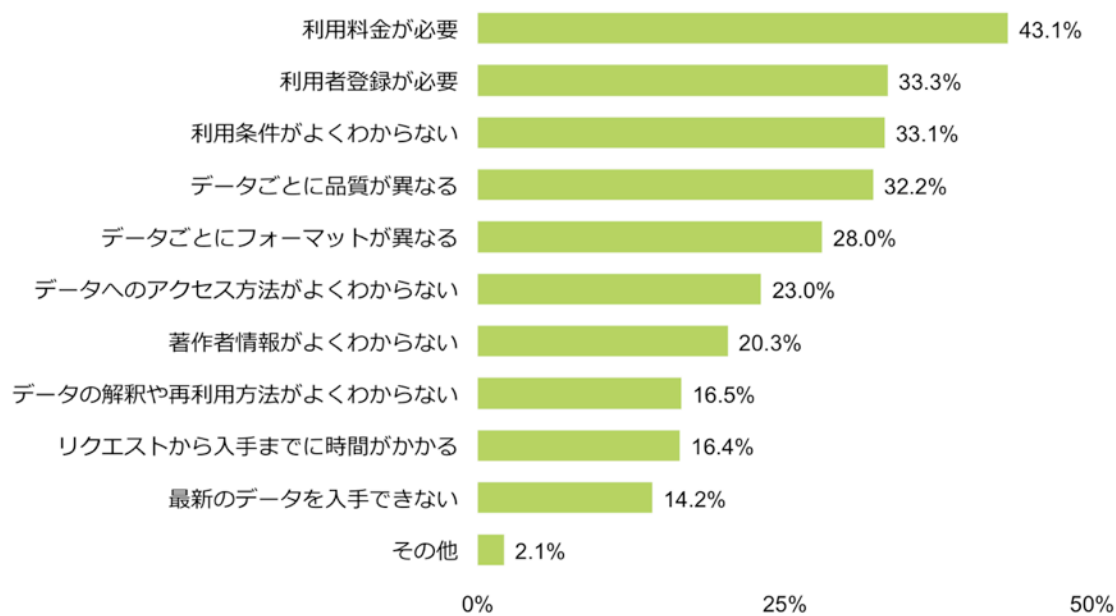


図 8 公開データ入手の障壁 (n=846)

本調査によって明らかになった日本の研究者によるデータ公開の実態や活用における課題について、学術機関、出版社、学協会、政策担当者、研究助成団体といった幅広いステークホルダーによる議論を誘発し、研究成果を効率的に最大化する仕組みを検討することが望まれる。また、本調査を原点調査と位置づけて、日本の研究者によるオープンサイエンスの実施と認識が今後どのように変化していくかを追跡するとともに、研究活動の推進のために必要とされている支援内容を明らかにしていきたい。

本編

1. 調査背景・目的

1.1 オープンサイエンスの隆盛と学術情報流通の変化

近年、科学研究の成果として論文と同様にその根拠となるデータを公開し、利活用を促進する動きが分野や地域を超えて広がっている¹。2007 年に OECD(経済開発協力機構)が『公的資金による研究データへのアクセスに関する OECD 原則とガイドライン』²を公開したことから各国・地域の政府組織や学術機関による対応がはじまり、さらに 2013 年の G8 科学技術大臣会合における研究データのオープン化に関する合意³を契機として、取組が加速した⁴。そして研究データの公開は、論文のオープンアクセス(以下、「OA」と記す)とあわせて「オープンサイエンス」と呼ばれるようになり⁵、その名を冠した政策やデータ公開基盤が次々と公開されている⁶。

オープンサイエンスの実現によって、研究の効率化⁷や研究不正への対策⁸、異分野データの統合による新たな成果の創出、産学連携による共同研究の推進、市民科学での活用⁹、イノベーションの創出などが期待されている。このため、各国の政府や助成機関は相次いでデータ公開を義務化している¹⁰。データ公開の義務化とは、研究資金を申請する際にデータ公開や保存の方法などを記したデータ管理計画(Data Management Plan, 以下「DMP」)の提出を求めることを指す。たとえば米国では、オバマ政権による 2013 年の OSTP(科学技術政策局)指令¹¹に呼応して、DOE(エネルギー省)や NASA(航空宇宙局)などが相次いで DMP を義務化している。

日本においてもオープンサイエンスに関する議論が盛んに行われるようになり、2015 年には内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」の報告書(以下、「内閣府報告書」)¹²、2016 年には『学術情報のオープン化の推進について(審議まとめ)』(文部科学省)¹³や『オープンイノベーションに資するオープンサイエンスのあり方に関する提言』(日本学術会議)¹⁴が公開された。第 5 期科学技術基本計画(2016 年度～2020 年度)¹⁵では、“国は、資金配分機関、大学等の研究機関、研究者と連携し、オープンサイエンスの推進体制を構築する”と述べられている。そして 2017 年 4 月には、科学技術振興機構が助成金を獲得した全てのプロクトを対象として、研究開始までに DMP を提出するよう求める『オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針』¹⁶を公開した。この方針は、助成金の申請時ではなく採択後に DMP を提出するよう求めている点や、データ公開を(必須ではなく)推奨している点が、国外でしばしばみられるデータ公開方針よりもゆるやかであるが、日本の研究者に研究データの公開や管理について意識喚起させるものとなっている。

さて、こうした政策文書では、研究データの公開を進める際には、国や分野の状況に応じた対応が重要であると繰り返し指摘されている。たとえば内閣府報告書においては、“各省庁等のステークホルダーは、オープンサイエンスを推進すべき領域、プロジェクトを選定し、研究活動上の利益・損失や研究途上の取扱い及び機微の判断など各分野の専門家・研究者、技術者の意見を十分に取り入れ、その分野の活動・研究成果が最大化されることを旨として、オープンサイエンス実施方針を定める”¹²と述べられている。また、データ公開に関する国際調査¹⁷を紹介した近藤は、国や地域によって結果に差異がみられることから、日本やアジアで“オープンリサーチデータを推進する際には、国際動向をふまえつつも、国・地域の事情に即した施策が必要”であると指摘している¹⁸。

学術界では、学術雑誌や出版者によるデータ公開要求が経年的に増加しており¹⁹、たとえば2016年12月、Springer Nature社は600以上の雑誌についてデータ共有ポリシーを適用すると発表した²⁰。一方、日本の研究者は論文の生産性が停滞していることなどが指摘されており²¹、第5期科学技術基本計画では、日本の総論文数を増やし、被引用回数トップ10%論文数の割合が10%となることを目指すとしている¹⁵。現状では学術雑誌によるデータ公開ポリシーは分野による差がみられる²²が、データ公開が盛んではない分野の研究者にとっては負担が大きく、論文投稿の障壁になる可能性がある。加えて、日本の研究者は研究時間や研究開発費が低迷していると指摘されている²³ことから、データ公開のための適切な支援体制の構築は、学術政策における喫緊の課題であるといえよう。

また、データ公開をデータ出版(data publishing)と呼び、論文の出版と同様に扱い、引用し、業績や評価の対象とする動向がみられる²⁴。データ出版については、公開データとデータに関する記述(data paperやdata descriptor)を査読して掲載するデータジャーナル²⁵がElsevierやNature Publishing Groupなどの大手学術出版社から相次いで刊行されており、研究成果の迅速な共有やビジビリティの向上を可能にしている。また、論文の引用と同様に、研究に用いたデータの典拠を示すデータ引用(data citation)を促す取組もみられる。2014年に国際イニシアティブのFORCE11が公開した『データ引用の共同原則』²⁶には学術出版社やデータアーカイブなどが賛同しており、2016年11月にはElsevier社が1,800を超える雑誌に採択した²⁷。データ引用を追跡するツールとして、2012年にThomson Reuter社(現Clarivate Analytics社)はData Citation Index(DCI)の提供を開始した。これによって、データの被引用回数や引用論文を調査することが可能となっている。そしてデータ公開の業績化について、NSF(米国国立科学財団)は2013年1月に業績記入欄の名称を“出版物(Publications)”から“生産物(Products)”に変更し、公開したデータやプログラムコードを記載することを可能にした²⁸。データを“学術の一級市民(first-class citizens of scholarship)”²⁹とみなす向きもあり、ゆくゆくはデータ引用が研究者や大学の評価指標として活用される可能性もある。こうした学術情報流通や評価の変化に迅速に対応し、日本の研究成果のプレゼンスを高めることも重要な課題であると考えられる。

オープンサイエンス政策や学術情報流通の新たな動向に対応するために、国外においては研究者を対象としたデータ公開の実態調査や課題の分析が行われている。一方、日本の研究者については、小野らによる地球科学・環境学などの研究者38名の回答を分析した意識調査³⁰や、日本学術会議による学協会を対象とした質問票調査¹⁴、デジタルリポジトリ連合(DRF)によるインタビュー調査³¹⁻³²、倉田らによるデータ管理に関する調査³³が行われているものの、データ公開に関する大規模な実態調査は行われていない。また、G7科学大臣会合は、オープンサイエンスを“幅広い分野の公的資金による研究成果に学術関係者だけでなく、民間企業や一般市民が、広く利用・アクセスできるようにするもの”³⁴としているが、国外の調査においても企業の研究者の回答率は低く、Tenopirら³⁵による調査では2.6%、Schmidtら¹⁷の調査においても2.6%であり、利活用の状況が十分に明らかにされているとは言い難い。そこで科学技術・学術政策研究所(NISTEP)科学技術予測センターは、2016年11月から12月にかけて、大学や公的機関、企業に所属する日本の研究者を対象とした、アンケートシステムによる調査を実施することとした。次項では、具体的な調査目的について述べる。

1.2 調査目的

オープンサイエンスの実現によって、学术界や産業界のみならず、社会全体における多様な効果が期待されていることから、政府機関や学術雑誌によるデータ公開の要求が高まり、評価に繋がる流れもできつつある。その重要性は十分に理解できる一方で、データ公開は多くの研究者にとって新たな取組であり、公開のためのエフォート、資金、人材、インフラなどの負担や公開への懸念があることが予想される。したがって、従来の研究活動をできるだけ妨げることなく、研究成果の発信力や効果を最大化するための適切な支援体制を検討する必要があると考えられる。そこでアンケートシステムによる調査を実施して、研究データ公開に関する実態を明らかにすることとした。

調査目的は、日本の研究者によるデータ公開と利用の実態、研究者が扱っているデータ、データ公開における課題、及びデータ公開・管理に対する認識を明らかにすることの4点である。比較対象として、データ公開よりも進展しているとされるOA論文⁶の状況もあわせて調査した。また、次章で述べる先行研究を質問票の参考とするとともに、結果の比較を行った。分析においては、内閣府報告書で指摘されている“各分野の専門家・研究者、技術者の意見を十分に取り入れ、その分野の活動・研究成果が最大化されること”¹²を目指すため、分野や所属機関ごとの差を明らかにした。

1.3 調査の意義と有効性

本調査の意義として、日本の研究者による研究データの公開状況や課題、研究者の認識を明らかにすることによって、オープンサイエンスに係る学術政策の議論における基礎資料を提供することが挙げられる。特に分野別のデータ公開状況や心理的な障壁について、数値による傾向の把握や比較を可能にした点に本調査の有効性があると考ええる。

物的資源の状況や人的支援の可能性に関する結果は、大学や研究機関、企業のマネジメント層、あるいは研究支援を行う技術職員やURA、図書館員などがデータ公開に関する適切な支援体制を検討する際に活用できると考えられる。研究者がデータを管理・公開しようとする際に何が不足しているのか、どこに懸念があるのか、どのような支援を求めているのかといった情報は、優先課題を特定するための判断材料となるのではないだろうか。

また、日本の研究者が扱うデータのプロフィールや入手における課題、関心の高いメトリクスなどは、データ公開用のリポジトリや検索システムといった研究基盤を構築する際の参考になると考えられる。研究者が必要とする機能や現状の問題点などを明らかにすることによって、基盤整備のための検討に資することを目指した。

2. 先行研究

2.1 データの公開状況

研究者を対象としたデータ公開に関する最近の質問紙調査について、調査者や調査年、主な回答者などを表 1 に示す。日本の回答者数が示されている場合は括弧に入れて示した。

表 1 データ公開に関する主な質問紙調査

| 調査者／調査主体 | 調査年 | 主な回答者 | 回答数 | 公開率 |
|-------------------------------------------|---------------|-----------------------------|----------------|--------------|
| Tenopir ら ³⁵ /DataONE | 2009 /2010 | 環境科学, 生態学等の研究者 | 1,329 | 36% |
| Huang ら ³⁶ | 2011 | 生物多様性分野の研究者 | 372 | 85% |
| Kim ら ³⁷ | 2012 /2013 | 米国の STEM 分野の研究者 | 1,317 | |
| Kim ら ³⁸ | 2012 /2013 | 米国の STEM 分野の研究者 | 1,298 | |
| Kim ら ³⁹ | 2012 /2013 | 米国の生物学分野の研究者 | 608 | |
| Danvad 社 ⁴⁰ | 2013 | ノルウェーの社会科学, 健康科学, 自然科学等の研究者 | 1,474 | 28% |
| Tenopir ら ⁴¹ /DataONE | 2013 /2014 | 環境科学, 生態学等の研究者 | 1,015 | |
| Kratz ら ⁴² | 2014 | 生物学や考古学等の研究者 | 249 | 68% |
| Ferguson ⁴³ /Wiley 社 | 2014 | 多分野の研究者 | 2,250+ (不明) | 52% (44%) |
| Schmidt ら ¹⁷ /Belmont Forum | 2014 | 地球科学・環境学等の研究者, 技術員 | 1,253 (26) | |
| Fecher ら ⁴⁴ | 2014 | 独国の自然科学, 社会科学等の研究者 | 1,564 | |
| 小野ら ³⁰ | 2015 | 日本の地球科学・環境学等の研究者 | 38 (38) | 8% (8%) |

※()内は、日本の研究者の数・比率

データの公開率は、調査によって尋ね方や回答者の属性が異なるため単純に比較することは難しいが、多分野の研究者を対象とした調査結果を 4.1 で本調査の結果と比較した。多分野の研究者を対象とした調査の結果について、Tenopirら(調査年 2009/2010, 以下では調査年のみを記す)³⁵によれば、“他者が簡単に自分のデータにアクセスできる”という質問に“強く同意する”または“同意する”を選んだ回答者は 1,329 名のうち 36%であった。ノルウェー研究評議会のための報告書⁴⁰で“すべての人”または“研究者”がデータを入手できるとした回答者は 1,474 名のうち 28%であった。Ferguson⁴³の報告では、“データを一般に公開している”回答者は 2,250 名以上のうち 52%であり、日本の回答者は 44%が公開していた。また、ライフサイエンスは 66%, 健康科学は 48%, 自然科学(physical sciences)は 45%, 人文・社会科学は 36%であり、分野による差がみられた。また、本研究と同様に 2016 年に実施された Berghmansら⁴⁵の報告は、多分野の研究者 1,162 名を対象としているが、論文の補足資料やデータリポジトリへの登録など、何らかの方法でデータを公開している回答者は 66%であった。

特定の分野やコミュニティを対象とした Huangら³⁶と Kratzら⁴²の調査では、データ公開率がより高かった。Huangらの調査は生物多様性分野の主要な 3 誌の著者を対象としているが、この 3 誌はデータ公開を求めていることもあり、論文に関連したデータを“常に”, “しばしば”, “時に”共有している回答者の合計は 85%であった。また、California Digital Library の Kratzらが SNS やブログで参加を呼びかけた調査では、生物学や考古学などの研究者 249 名のうち 68%がデータ共有経験を有していた。本調査は、Huangらと Kratzらの質問項目を参考としたが、データ公開率については、回答者がデータ公開を求められている特定分野、及びデータ公開に関心があるコミュニティであったため、比較対象とはしなかった。

2.2 データ公開の方法

データの公開方法について、先行研究で回答率が高かった方法を 3 件ずつ示す。Huangら³⁶によれば、論文の補足資料(52%), パブリックデータベース(38%), 機関のサイト(25%)の順であり、Ferguson⁴³の報告では、雑誌の補足資料(67%), 個人・機関・プロジェクトのウェブページ(37%), 機関リポジトリ(26%), Kratzら⁴²では、直接コンタクト(87%), リポジトリ(54%), 雑誌の補足資料(37%), そして Berghmansら⁴⁵では、出版物の付録・補足資料(33%), データジャーナルでの出版(28%), データリポジトリ(13%)であった。以上の結果から、リポジトリやパブリックデータベースによるデータ公開よりも、雑誌の補足資料や特定のウェブサイトへの掲載、あるいは個人的なやり取りによるデータ共有の方が行われている傾向にあることがうかがえた。

2.3 データの公開理由と障壁

先行研究では、データを公開する理由やデータ公開の際の障壁について、さまざまな結果が示されている。そこで、表 1 に示した質問紙調査、及び Wallisら⁴⁶, Van den Eyndenら⁴⁷, Nature news チーム⁴⁸によるインタビュー調査、Wichertsら⁴⁹による公開データの分析、Fecherら⁵⁰によるシステムティックレビューで明らかにされたデータの公開理由と障壁のうち、複数の研究で指摘されていた主要な事項を整理して表 2 に示す。その際、類似の内容は統合して示した。たとえば Schmidt らはデータ公開の障壁として、“誤解や誤用の懸念”と“データが独り歩きすることへの懸

念”を挙げているが、後者はデータが文脈を無視して誤った解釈で利用されることへの懸念であり、広義には誤用に対する懸念であると考えて“誤用・悪用への懸念”に統合した。表 2 のうち資源については、充足している場合は公開理由となるが、不足している場合は障壁となる。

表 2 先行研究によるデータの公開理由と障壁

| | 公開理由 | 障壁 |
|--------------|-------------------------------------------------------------|--------------------------------------------|
| 心理的・ 内的要因 | 研究成果の認知度の向上 科学的利他性 科学研究の発展・加速 オープンデータへの貢献 業績・評価 | 引用せずに利用される可能性 研究の誤りを発見される可能性 データの難しさ |
| 規範 | 分野・コミュニティの文化 科学的利他性 | データの誤用・悪用への懸念 先取権の喪失・盗用への懸念 |
| ポリシー | 助成機関 学術雑誌 所属機関 | |
| 法・倫理 | | 機密・プライバシー情報 商用利用への懸念 知的財産権 |
| 資源※ | 公開のための時間・資金・人材 公開のためのリポジトリ | |

※充足している場合は公開理由に、不足している場合は障壁となる。

3. 調査・分析方法

Web アンケートシステム Questant を用いて調査を実施した。本章では、質問票に用いた主要な概念・用語の定義を示した後、調査対象、プレテスト、質問票の構成などについて述べる。

3.1 主要な概念の定義

本報告書及び質問票で使用した概念や用語の定義は、以下の通りである。

データ

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もので、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。

カレントデータ

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・観測したデータを指す。プレテストの結果、研究ごとに扱うデータの種類や量が異なる回答者が存在することが確認されたため、対象を限定して回答できるように定義した。

データの提供

E-mail や USB フラッシュメモリ、クラウドサービス (Dropbox や Google Drive) などを使って、共同研究者を除く他者に渡す (共有する) ことを指す。特定の人以外はアクセスできない状態として、「データ公開」と区別する。

データ公開

データ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して、インターネットでアクセスして利用できる状態を指す。利用料金や利用者登録が必要な場合も含める。また、このような状態で公開されているデータを「公開データ」と呼ぶ。

オープンアクセス (OA)

論文がインターネットで公開され、読者は無料で読むことができる状態とする。いわゆるゴールド OA、グリーン OA の別は問わない。たとえば、OA の雑誌で出版する場合や雑誌の OA オプションを選択した場合、雑誌等が一定期間経過後に論文を OA にする場合、著者が機関リポジトリやプレプリントサーバで論文を公開する場合を含む。

オープンサイエンス

データ公開とオープンアクセスを含む概念とする。オープンエデュケーションやオープンピアレビュー等は本調査では対象としない。

3.2 調査項目

4 つの調査目的を達成するため、先行研究に基づいて 6 つの調査項目を設定した(表 3)。先行研究と比較を行う際には、先行研究と同じ選択肢や質問方式を用いることが望ましいが、回答者の負担を軽減するために全体の質問数を抑制すること、表 2 と同様に似たような選択肢を統合すること、尺度を尋ねる質問は合計 30 以下として適宜複数選択方式に変更することを基本方針とした。また、無回答や適当な回答を避けるために、「わからない」という選択肢を加えた。

表 3 調査目的と調査項目

| 調査目的 | 調査項目 |
|-------------------|----------------------------------------------------|
| 1. データ公開と利用の実態 | (1)データの公開状況 (2)データの提供・被提供状況 (3)公開データの利用状況と課題 |
| 2. 研究者が扱うデータの把握 | (4)データのプロフィール |
| 3. データ公開における課題 | (5)データを公開する場合の障壁 |
| 4. データ公開・管理に対する認識 | (6)データの管理・公開に対する関心と専門性 |

以下では各調査項目の概要を述べ、表 4 に調査項目ごとの質問と、それぞれの質問を作成する際に参考とした先行研究の一覧を示す。なお、研究分野は、米国科学審議会 (National Science Board) の科学工業指標 (Science and Engineering Indicators) の分類から「その他生命科学」を削除して「人文学」を追加した 13 分野、及び「その他」を提示した。分野を尋ねる質問は、Q1, Q6, Q14, Q17 である。

1. データ公開と利用の実態

(1) データの公開状況

研究データの公開経験の有無、及びその方法を確認した。表 2 の「公開理由」を参考に、公開・非公開の理由をそれぞれ尋ねるとともに、非公開理由が解決された場合の公開意思を確認した。また、OA 論文についても同様に尋ねて比較した。

(2) データの提供・被提供状況

個人的にデータの提供や被提供(データ共有)を行っている研究者は、研究の効率化や共同研究といったメリットを理解しているためデータを公開する傾向にあるのではないかと、また、データ共有を行ってきた分野や研究者はデータ公開を行う傾向にあるのではないかとという仮説に基づき、データの提供・被提供経験を尋ね、データ公開経験との関連を明らかにした。

(3) 公開データの利用状況と課題

公開データの利用経験によって、自身のデータも公開しようとする利他的な動機が醸成されるのではないかと考えられる。一方、公開されたデータは利用したいが、自身のデータは公開していない(公開できない)とする調査結果もみられる^{35 51}。そこで、公開データの利用状況について尋ね、データ公開経験との関連を明らかにした。あわせて、公開デー

タの入手における問題や、入手したデータの利用目的、分野、今後利用してみたい分野も尋ねた。

また、公開データの流通環境の整備に資するために、公開データの探索方法や、入手したデータの信頼性の判断基準について尋ねた。参考のため、論文についても同様に尋ねて比較を行った。

2. 研究者が扱うデータの把握

(4) データのプロファイル

研究者が扱うデータは分野や研究テーマによって多種多様であると考えられる。データの量やデータに含まれる機密情報などがデータ公開の障壁となっているのではないかという仮説に基づき、研究者が扱っているデータについて尋ね、データ公開経験との関連を明らかにした。

なお、同じ研究者であっても研究によって扱うデータの量や種類、性質が異なる場合があるため、前節に示したように、“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”を「カレントデータ」と定義して、当該データについて回答を求めた。

3. データ公開における課題

(5) データを公開する場合の障壁

先行研究でデータ公開を阻害する心理的な障壁や資源不足などが明らかにされている。そのうち日本の研究者にとって重要な障壁を特定するために、表 2 に示した障壁を参考として、それぞれに対する懸念の強さや充足度を明らかにした。

この設問は、データ公開経験の有無にかかわらず全ての回答者に尋ねることとして、「カレントデータ」を公開する場合を想定して回答を求めた。データ公開経験が「ない」または「わからない」とした回答者には公開することを想定してもらうため、例として“研究 1 件のために質問紙調査とインタビュー調査を実施した場合”を示した。この場合、対象となるカレントデータは、“質問紙の回答を入力したスプレッドシート、回答を分析するための R のコード、インタビューを録音した音声データ、インタビューを書き起こしたテキストデータなど”であり、データの公開とは、“質問紙調査の回答データを第三者が再利用できるように、調査概要の説明を作成、回答を入力したスプレッドシートから個人情報情報を削除、項目に見出しをつけるといった処理を行った上で、質問紙や分析のためのコードとともにリポジトリに登録する”とした。

4. データ公開・管理に対する認識

(6) データの管理・公開に対する関心と専門性

国外の大学・研究図書館ではデータ公開やデータ管理の支援を行っており、『学術情報のオープン化の推進について(審議まとめ)』¹³においても、“技術職員、URA 及び大学図書館職員等を中心としたデータ管理体制を構築し、研究者への支援に資する”と述べられている。そこで、データを公開した場合のメトリクスや、データの管理・公開のためのリテラシーへの関心、及びデータを扱う際の専門性について尋ねた。

表 4 質問項目と参考文献

| データ | 論文 | 質問項目 | 参考文献* |
|------------------------|----|---------------------|-----------|
| Q1 | | 研究分野 | 科学工業指標 |
| (1)データの公開状況 | | | |
| Q18 | Q2 | データの公開経験 | F, H, K1 |
| Q19 | Q3 | データの公開理由 | 表 2 |
| Q20 | Q4 | データの非公開理由 | 表 2, T |
| Q21 | Q5 | 非公開理由が解決された場合の公開意思 | |
| (2)データの提供・被提供状況 | | | |
| Q9 | | データの提供頻度 | |
| Q10 | | データの被提供頻度 | |
| (3)公開データの利用状況と課題 | | | |
| Q11 | | 公開データの入手経験 | F, H, K1 |
| Q12 | | 公開データ入手の障壁 | S |
| Q13 | | 公開データの利用目的 | K2, P |
| Q14 | Q6 | 利用したことがある公開データの分野 | T |
| Q15 | Q7 | 公開データの検索ツールと情報源 | K2, S |
| Q16 | Q8 | 公開データの信頼性の判断基準 | K2 |
| Q17 | | 今後利用してみたい公開データの分野 | T |
| (4)データのプロフィール | | | |
| Q22 | | 所属機関のデータ保存期間規定の有無 | |
| Q23 | | カレントデータの量 | |
| Q24 | | カレントデータの所有権 | 表 2, O |
| Q25 | | カレントデータの機密情報 | O |
| Q26 | | カレントデータの望ましい保存期間 | |
| (5)データを公開する場合の障壁 | | | |
| Q27 | | カレントデータ公開の有無 | F, H, K1 |
| Q28 | | カレントデータの公開に必要な資源の状況 | 表 2, S, T |
| Q29 | | カレントデータ公開の懸念 | 表 2, S, T |
| Q30 | | カレントデータの理解(自分分野) | |
| Q31 | | カレントデータの理解(異分野) | |
| (6)データの管理・公開に対する関心と専門性 | | | |
| Q32 | | メトリクスへの関心 | K2 |
| Q33 | | カレントデータの公開に関する知識・関心 | J |
| Q34 | | カレントデータの管理・公開の専門性 | J |
| Q35 | | 自由回答 | |

*F=Ferguson⁴³, H=Huang ら³⁶, J=Jones ら, K1=Kratz ら⁴², K2=Kratz ら⁵², O=OECD⁶, P=Piwowar²⁸, S=Schmidt ら¹⁷, T=Tenopir ら(2009/2010)³⁵ 及び Tenopir ら(2013/2014)⁴¹

3.3 プレテスト

プレテストは3回実施して、質問の順序やワーディング等の修正を行った。プレテストの第1回(2016年8月26日～30日)は紙で実施し、第2回(11月4日～7日)と第3回(11月21日～24日)は本調査と同じアンケートシステムを用いて実施した。対象は大学や企業の研究者であり、それぞれ15名、8名、12名の協力を得た。

プレテストによる主な改善点は次の2点である。(1)研究によって扱うデータの種類や量が異なるという指摘から、“論文などの成果物を出版済みの最新の主な研究1件のためのデータ”を「カレントデータ」と定義した上で、質問を行う。(2)データ公開にはなじみのない回答者が多いという指摘や、“論文もデータも公開していないと答え難い”という意見から、質問の順番を入れ替えて論文に関する質問を冒頭にまとめた。また、依頼文に“論文やデータの公開経験がない、というご回答も参考になりますので、ぜひ、率直なご意見をお聞かせ下さい。”という一文を追加した。

最終的な質問数は35問、うち尺度項目は7問(19件)として、以下の7セクションに分けて尋ねた。回答者が事前に全ての質問を確認できるように、アンケートのトップページにPDF版の質問票を掲載した。原則としてアンケートシステムでの回答を求めたが、問題が生じた場合は、Wordファイル等による提出も受け付けることとした。

1. 研究分野について(Q1, 1問)
2. 学術論文について(Q2～Q8, 7問)
3. 研究データの提供について(Q9～Q10, 2問)
4. 公開データの利用について(Q11～Q17, 7問)
5. 研究データの公開について(Q18～Q22, 5問)
6. 最近の研究のためのデータについて(Q23～Q34, 12問)
7. 自由回答(Q35, 1問)

3.4 調査対象

調査対象は、科学技術予測センターが運営している「科学技術専門家ネットワーク」⁵³である。科学技術専門家ネットワークとは、産学官の研究者、技術者、マネージャ等を含む2,000人規模の専門家集団であり⁵⁴、多分野かつ幅広い年齢層の回答者による意見を収集することができる。2016年度の科学技術専門家ネットワーク構成員の所属の比率は、大学60.4%、企業22.7%、公的機関・団体16.0%、その他0.9%であった。

3.5 実施期間

調査期間は、2016年11月30日から12月9日とした。アンケートへの協力依頼は、11月30日にE-mailで科学技術専門家ネットワークの各位に送信した。多重回答を防ぐため、回答者ごとの個別URLを作成した上で、回答完了後には再度回答が行えないよう設定した。リマインダは、未回答者を対象として12月2日と7日に送信した。なお、第2回目のリマインダでは、回答率が低い分野を示して協力を仰いだ。12月10日以降も回答入力があったため、最終的に12月14

日の回答までを結果に含めた。回答のうち 1 件は Word ファイルによる提出であったため、アンケートシステムに調査者が登録した上で、他の回答とあわせて出力した。

3.6 分析方法

分析に先立って、回答のクリーニングを行った。まず、選択肢「その他」に入力された記述のうち、適切な選択肢があると判断できる場合は当該選択肢を選んだものとした。たとえば、研究分野は 19 名が「その他」を選択していたが、16 名分は該当する選択肢に修正した。ただし、判断に迷う場合は修正を行わなかった。また、「その他」や「自由回答」における記述の誤字・脱字は適宜修正した。

各回答は、所属、分野別に差があるかどうかを分析した。まず、クロス集計表を作成して、名義尺度はカイ二乗検定を、順序尺度は Kruskal-Wallis の検定を行った。続いて、データ公開経験の有無との関連を確認するために、名義尺度はカイ二乗検定を、順序尺度は Mann-Whitney の U 検定を行った。所属と分野の「その他」、及び回答の「わからない」は欠損値とした。年齢は Schmidt ら¹⁷の調査結果と比較するために 5 年ごとにまとめたが、21～25 歳の回答者は 1 名であったため、21～30 歳としてまとめた。有意水準 p は 5%として、有意差がみられた場合は適宜 p 値を示した。ソフトウェアは SPSS Ver.24 を用いた。次章の結果では、まず全体の結果について示した上で、差がみられた項目や先行研究と比較した結果について述べる。

3.7 回答率と回答者の属性

調査依頼の送付数は 1,983 名、最終回答数は 1,406 名(回答率 70.9%)であった。以下では、所属、年齢、分野別に回答者の構成比率について述べる。なお、図表の「 n 」は回答者数を示す。

回答者の所属は、大学が 844 名(全体の 60.0%)で最も多く、次いで企業 330 名(23.5%)、公的機関 188 名(13.4%)の順であった(図 9)。科学技術専門家ネットワークの所属別人数に対する回答率は、「その他」を除くと 70.4%～73.2%であり、特に回答率が低い所属はなかった。

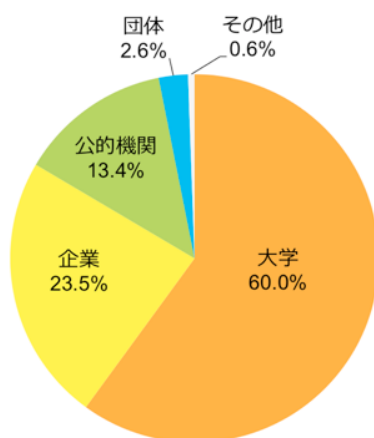


図 9 回答者の所属 (n=1,406)

年齢は、31～40 歳が 649 名（全体の 46.2%）で最も多く、次いで 41～50 歳が 414 名（29.4%）、51～60 歳が 188 名（13.4%）の順であった（図 10）。回答率は若年層ほど高いものの、特に低い年齢層はなかった。

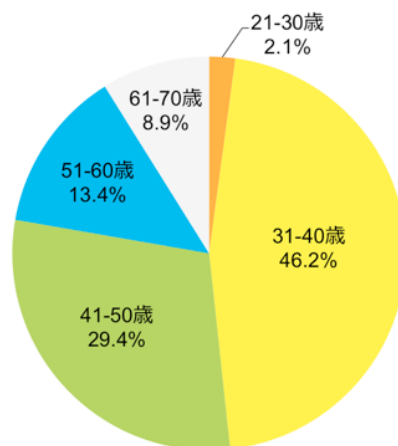
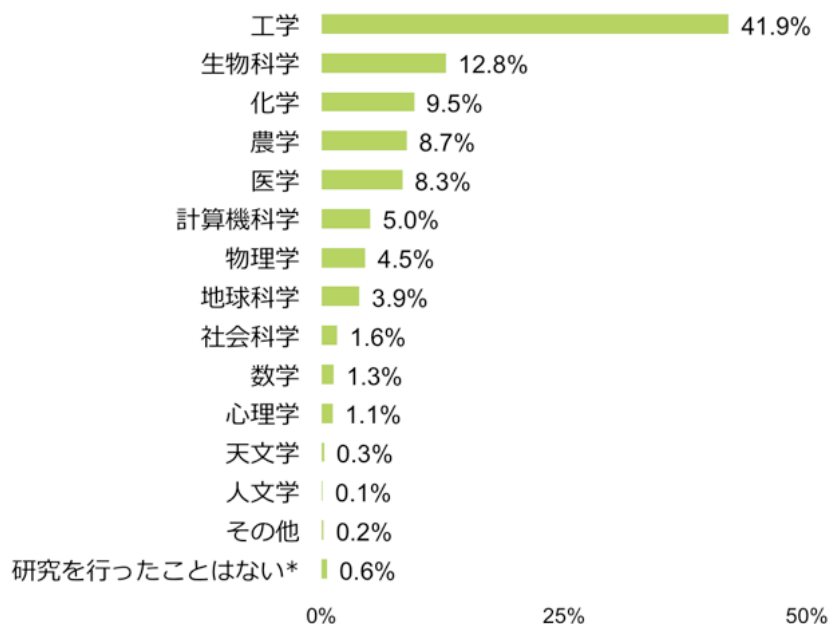


図 10 回答者の年齢（n=1,406）

Q1 で尋ねた分野別の回答率を図 11 に示す。工学が 589 名（41.9%）で最も多く、次いで生物科学が 180 名（12.8%）、化学が 134 名（9.5%）であった。回答者数が 10 名以下であった天文学（4 名）と人文学（2 名）については、それぞれ物理学、社会科学とあわせて「物理学・天文学」（合計 67 名、4.8%）、「人文社会科学」（合計 25 名、1.8%）として分析を行った。なお、「計算機科学」は質問票では「コンピュータサイエンス（CS）」と記した。



*口頭発表や論文出版の経験はない

図 11 専門分野（n=1,406）

4. 調査結果

調査結果について、「データと論文の公開状況」、データ公開の前提であると仮定した「データの提供・被提供状況」及び「公開データの利用状況と課題」、研究者が扱っている「データのプロフィール」、データ公開や管理における問題点を明らかにするための「データ公開における障壁」及び「データの管理状況と支援の可能性」、そして「自由回答」の順に示す。それぞれの結果に関連する自由回答は、結果とあわせて示す。また、データ公開経験の有無、及び分野、属性(所属や年齢)による差がみられた場合はそれぞれ示し、本章の最後(4.8 節)に一覧表としてまとめる。

なお、Q1 から Q34 までの単純集計結果は資料として掲載し(p.91～)、Q35 の自由回答は別途 Web 上に電子付録として掲載した(<http://doi.org/10.15108/rm268>)。

4.1 データと論文の公開状況

(1) データと論文の公開経験

データと論文の公開経験と方法を確認するため、「Q18. これまでに、研究データを以下の方法で公開したご経験はありますか?」と尋ねた(論文は Q2)。選択肢として 7 種類の公開方法を複数選択方式で示すとともに、同時に選べない排他的選択肢として「公開したことはない」、「わからない」、「研究にデジタルデータは用いない」を示した。

その結果、データの公開経験は 713 名(51.0%)、OA の論文がある回答者は 991 名(70.9%)であり、論文の公開率の方が高かった。また、データと論文両方の公開経験がある回答者は 568 名(40.6%)、いずれもない回答者は 214 名(15.3%)であり、OA 論文の有無とデータの公開経験の有無には関連がみられた($p<0.001$)。

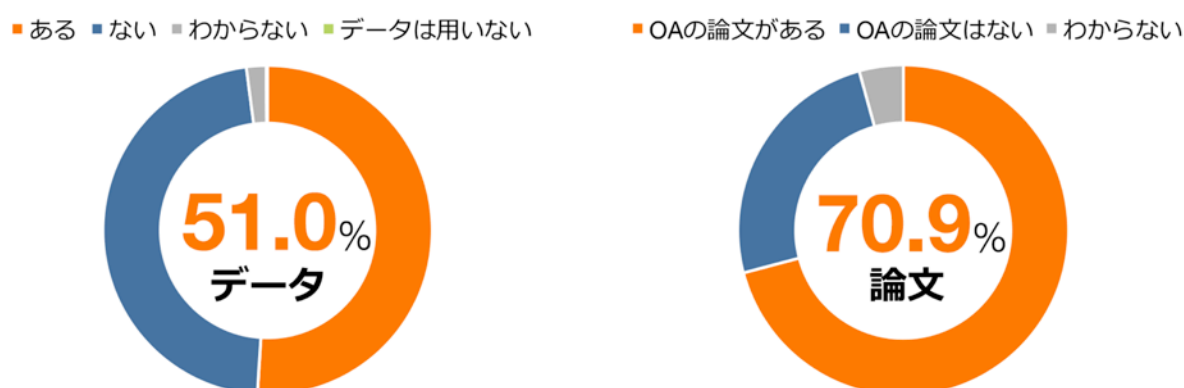


図 12 公開データと OA 論文の有無 (いずれも n=1,398)

データ公開は分野によって差があることが指摘されているが、分野別に公開率を比較すると、生物科学(66.7%)、農学(64.2%)、計算機科学(63.4%)の公開率が高く、心理学(37.5%)、人文社会科学(40.0%)、工学(40.4%)は半数未満であり、本調査においても差がみられた($p<0.01$)。

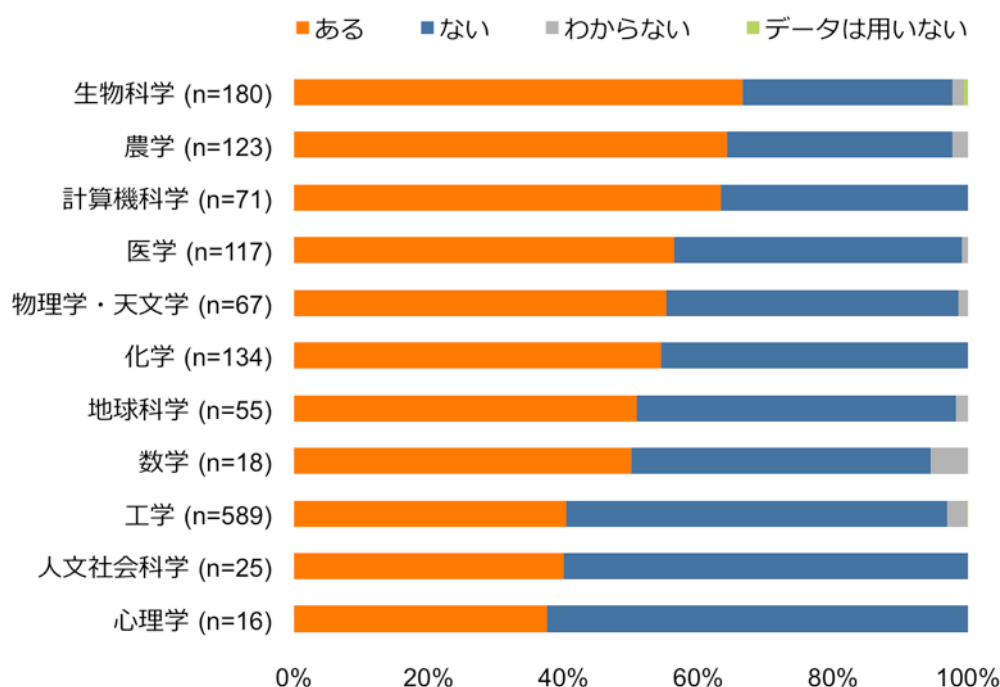


図 13 分野別データ公開経験 (n=1,395)

所属別にデータの公開経験を確認すると、公的機関(58.5%)や大学(56.9%)の公開率が比較的高かったが、企業の研究者も3割以上(32.4%)がデータ公開経験を有していた。オープンサイエンスは、公的資金を財源とした研究成果を市民に還元することが目的の一つとなっているが、これにあてはまらない企業の研究者によるデータ公開も行われていることが明らかになった。

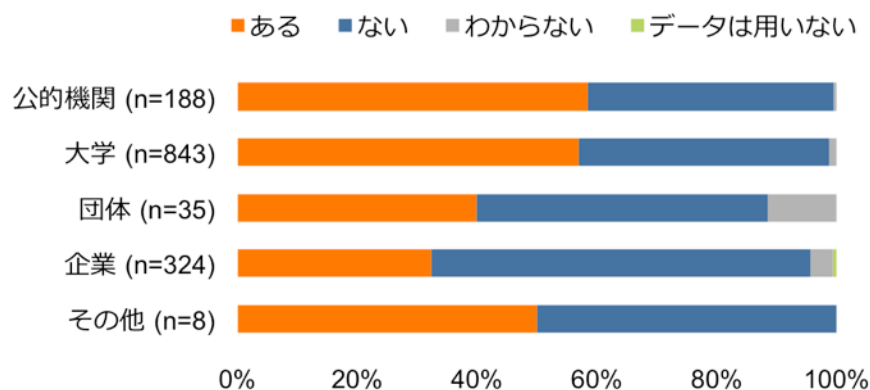


図 14 所属別データ公開経験 (n=1,398)

研究者によるデータ公開率を調査した先行研究のうち、複数の国・地域、かつ複数の分野を対象とした調査の結果と本調査の結果を比較する(図 15)。括弧内の数字は、調査年を示す。

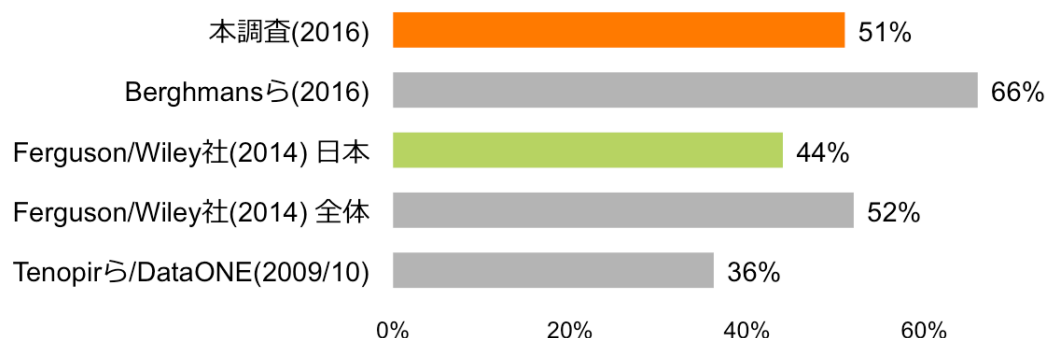


図 15 先行研究と本研究の調査年・データ公開率

Ferguson⁴³ の調査によるデータ公開率は、対象国全体が 52%であるのに対して日本は 44%であったこと、本調査と同じ年に実施された Berghmanら⁴⁵ の調査では全体が 66%であったことから、本調査の 51%という結果はおおむね妥当であると考えられる。現時点での日本の研究者によるデータ公開率は、国や助成機関による本格的な義務化が行われていないにもかかわらず、遜色がないと言えよう。ただし、より正確に日本と国外の研究者のデータ公開状況を比較するためには、同一内容の調査を同一時期に実施することが望ましい。この点は、今後の課題としたい。

(2) データの公開方法

データの公開方法は、個人や研究室のウェブサイトへの掲載や論文の補足資料 (supplementary materials) が多く、所属機関や特定分野のリポジトリは公開経験「あり」の回答者の 34.2%, 16.4%にとどまった。この傾向は先行研究^{36 42 43 45}と同様であり、リポジトリの整備と活用が今後の課題であると考えられる。「その他」には、“プロジェクト研究の Web サイト”、“YouTube”などの記述があった。

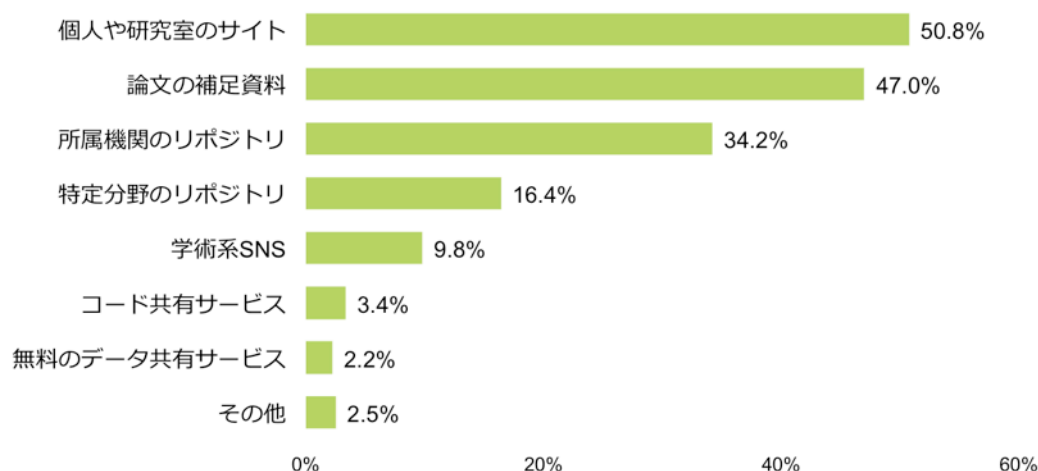


図 16 データの公開方法 (n=713)

分野別に「特定分野のリポジトリ」の選択率を確認すると、生物科学 (22.8%), 農学 (21.1%) から工学 (2.5%) まで差がみられた ($p < 0.01$)。生物科学, 農学はデータの公開率も 1 位, 2 位であり、分野リポジトリを活用してデータ公開を行っていると考えられる。

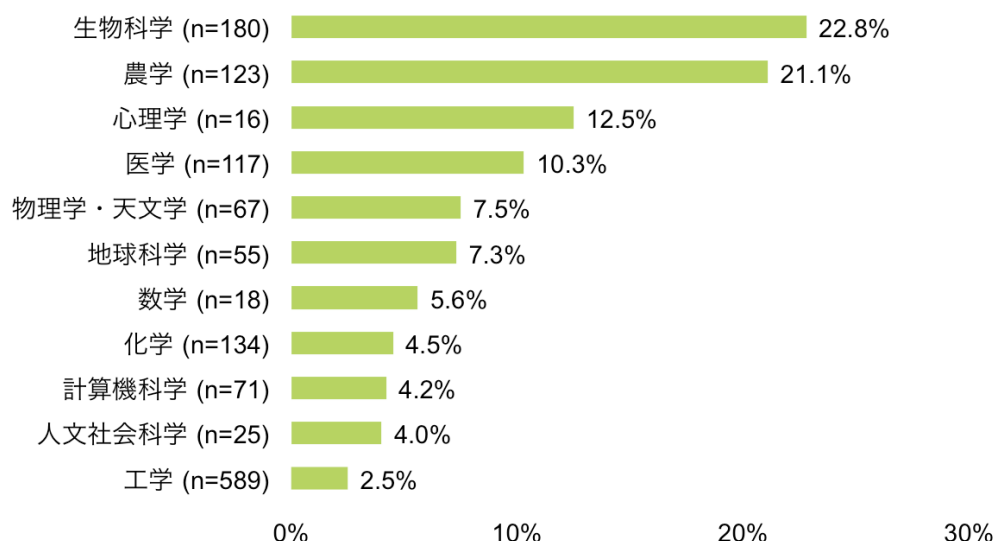


図 17 分野別「特定分野のリポジトリ」選択率 (n=1,395)

所属機関のリポジトリ(以下、「機関リポジトリ」)の選択率を所属別に確認すると、団体、公的機関の順に多く、差がみられた($p<0.05$)。

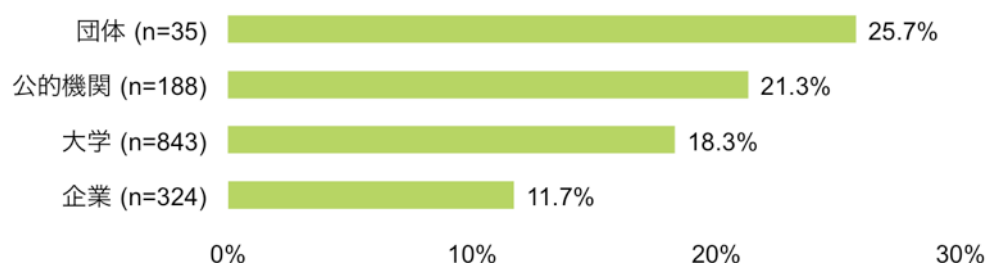


図 18 所属別「所属機関のリポジトリ」選択率 (n=1,390)

国立情報学研究所(NII)によれば、2017年5月の時点で大学など712機関が機関リポジトリを構築済みであり、77機関がJAIRO Cloud(NIIが提供する機関リポジトリ共同サービス)で構築中である⁵⁵。NIIがハーベストしている機関リポジトリ645件のコンテンツをみると、内容がある2,038,019件のうち、データやデータセットは53,735件(2.6%)、ソフトウェアは46件(0.0%)である⁵⁶。現状では出版物が大半を占めるが、オープンアクセス推進協議会(JPCOAR)によってデータも対象としたメタデータスキーマの検討が進められており⁵⁷、データ公開のプラットフォームとしての活用が期待される。

(3) データと論文の公開理由

続いてデータ公開や OA 論文の経験がある回答者に、それぞれの理由を複数選択方式で尋ねた(データ公開は Q19, OA 論文は Q3)。結果をデータ公開の理由の選択率が高い順に図 19 に示す。

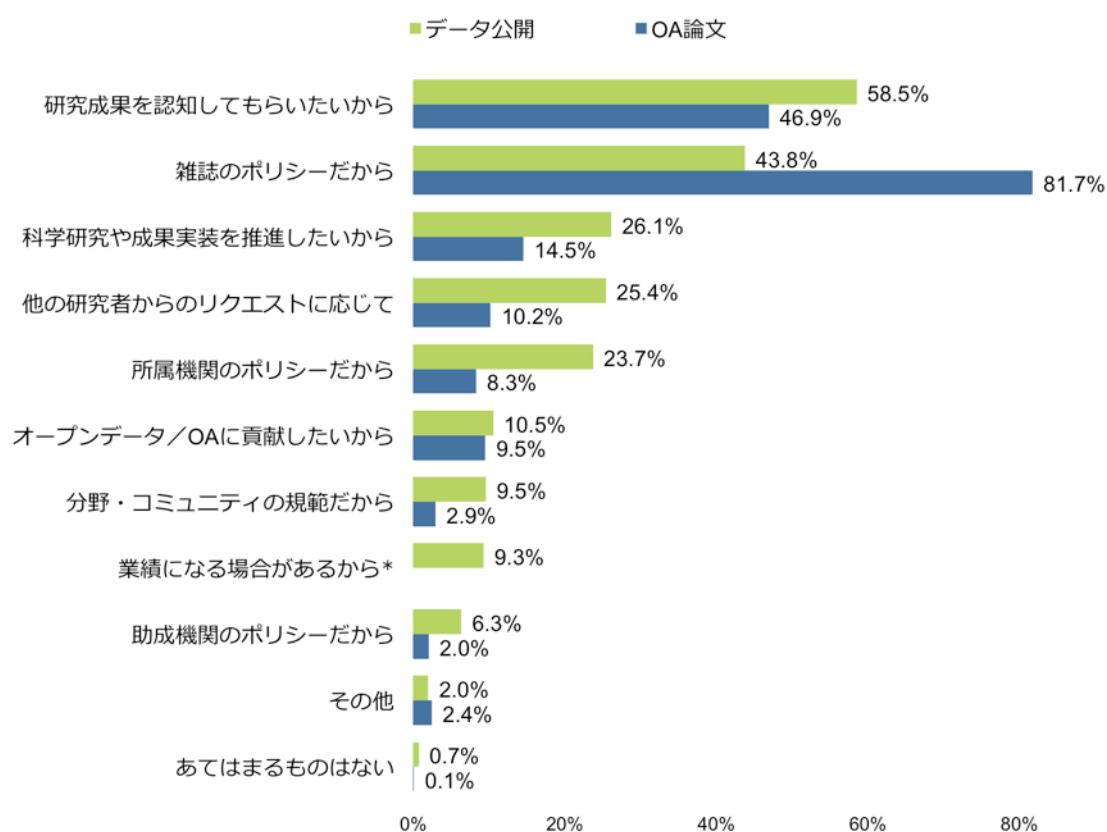


図 19 データ公開と OA 論文の理由 (データ n=713 論文 n=991)

OA 論文の理由は、「雑誌のポリシーだから」(81.7%)の選択率が突出しており、続いて「研究成果を認知してもらいたいから」(46.9%)が高かった。3 位以下はデータとほぼ同じ順位であるものの、いずれも選択率が低かった。

一方、データ公開は OA 論文のような突出した理由がなかった。以下では、約半数の回答者が選択した理由、約 1/4 の回答者が選択した理由、あまり選択されなかった理由(約 1 割)について、それぞれ述べる。

約半数が選択した理由は、「研究成果を認知してもらいたいから」(58.5%)と「雑誌のポリシーだから」(43.8%)であった。先行研究では助成機関のポリシーが主な公開理由となっている場合がみられたが、助成機関による義務化が行われていない現時点の日本においては、この 2 つが主な公開理由であった。約 1/4 が選択した理由は、科学的な利他性に関連する「科学研究や成果実装を推進したいから」(26.1%)や「他の研究者からのリクエストに応じて」(25.4%)、及び「所属機関のポリシーだから」(23.7%)であった。先行研究では「分野・コミュニティの規範」が主要な公開理

由の1つであったが、本調査では 9.5%にとどまった。また、「業績になる場合があるから」(9.3%)も同程度の選択率であり、かつ、分野別の選択率を確認すると、それぞれ同程度であった(図 20)。生物科学は規範が 15.0%であるのに対して業績が 8.3%であり、やや乖離がみられたが、全体としては分野やコミュニティの規範となっている場合は業績として評価もされている傾向にあった(Pearson の相関係数 $r=0.833$, $p<0.01$)。

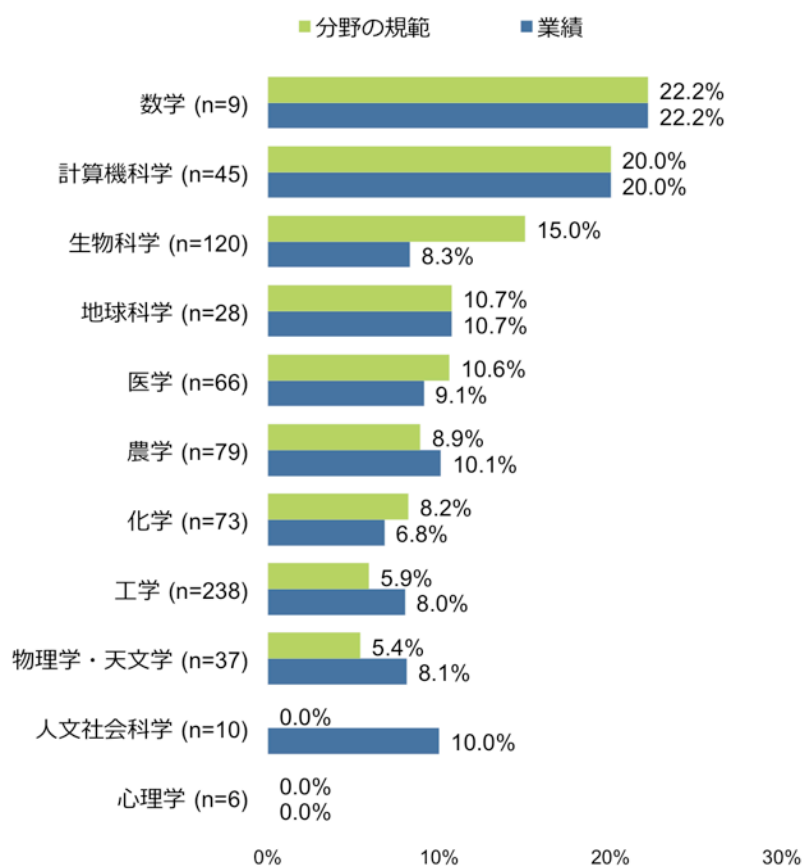


図 20 分野別「分野・コミュニティの規範」と「業績になる場合がある」の選択率 (n=711)

「その他」に記述されたデータ公開と OA 論文の理由を、積極的・自発的なものと消極的・強制的なものにわけて表 5 に示す。

データ公開の積極的な理由として、研究の信頼性の担保(2 名)、結果の検証を可能にする(2 名)といった研究の透明性の向上に関する記述がみられた。また、公開のインセンティブとなりうる“次のポジションのための成果・技術アピール”、“データを公開すると実験設備の利用料が安くなる”、“新たな研究者とのつながりの構築”といった記述もみられた。OA 論文については業績に直接関わるインセンティブとして、査読期間が短い(2 名)、先取権の確保、“OA の雑誌のインパクトファクターが上昇”、“引用が増えると考えられたから”、“多くの分野の学者や専門家と対話を進める為”といった記述がみられた。

データ公開の消極的な理由として、査読者・編集者に勧められたという記述がみられた(それぞれ 1 名)。OA 論文については、OA 論文の対象に選出された(2 名)、“知らない間に OA になっていた”といった記述がみられた。

表 5 データ公開・OA 論文の理由（自由記述）

| 理由 | Q19. データ公開 | Q3. OA 論文 |
|------------|---------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| 積極的 自発的 | 研究の信頼性の担保(2 名) 結果の検証を可能にする(2 名) “次のポジションのための成果・技術アピール” “データを公開すると実験設備の利用料が安くなる” “新たな研究者とのつながりの構築” | 査読期間が短い(2 名) 先取権の確保 “OA の雑誌のインパクトファクターが上昇” “引用が増えると考えられたから” “多くの分野の学者や専門家と対話を進める為” |
| 消極的 強制的 | 査読者に勧められた 編集者に勧められた | OA 論文に選出された(2 名) “知らない間に OA になっていた” |

(4) データと論文の非公開理由

データ公開及び OA 論文の経験が「なし」、「わからない」を選択した回答者に、それぞれの非公開理由を複数選択方式で尋ねた(図 21)。その結果、データ公開、OA 論文ともに1位は「雑誌のポリシーではない」ことであった。ただし、選択率には差があり、データ公開は 26.4%、OA 論文は 60.4%であった。また、データ公開は1位から5位までの選択率が20%台である一方で、OA 論文は1位の「雑誌のポリシーではない」(60.4%)と2位の「資金が必要だから」(39.6%)に集中している。OA 論文のための資金、すなわち論文出版加工料(Article Processing Charge, APC)が負担となっていることについては、自由回答においても多くの記述がみられた。

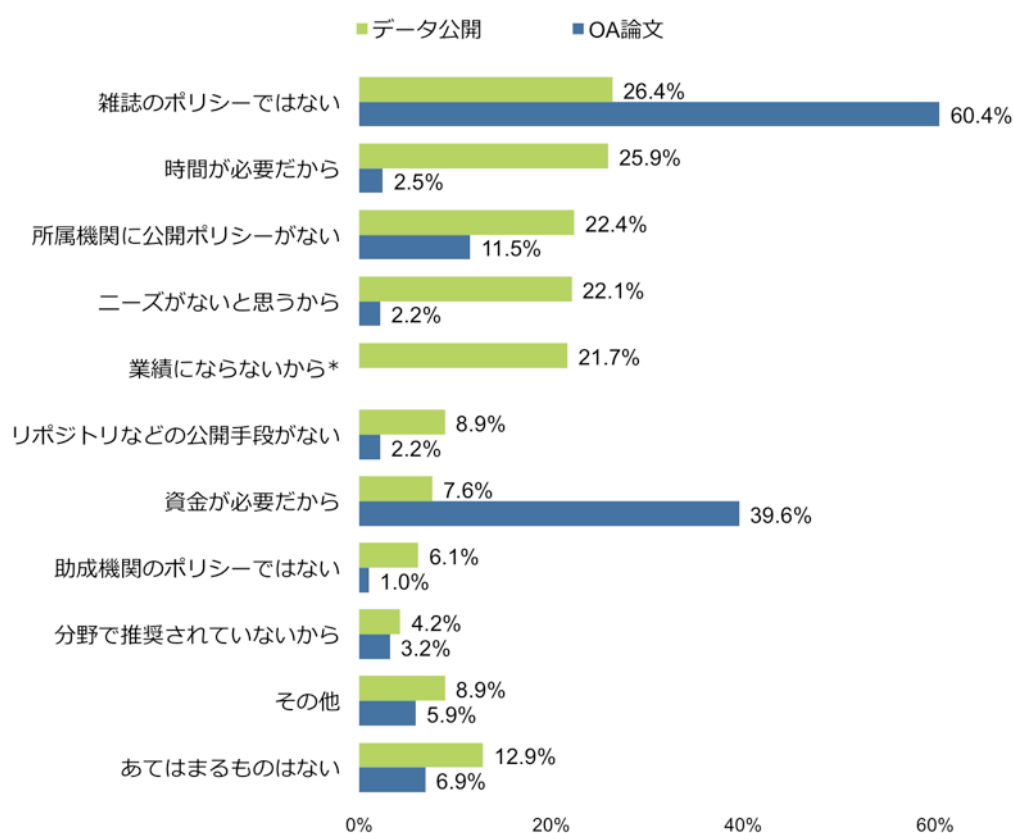


図 21 データと論文の非公開理由 (データ n=683 論文 n=407)

「その他」では、データ公開は不要(9 件)、公開するデータがない(4 件)といった回答がみられた。また、自由回答では、“大掛かりな実験をしなければ再現できない実験系の分野と、読者が自分の頭で考えて再現できる理論系の分野では事情が全く異なると思われる”、“数学は純粋に理論的な研究分野であるため、実証研究的な意味合いでのデータというものは存在せず、(数表・分類表などのように)特定の数学的事実を網羅的にまとめたものや、抽象的・一般的事実の具体例を提示するものとしてのみデータが扱われることになる”、“(物性物理学分野は)生データを公開しても他研究者に理解することは難しい。理解できるように整えるのは、時間的にかなり負担である。また、結果を分かりやすくまとめたものが論文であり、生データを公開する意味が分からない

(この点は、分野によって事情が異なるのかもしれませんが)”といった指摘もみられた。OA 論文については、OA 誌の質への懸念や IF の低さ(3 名)、偶然(2 名)、“学术论文を無料で読めるようにすることには違和感がある”、“意図しない利用のされ方を恐れているため”、“適切なジャーナルを知らない”といった記述がみられた。

「あてはまるものはない」の選択率は、データが 12.9%，論文が 6.9%であり、特にデータが高かった。データの非公開理由については、インタビュー調査などによる詳細な調査が必要であると考えられる。

(5) データと論文の公開意思

データと論文の非公開理由を 1 つ以上選択した回答者に、「Q21. Q20 の理由が解決された場合、研究データを公開したいと思われますか?」、「Q4 の理由が解決された場合、論文をオープンアクセスにしたいと思われますか?」と尋ねた。

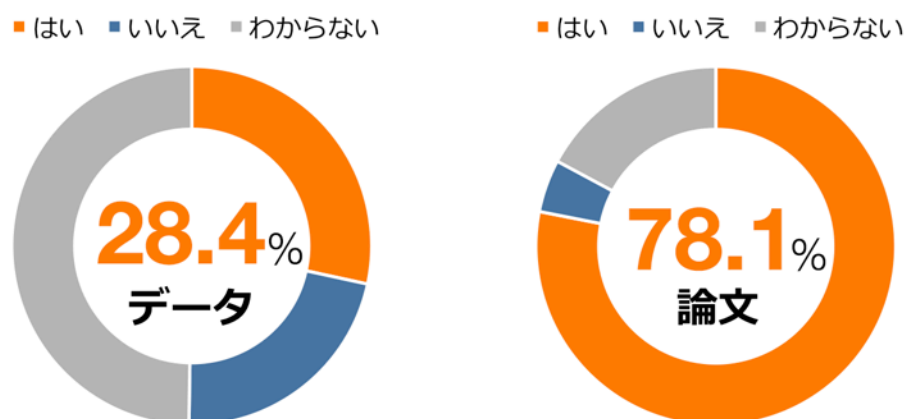


図 22 非公開理由が解決した場合のデータと論文の公開意思(データ n=595, 論文 n=379)

OA 論文は「はい」の比率が 78.1%と高かったが、データ公開については「はい」が 28.4%と低く、「わからない」が約半数(49.7%)を占めた(図 22)。「はい」を選んだ回答者の非公開理由、あるいは「いいえ」を選んだ回答者の非公開理由を確認するため、「はい」、「いいえ」、「わからない」を選んだ回答者ごとに非公開理由を集計したが、大きな偏りはみられなかった。つまり、特定の非公開理由が解決されればデータを公開する、あるいは解決されたとしても公開しない、とはいえないことがわかった。分野による差は検定できず、属性による差もみられなかった。

4.2 データの提供・被提供状況

データの提供及び被提供(共有)経験を確認するため、「Q9. 共同研究者を除く他の研究者にデータを提供したご経験はありますか?」、「Q10. 共同研究者を除く他の研究者からデータの提供を受けたご経験はありますか?」と4件法で尋ねた。その結果、「よくある」と「たまにある」を選択した回答者は、それぞれ44.0%と38.7%であり、提供経験の方がやや多い結果となった(図23)。両者の相関は高く(0.77, $p<0.001$)、データを提供している回答者は、提供を受けてもいる傾向が確認された。

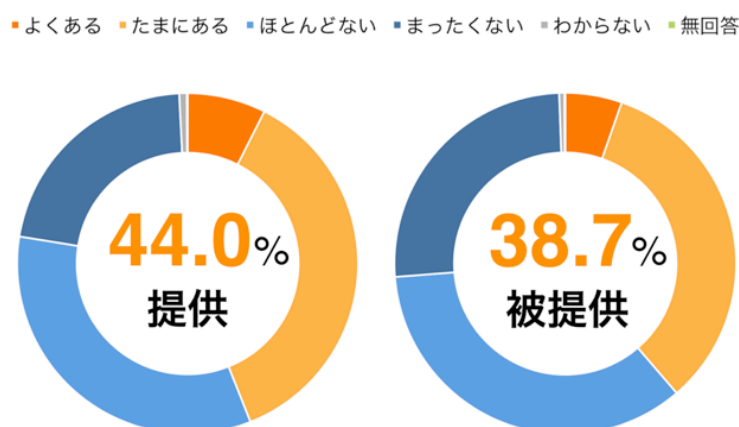


図23 データの提供・被提供経験 (n=1,398)

データの提供・被提供経験とデータ公開経験との関連をみるため、頻度別のデータ公開経験を確認した(図24)。その結果、データ提供経験、被提供経験ともに頻度が高い回答者ほどデータ公開経験がある傾向が明らかになった(いずれも $p<0.001$)。ただし、データの提供や被提供をまったく行ったことがなくてもデータ公開経験をもつ回答者の存在も確認された。

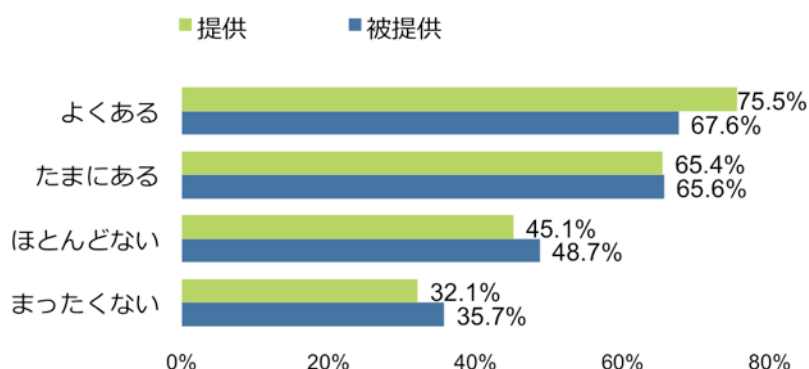


図24 データ提供・被提供頻度別データ公開経験 (提供 n=1,364, 被提供 n=1,365)

分野別に、データの提供・被提供経験を「よくある」=4,「たまにある」=3,「ほとんどない」=2,「まったくない」=1で重み付けして比較すると、提供・被提供頻度はともに地球科学, 数学, 計算機科

学が高く、心理学、医学、化学は低かった($p<0.001$)。なお、計算機科学のみ提供経験よりも被提供経験の方が多かった(図 25)。

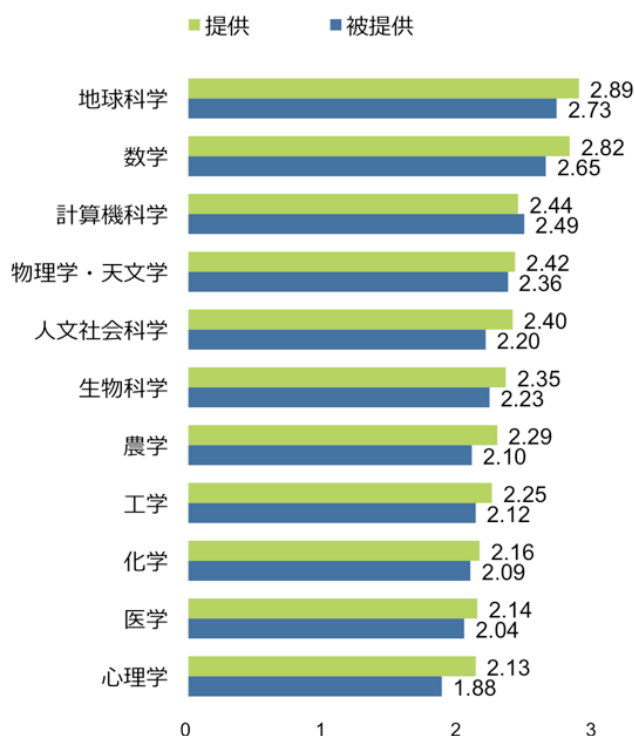


図 25 分野別データの提供・被提供経験 (提供 n=1,384 被提供 n=1,387)

所属別では、データの提供経験のみ差がみられた($p<0.01$)。図 26 に示すように、公的機関は「よくある」(11.2%),「たまにある」(36.6%)の比率が高く、企業は「まったくない」(36.8%),「ほとんどない」(26.0%)の比率が高かった。

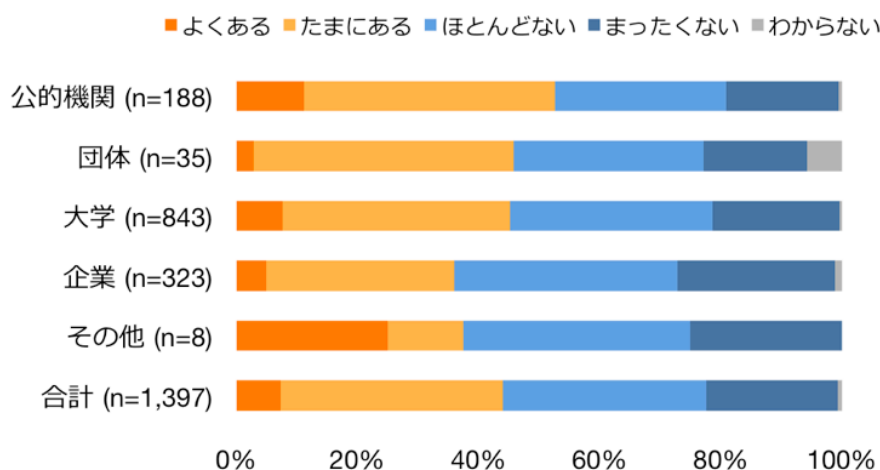


図 26 所属別データの提供経験

4.3 公開データの利用状況と課題

(1) 公開データの入手経験

公開データの利用状況を確認するため、「Q11. これまでに、公開データを以下の公開先から入手したご経験はありますか？」と尋ねた。選択肢に挙げた入手方法を1つ以上チェックした回答者は1,060名(75.8%)、排他的選択肢である「ない」は318名(22.7%)、「わからない」は20名(1.4%)であり、約3/4は公開データの入手経験があることが明らかになった(図27)。

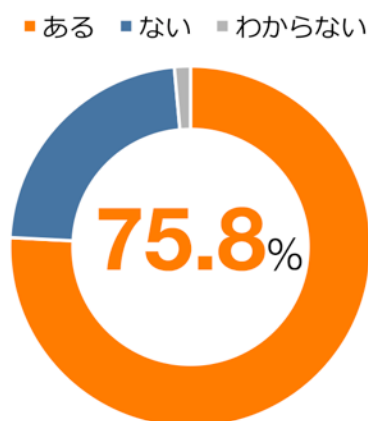


図27 公開データの入手経験の有無 (n=1,398)

公開データの入手方法の順位は、データ公開方法の順位と一致していた(図28)。このうち、半数以上が選択していたのは「個人や研究室のウェブサイト」(64.8%)、「論文の補足資料」(53.1%)、「学術機関のリポジトリ(大学やNASAのリポジトリなど)」(50.0%)であった。機関リポジトリは公開先としては34.2%にとどまったが、データの入手先としては比較的良好に活用されていた。「コード共有サービス(GitHubなど)」や、「無料のデータ共有サービス(figshare, zenodoなど)」は、いずれも選択率が低かった。「その他」では、官公庁や自治体が公開しているオープンデータ(6名)、大学・学会図書館(2名)、YouTubeなどが挙げられていた。

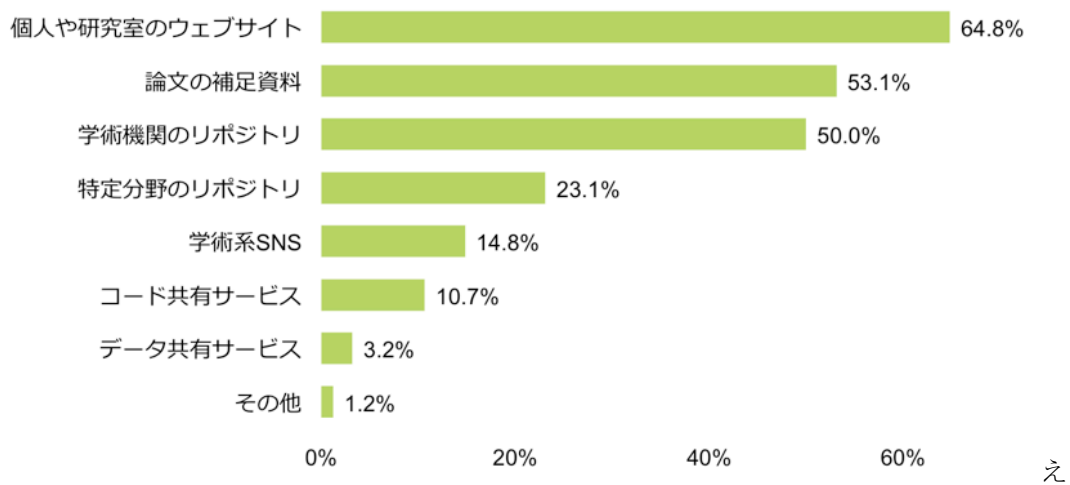


図 28 公開データの入手方法 (n=1,060)

公開データの入手経験とデータ公開経験には関連がみられたが($p<0.001$), 所属や分野については検定できなかった。そこで機関リポジトリと分野リポジトリについて, 分野別の入手状況を確認したところ, 分野リポジトリは分野による差がみられた($p<0.001$)。入手先として選択した回答者の比率が高かったのは, 生物科学(41.1%), 農学(32.5%)であり, 人文社会科学(4.0%), 物理学・天文学(6.0%)は低かった(図 29)。

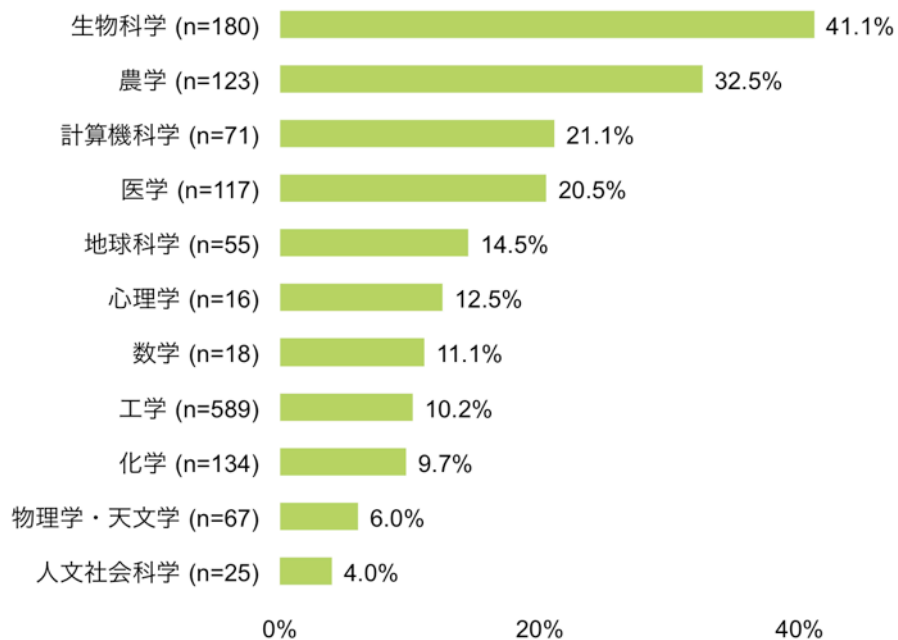


図 29 分野別「分野リポジトリ」選択率 (n=1,395)

(2) 公開データ入手の障壁

「Q12. 公開データの入手にあたって、問題だと感じたことがある項目をお選び下さい。」と尋ねたところ、排他的選択肢である「問題を感じたことはない」を選んだ回答者は 214 名 (20.2%), 1 つ以上の問題を選択した回答者は 846 名 (79.8%) であった (図 30)。

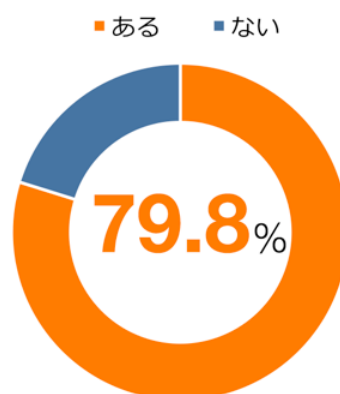


図 30 データ入手の障壁の有無 (n=1,060)

項目別にみると、「利用料金が必要」(43.1%) が最も多く、次いで「利用者登録が必要」(33.3%), 「利用条件がよくわからない」(33.1%), 「データごとに品質が異なる」(32.2%) と続く (図 31)。なお、半数以上の回答者が選択した項目はなかった。先行研究¹⁷ではデータの利活用における障壁を感じる回答者が多かったが、本調査ではデータの入手時点で障壁を感じる回答者が多かった。「その他」では、“データサイズが大きすぎてダウンロードできないことがある”など、データ量に関する問題(3 名)や、“検索が一度にできずに、それぞれのデータベースに入る必要がある”など、統合検索システムの欠如(2 名)についての記述がみられた。

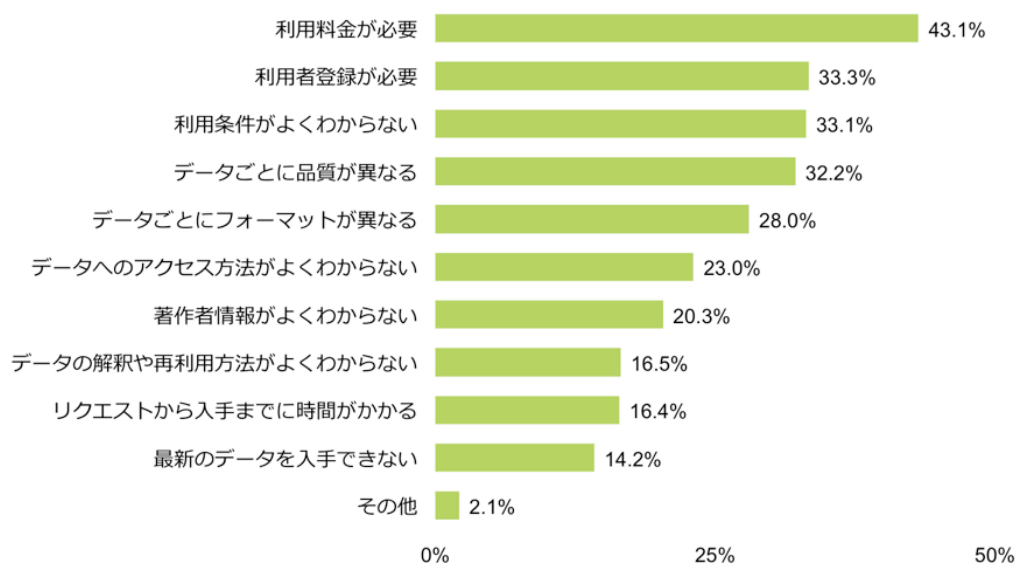


図 31 データ入手の障壁 (n=846)

(3) 公開データの利用目的

公開データの利用目的を、「自身の研究のアイデアや仮説の参考にする」、「再分析・再利用して自身の研究を行う」、「(他者の)研究を再現・追試する」に分けて、それぞれの頻度を尋ねた(Q13)。その結果、先行研究⁵²と同様に「研究の参考」、「研究に再利用」、「研究を再現」の順に頻度が高かった(図 32)。

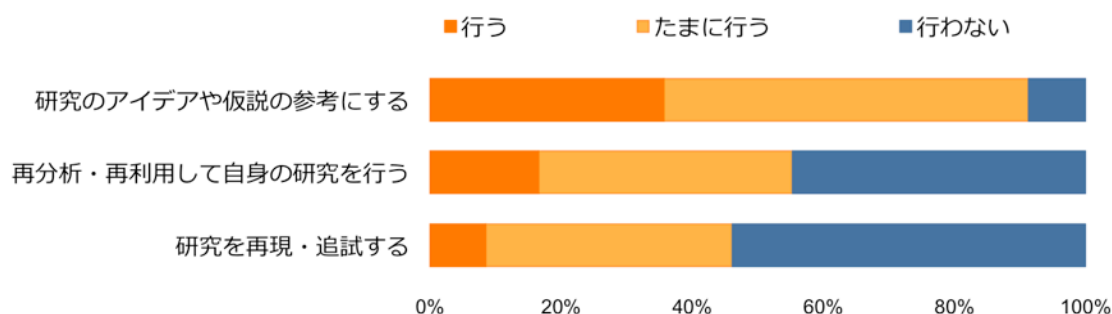


図 32 公開データの利用目的 (n=1,060)

データ公開経験とは、すべて関連がみられた(それぞれ $p<0.01$, $p<0.01$, $p<0.05$)。分野別では、「再分析・再利用」と「研究を再現・追試」について差がみられた(それぞれ $p<0.001$, $p<0.05$)。再分析・再利用して自身の研究を「行う」または「たまに行う」を選択した回答者の比率が高い順に図 33 に示す。再利用を「行う」または「たまに行う」比率が高いのは地球科学(分野内 81.3%), 計算機科学(74.6%), 人文社会科学(62.5%)であり、「行わない」比率が高いのは心理学(88.9%), 医学(52.9%), 化学(50.9%)であった。

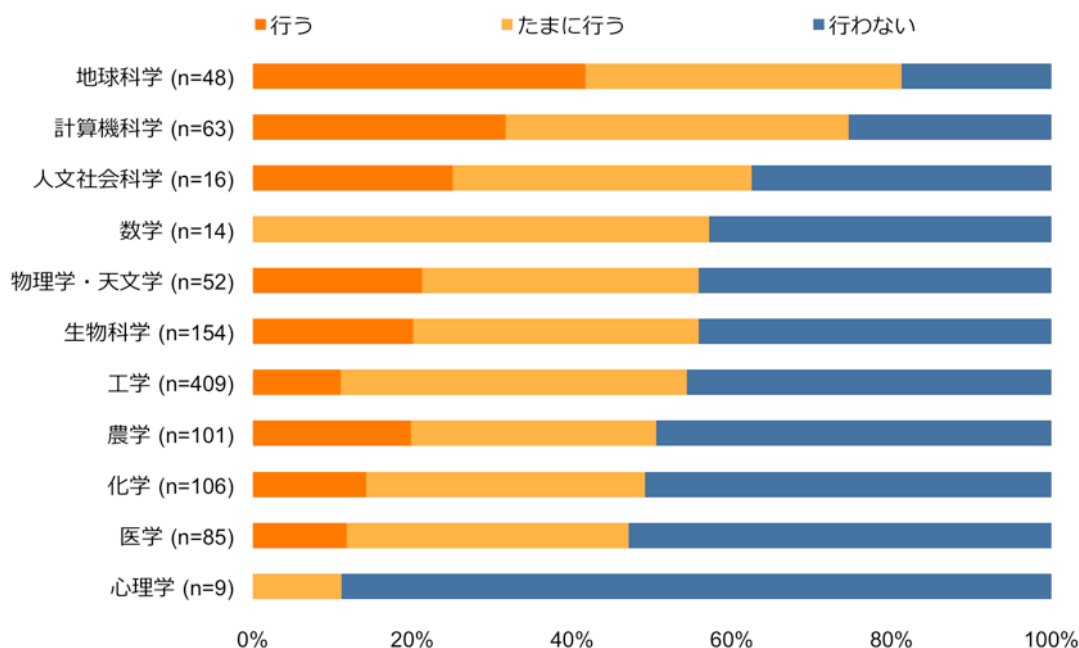


図 33 分野別データの再分析・再利用状況 (n=1,057)

また、図 34 に示すように、研究の再現・追試を「行う」または「たまに行う」比率が高いのは計算機科学(61.9%)、物理学・天文学(57.7%)、心理学(55.6%)であり、「行わない」比率が高いのは数学(71.4%)、地球科学(62.5%)、農学(61.4%)であった。図 33 と図 34 を比較すると、地球科学や数学分野の研究者は、データを再分析や再利用して自身の研究を行う比率が高い一方で、再現や追試については行わない比率が高いことが明らかになった。

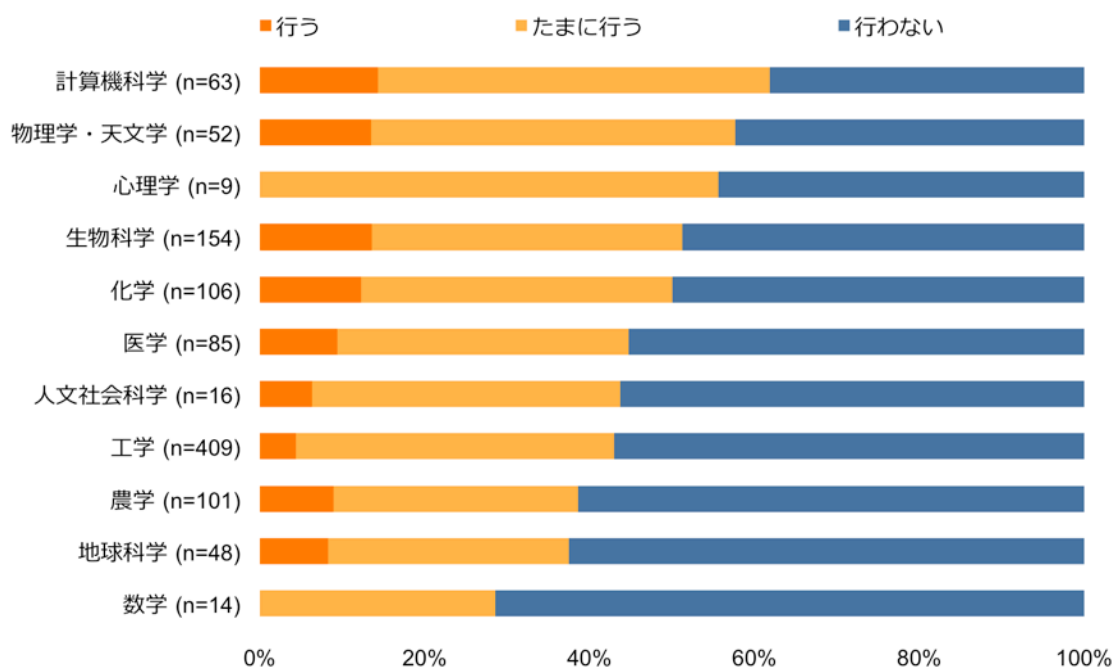


図 34 分野別データの再現・追試状況 (n=1,057)

所属も同じく「再分析・再利用」と「研究を再現・追試」について差がみられた(それぞれ $p<0.05$, $p<0.01$)。再利用を「行う」比率が高いのは公的機関(23.8%)であり、「行わない」比率が高いのは大学(47.9%)であった。再現を「行う」比率が高いのは企業(14.0%)であり、「行わない」比率が高いのは団体(69.2%)であった。

(4) データと論文の利用分野

公開データの利用状況や、今後の利用意思を確認するために、「Q14. 研究に利用したことがある(参考にしたたり再分析した経験がある)公開データの分野をお選び下さい。(n=1,060)」「Q17. 今後、利用してみたいと思う公開データの分野をお選び下さい。(n=1,398)」という質問を行った。また、論文の利用状況と比較するために、「Q6. 研究に利用したことがある(アイデアの参考にしたたり引用した経験がある)論文の分野をお選び下さい。」という質問を行った。いずれも複数回答方式である。

Q14で1つ以上の分野を選択した回答者は1,034名(97.5%)、排他的選択肢である「公開データを入手しても、研究に利用したことはない」を選択した回答者は26名(2.5%)であった。また、Q17で1つ以上の分野を選択した回答者は1,357名(97.1%)、排他的選択肢である「他の研究者の公開データを利用したいとは思わない」を選択した回答者は41名(2.9%)であった(図35)。

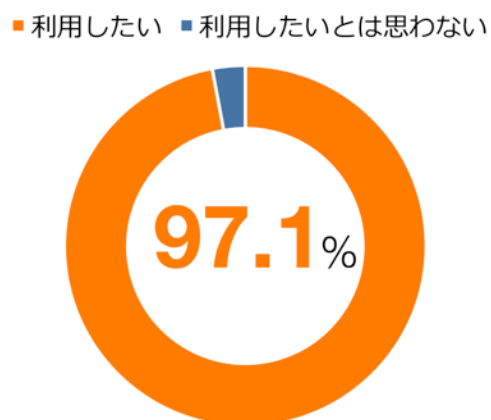


図 35 データ利用希望の有無 (n=1,398)

論文とデータの利用及び利用希望分野について検討するため、分野ごとの集計結果と合計選択数、及び合計を回答者数で除した一人あたりの選択数(平均値)を表6に示す。論文(2.67)と公開データ(2.03)を比較すると、論文を利用したことがある分野の方が平均的に多く選択されていたが、公開データを利用してみたい分野(2.79)が最も多く選択されていた。

表 6 研究に利用したことがある分野と利用してみたい分野（複数回答）

| | Q6. 論文を 利用したことがある分野 | | Q14. 公開データを 利用したことがある分野 | | Q17. 公開データを 利用してみたい分野 | |
|-------|------------------------|-------|----------------------------|-------|--------------------------|-------|
| 工学 | 847 | 60.6% | 518 | 50.1% | 767 | 56.5% |
| 生物科学 | 511 | 36.6% | 348 | 33.7% | 515 | 38.0% |
| 化学 | 473 | 33.8% | 271 | 26.2% | 434 | 32.0% |
| 物理学 | 431 | 30.8% | 208 | 20.1% | 394 | 29.0% |
| 医学 | 381 | 27.3% | 202 | 19.5% | 400 | 29.5% |
| 農学 | 287 | 20.5% | 160 | 15.5% | 303 | 22.3% |
| 計算機科学 | 268 | 19.2% | 163 | 15.8% | 342 | 25.2% |
| 地球科学 | 149 | 10.7% | 91 | 8.8% | 171 | 12.6% |
| 数学 | 140 | 10.0% | 48 | 4.6% | 136 | 10.0% |
| 社会科学 | 106 | 7.6% | 39 | 3.8% | 143 | 10.5% |
| 心理学 | 71 | 5.1% | 21 | 2.0% | 89 | 6.6% |
| 天文学 | 33 | 2.4% | 19 | 1.8% | 42 | 3.1% |
| 人文学 | 24 | 1.7% | 14 | 1.4% | 39 | 2.9% |
| その他 | 5 | 0.4% | 2 | 0.2% | 8 | 0.6% |
| 合計 | 3,726 | | 2,104 | | 3,783 | |
| n | 1,398 | | 1,034 | | 1,357 | |
| 平均 | 2.67 | | 2.03 | | 2.79 | |

※Q6 を基準として順位が入れ替わっている箇所を斜体で示す。

(5) データと論文の探索方法

公開データと論文の探索方法を比較するために、「Q15 公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい」と複数選択方式で尋ねた(論文は Q7)。データの探索方法として選択率が高い項目順に図 36 に示す。

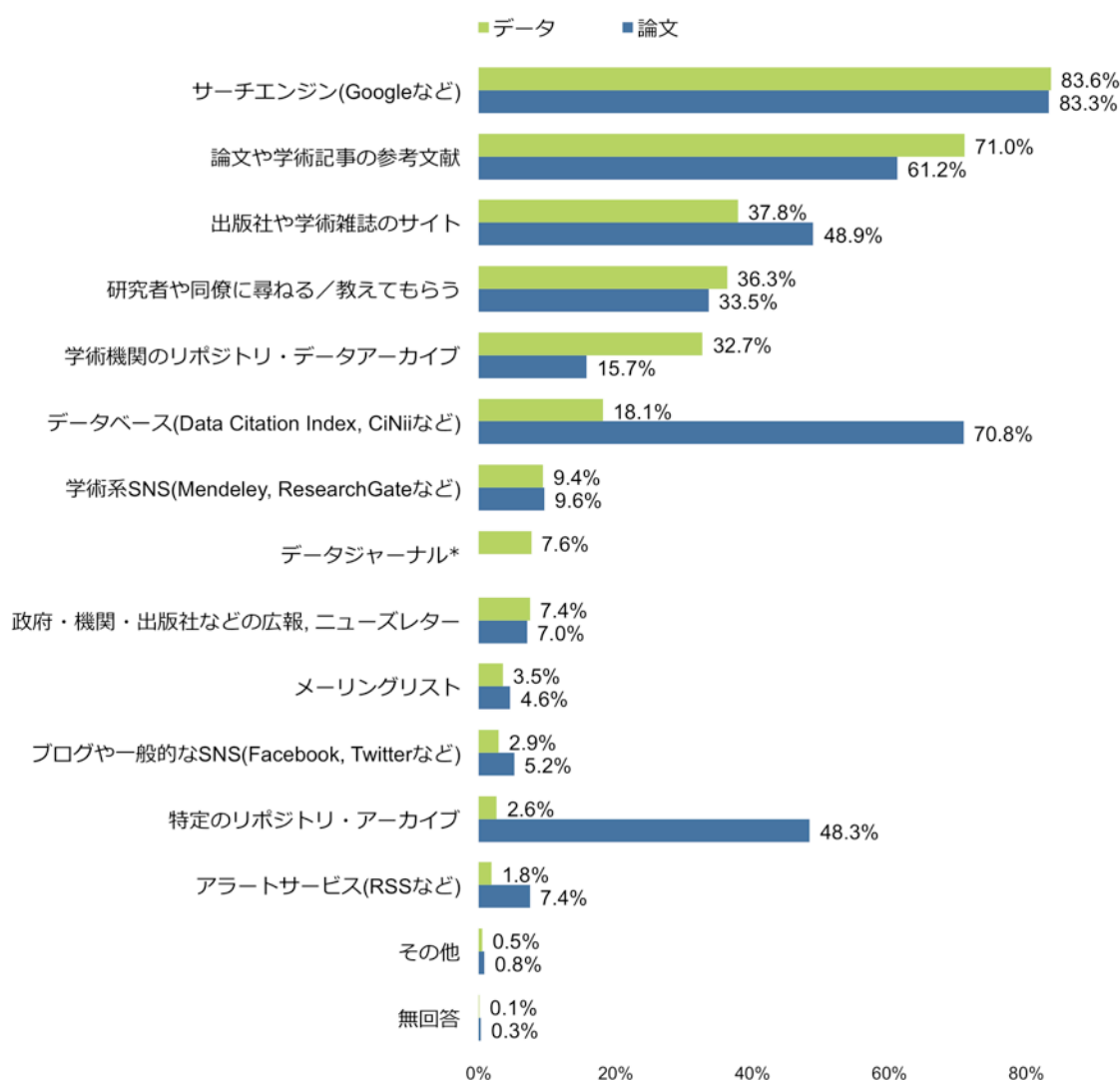


図 36 データと論文の探索方法 (データ n=1,034 論文 n=1,398)

データ、論文ともに最も使われているのは Google や Google Scholar などの「サーチエンジン」であった。データは 2 位が「論文や学術雑誌の参考文献」、3 位が「出版者や学術雑誌のサイト」であるのに対して、論文の 2 位は Web of Science, Scopus, CiNii Articles など「論文情報のデータベース」であることが大きく異なっている(データは 18.1%)。また、4 位の「論文やプレプリントのサーバ(PubMed Central, arXiv, J-Stage など)」を約半数(48.3%)が選択しているのに対して、「特定のデータリポジトリ」は 2.6%にとどまった。

(6) 信頼性の判断基準

公開データと論文の信頼性の判断基準を比較するために、「Q16. 公開データを利用する際に、信頼性の判断基準としている項目をお選び下さい。」と複数選択方式で尋ねた(論文は Q8)。

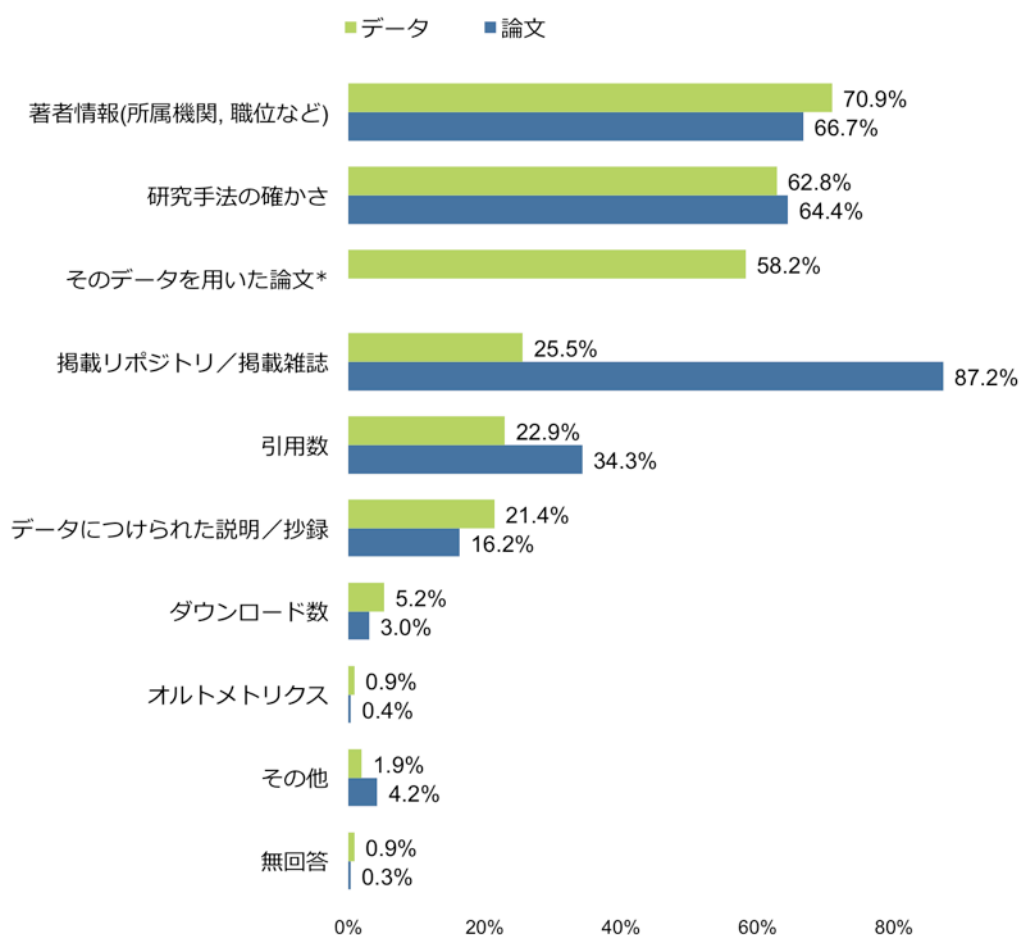


図 37 データと論文の信頼性の判断基準 (データ n=1,034 論文 n=1,398)

論文は「掲載雑誌」がもっとも多く、回答者の 87.2%が選択している。データを掲載しているリポジトリは、論文の掲載雑誌に相当すると仮定して選択肢に挙げたが、これを選択した回答者は全体の 25.5%にとどまった(図 37)。

「その他」では、「データそのものを解析してみる」、「内容そのもの」、「データ間での比較による」など、データを直接確認するという回答が 4 件、「著者の過去論文、過去の研究手法、研究内容」、「当該研究グループの過去の論文」など、著者の評価に関する回答が 6 件みられた。間接的な評価として、「研究者や同僚からのコメント」、「研究者コミュニティでの評価」の 2 件があった。自由回答では、公開されたデータの信頼性に対する指摘がみられた(12 件)。具体的には、“データを利用する立場で考えても、論文化されていないデータはデータの質に不安があって、利用したくない”、“review プロセスを経ていないので、データの信頼性に問題が残ると思われる”といった記述があった。

4.4 カレントデータのプロフィール

(1) データの保存期間規定

所属機関によるデータの保存期間に関する規定が「定められている」という回答は 48.1%, 「定められていない」は 27.2%, 「わからない」は 24.6%であった(図 38)。

■ 定められている ■ 定められていない ■ わからない

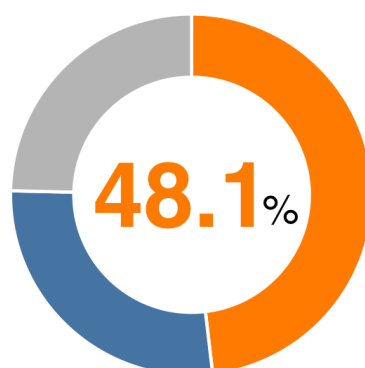


図 38 データの保存期間規定の有無 (n=1,396)

保存期間規定の有無とデータ公開経験には関連がみられなかったが、分野による差はみられた($p<0.001$)。「定められている」は医学(62.4%), 生物科学(57.0%), 心理学(58.3%), 化学(56.0%), 「定められていない」は人文社会科学(56.0%), 「わからない」は数学(50.0%), 計算機科学(38.0%)の回答者の比率がそれぞれ高かった(図 39)。

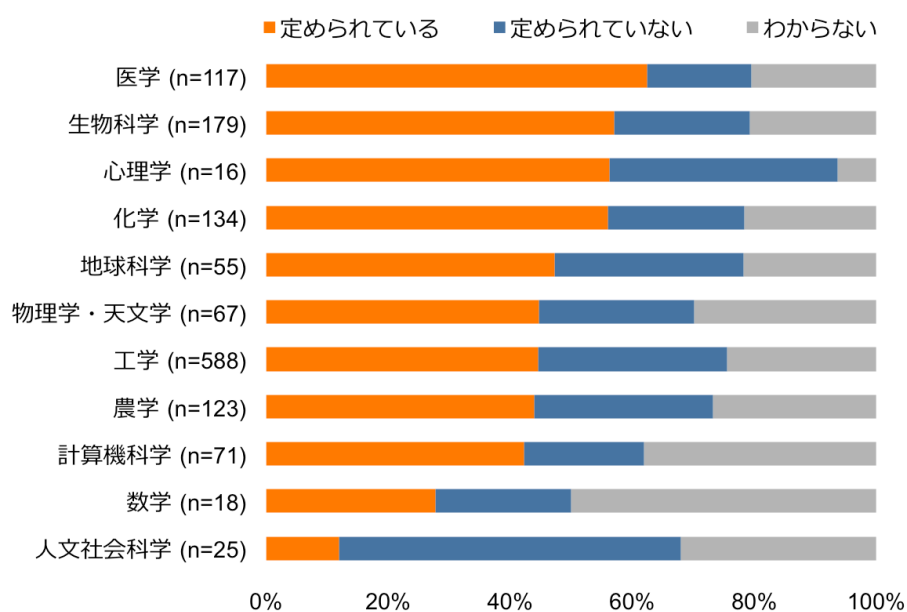


図 39 分野別データの保存期間規定 (n=1,393, 無回答=0)

所属別の検定は行えなかったが、「定められている」と回答した比率は公的機関(64.4%)や大学(48.9%)が高かった。また、「わからない」という回答は、大学(28.6%)と団体(28.6%)の比率が高かった。

(2) データ量

研究に用いるデータの量を確認するため、「Q23. カレントデータの総量は、およそどれくらいでしたか？ 論文などには使用しなかったデータも含めてあてはまる単位をお選びください。」と尋ねた。図 40 に示す通り、GB(ギガバイト)が最も多く、4 名(0.3%)は PB(ペタバイト)以上を選択していた。

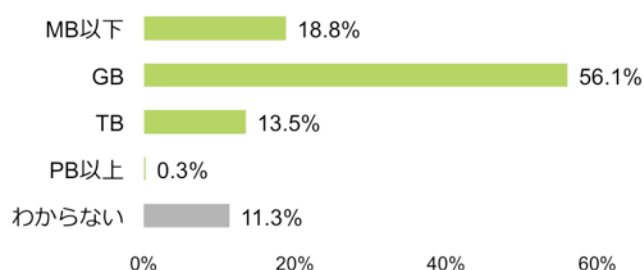


図 40 カレントデータの量 (n=1,396)

カレントデータ量の多寡とデータ公開経験には関連がみられなかった。データ量が比較的多い分野は、計算機科学、地球科学、物理学・天文学であり、少ない分野は心理学、人文社会科学、数学であった($p<0.001$)。所属による差はみられなかった。

(3) 所有権

「Q24. ご自身以外で、カレントデータの所有権をもつ人・組織をお選び下さい。」と尋ねたところ、回答者の83.5%がカレントデータの所有権をもつ人や組織を1つ以上選択していた(図 41)。自身以外の所有権の有無とデータ公開経験には関連がみられなかった。

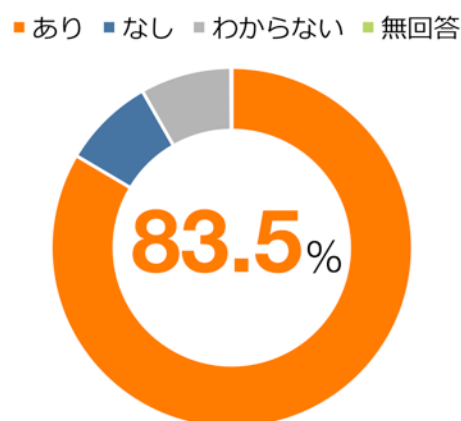


図 41 自身以外に所有権をもつ人・組織の有無 (n=1,396)

分野別の検定は行えなかったが、物理学・天文学(91.0%)、地球科学(89.1%)、化学(88.1%)が多く、人文社会科学(60.0%)、数学(61.1%)は少なかった(図 42)。

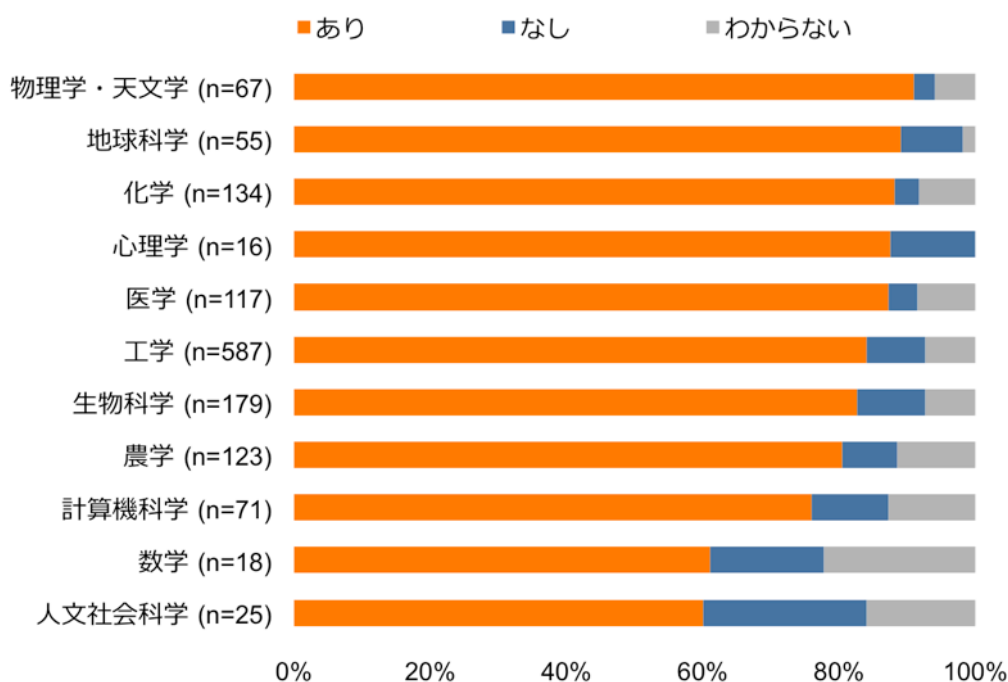


図 42 分野別自分以外に所有権をもつ人・組織の有無 (n=1,392)

所属別にみると、自身以外に所有権をもつ人・組織を1つ以上選択した回答者の比率は、企業（88.8%）、公的機関（87.8%）の順に多かった（ $p<0.05$ ）（図 43）。

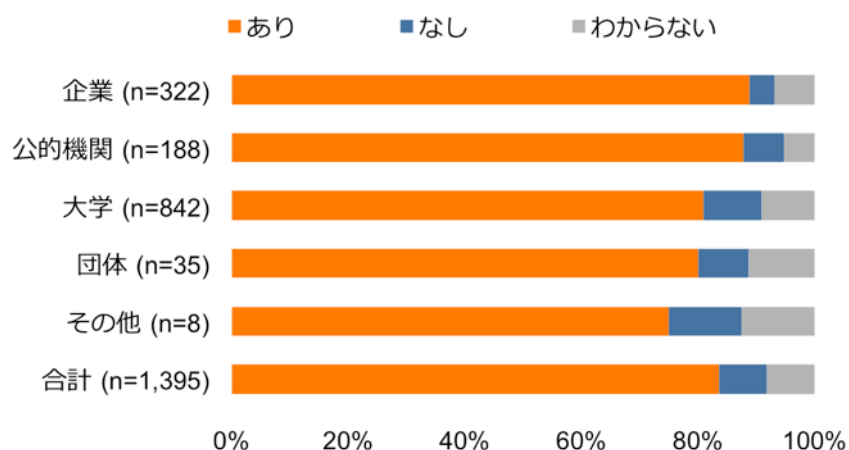


図 43 所属別自身以外に所有権をもつ人・組織の有無

選択肢のうち、最も多かったのは共同研究者（82.1%）、次いで所属機関（58.2%）であり、所属機関以外の研究機関や企業、研究助成機関はそれぞれ1割未満であった（図 44）。その他には、「データ提供した実験参加者個々人」、「公開されているデータベースの権利者」、「研究プロジェクトの発注機関」、「出版社に譲渡している中に含まれているものと考えていた」といった記述がみられた。

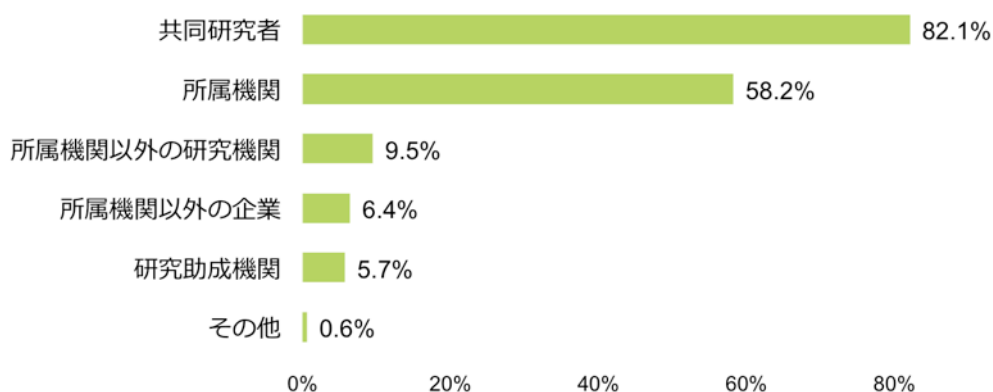


図 44 カレントデータの所有権をもつ組織・人（n=1,166）

(4) 機密情報

「Q25. カレントデータには、以下の情報が含まれていますか？」と尋ねたところ、回答者の約半数(48.9%)が機密情報を 1 つ以上選択していた(図 41)。機密情報の有無とデータ公開経験には関連がみられなかった。

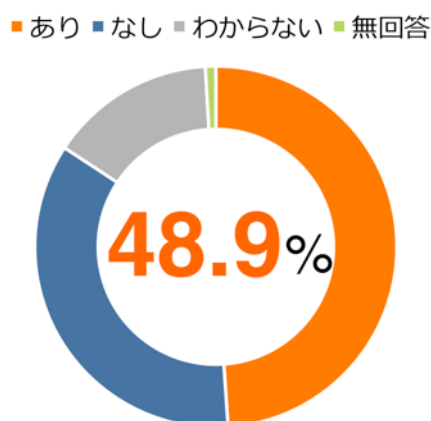


図 45 機密情報の有無 (n=1,396)

分野別では、医学(60.7%)、心理学(56.3%)、工学(55.6%)、化学(54.1%)は 1 つ以上を選択する回答者の比率が高く、数学(16.7%)、農学(35.2%)、生物科学(36.9%)は低かった($p<0.01$) (図 46)。

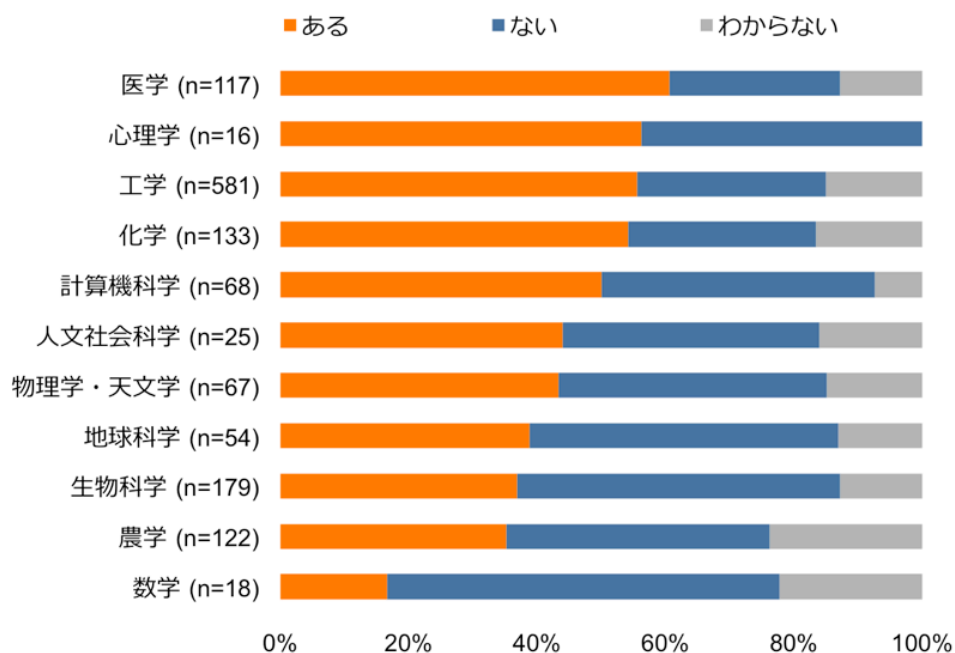


図 46 分野別機密情報の有無 (n=1,380)

所属別では企業(70.5%), 団体(62.9%)は機密情報ありの比率が高く, 大学(42.0%), 公的機関(41.9%)は低かった($p<0.001$) (図 47)。

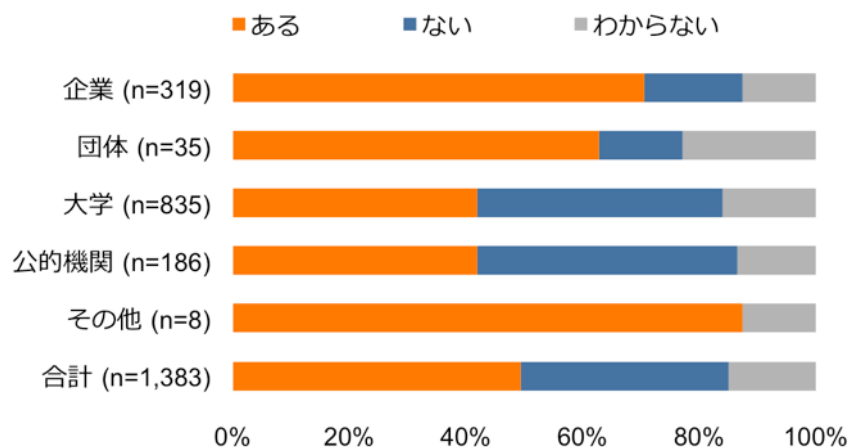


図 47 所属別機密情報の有無

図 48 に選択肢ごとの集計結果を示す。OECD の報告書⁶等で機密情報として挙げられている「個人情報」や「企業・商業上の機密情報」などを挙げたが, 「その他, 守秘義務がある情報」がもっとも多い結果となった(58.4%)。なお, 機密に関わるため, この質問のみ「その他」で具体的な記述を求めなかった。選択肢以外にどのような機密があるかという点については, 今後の課題とした。

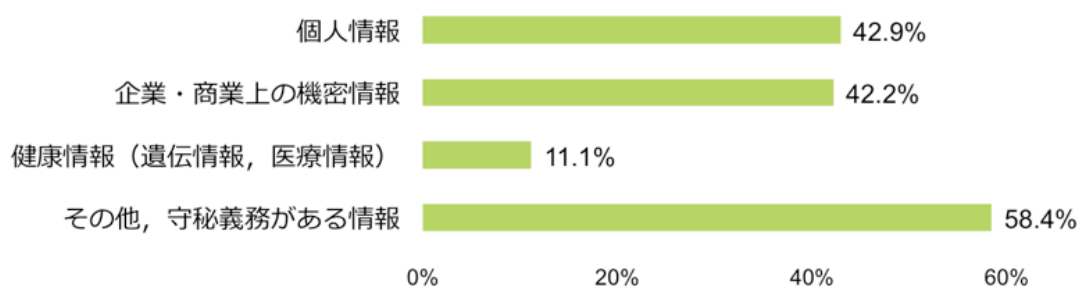


図 48 カレントデータに含まれる情報 (n=683)

(5) 望ましい保存期間

研究者が望ましいと考えるデータの保存期間を確認するために、「Q26. カレントデータは何年くらい保存する必要があると思われますか？ 年数を数字で入力して下さい。保存の必要はないと考える場合は「0」、永久保存の場合は「999」と入力して下さい。」と自由記入方式で尋ねた。永久保存を除いた平均は 9.9 年であり、最も多い回答は 10 年(432 名, 30.9%), 次いで 5 年(422 名, 30.2%)であった。図 49 に集計結果を示す。

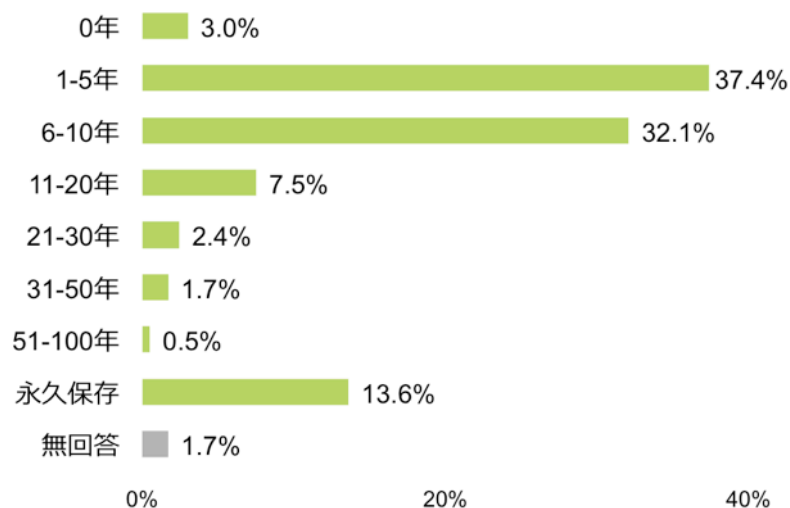


図 49 カレントデータの望ましい保存期間 (n=1,396)

分野別の平均年数を示す。医学(12.2 年), 化学(11.9 年)の順に長く, 数学(3.3 年), 計算機科学(7.1 年)は短かった(図 50)。

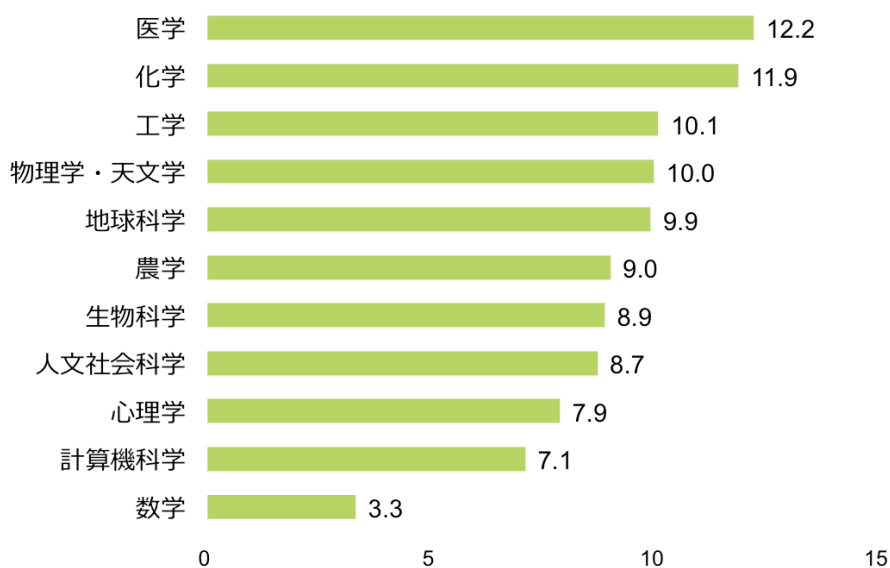


図 50 分野別カレントデータの望ましい保存期間 (n=1,179)

4.5 カレントデータを公開する場合の障壁

(1) データ公開に必要な資源の状況

データ公開に必要な資源の充足状況を確認するため、「Q28. カレントデータを管理・公開しようとする場合、次の資源は十分に整っていますか(いましたか)?」と「不十分」、「やや不十分」、「ほぼ充分」、「充分」の4件法で尋ねた。結果を「不十分」と「やや不十分」の合計選択率が高い順に示す(図 51)。

全体的に不十分であるという認識の回答者が多く、特にデータ整備・公開のための人材、資金、時間が不十分であるとする回答が多い。最も充足度が高い「研究中のデータ用ストレージ」でも、「充分」と「ほぼ充分」をあわせて 38.1%であった。また、「データ公開用のリポジトリ」については、「わからない」の比率が 27.1%と高かった。データ公開経験との関連は、「データ保存用ストレージ」と「研究中のデータ用ストレージ」以外にみられた(すべて $p<0.001$)。

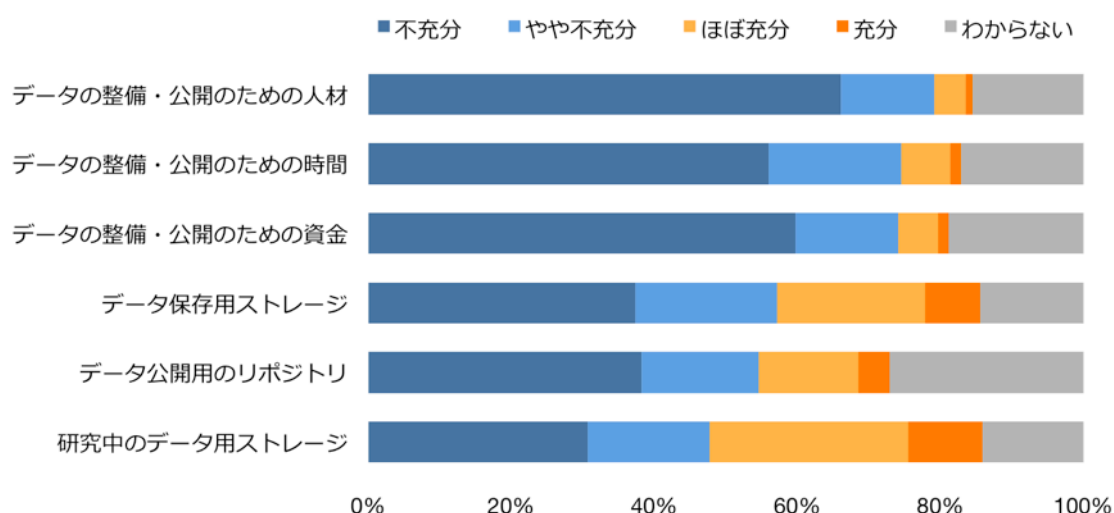


図 51 データ公開に関する資源の充足度 (n=1,396)

分野別では、最も充足度が低い人材と研究中のデータストレージ以外に差がみられた。図 52 に示すように、データ公開に関する人材は、分野を問わず不足していることがわかる(数学のみ「わからない」を選択した回答者が多く、50%を占めた)。

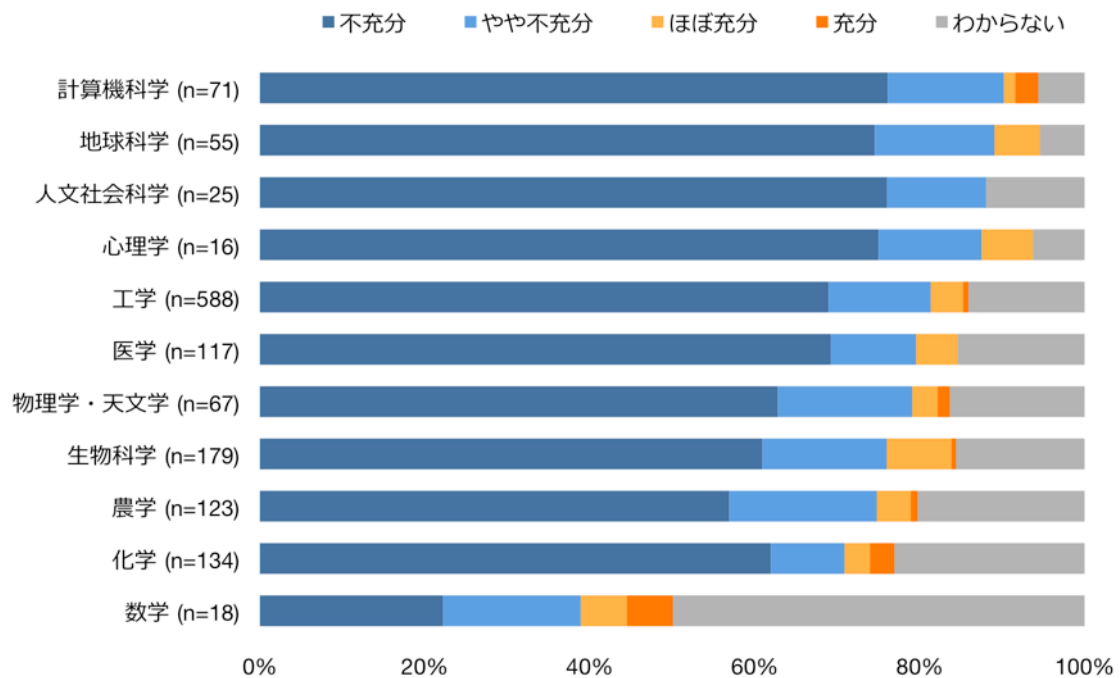


図 52 分野別データ公開に関する人材の充足度 (n=1,393)

所属別では、人材、資金、時間に関連がみられた(すべて $p<0.01$)。人材と時間は大学と企業の回答者が「不十分」、「やや不十分」を選択する比率が高く、資金は大学と公的機関・団体の回答者が「不十分」、「やや不十分」を選択する比率が高かった。

(2) データ公開に対する懸念

先行研究で示されているデータ公開に関する懸念の強さを確認するため、「Q29. カレントデータを論文の発表前に公開しようとする場合、次の点は問題(懸案)となりますか(なりましたか)?」と「問題である」、「やや問題である」、「あまり問題ではない」、「問題ではない」の4件法で尋ねた。結果を「問題である」と「やや問題である」の合計選択率が高い順に示す(図 53)。なお、データ公開経験と関連がみられたのは、「機密・プライバシー情報」と「商用利用される可能性」であった(いずれも $p<0.01$)。

最も懸念が強い項目は、「引用せずに利用される可能性」であった。自由回答でも、引用せずに利用される可能性について、剽窃や盗用、無断引用など多くの問題が指摘されていた。さらに、「悪意のある引用」、「揚げ足取り」、「公開したことで血祭りにあげられるような事象」、「実験条件の差異により、公開データが間違っていると揶揄される」、「データの扱い、論文の剽窃等、意識の面でグローバル・スタンダードに達しておらず、そのことが日本の学術に関する信用性を著しく低めていると感じています」など利用者の問題や不適切な批判を懸念する記述もみられた。

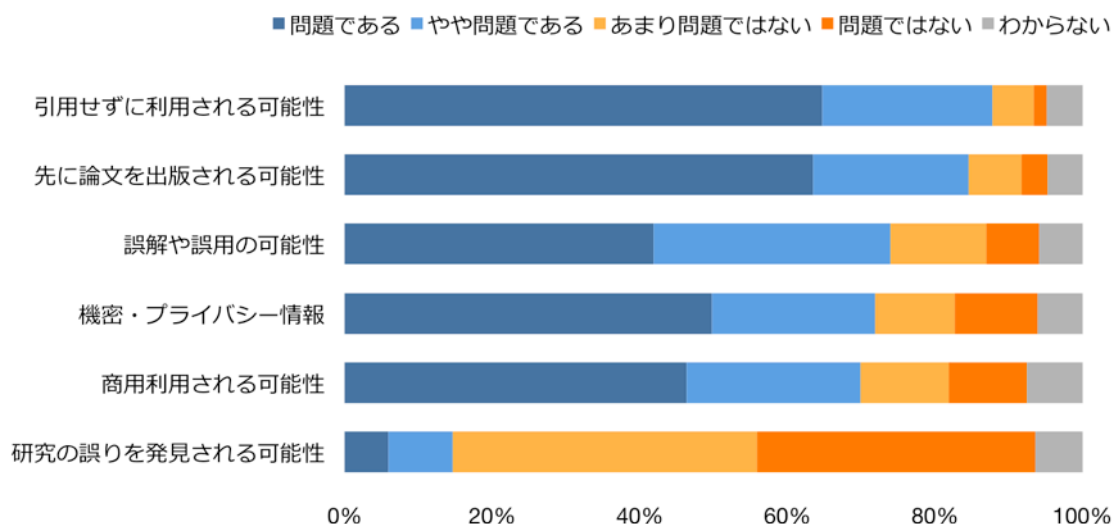


図 53 カレントデータの公開に関する懸念 (n=1,396)

分野別では、全ての項目について差がみられた。「引用せずに利用される可能性」については、人文社会科学、心理学、化学の順に懸念が強く、化学は「問題である」とする回答者の比率が最も高かった。数学は比較的懸念が弱いものの、「わからない」とする回答者も多かった(図 54)。

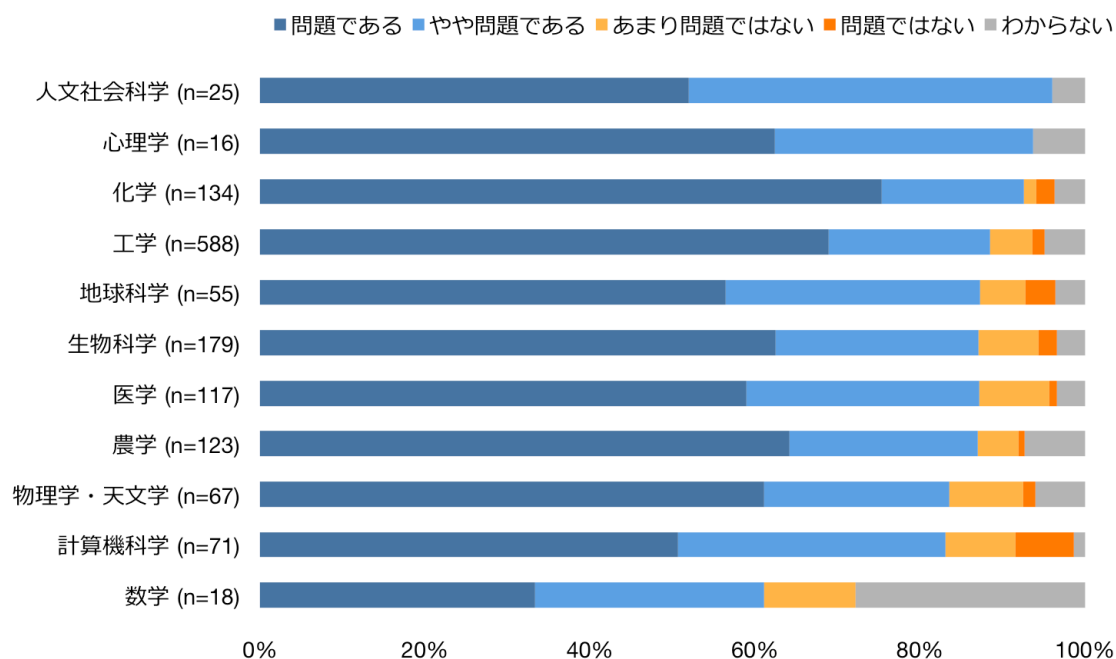


図 54 分野別「引用せずに利用される可能性」への懸念 (n=1,393)

次いで懸念が強い項目は、「先に論文を出版される可能性」であった。分野別にみると、化学、生物科学、医学の順に懸念が強い。数学は懸念が弱く、「わからない」とする回答者も多かった(図 55)。

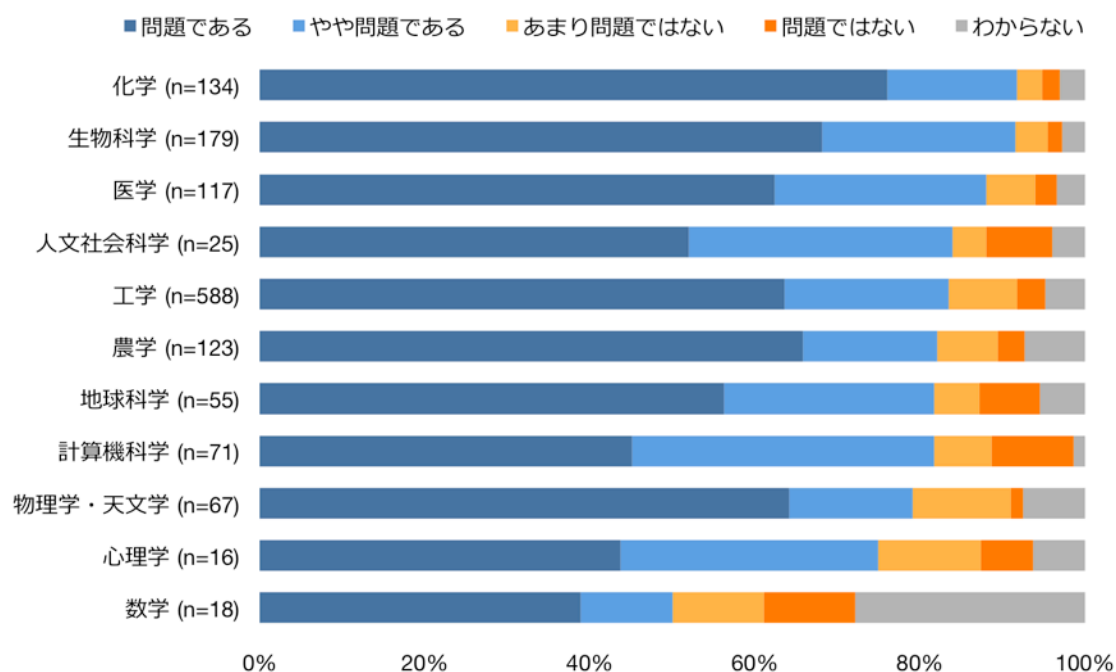


図 55 分野別「先に論文を出版される可能性」への懸念 (n=1,393)

「先に論文を出版される可能性」は, Schmidt らによる調査の「データを公開する前に結果を出版したいという願望」を意識した質問であり, 当該調査では若年層ほど重大な障壁となっていることが明らかにされていた(図 56)¹⁷。

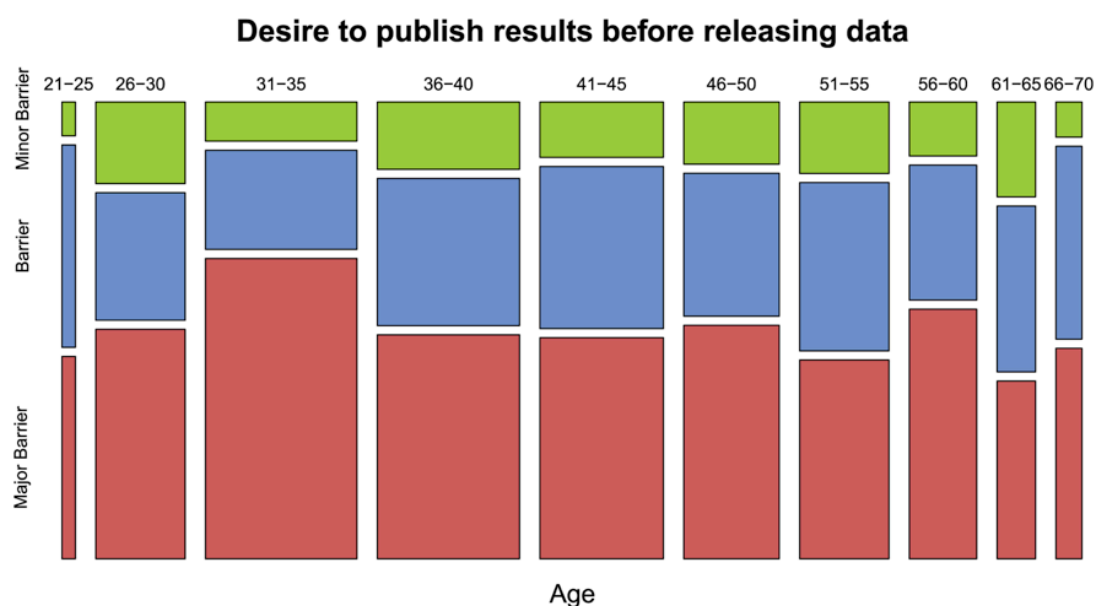


図 56 データを公開する前に結果を出版したいという願望¹⁷

<https://doi.org/10.1371/journal.pone.0146695.g008>

本調査の年齢別の結果を図 57 に示す。Schdmit らの調査とは選択肢の表現やスケールが異なるものの、同じく、これからキャリアを重ねると考えられる若年層ほど懸念が強かった ($p<0.01$)。

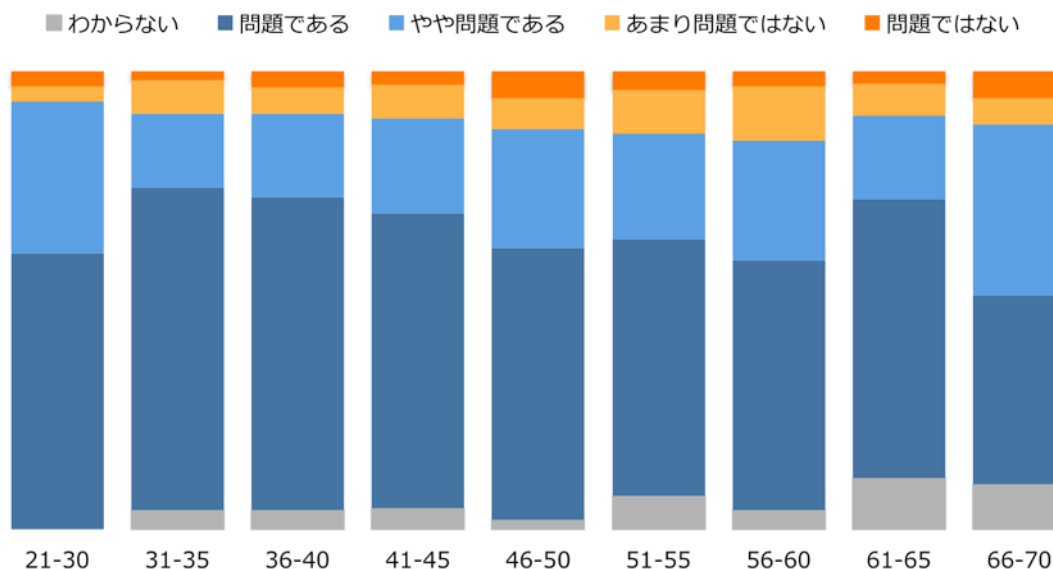


図 57 年齢別先に論文を出版される可能性への懸念 (n=1,393)

注) 図 56 は棒グラフの太さが回答者数を示しているが、図 57 は回答者数を反映していない。

Tenopir ら(2013/2014)の調査においても、11 項目の障壁のうち“(データよりも)先に(論文を)出版する必要がある”の選択率が最も高く(43.5%)、かつ、最も若いグループの選択率が高かった⁴¹。データ公開の議論や政策文書では、論文の出版より前にデータを公開することが想定されている場合が多いが、公開のタイミングは慎重に検討する必要があるだろう。

自由回答では、論文の出版前のデータ公開は、特許を取るために不可能であるとする意見や、先取権、業績や採用、昇進に関連した意見がみられた。たとえば、“アカデミックな論文に関しては、短期的な業績が求められ、かつ業績の質を掲載誌のインパクトファクターで測るような状態である限り、論文の出版前にデータを公開することは自殺行為であるので不可能と考える”といった記述があった。さらに、論文出版後であっても、一旦公開したデータを別の論文のために再利用するのをためらう意見もみられた。データ公開のタイミング、及びデータ公開者の先取権やオリジナリティを担保するシステムの開発を検討する必要があると考えられる。

(3) 他の研究者によるデータの理解

データの再利用においては、そのデータを解釈可能かどうかが重要であり、先行研究で誤用が阻害要因となっていたことから、カレントデータを同じ分野の研究者や異分野の研究者が理解できるかどうかを尋ねた(Q30, Q31)。その結果、同じ分野の研究者については 60.0%が「できると思う」を選択しているのに対して、異分野の研究者については 13.2%にとどまった(図 58)。

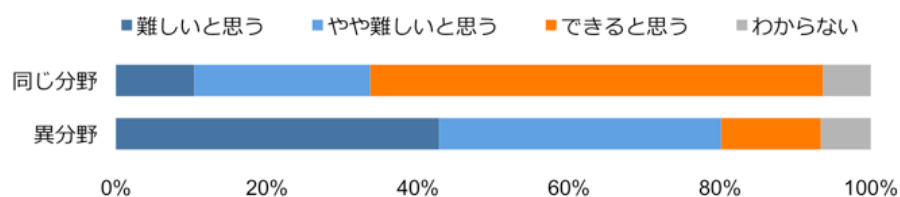


図 58 他の研究者によるデータの理解 (n=1,396)

異分野の研究者による理解については、分野による差がみられた($p<0.001$)。図 59 に、分野別の結果を「難しいと思う」と「やや難しいと思う」を選択した回答者の比率が高い順に示す。ただし、この順位は「できると思う」の比率の低い順位とは一致していない。たとえば、数学は「難しいと思う」、「やや難しいと思う」の合計は 77.8%であり比較的少ないが、「できると思う」は 0%であった。

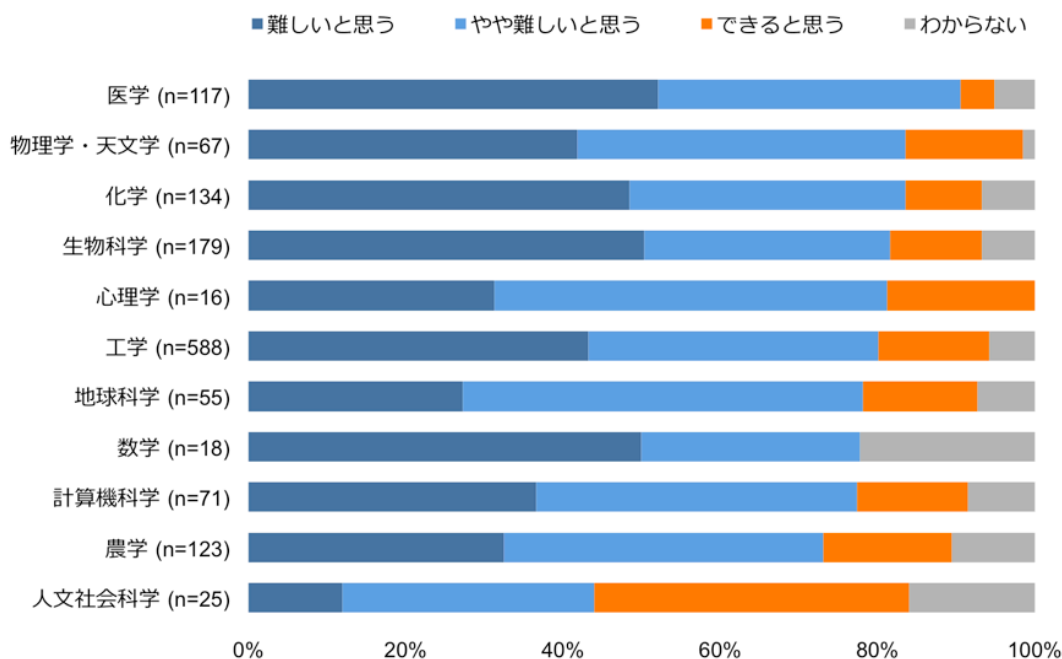


図 59 分野別異分野の研究者によるデータの理解 (n=1,393)

4.6 データ管理・公開に対する関心と専門性

(1) メトリクスへの関心

「Q32. カレントデータを公開した場合（公開したカレントデータについて）、ご関心がある項目をお選び下さい。」と複数選択方式で尋ねたところ、メトリクスを1つ以上選択した回答者は85.8%であった（図 60）。データ公開経験との関連はみられなかった。

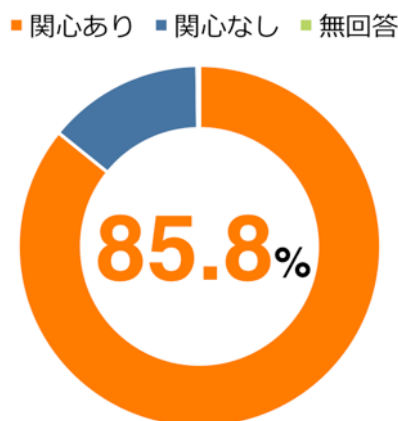
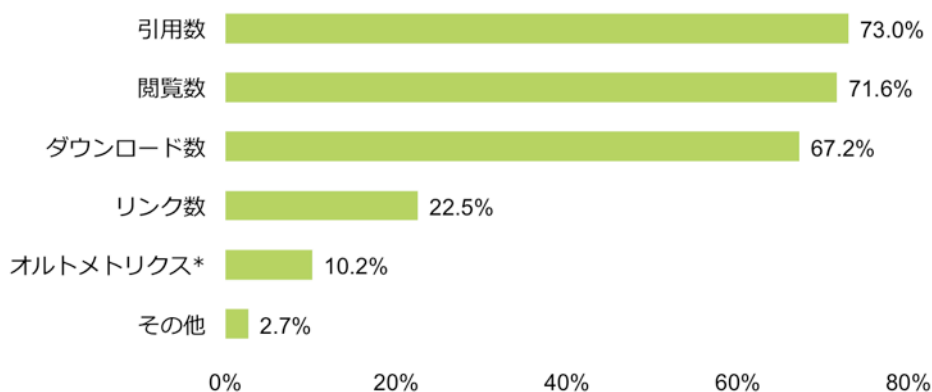


図 60 メトリクスへの関心の有無（n=1,396）

選択率が高かった項目は、引用数（データを利用した文献数）、閲覧数、ダウンロード数であり、オルトメトリクス（SNS 等での言及数）やリンク数は選択率が低かった（図 61）。「その他」では、利用者（15 名）や利用例（4 名）、利用目的（2 名）などが挙げられており、公開したデータがどのように再利用されたのかを追跡したいという認識がみられた。



*SNS 等での言及数

図 61 関心のあるメトリクス（n=1,198）

(2) データリテラシー教育への関心の有無

データの管理や公開に関するリテラシーへの関心を把握するため、「Q33. カレントデータを整備・公開する上で、より詳しく知りたいと思われる項目をお選び下さい。」複数選択方式で尋ねたところ、1 つ以上の項目を選択した回答者は 90.0%であった(図 62)。

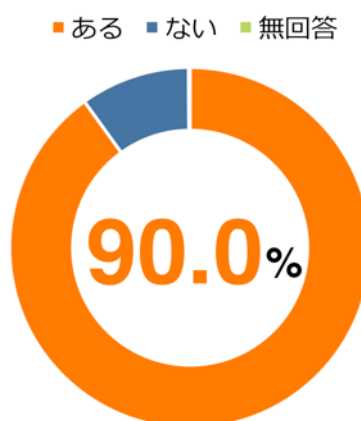


図 62 データリテラシー教育への関心の有無 (n=1,396)

選択率が高かった項目は、「知的財産権やライセンス」、「データの安全な管理方法」、「適切なデータ形式」であった(図 63)。「その他」では、問題が起きた場合の対応について、“商用盗用されたときの対応”、“クレームも含めた利用者への対応方法”などの記述がみられた。また、“分野間に渡る研究者が使用できるよう、適切なデータの活用法”、“データの削除方法とそのルール”などの記述もみられた。

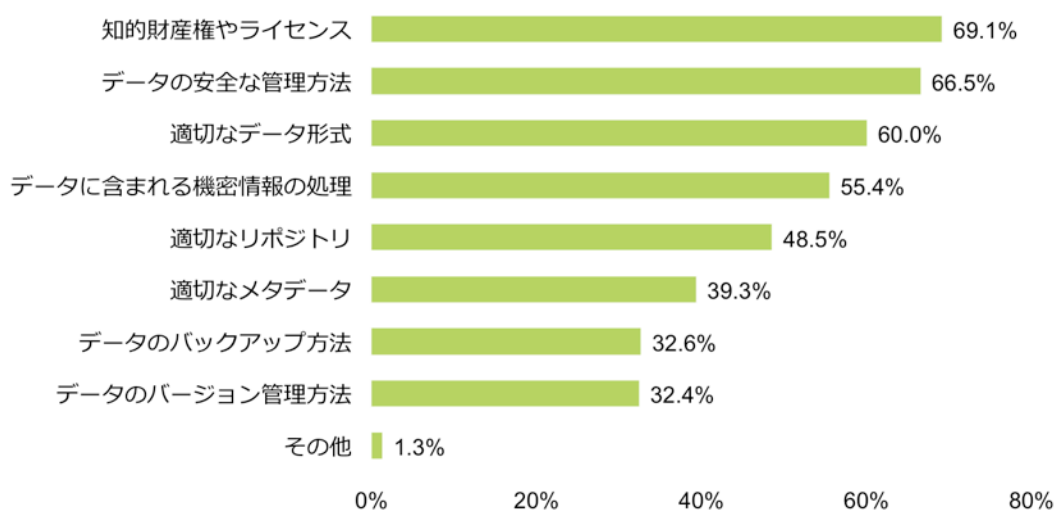


図 63 関心のある項目 (n=1,257)

(3) データ管理・公開の専門性

「Q34. カレントデータの整備や公開を、ご自身や共同研究者にかわって図書館員やデータキュレーターなどの第三者が行う場合、分野の知識や専門性が必要であると考えられる項目をお選び下さい。カレントデータが多様な場合は、もっとも難しいと考えられるデータについてお答え下さい」と複数選択方式で尋ねたところ、1 つ以上の項目を選択した回答者は 93.3%であった(図 64)。専門性の有無とデータ公開経験には関連がみられた($p<0.01$)。

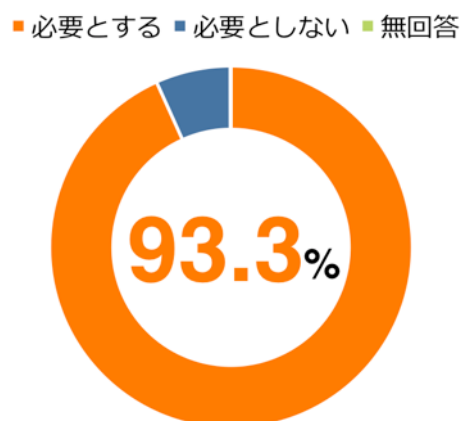


図 64 専門性の有無 (n=1,396)

半数以上の回答者が選択した項目は、「適切なデータ形式への変換」、「データを再利用しやすいように整える」、「適切なメタデータ標準の選択」であった(図 65)。「その他」では、“コード整備”、“適切なグラフ表示”、“クレームも含めた利用者への対応”など具体的な記述がみられた。比較的、選択率が低かったのは「機関リポジトリでの公開」や「異分野の研究者への紹介」、「メタデータの作成」であった。

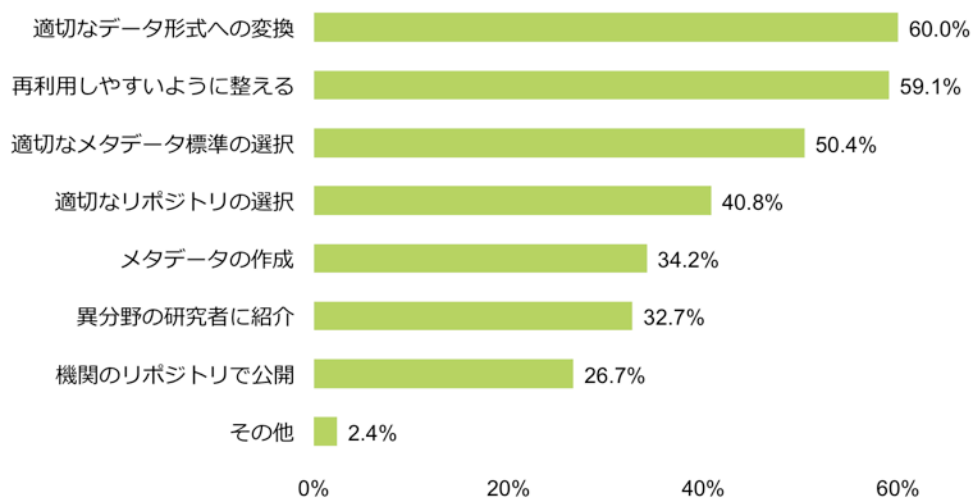


図 65 専門性を必要とする項目 (n=1,302)

自由回答では、“サポートスタッフがいないと継続的に公開することは不可能と思われる”という記述がみられた一方で、第三者による支援は不可能・困難であるとする意見が 11 件あり、“他人に行わせることはない”、“公開のためのデータ整理は、論文作成に近い作業であり、それを支援してもらおうというのは、論文を他人に書いてもらうのに近いことである”といった指摘もみられた。人的支援については、より詳細な調査や議論が必要であると考えられる。

4.7 自由回答

Q35 の自由回答では、アンケートに対する謝辞などを除くと 400 名が何らかの記述を行っていた。ここでは、(1)データ公開のインセンティブ、(2)データの保存と公開のインフラ、(3)データ公開におけるその他の問題について、主要な意見を示す。なお、全ての自由回答は Web 上に電子付録として掲載した(<http://doi.org/10.15108/rm268>)。

(1) データ公開のインセンティブ

データ公開に関する問題として、最も多くの指摘があったのはインセンティブの欠如であった(18 件)。このことは、“公開した人は公開の労力が必要なのにに対してメリットが無く、そのデータを利用する側にメリットがあるため、公開する側に大きなメリットがある仕組み作りが必要”という記述に端的に示されている。インセンティブの例としては、所属機関や助成機関による業績化とその評価(5 件)、引用(3 件)、公開データから収入を得ること、データ利用者からのフィードバックなどが挙げられている。たとえば、“データ公開が業績評価の対象になるのであれば、多少は力を入れることも可能になります”、“データの公開は、誰がどういう目的でということがわかった上で、論文化時に引用等してもらえらるなら歓迎します”といった記述がみられた。一方、“データを得ることは研究の一部ではあってもそれを得ただけで業績として扱われるのは、問題であると思う”という見解も示されていた。

(2) データの保存と公開のためのインフラ

データの(長期)保存に関しては、9 件の記述がみられた。“研究不正防止の観点から、生データの一定期間保存は重要”とする意見がある一方で、“データの公開以前に、保存についての段階で既に、個々のラボのキャパシティを越えていると思われます。これは現在の化学に於いて一回の解析で出てくるデータ量が莫大になる傾向が強まっているためです(とくにオミクス解析や画像データの多い生物学)”など保存コストの問題が指摘されている。また、研究者の流動性とデータの保管場所について、“研究者は異動することがあるため、前所属機関のサーバで公開していたデータが、果たしていつまで保存されているのか全く分からないことである。個人がいわゆる google などでデータを保存するのは容量的に無理があるため、公開するためのデータを保存できる研究者専用の公的なサーバやリポジトリが必要かもしれない。”とする意見もみられた。さらに、データを保存・公開するサーバについて、“データベースが適切に構築されるのか、つまりメタデータの出来・不出来でデータが生きるか・死ぬかが決まると思う。専門的なデータでさえデータベース化は難しい。多様な専門分野からのデータを効率よくデータベース化する方法はあるのだろうか？ 他者が望むデータを効率よく探し出せるのだろうか？ データベースが巨大化するほど難易度が上がる。そういった分野の発展にも期待する。”との見解も示された。

(3) データ公開におけるその他の問題

データを公開する際の形式について、標準化すべきであるとする意見と標準化によるコストを問題視する意見の両方がみられた。“統一フォーマットデータの提供を行えば、データのリサイクルやメタ解析分野の研究が進む”という意見がみられる一方で、標準化したとしてもその“耐用年数”は短いのではないかとする意見や標準化されたデータの価値に対する疑念が示されている。また、データを公開する国と活用する国の不均衡(2件)や、データベースのセキュリティ(2件)、“失敗したデータ(と失敗した原因)”の公開といった問題点が指摘された。

4.8 質問項目とデータ公開経験・属性との関連

表 7 に、本章で示した結果のまとめとして、質問項目ごとのデータ公開経験及び属性との関連を示す。「○」は関連がみられたこと、「なし」は関連がみられなかったこと、「不可」は検定が行えなかったことを表す。

表 7 質問項目とデータ公開経験・属性との関連

| 質問項目 | | 公開 | 分野 | 所属 | 年齢 |
|------------------------|-----------|----|----|----|----|
| (1)データの公開状況 | | | | | |
| Q18 データの公開経験 | | — | ○ | ○ | なし |
| Q21 非公開理由が解決された場合の公開意思 | | — | 不可 | なし | なし |
| (2)データの提供・被提供状況 | | | | | |
| Q9 データの提供頻度 | | ○ | ○ | ○ | ○ |
| Q10 データの被提供頻度 | | ○ | ○ | なし | ○ |
| (3)公開データの利用状況と課題 | | | | | |
| Q11 公開データの入手経験 | | ○ | 不可 | 不可 | 不可 |
| Q12 公開データ入手の障壁 | | なし | ○ | ○ | なし |
| Q13 公開データの利用目的 | 研究の参考 | ○ | なし | なし | なし |
| | 再分析・再利用 | ○ | ○ | ○ | なし |
| | 再現・追試 | ○ | ○ | ○ | なし |
| (4)データのプロフィール | | | | | |
| Q22 所属機関のデータ保存期間規定の有無 | | なし | ○ | 不可 | ○ |
| Q23 カレントデータの量 | | なし | ○ | なし | ○ |
| Q24 カレントデータの所有権 | | なし | 不可 | ○ | なし |
| Q25 カレントデータの機密情報 | | なし | ○ | ○ | ○ |
| (5)データを公開する場合の障壁 | | | | | |
| Q28 データ公開の資源 | 研究中のストレージ | なし | ○ | なし | なし |
| | 公開用リポジトリ | ○ | ○ | なし | なし |
| | 保存用ストレージ | なし | なし | なし | なし |
| | 時間 | ○ | ○ | ○ | なし |
| | 人材 | ○ | なし | ○ | なし |
| | 資金 | ○ | ○ | ○ | なし |
| Q29 データ公開の懸念 | 商用利用 | ○ | ○ | ○ | なし |
| | 機密・プライバシー | ○ | ○ | ○ | なし |
| | 誤解や誤用 | なし | ○ | なし | なし |
| | 引用せずに利用 | なし | ○ | ○ | ○ |
| | 研究の誤りを発見 | なし | ○ | ○ | なし |
| | 先に論文を出版 | なし | ○ | ○ | ○ |
| Q30 カレントデータの理解(自分野) | | なし | なし | なし | なし |
| Q31 カレントデータの理解(異分野) | | ○ | なし | なし | なし |
| (6)データの管理・公開に対する関心と専門性 | | | | | |
| Q32 メトリクスへの関心 | | なし | ○ | なし | なし |
| Q34 カレントデータの管理・公開の専門性 | | ○ | 不可 | なし | なし |

※「不可」は、n が少ないために検定できなかった項目を示す。

5. まとめ

本調査は、国内外でオープンサイエンスの政策議論が盛んに行われている現時点において、日本の研究者によるデータ公開の実態と課題、支援の可能性について、国外の調査を参考にしつつ、分野や所属機関による違いやOA論文との比較を行いながら明らかにすることを目指した。OA論文は雑誌のポリシーや資金の問題が解決されることによって進展することが示唆されたが、データ公開については、公開に対するさまざまな懸念が強く、公開のための資源も不足しており、一部の問題に対応するだけでは進展させることが困難であることが示唆された。安全・安心なデータの公開・入手環境をインフラ・制度ともに整備することが重要であると考えられる。そのための新たな課題と期待される効果を表8に示す。なお、本報告書は調査資料として主要な結果を示したが、考察や分析については報告者らによる紹介記事⁵⁸や論文⁵⁹をご参照いただきたい。

表8 データ公開の障壁と新たな課題及び期待される効果

| 障壁 | 新たな課題 | 期待される効果 |
|---------------------------------------------------------|------------------------------------------------------|---------------------------------------------------|
| キャリアリスク 引用せずに利用される可能性 先に論文を出版される可能性 インセンティブの欠如 | データ引用の標準化と徹底 データ公開の業績化 データ公開時期の（再）検討 先取権の保証 | キャリアベネフィット 論文の引用増加 国際共同研究の契機 採用や昇進 |
| データ公開のための資金 データ公開用のリポジトリ 公開データの信頼性 | インフラへの投資と整備 発見可能性の向上 相互運用性の向上 データの評価・品質管理 | 研究成果の認知向上 公開データの利用促進 科学的利他性の醸成 低コストでの利活用 |
| 法的問題（知的財産権） 倫理的問題（機密等） 誤解・誤用・不正利用の懸念 | リテラシー教育 ガイドラインの策定 | 再利用に関する懸念の低減 永続性の確保 |
| データ公開のための人材 データ公開のための時間 | データ管理支援体制の構築 人材育成とキャリアパス | 研究コストの削減 時間や資金の節減 |

調査実施時点の2016年から現在までの間にも、Springer Nature社によるデータ公開ポリシーの公開⁶⁰や科学技術振興機構によるオープンサイエンス方針の公開¹⁶など、オープンサイエンスを取り巻く環境はめまぐるしく進展している。こうした状況に応じて、研究者のオープンサイエンスに対する認識や研究を阻害する要因も変化してゆくだろう。研究のための資金や時間といった資源を最大限に活用するためにも、日本の研究者の直面している問題を継続的に明らかにした上で、学術政策に反映させる必要があるのではないだろうか。また、日本の研究者の状況をより正確に把握するためには、国外の研究者を含めた調査を実施することが望ましいと考えられる。

謝辞

本調査及びプレテストにご協力を賜りました皆様に、心よりお礼申し上げます。質問数の多い調査であったにもかかわらず、大勢の皆様からのご協力をいただきましたこと、必須ではない質問においても無回答の項目が極めて少数であったこと、自由回答に率直なご意見を頂戴しましたことに重ねて感謝申し上げます。

参考文献

- ¹ 林和弘, 村山泰啓. オープンサイエンスをめぐる新しい潮流(その3): 研究データ出版の動向と論文の根拠データの公開促進に向けて. 科学技術動向研究. 2015, vol. 148, p. 4-9.
<http://hdl.handle.net/11035/2999>, (accessed 2017-12-12).
- ² OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD Publications, 2007, 23p. <http://www.oecd.org/science/sci-tech/38500813.pdf>, (accessed 2017-12-12).
- ³ Foreign & Commonwealth Office. G8 Science Ministers Statement. GOV.UK. 2013-6-13,
<https://www.gov.uk/government/news/g8-science-ministers-statement>, (accessed 2017-12-12).
- ⁴ 村山泰啓, 林和弘. オープンサイエンスをめぐる新しい潮流(その1): 科学技術・学術情報共有の枠組みの国際動向と研究のオープンデータ. 科学技術動向. 2014, no. 146, p. 12-17.
- ⁵ Validation of the Results of the Public Consultation on Science 2.0: Science in Transition. European Commission, 2015, 38p.
<https://ec.europa.eu/digital-single-market/news/final-report-science-20-public-consultation>, (accessed 2017-12-12).
- ⁶ OECD. Making Open Science a Reality. OECD Publishing, 2015, OECD Science, Technology and Industry Policy Papers, no. 25, 112p. <http://doi.org/10.1787/5jrs2f963zs1-en>, (accessed 2017-12-12).
- ⁷ The Royal Society. Science as an Open Enterprise. The Royal Society Science Policy Centre, 2012, 104p.
http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf, (accessed 2017-12-12).
- ⁸ 林和弘. オープンサイエンス時代の研究公正. 情報の科学と技術. 2016, vol. 66, no. 3, p. 98-102. http://doi.org/10.18919/jkg.66.3_98, (accessed 2017-12-12).
- ⁹ 林和弘. オープンサイエンスをめぐる新しい潮流(その5): オープンな情報流通が促進するシチズンサイエンス(市民科学)の可能性. 科学技術動向. 2015, vol. 150, p. 21-25.
<http://hdl.handle.net/11035/3097>, (accessed 2017-12-12).
- ¹⁰ 林和弘. 動向レビュー: 世界のオープンアクセス, オープンサイエンス政策の動向と図書館の役割. カレントアウェアネス. 2015, no. 324, p. 15-18. <http://doi.org/10.11501/9396326>, (accessed 2017-12-12).
- ¹¹ Executive Office of the President: Office of Science and Technology Policy. “Increasing Access to the Results of Federally Funded Scientific Research”. 2013-2-22.
http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.p

df, (accessed 2017-12-12).

- ¹² 国際的動向を踏まえたオープンサイエンスに関する検討会. 我が国におけるオープンサイエンス推進のあり方について～サイエンスの新たな飛躍の時代の幕開け～. 内閣府, 2015, 23p. <http://www8.cao.go.jp/cstp/sonota/openscience/>, (accessed 2017-12-12).
- ¹³ 文部科学省科学技術・学術審議会学術分科会第8期学術情報委員会. 学術情報のオープン化の推進について(審議まとめ). 2016, 26p. http://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/afieldfile/2016/04/08/1368804_1_1_1.pdf, (accessed 2017-12-12).
- ¹⁴ 日本学術会議. オープンイノベーションに資するオープンサイエンスのあり方に関する提言. 2016, 28p. <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t230.pdf>, (accessed 2017-12-12).
- ¹⁵ 内閣府. 第5期科学技術基本計画, 2016. 53p. <http://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>, (accessed 2017-12-12).
- ¹⁶ “オープンサイエンス促進に向けた研究成果の取扱いに関する JST の基本方針”. 科学技術振興機構, 2017. <https://www.jst.go.jp/pr/intro/openscience/index.html>, (accessed 2017-12-12)
- ¹⁷ Schmidt, Birgit; Gemeinholzer, Birgit; Treloar, Andrew. Open data in global environmental research: The Belmont Forum’s open data survey. PLOS ONE. 2016, vol. 11, no. 1, e0146695. <http://doi.org/10.1371/journal.pone.0146695>, (accessed 2017-12-12).
- ¹⁸ 近藤康久. レポート紹介『地球環境研究におけるオープンデータ: ベルモント・フォーラムによるオープンデータ調査』. 情報管理. 2016, vol. 59, no. 4, p. 250-258. <http://doi.org/10.1241/johokanri.59.250>, (accessed 2017-12-12).
- ¹⁹ Stodden, Victoria; Guo, Peixuan; Ma, Zhaokun. Toward reproducible computational research: An empirical analysis of data and code policy adoption by journal. PLOS ONE. 2013, vol. 8, no. 6, e67111. <http://doi.org/10.1371/journal.pone.0067111>, (accessed 2017-12-12).
- ²⁰ “Over 600 Springer Nature journals commit to new data sharing policies”. Springer Nature. 2016-12-6. <http://www.springernature.com/br/group/media/press-releases/over-600-springer-nature-journals-commit-to-new-data-sharing-policies/11111248>, (accessed 2017-12-12).
- ²¹ Phillips, Nicky. Striving for a research renaissance. Nature. 2017, vol. 543, no. 7646, S7 (Nature Index). <http://doi.org/10.1177/10.1038/543S7a>, (accessed 2017-12-12).
- ²² 池内有為, 逸村裕. 学術雑誌によるデータ共有ポリシー: 分野間比較と特徴分析. 日本図書館情報学会誌. 2016, vol. 62, no. 1, p. 20-37. http://doi.org/10.20651/jslis.62.1_20, (accessed 2017-12-12).
- ²³ 科学技術・政策研究所. 科学技術の状況に係る総合的意識調査(NISTEP 定点調査 2015) 報告書. 科学技術政策研究所, 2016, NR-166, 182p. <http://doi.org/10.15108/nr166>, (accessed 2017-12-12).

- 2017-12-12).
- ²⁴ 林和弘ほか. 研究データ出版の動向と論文の根拠データの公開促進に向けて. 科学技術動向, 2015, 148, p. 4-9. <http://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-STT148J-4.pdf>, (accessed 2017-12-12).
- ²⁵ 南山泰之. データジャーナル: 研究データ管理の新たな試み. カレントアウェアネス. 2015, no. 325, p. 19-22. <http://doi.org/10.11501/9497651>, (accessed 2017-12-12).
- ²⁶ “Joint Declaration of Data Citation Principles”. FORCE11. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>, (accessed 2017-12-12).
- ²⁷ “Data Citation”. Elsevier. <https://www.elsevier.com/about/open-science/research-data/data-citation>, (accessed 2017-12-12).
- ²⁸ Piwowar, Heather. Altmetrics: Value all research products. Nature. 2013, vol. 493, no. 7431, p. 159. <http://doi.org/10.1038/493159a>, (accessed 2017-12-12).
- ²⁹ Martone, M. E. Brain and Behavior: We want you to share your data. Brain and Behavior. 2014, vol. 4, no. 1, p. 1-3. <http://doi.org/10.1002/brb3.192>, (accessed 2017-12-12).
- ³⁰ 小野雅史, 小池俊雄, 柴崎亮介. 地球環境情報分野における研究データ共有に関する意識調査: 研究現場の実態. 情報管理. 2016, vol. 59, no. 8, p. 514-525. <http://doi.org/10.1241/johokanri.59.514>, (accessed 2017-12-12).
- ³¹ 平成 27 年度機関リポジトリ担当者オンラインワークショップ「研究データから研究プロセスを知る」. デジタルリポジトリ連合. <http://drf.lib.hokudai.ac.jp/drf/index.php?onlineworkshop2015>, (accessed 2017-12-12).
- ³² 平成 28 年度機関リポジトリ担当者オンラインワークショップ「研究データから研究プロセスを知る」. デジタルリポジトリ連合. <http://drf.lib.hokudai.ac.jp/drf/index.php?onlineworkshop2016>, (accessed 2017-12-12).
- ³³ 倉田敬子, 松林麻実子, 武田将季. 日本の大学・研究機関における研究データの管理, 保管, 公開: 質問紙調査に基づく現状報告. 情報管理. 2017, vol. 60, no. 2, p. 119-127. <http://doi.org/10.1241/johokanri.60.119>, (accessed 2017-12-12).
- ³⁴ G7 茨城・つくば科学技術大臣会合. つくばコミュニケ(共同声明). 内閣府. 2016, p. 9. http://www8.cao.go.jp/cstp/kokusaiteki/g7_2016/2016communique.html, (accessed 2017-12-12).
- ³⁵ Tenopir, Carol et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011, vol. 6, no. 6, e21101. <http://doi.org/10.1371/journal.pone.0021101>, (accessed 2017-12-12).
- ³⁶ Huang, Xiaolei et al. Willing or unwilling to share primary biodiversity data: Results and implications of an international survey. Conservation Letters. 2012, vol. 5, p. 399-406.

- ³⁷ Kim, Youngseek; Stanton, Jeffrey M. Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*. 2015, vol. 67, no. 4, p. 776-799. <http://doi.org/10.1002/asi.23424>, (accessed 2017-12-12).
- ³⁸ Kim, Youngseek; Zhang, Ping. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*. 2015, vol. 37, no. 3, p. 189-200. <http://doi.org/10.1016/j.lisr.2015.04.006>, (accessed 2017-12-12).
- ³⁹ Kim, Youngseek; Burns, C. Sean. Norms of data sharing in biological sciences: The roles of metadata, data repository, and journal and funding requirements. *Journal of Information Science*. 2016, vol. 42, no. 2, p. 230-245. <http://doi.org/10.1177/0165551515592098>, (accessed 2017-12-12).
- ⁴⁰ Sharing and Archiving of Publicly Funded Research Data: Report to the Research Council of Norway. Danvad, 2014, 74p.
- ⁴¹ Tenopir, Carol et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE*. 2015, vol. 10, no. 8, e0134826. <http://doi.org/10.1371/journal.pone.0134826>, (accessed 2017-12-12).
- ⁴² Kratz, John Ernest; Strasser, Carly. Researcher perspectives on publication and peer review of data. *PLOS ONE*. 2015, vol. 10, no. 2, e0117619. <http://doi.org/10.1371/journal.pone.0117619>, (accessed 2017-12-12).
- ⁴³ Ferguson, Liz. "How and why researchers share data (and why they don't)". Wiley exchanges: Our ideas, research and discussion blog. 2014-11-3. <https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges>, (accessed 2017-12-12).
- ⁴⁴ Fecher, Benedikt et al. A reputation economy: Results from an empirical survey on academic data sharing. *arXiv.org*, 2015, 2015arXiv150300481F. <https://arxiv.org/abs/1503.00481>
- ⁴⁵ Berghmans, Stephane et al. Open Data: The Researcher Perspective. Centre for Science and Technology Studies (CWTS), 2017, 48p. <https://www.cwts.nl/download/f-53w2.pdf>, (accessed 2017-12-12).
- ⁴⁶ Wallis, Jillian C.; Rolando, Elizabeth; Borgman, Christine L. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLOS ONE*. 2013, vol. 8, no. 7, e67332. <http://doi.org/10.1371/journal.pone.0067332>, (accessed 2017-12-12).
- ⁴⁷ Van den Eynden, Veerle; Bishop, Libby. Sowing the Seed: Incentives and Motivations for Sharing Research Data, a Researcher's Perspective. Knowledge Exchange, 2014, 48p. http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf, (accessed 2017-12-12).

- ⁴⁸ Van Noorden, Richard. Confusion over publisher's pioneering open-data rules. *Nature*. 2014, vol. 515, no. 7528, p. 478. <http://doi.org/10.1038/515478a>, (accessed 2017-12-12).
- ⁴⁹ Wicherts, Jelte M.; Bakker, Marjan; Molenaar, Dylan. Willingness to share research data is related to the strength of the evidence and quality of reporting of statistical results. *PLOS ONE*. 2011, vol. 6, no. 11, e26828. <http://doi.org/10.1371/journal.pone.0026828>, (accessed 2017-12-12).
- ⁵⁰ Fecher, Benedikt; Friesike, Sascha; Hebing, Marcel. What drives academic data sharing? *PLOS ONE*. 2015, vol. 10, no. 2, e0118053. <http://doi.org/10.1371/journal.pone.0118053>, (accessed 2017-12-12).
- ⁵¹ Challenges and opportunities. *Science*. 2011, vol. 331, no. 6018, p. 692-693. <http://doi.org/10.1126/science.331.6018.692>, (accessed 2017-12-12).
- ⁵² Kratz, John E.; Strasser, Carly. Making data count. *Scientific Data*. 2015, vol. 2, 150039 (2015). <http://doi.org/10.1038/sdata.2015.39>, (accessed 2017-12-12).
- ⁵³ <http://www.nistep.go.jp/activities/st-experts-network>
- ⁵⁴ 矢野幸子. 科学技術予測活動におけるウェブメディア双方向性機能強化の検討. *STI Horizon*. 2017, vol. 3, no. 2, p. 32-35. <http://doi.org/10.15108/stih.00081>, (accessed 2017-12-12).
- ⁵⁵ “機関リポジトリ公開数とコンテンツ数の推移”. 国立情報学研究所:学術機関リポジトリ構築連携支援事業. <https://www.nii.ac.jp/irp/archive/statistic/>, (accessed 2017-12-12).
- ⁵⁶ “コンテンツ増減”. 国立情報学研究所:IRDB コンテンツ分析. <http://irdb.nii.ac.jp/analysis/index.php>, (accessed 2017-12-12).
- ⁵⁷ 大園隼彦. オープンサイエンスの最新情報 : メタデータの相互運用性を中心に. 薬学図書館. 2017, vol. 62, no. 1, p. 40-47.
- ⁵⁸ 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ーオープンサイエンスの課題と展望ー. *STI Horizon*. 2017, vol. 3, no. 4, p. 27-32.
- ⁵⁹ 池内有為. 日本における研究データ公開の状況と推進要因, 阻害要因の分析. *Library and Information Science*. (2018 年刊行予定)
- ⁶⁰ “Research data policies”. Springer Nature. <http://www.springernature.com/gp/group/data-policy/>, (accessed 2017-12-12).

資料

資料

(1) 質問票

論文と研究データの公開に関する実態調査：オープンサイエンス支援体制の構築に向けて

<調査の趣旨>

近年、科学研究の成果として論文や、その根拠となるデータをインターネットで公開するオープンサイエンスの動きが国内外で盛んになっています。2016年5月の[G7科学技術大臣会合](#)ではオープンサイエンスが分野横断的な課題として位置づけられ、[第5期科学技術基本計画（2016～2020年度）](#)では“国は、資金配分機関、大学等の研究機関、研究者と連携し、オープンサイエンスの推進体制を構築する”と述べられています。

研究データの公開については、データの再利用によって研究を効率的に進めることや、異分野データを融合して産学官におけるイノベーションを創出することが期待されており、論文の投稿時にデータ公開を要求する学術雑誌も増加しています。一方、研究データを公開するためには人的・物的資源が必要であるため、研究のための時間や資金に限りがある産学官の皆様への支援体制の構築は喫緊の課題であるとされています。

そこで、オープンサイエンスに関する適切な支援体制の構築に向けて、日本における現状と課題、支援のニーズを明らかにするための質問紙調査を実施致します。**論文や研究データを公開したご経験がないという回答も参考になりますので、ぜひ率直なご意見をお聞かせ下さい。**

調査結果は、NISTEPの刊行物として公表する予定です。また、ご回答の一部は、担当者が他の出版物（学術論文及び博士論文等）に使用させていただきます。ご回答は個人が特定できないよう統計的に処理し、自由回答を引用する場合は匿名性を確保致します。

お忙しいところ誠に恐縮ですが、ご協力のほど何卒よろしくお願い申し上げます。

2016年11月30日

文部科学省

科学技術・学術政策研究所

科学技術予測センター

上席研究官 林 和弘

客員研究官 池内有為

<回答要領>

(1)2016年12月9日（金）までにご回答下さいますよう、お願い申し上げます。

(2)質問は合計35問で、回答には30分程度を要します。

なお、必須項目（＊印）は20問です。

また、選択肢で「その他」を選んだ場合は具体的な内容の入力が必須となります。

(3)あらかじめ質問をご覧になりたい場合は、PDF形式の調査票を[こちら](#)からダウンロードできます。

(4)ご回答は、文部科学省科学技術・学術政策研究所において厳正に管理します。

(5)ご不明な点がある場合には、下記担当までご連絡下さい。

<留意事項>

- ・アンケートは画面下の「回答する」をクリックして進めてください。
- ・前に戻る場合は、Webブラウザの「戻る」（「←」ボタン等）で戻ってください。

- ・ 回答を一時的に保存したり、再開する機能はご用意していません。
途中で回答を中止する際には、再度最初の質問から回答し直して頂く必要があります。
- ・ なお、回答画面（回答用URL）には、タイムアウトによるセッションの切断はないため、ブラウザを閉じない限りは、回答状況が維持されます。
- ・ 回答中にページを再読み込み（リロード）した場合、アンケート開始ページに遷移し、それまでの回答は破棄されますのでご注意ください。
- ・ 回答データは、「送信」を押さない限り送信されないため、回答終了後、忘れずに「送信」ボタンを押してください。なお、「送信」後の修正は出来ません。

<アンケート担当>

科学技術予測センター

林 宣子

Mail: stfc@nistep.go.jp

回答する

1. 研究分野について

Q1. ご自身の研究分野に最も近いものをお選び下さい。

*

- (1) 工学
- (2) 天文学
- (3) 化学
- (4) 物理学
- (5) 地球科学
- (6) 数学
- (7) コンピュータサイエンス
- (8) 農学
- (9) 生物科学
- (10) 医学
- (11) 心理学
- (12) 社会科学
- (13) 人文学
- (14) その他 [具体的に]

(15) 研究を行ったことはない（口頭発表や論文出版の経験はない） **【→Q35へ】**

次へ

2. 学術論文について

ご自身の論文のオープンアクセス（公開）状況や、論文の利用についてお伺いします。

なお、本調査における「オープンアクセス」とは、論文がインターネットで公開され、読者は無料で読むことができる状態とします。

【オープンアクセスの例】

- ・ オープンアクセスの雑誌で出版する（PLOSなど）
- ・ 雑誌のオープンアクセスオプションを選択する（Springer Open Choiceなど）
- ・ 雑誌等が一定期間経過後に論文をオープンアクセスにする
- ・ 機関リポジトリやプレプリントサーバで論文を公開する（arXivなど）

Q2. ご自身の論文について、あてはまるものをお選び下さい。

*

- (1) オープンアクセスの論文がある
- (2) オープンアクセスの論文はない **【→Q4へ】**
- (3) わからない **【→Q4へ】**

次へ

Q3. 論文をオープンアクセスにした理由として、あてはまるものをお選び下さい。

* (複数選択)

- (1) 他の研究者からのリクエストに応じて
- (2) 論文を投稿した雑誌がオープンアクセスだから
- (3) 助成機関のポリシー（助成条件）だから
- (4) 所属機関のポリシーだから
- (5) 分野・コミュニティの規範だから
- (6) 研究成果を広く認知してもらいたいから
- (7) 科学研究や成果実装を推進したいから
- (8) オープンアクセスに貢献したいから
- (9) その他〔具体的に〕

- (10) あてはまるものはない

【すべて→Q6へ】

次へ

Q4. 論文がオープンアクセスではない理由として、あてはまるものをお選び下さい。

* (複数選択)

- (1) 資金が必要だから
- (2) 時間が必要だから
- (3) ニーズがないと思うから
- (4) 投稿したい雑誌がオープンアクセスではないから
- (5) 助成機関のポリシー（助成条件）ではないから
- (6) 所属機関にオープンアクセス方針がないから
- (7) 分野・コミュニティで推奨されていないから
- (8) リポジトリなどの公開手段がないから
- (9) その他〔具体的に〕

- (10) あてはまるものはない **【→Q6へ】**

次へ

Q5. Q4の理由が解決された場合、論文をオープンアクセスにしたいと思われますか？

*

- (1) はい
- (2) いいえ
- (3) わからない

Q6.

研究に利用したことがある（アイデアの参考にしたり引用した経験のある）論文の分野をお選び下さい。

* (複数選択)

- (1) 工学
- (2) 天文学
- (3) 化学
- (4) 物理学
- (5) 地球科学
- (6) 数学
- (7) コンピュータサイエンス
- (8) 農学
- (9) 生物科学
- (10) 医学
- (11) 心理学
- (12) 社会科学
- (13) 人文学
- (14) その他 [具体的に]

Q7. 論文を探す際に、よく利用する検索ツールや情報源をお選び下さい。

(複数選択)

- (1) サーチエンジン (Google, Google Scholarなど)
- (2) 論文情報のデータベース (Web of Science, Scopus, CiNii Articlesなど)
- (3) 出版社や学術雑誌のサイト (Elsevier, Wileyなど)
- (4) 学術機関のリポジトリ・アーカイブ (大学やNASAの機関リポジトリなど)
- (5) 論文やプレプリントのサーバ (PubMed Central, arXiv, J-Stageなど)
- (6) 論文や学術記事の参考文献
- (7) 政府・機関・出版社などの広報、ニュースレター
- (8) ブログや一般的なSNS (Facebook, Twitterなど)
- (9) 学術系SNS (Mendeley, ResearchGateなど)
- (10) アラートサービス (RSSなど)
- (11) メーリングリスト
- (12) 研究者や同僚に尋ねる／教えてもらう
- (13) その他 [具体的に]

Q8. 論文を利用する際に、信頼性の判断基準としている項目をお選び下さい。

(複数選択)

- (1) 著者情報 (所属機関, 職位など)
- (2) 掲載雑誌
- (3) 引用数 (その論文を引用した文献数)
- (4) オルトメトリクス (SNSやブログの言及数)
- (5) ダウンロード数
- (6) 抄録
- (7) 研究手法の確かさ
- (8) その他 [具体的に]

3. 研究データの提供について

本調査における「データ」とは、研究のために収集・作成・観測した**デジタルデータ**を指します。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しません。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、内容も限定しません。

データの「提供」とは、E-mailやUSBフラッシュメモリ、クラウドサービス（DropboxやGoogle Drive）などを使って、共同研究者を除く他者に渡す（共有する）ことを指します。特定の人以外はアクセスできない状態です。

Q9. 共同研究者を除く他の研究者にデータを提供したご経験はありますか？

- (1) よくある
- (2) たまにある
- (3) ほとんどない
- (4) まったくない
- (5) わからない

Q10. 共同研究者を除く他の研究者からデータの提供を受けたご経験はありますか？

- (1) よくある
- (2) たまにある
- (3) ほとんどない
- (4) まったくない
- (5) わからない

4. 公開データの利用について

「公開データ」とは、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します。利用料金や利用者登録が必要な場合も含まれます。

Q11. これまでに、公開データを以下の公開先から入手したご経験はありますか？

* (複数選択)

- (1) 個人や研究室のウェブサイト
- (2) 学術機関のリポジトリ・データアーカイブ (大学やNASAのリポジトリなど)
- (3) 特定分野のリポジトリ・データアーカイブ (DDBJやICPSRなど)
- (4) データ共有サービス (figshare, zenodoなど)
- (5) コード共有サービス (GitHubなど)
- (6) 論文の補足資料 (supplementary materials)
- (7) 学術系SNS (Mendeley, ResearchGateなど)
- (8) その他 [具体的に]

- (9) ない **【→Q17へ】**
- (10) わからない **【→Q17へ】**

次へ

4. 公開データの利用について

「公開データ」とは、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します。利用料金や利用者登録が必要な場合も含まれます。

Q11. これまでに、公開データを以下の公開先から入手したご経験はありますか？

* (複数選択)

- (1) 個人や研究室のウェブサイト
- (2) 学術機関のリポジトリ・データアーカイブ (大学やNASAのリポジトリなど)
- (3) 特定分野のリポジトリ・データアーカイブ (DDBJやICPSRなど)
- (4) データ共有サービス (figshare, zenodoなど)
- (5) コード共有サービス (GitHubなど)
- (6) 論文の補足資料 (supplementary materials)
- (7) 学術系SNS (Mendeley, ResearchGateなど)
- (8) その他 [具体的に]

- (9) ない **【→Q17へ】**
- (10) わからない **【→Q17へ】**

次へ

Q12. 公開データの入手にあたって、問題だと感じたことがある項目をお選び下さい。

* (複数選択)

- (1) 利用料金が必要
- (2) 利用者登録が必要
- (3) リクエストから入手までに時間がかかる
- (4) 最新のデータを入手できない
- (5) データごとにフォーマットが異なる
- (6) データごとに品質が異なる
- (7) データの解釈や再利用方法がよくわからない
- (8) 利用したいデータへのアクセス方法がよくわからない
- (9) 利用条件がよくわからない
- (10) 著作者情報がよくわからない
- (11) その他 [具体的に]

- (12) 問題を感じたことはない

Q13. 入手した公開データの利用目的について、あてはまるものをそれぞれお選び下さい。

*

行わない 1

たまに行う 2

行う 3

自身の研究のアイデアや仮説の参考にする

再分析・再利用して自身の研究を行う

研究を再現・追試する

Q14.

研究に利用したことがある（参考にしたたり再分析した経験がある）公開データの分野をお選び下さい。

* (複数選択)

- (1) 工学
- (2) 天文学
- (3) 化学
- (4) 物理学
- (5) 地球科学
- (6) 数学
- (7) コンピュータサイエンス
- (8) 農学
- (9) 生物科学
- (10) 医学
- (11) 心理学
- (12) 社会科学
- (13) 人文学
- (14) その他 [具体的に]

- (15) 公開データを入手しても、研究に利用したことはない [【→Q17へ】](#)

次へ

Q15. 公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい。

(複数選択)

- (1) サーチエンジン (Googleなど)
- (2) データ情報のデータベース (Data Citation Indexなど)
- (3) 学術機関のリポジトリ・データアーカイブ (大学やNASAのリポジトリなど)
- (4) 出版社や学術雑誌のサイト (Elsevier, Wileyなど)
- (5) 特定のデータリポジトリ [具体的に]

- (6) 論文や学術記事の参考文献
- (7) データジャーナル (簡易なデータ記述とデータへのリンクを掲載した雑誌)
- (8) 政府・機関・出版社などの広報, ニュースレター
- (9) ブログや一般的なSNS (Facebook, Twitterなど)
- (10) 学術系SNS (Mendeley, ResearchGateなど)
- (11) アラートサービス (RSSなど)
- (12) メーリングリスト
- (13) 研究者や同僚に尋ねる／教えてもらう
- (14) その他 [具体的に]

Q16. 公開データを利用する際に、信頼性の判断基準としている項目をお選び下さい。

(複数選択)

- (1) 著者情報 (所属機関, 職位など)
- (2) 掲載リポジトリ
- (3) 引用数 (そのデータを引用した文献数)
- (4) オルトメトリクス (SNSやブログの言及数)
- (5) ダウンロード数

- (6) 研究手法の確かさ
- (7) そのデータを用いた論文
- (8) データにつけられた説明（Readmeファイルなど）
- (9) その他〔具体的に〕

Q17. 今後、利用してみたいと思う公開データの分野をお選び下さい。

* (複数選択)

- (1) 工学
- (2) 天文学
- (3) 化学
- (4) 物理学
- (5) 地球科学
- (6) 数学
- (7) コンピュータサイエンス
- (8) 農学
- (9) 生物科学
- (10) 医学
- (11) 心理学
- (12) 社会科学
- (13) 人文学
- (14) その他

- (15) 他の研究者の公開データを利用したいとは思わない

次へ

5. 研究データの公開について

これまでに、ご自身の研究データをインターネットで公開したご経験についてお伺いします。

(再掲) 本調査における「データ」とは、研究のために収集・作成・観測した**デジタルデータ**を指します。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しません。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、内容も限定しません。

Q18. これまでに、研究データを以下の方法で公開したご経験はありますか？

* (複数選択)

- (1) 個人や研究室のウェブサイト
- (2) 所属機関のリポジトリ・データアーカイブ
- (3) 特定分野のリポジトリ・データアーカイブ (DDBJやICPSRなど)
- (4) 無料のデータ共有サービス (figshare, zenodoなど)
- (5) コード共有サービス (GitHubなど)
- (6) 論文の補足資料 (supplementary materials)
- (7) 学術系SNS (Mendeley, ResearchGateなど)
- (8) その他 [具体的に]

- (9) 公開したことはない **【→Q20へ】**
- (10) わからない **【→Q20へ】**
- (11) 研究にデジタルデータは用いない **【→Q35へ】**

次へ

Q19. 研究データを公開した理由として、あてはまるものをお選び下さい。

* (複数選択)

- (1) 他の研究者からのリクエストに応じて
- (2) 論文を投稿した雑誌のポリシー（投稿規定）だから
- (3) 助成機関のポリシー（助成条件）だから
- (4) 所属機関のポリシーだから
- (5) 分野・コミュニティの規範だから
- (6) 業績になる場合があるから
- (7) 研究成果を広く認知してもらいたいから
- (8) 科学研究や成果実装を推進したいから
- (9) オープンデータに貢献したいから
- (10) その他【具体的に】

- (11) あてはまるものはない

【すべて→Q22へ】

次へ

Q20. 研究データを公開していない理由として、あてはまるものをお選び下さい。

* (複数選択)

- (1) 資金が必要だから
- (2) 時間が必要だから
- (3) ニーズがないと思うから
- (4) 論文を投稿した雑誌のポリシー（投稿条件）ではないから
- (5) 助成機関のポリシー（助成条件）ではないから
- (6) 所属機関にデータ公開ポリシーがないから
- (7) 分野・コミュニティで推奨されていないから
- (8) 業績にならないから
- (9) リポジトリなどの公開手段がないから
- (10) その他【具体的に】

- (11) あてはまるものはない [【→Q22へ】](#)

次へ

Q21. Q20の理由が解決された場合、研究データを公開したいと思われますか？

*

- (1) はい
- (2) いいえ
- (3) わからない

Q22. 所属機関では、研究データの保存期間は定められていますか？

- (1) 定められている
- (2) 定められていない
- (3) わからない

次へ

6. 最近の研究のためのデータについて

ここからは、研究データの整備や公開に必要な支援について検討するための質問です。

研究ごとに扱うデータの種類や量が異なる場合が多いと考えられるため、「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」についてお伺いします（以下、「**カレントデータ**」と記します）。

データ公開のご経験がない場合は、カレントデータのうち、論文等の根拠となる部分を公開すると想定してお答え下さい。

【例：研究1件のために質問紙調査とインタビュー調査を実施した場合】

<カレントデータ>

- ・質問紙の回答を入力したスプレッドシート
- ・回答を分析するためのRのコード
- ・インタビューを録音した音声データ
- ・インタビューを書き起こしたテキストデータなど

<データ公開>

質問紙調査の回答データを第三者が再利用できるように、調査概要の説明を作成、回答を入力したスプレッドシートから個人情報を削除、項目に見出しをつけるといった処理を行った上で、質問紙や分析のためのコードとともにリポジトリに登録する。

Q23.

カレントデータの総量は、およそどれくらいでしたか？ 論文などには使用しなかったデータも含めてあてはまる単位をお選び下さい。

- (1) MB（メガバイト）以下
- (2) GB（ギガバイト）
- (3) TB（テラバイト）
- (4) PB（ペタバイト）以上
- (5) わからない

Q24. ご自身以外で、カレントデータの所有権をもつ人・組織をお選び下さい。

(複数選択)

- (1) 共同研究者
- (2) 所属機関
- (3) 所属機関以外の研究機関
- (4) 所属機関以外の企業
- (5) 研究助成機関
- (6) その他〔具体的に〕

- (7) なし（自分のみ）
- (8) わからない

Q25. カレントデータに含まれている情報をお選び下さい。

(複数選択)

- (1) 個人情報
- (2) 健康情報（遺伝情報、医療情報）
- (3) 企業・商業上の機密情報
- (4) その他、守秘義務がある情報
- (5) なし
- (6) わからない

Q26.

カレントデータの保存期間として適切だと思われる年数を数字で入力して下さい。
（保存の必要はないと考える場合は「0」、永久保存の場合は「999」と入力して下さい）

年数

 年

0文字 (数字)

Q27. 現在までに、カレントデータ（データの一部）を公開しましたか？

*

- (1) はい：以下の質問は、（カッコ）をお読みの上お答え下さい
- (2) いいえ：以下の質問は、公開することを想定してお答え下さい

(3) わからない：以下の質問は、公開することを想定してお答え下さい

Q28.

カレントデータを**整備・公開しようとする場合**、次の資源は十分に整っていますか（いましたか）？

*

わからない 0 不十分 1 やや不十分 2 ほぼ充分 3 充分 4

研究中のデータ用ストレージ

データ公開用のリポジトリ

研究終了後のデータ保存用ストレージ

データの整備・公開のための時間

データの整備・公開のための人材

わからない 0 不十分 1 やや不十分 2 ほぼ充分 3 充分 4

データの整備・公開のための資金

Q29.

カレントデータを**論文の発表前**に公開しようとする場合、次の点は問題や懸案となりますか（なりましたか）？

*

わからない 0 問題ではない 1 あまり問題ではない 2 やや問題である 3 問題である 4

商用利用される可能性

機密・プライバシー情報

誤解や誤用の可能性

引用せずに利用される可能性

研究の誤りを発見される可能性

わからない 0 問題ではない 1 あまり問題ではない 2 やや問題である 3 問題である 4

公開したデータを使って自分より先に論文を出版される可能性

Q30.

カレントデータを公開した場合（公開したカレントデータは）、ご自身と同じ分野、すなわちQ1で選択した分野の研究者の多くが理解できると思われますか？

*

- (1) できると思う：データにつけた説明（Readmeファイルなど）を読めば理解できると思う
- (2) やや難しいと思う
- (3) 難しいと思う
- (4) わからない

Q31.

カレントデータを公開した場合（公開したカレントデータは）、異分野の研究者の多くが理解できると思われますか？

*

- (1) できると思う：データにつけた説明（Readmeファイルなど）を読めば理解でき
ると思う
- (2) やや難しいと思う
- (3) 難しいと思う
- (4) わからない

Q32.

カレントデータを公開した場合（公開したカレントデータについて）、ご関心がある項目をお選び
下さい。

(複数選択)

- (1) 閲覧数
- (2) ダウンロード数
- (3) リンク数（別のウェブサイトからのリンク）
- (4) オルトメトリクス（SNSやブログの言及数）
- (5) 引用数（データを利用した文献数）
- (6) その他〔具体的に〕

- (7) あてはまるものはない（関心なし）

Q33.

カレントデータの整備や公開について、より詳しく知りたいと思われる項目をお選び下さい。

(複数選択)

- (1) 適切なデータ形式
- (2) 適切なリポジトリ
- (3) 適切なメタデータ（作成者やキーワードなど、データを検索するために付与する
データの要約情報）標準（例：ジャパンリンクセンターによる[研究データのメタデ
ータ仕様 \(PDF\)](#)）
- (4) データのバージョン管理方法
- (5) データのバックアップ方法
- (6) データの安全な管理方法（セキュリティ）
- (7) データに含まれる機密情報の処理
- (8) 知的財産権やライセンス
- (9) その他〔具体的に〕

(10) あてはまるものはない

Q34.

カレントデータの整備や公開を、ご自身や共同研究者にかわって図書館員やデータキュレーターなどの第三者が行う場合、分野の知識や専門性が必要であると考えられる項目をお選び下さい。カレントデータが多様な場合は、もっとも難しいと考えられるデータについてお答え下さい。

(複数選択)

- (1) 適切なデータ形式への変換
- (2) 適切なリポジトリの選択
- (3) 適切なメタデータ（作成者やキーワードなど、データの要約情報）標準の選択
- (4) メタデータの作成
- (5) データを再利用しやすいように整える
- (6) 機関のリポジトリによるデータ公開
- (7) データを異分野の研究者に紹介する
- (8) その他【具体的に】

(9) あてはまるものはない（いずれも専門性を必要としない）

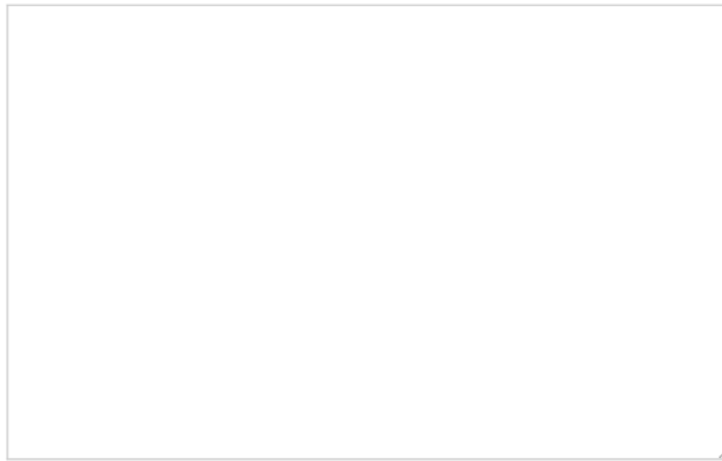
次へ

7. 自由回答

Q35.

論文やデータの公開，オープンサイエンス，および調査に関するご意見やご感想がありましたら，ご自由にお書き下さい。

(最大1,000文字)



0文字

質問は以上です。

修正する場合は、ブラウザの「←」（戻る）で質問に戻ってください。
終了する場合は、最後に「送信」を押して下さい。

送信

(2) 単純集計結果

回答者の属性

表 9 所属機関

| | 回答者数 | 比率 |
|------|-------|--------|
| 大学 | 844 | 60.0% |
| 企業 | 330 | 23.5% |
| 公的機関 | 188 | 13.4% |
| 団体 | 36 | 2.6% |
| その他 | 8 | 0.6% |
| 合計 | 1,406 | 100.0% |

表 10 年齢

| | 回答者数 | 比率 |
|---------|-------|--------|
| 21-30 歳 | 30 | 2.1% |
| 31-40 歳 | 649 | 46.2% |
| 41-50 歳 | 414 | 29.4% |
| 51-60 歳 | 188 | 13.4% |
| 61-70 歳 | 125 | 8.9% |
| 合計 | 1,406 | 100.0% |

1. 研究分野について

Q1. ご自身の研究分野に最も近いものをお選び下さい。 *

表 11 回答者の研究分野

| 研究分野 | 人数 | 比率 |
|------------------------------|-------|--------|
| 工学 | 589 | 41.9% |
| 生物科学 | 180 | 12.8% |
| 化学 | 134 | 9.5% |
| 農学 | 123 | 8.7% |
| 医学 | 117 | 8.3% |
| 計算機科学 | 71 | 5.0% |
| 物理学 | 63 | 4.5% |
| 地球科学 | 55 | 3.9% |
| 社会科学 | 23 | 1.6% |
| 数学 | 18 | 1.3% |
| 心理学 | 16 | 1.1% |
| 天文学 | 4 | 0.3% |
| 人文学 | 2 | 0.1% |
| その他 | 3 | 0.2% |
| 研究を行ったことはない(口頭発表や論文出版の経験はない) | 8 | 0.6% |
| 合計 | 1,406 | 100.0% |

2. 学術論文について

Q2. ご自身の論文について、あてはまるものをお選び下さい。 *

表 12 オープンアクセスの状況

| | | |
|-----------|-------|--------|
| OA の論文がある | 991 | 70.9% |
| OA の論文はない | 347 | 24.8% |
| わからない | 60 | 4.3% |
| 合計 | 1,398 | 100.0% |

Q3. 論文をオープンアクセスにした理由として、あてはまるものをお選び下さい。[複数選択可] *

表 13 論文を OA にした理由 (n=991, 複数回答)

| 理由 | 人数 | 比率 |
|-----------------------|-----|-------|
| 論文を投稿した雑誌がオープンアクセスだから | 810 | 81.7% |
| 研究成果を広く認知してもらいたいから | 465 | 46.9% |
| 科学研究や成果実装を推進したいから | 144 | 14.5% |
| 他の研究者からのリクエストに応じて | 101 | 10.2% |
| オープンアクセスに貢献したいから | 94 | 9.5% |
| 所属機関のポリシーだから | 82 | 8.3% |
| 分野・コミュニティの規範だから | 29 | 2.9% |
| 助成機関のポリシー(助成条件)だから | 20 | 2.0% |
| その他 | 24 | 2.4% |
| あてはまるものはない | 1 | 0.1% |

Q4. 論文がオープンアクセスではない理由として、あてはまるものをお選び下さい。[複数選択可] *

表 14 論文が OA ではない理由 (n=407, 複数回答)

| 理由 | 人数 | 比率 |
|------------------------|-----|-------|
| 投稿したい雑誌がオープンアクセスではないから | 246 | 60.4% |
| 資金が必要だから | 161 | 39.6% |
| 所属機関にオープンアクセス方針がないから | 47 | 11.5% |
| 分野・コミュニティで推奨されていないから | 13 | 3.2% |
| 時間が必要だから | 10 | 2.5% |
| リポジトリなどの公開手段がないから | 9 | 2.2% |
| ニーズがないと思うから | 9 | 2.2% |
| 助成機関のポリシー(助成条件)ではないから | 4 | 1.0% |
| その他 | 24 | 5.9% |
| あてはまるものはない | 28 | 6.9% |

Q5. Q4 の理由が解決された場合、論文をオープンアクセスにしたいと思われますか? *

表 15 オープンアクセスの意思

| | | |
|-------|-----|--------|
| はい | 296 | 78.1% |
| いいえ | 18 | 4.7% |
| わからない | 65 | 17.2% |
| 合計 | 379 | 100.0% |

Q6. 研究に利用したことがある（アイデアの参考にしたり引用した経験がある）論文の分野をお選び下さい。[複数選択可] *

表 16 研究に利用したことがある分野 (n=1,398, 複数回答)

| 研究分野 | 人数 | 比率 |
|-------|-----|-------|
| 工学 | 847 | 60.6% |
| 生物科学 | 511 | 36.6% |
| 化学 | 473 | 33.8% |
| 物理学 | 431 | 30.8% |
| 医学 | 381 | 27.3% |
| 農学 | 287 | 20.5% |
| 計算機科学 | 268 | 19.2% |
| 地球科学 | 149 | 10.7% |
| 数学 | 140 | 10.0% |
| 社会科学 | 106 | 7.6% |
| 心理学 | 71 | 5.1% |
| 天文学 | 33 | 2.4% |
| 人文学 | 24 | 1.7% |
| その他 | 5 | 0.4% |

Q7. 論文を探す際に、よく利用する検索ツールや情報源をお選び下さい。[複数選択可]

表 17 論文の探索ツール (n=1,398, 複数回答)

| 論文の探索ツール | 人数 | 比率 |
|--------------------------------------------------------|-------|-------|
| サーチエンジン(Google, Google Scholar など) | 1,164 | 83.3% |
| 論文情報のデータベース(Web of Science, Scopus, CiNii Articles など) | 990 | 70.8% |
| 論文や学術記事の参考文献 | 855 | 61.2% |
| 出版社や学術雑誌のサイト(Elsevier, Wiley など) | 683 | 48.9% |
| 論文やプレプリントのサーバ(PubMed Central, arXiv, J-Stage など) | 675 | 48.3% |
| 研究者や同僚に尋ねる／教えてもらう | 469 | 33.5% |
| 学術機関のリポジトリ・アーカイブ(大学や NASA の機関リポジトリなど) | 220 | 15.7% |
| 学術系 SNS(Mendeley, ResearchGate など) | 134 | 9.6% |
| アラートサービス(RSS など) | 104 | 7.4% |
| 政府・機関・出版社などの広報, ニュースレター | 98 | 7.0% |
| ブログや一般的な SNS(Facebook, Twitter など) | 72 | 5.2% |
| メーリングリスト | 64 | 4.6% |
| その他 | 11 | 0.8% |
| 無回答 | 4 | 0.3% |

Q8. 論文を利用する際に、その信頼性の判断基準としている項目をお選び下さい。[複数選択可]

表 18 論文の信頼性の判断基準 (n=1,398, 複数選択)

| 判断基準 | 人数 | 比率 |
|------------------------|-------|-------|
| 掲載雑誌 | 1,219 | 87.2% |
| 著者情報(所属機関, 職位など) | 932 | 66.7% |
| 研究手法の確かさ | 901 | 64.4% |
| 引用数(その論文を引用した文献数) | 479 | 34.3% |
| 抄録 | 227 | 16.2% |
| ダウンロード数 | 42 | 3.0% |
| オルトメトリクス(SNS やブログの言及数) | 5 | 0.4% |
| その他 | 59 | 4.2% |
| 無回答 | 4 | 0.3% |

3. 研究データの提供について

Q9. 共同研究者を除く他の研究者にデータを提供したご経験はありますか？

表 19 データ提供経験

| | | |
|--------|-------|--------|
| よくある | 104 | 7.4% |
| たまにある | 511 | 36.6% |
| ほとんどない | 469 | 33.5% |
| まったくない | 303 | 21.7% |
| わからない | 10 | 0.7% |
| 無回答 | 1 | 0.1% |
| 合計 | 1,398 | 100.0% |

Q10. 共同研究者を除く他の研究者からデータの提供を受けたご経験はありますか？

表 20 データ被提供経験

| | | |
|--------|-------|--------|
| よくある | 75 | 5.4% |
| たまにある | 466 | 33.3% |
| ほとんどない | 490 | 35.1% |
| まったくない | 359 | 25.7% |
| わからない | 7 | 0.5% |
| 無回答 | 1 | 0.1% |
| 合計 | 1,398 | 100.0% |

4. 公開データの利用について

Q11. これまでに、公開データを以下の公開先から入手したご経験はありますか？〔複数選択可〕 *

表 21 公開データの入手経験

| | | |
|------------|-------|--------|
| ある(チェックあり) | 1,060 | 75.8% |
| ない | 318 | 22.7% |
| わからない | 20 | 1.4% |
| 合計 | 1,398 | 100.0% |

表 22 公開データの入手方法 (n=1,060, 複数回答)

| 入手方法 | 人数 | 比率 |
|----------------------------------------|-----|-------|
| 個人や研究室のウェブサイト | 687 | 64.8% |
| 論文の補足資料 | 563 | 53.1% |
| 学術機関のリポジトリ・データアーカイブ(大学や NASA のリポジトリなど) | 530 | 50.0% |
| 特定分野のリポジトリ・データアーカイブ(DDBJ や ICPSR など) | 245 | 23.1% |
| 学術系 SNS(Mendeley, ResearchGate など) | 157 | 14.8% |
| コード共有サービス(GitHub など) | 113 | 10.7% |
| データ共有サービス(figshare, zenodo など) | 34 | 3.2% |
| その他 | 13 | 1.2% |

Q12. 公開データの入手にあたって、問題だと感じたことがある項目をお選び下さい。[複数選択可] *

表 23 公開データ入手における問題の有無

| | | |
|-------------|-------|--------|
| ある(チェックあり) | 846 | 79.8% |
| 問題を感じたことはない | 214 | 20.2% |
| 合計 | 1,060 | 100.0% |

表 24 公開データを手に入れる際の障壁 (n=846, 複数回答)

| 公開データの入手障壁 | 人数 | 比率 |
|--------------------------|-----|-------|
| 利用料金が必要 | 365 | 43.1% |
| 利用者登録が必要 | 282 | 33.3% |
| 利用条件がよくわからない | 280 | 33.1% |
| データごとに品質が異なる | 272 | 32.2% |
| データごとにフォーマットが異なる | 237 | 28.0% |
| 利用したいデータへのアクセス方法がよくわからない | 195 | 23.0% |
| 著作者情報がよくわからない | 172 | 20.3% |
| データの解釈や再利用方法がよくわからない | 140 | 16.5% |
| リクエストから入手までに時間がかかる | 139 | 16.4% |
| 最新のデータを手に入れない | 120 | 14.2% |
| その他 | 18 | 2.1% |

Q13. 入手した公開データの利用目的について、あてはまるものをそれぞれお選び下さい。
*

表 25 公開データの利用目的 (n=1,060)

| | 行う | | たまに行う | | 行わない | |
|-------------------|-----|-------|-------|-------|------|-------|
| | 人数 | 比率 | 人数 | 比率 | 人数 | 比率 |
| 研究のアイデアや仮説の参考にする | 378 | 35.7% | 589 | 55.6% | 93 | 8.8% |
| 再分析・再利用して自身の研究を行う | 176 | 16.6% | 410 | 38.7% | 474 | 44.7% |
| 研究を再現・追試する | 90 | 8.5% | 399 | 37.6% | 571 | 53.9% |

Q14. 研究に利用したことがある（参考にしたり再分析した経験がある）公開データの分野をお選び下さい。[複数選択可] *

表 26 公開データを研究に利用した経験の有無

| | | |
|--------------------------|-------|--------|
| ある（チェックあり） | 1,034 | 97.5% |
| 公開データを入手しても、研究に利用したことはない | 26 | 2.5% |
| 合計 | 1,060 | 100.0% |

表 27 研究に利用したことがある公開データの分野（n=1,034，複数回答）

| 研究分野 | 人数 | 比率 |
|-------|-----|-------|
| 工学 | 518 | 50.1% |
| 生物科学 | 348 | 33.7% |
| 化学 | 271 | 26.2% |
| 物理学 | 208 | 20.1% |
| 医学 | 202 | 19.5% |
| 計算機科学 | 163 | 15.8% |
| 農学 | 160 | 15.5% |
| 地球科学 | 91 | 8.8% |
| 数学 | 48 | 4.6% |
| 社会科学 | 39 | 3.8% |
| 心理学 | 21 | 2.0% |
| 天文学 | 19 | 1.8% |
| 人文学 | 14 | 1.4% |
| その他 | 2 | 0.2% |

Q15. 公開データを探す際に、よく利用する検索ツールや情報源をお選び下さい。[複数選択可]

表 28 公開データの探索ツール (n=1,034, 複数回答)

| データの探索ツール | 人数 | 比率 |
|--------------------------------------|-----|-------|
| サーチエンジン(Google など) | 864 | 83.6% |
| 論文や学術記事の参考文献 | 734 | 71.0% |
| 出版社や学術雑誌のサイト(Elsevier, Wiley など) | 391 | 37.8% |
| 研究者や同僚に尋ねる／教えてもらう | 375 | 36.3% |
| 学術機関のリポジトリ・データアーカイブ | 338 | 32.7% |
| データ情報のデータベース(Data Citation Index など) | 187 | 18.1% |
| 学術系 SNS(Mendeley, ResearchGate など) | 97 | 9.4% |
| データジャーナル(簡易なデータ記述とデータへのリンクを掲載した雑誌) | 79 | 7.6% |
| 政府・機関・出版社などの広報, ニュースレター | 77 | 7.4% |
| メーリングリスト | 36 | 3.5% |
| ブログや一般的な SNS | 30 | 2.9% |
| 特定のデータリポジトリ | 27 | 2.6% |
| アラートサービス(RSS など) | 19 | 1.8% |
| その他 | 5 | 0.5% |
| 無回答 | 1 | 0.1% |

Q16. 公開データを利用する際に、信頼性の判断基準としている項目をお選び下さい。[複数選択可]

表 29 公開データの信頼性の判断基準 (n=1,034, 複数回答)

| 判断基準 | 人数 | 比率 |
|----------------------------|-----|-------|
| 著者情報(所属機関, 職位など) | 733 | 70.9% |
| 研究手法の確かさ | 649 | 62.8% |
| そのデータを用いた論文 | 602 | 58.2% |
| 掲載リポジトリ | 264 | 25.5% |
| 引用数(そのデータを引用した文献数) | 237 | 22.9% |
| データにつけられた説明(Readme ファイルなど) | 221 | 21.4% |
| ダウンロード数 | 54 | 5.2% |
| オルトメトリクス(SNS やブログの言及数) | 9 | 0.9% |
| その他 | 20 | 1.9% |
| 無回答 | 9 | 0.9% |

Q17. 今後、利用してみたいと思う公開データの分野をお選び下さい。[複数選択可] *

表 30 利用したい公開データの有無

| | | |
|-------------------------|-------|--------|
| ある（チェックあり） | 1,357 | 97.1% |
| 他の研究者の公開データを利用したいとは思わない | 41 | 2.9% |
| 合計 | 1,398 | 100.0% |

表 31 今後、利用してみたいと思う公開データの分野（n=1,357、複数回答）

| 研究分野 | 人数 | 比率 |
|-------|-----|-------|
| 工学 | 767 | 56.5% |
| 生物科学 | 515 | 38.0% |
| 化学 | 434 | 32.0% |
| 医学 | 400 | 29.5% |
| 物理学 | 394 | 29.0% |
| 計算機科学 | 342 | 25.2% |
| 農学 | 303 | 22.3% |
| 地球科学 | 171 | 12.6% |
| 社会科学 | 143 | 10.5% |
| 数学 | 136 | 10.0% |
| 心理学 | 89 | 6.6% |
| 天文学 | 42 | 3.1% |
| 人文学 | 39 | 2.9% |
| その他 | 8 | 0.6% |

5. 研究データの公開について

Q18. これまでに、研究データを以下の方法で公開したご経験はありますか？〔複数選択可〕 *

表 32 データ公開経験の有無

| | | |
|------------|-------|--------|
| ある(チェックあり) | 713 | 51.0% |
| ない | 656 | 46.9% |
| わからない | 27 | 1.9% |
| データは用いない | 2 | 0.1% |
| 合計 | 1,398 | 100.0% |

表 33 データを公開した方法 (n=713, 複数回答)

| 公開方法 | 人数 | 比率 |
|--------------------------------------|-----|-------|
| 個人や研究室のウェブサイト | 362 | 50.8% |
| 論文の補足資料(supplementary materials) | 335 | 47.0% |
| 所属機関のリポジトリ・データアーカイブ | 244 | 34.2% |
| 特定分野のリポジトリ・データアーカイブ(DDBJ や ICPSR など) | 117 | 16.4% |
| 学術系 SNS(Mendeley, ResearchGate など) | 70 | 9.8% |
| コード共有サービス(GitHub など) | 24 | 3.4% |
| データ共有サービス(figshare, zenodo など) | 16 | 2.2% |
| その他 | 18 | 2.5% |

Q19. 研究データを公開した理由として、あてはまるものをお選び下さい。[複数選択可]

*

表 34 データを公開した理由 (n=713, 複数回答)

| 理由 | 人数 | 比率 |
|-------------------------|-----|-------|
| 研究成果を広く認知してもらいたいから | 417 | 58.5% |
| 論文を投稿した雑誌のポリシー(投稿規定)だから | 312 | 43.8% |
| 科学研究や成果実装を推進したいから | 186 | 26.1% |
| 他の研究者からのリクエストに応じて | 181 | 25.4% |
| 所属機関のポリシーだから | 169 | 23.7% |
| オープンデータに貢献したいから | 75 | 10.5% |
| 分野・コミュニティの規範だから | 68 | 9.5% |
| 業績になる場合があるから | 66 | 9.3% |
| 助成機関のポリシー(助成条件)だから | 45 | 6.3% |
| その他 | 14 | 2.0% |
| あてはまるものはない | 5 | 0.7% |

Q20. 研究データを公開していない理由として、あてはまるものをお選び下さい。[複数選択可] *

表 35 データを公開していない理由 (n=683, 複数回答)

| 理由 | 人数 | 比率 |
|----------------------------|-----|-------|
| 論文を投稿した雑誌のポリシー(投稿条件)ではないから | 180 | 26.4% |
| 時間が必要だから | 177 | 25.9% |
| 所属機関にデータ公開ポリシーがないから | 153 | 22.4% |
| ニーズがないと思うから | 151 | 22.1% |
| 業績にならないから | 148 | 21.7% |
| リポジトリなどの公開手段がないから | 61 | 8.9% |
| 資金が必要だから | 52 | 7.6% |
| 助成機関のポリシー(助成条件)ではないから | 42 | 6.1% |
| 分野・コミュニティで推奨されていないから | 29 | 4.2% |
| その他 | 61 | 8.9% |
| あてはまるものはない | 88 | 12.9% |

Q21. Q20 の理由が解決された場合、研究データを公開したいと思われませんか？ *

表 36 データ公開意思

| | | |
|-------|-----|--------|
| はい | 169 | 28.4% |
| いいえ | 130 | 21.8% |
| わからない | 296 | 49.7% |
| 合計 | 595 | 100.0% |

※データを公開していない回答者とわからない回答者(683名)のうち、Q20で「あてはまるものはない」を選択しなかった回答者 88 名を除いた。

Q22. 所属機関では、研究データの保存期間は定められていますか？

表 37 データ保存期間規定の有無

| | | |
|----------|-------|--------|
| 定められている | 672 | 48.1% |
| 定められていない | 380 | 27.2% |
| わからない | 344 | 24.6% |
| 合計 | 1,396 | 100.0% |

6. 最近の研究のためのデータについて

Q23. カレントデータの総量は、およそどれくらいでしたか？ 論文などには使用しなかったデータも含めてあてはまる単位をお選びください。

表 38 カレントデータの量

| | | |
|-------------|-------|--------|
| MB(メガバイト)以下 | 262 | 18.8% |
| GB(ギガバイト) | 781 | 55.9% |
| TB(テラバイト) | 188 | 13.5% |
| PB(ペタバイト)以上 | 4 | 0.3% |
| MB(メガバイト)以下 | 158 | 11.3% |
| 無回答 | 3 | 0.2% |
| 合計 | 1,396 | 100.0% |

Q24. ご自身以外で、カレントデータの所有権をもつ人・組織をお選び下さい。[複数選択可]

表 39 カレントデータの所有権をもつ人・組織の有無

| | | |
|------------|-------|--------|
| ある(チェックあり) | 1,166 | 83.5% |
| なし(自分のみ) | 115 | 8.2% |
| わからない | 114 | 8.2% |
| 無回答 | 1 | 0.1% |
| 合計 | 1,396 | 100.0% |

表 40 カレントデータの所有権をもつ人・機関 (n=1,166, 複数回答)

| 所有権をもつ人・機関 | 人数 | 比率 |
|-------------|-----|-------|
| 共同研究者 | 957 | 82.1% |
| 所属機関 | 679 | 58.2% |
| 所属機関以外の研究機関 | 111 | 9.5% |
| 所属機関以外の企業 | 75 | 6.4% |
| 研究助成機関 | 67 | 5.7% |
| その他 | 7 | 0.6% |

Q25. カレントデータには、以下の情報が含まれていますか？ [複数選択可]

表 41 カレントデータの機密情報の有無

| | | |
|------------|-------|--------|
| ある(チェックあり) | 683 | 48.9% |
| なし | 493 | 35.3% |
| わからない | 207 | 14.8% |
| 無回答 | 13 | 0.9% |
| 合計 | 1,396 | 100.0% |

表 42 カレントデータに含まれる機密情報 (n=683, 複数回答)

| 機密情報 | 人数 | 比率 |
|------------------|-----|-------|
| 個人情報 | 293 | 42.9% |
| 企業・商業上の機密情報 | 288 | 42.2% |
| 健康情報(遺伝情報, 医療情報) | 76 | 11.1% |
| その他, 守秘義務がある情報 | 399 | 58.4% |

Q26. カレントデータは何年くらい保存する必要があると思われますか？ 年数を数字で入力して下さい。（保存の必要はないと考える場合は「0」、永久保存の場合は「999」と入力して下さい）

表 43 カレントデータの望ましい保存期間

| | | |
|---------------|-------|--------|
| 0 年(保存の必要はない) | 42 | 3.0% |
| 1 年 | 6 | 0.4% |
| 2 年 | 14 | 1.0% |
| 3 年 | 78 | 5.6% |
| 4 年 | 2 | 0.1% |
| 5 年 | 422 | 30.2% |
| 6 年 | 0 | 0.0% |
| 7 年 | 11 | 0.8% |
| 8 年 | 4 | 0.3% |
| 9 年 | 1 | 0.1% |
| 10 年 | 432 | 30.9% |
| 12 年 | 1 | 0.1% |
| 15 年 | 23 | 1.6% |
| 20 年 | 81 | 5.8% |
| 25 年 | 5 | 0.4% |
| 30 年 | 29 | 2.1% |
| 40 年 | 1 | 0.1% |
| 50 年 | 23 | 1.6% |
| 99 年 | 2 | 0.1% |
| 100 年 | 5 | 0.4% |
| 永久保存 | 190 | 13.6% |
| 無回答 | 24 | 1.7% |
| 合計 | 1,396 | 100.0% |

Q27. 現在までに、カレントデータ（データの一部）を公開しましたか？ *

表 44 カレントデータ公開の有無

| | | |
|-------|-------|--------|
| はい | 254 | 18.2% |
| いいえ | 1,005 | 72.0% |
| わからない | 137 | 9.8% |
| 合計 | 1,396 | 100.0% |

Q28. カレントデータを管理・公開しようとする場合、次の資源は十分に整っていますか（いましたか）？ *

表 45 カレントデータの管理資源の充実度（n=1,396）

| | 0 | 1 | 2 | 3 | 4 |
|-------------------|-----|-----|-----|-----|-----|
| 研究中のデータ用ストレージ | 197 | 429 | 238 | 387 | 145 |
| データ公開用のリポジトリ | 378 | 534 | 228 | 195 | 61 |
| 研究終了後のデータ保存用ストレージ | 201 | 521 | 277 | 289 | 108 |
| データの整備・公開のための時間 | 239 | 782 | 258 | 96 | 21 |
| データの整備・公開のための人材 | 216 | 922 | 183 | 61 | 14 |
| データの整備・公開のための資金 | 262 | 834 | 201 | 77 | 22 |

0=わからない, 1=不十分, 2=やや不十分, 3=ほぼ充分, 4=充分

Q29. カレントデータを論文の発表前に公開しようとする場合、次の点は問題（懸案）となりますか（なりましたか）？ *

表 46 カレントデータを公開する際の障壁（n=1,396）

| | 0 | 1 | 2 | 3 | 4 |
|------------------------------|-----|-----|-----|-----|-----|
| 商用利用される可能性 | 106 | 148 | 166 | 330 | 646 |
| 機密・プライバシー情報 | 85 | 156 | 151 | 309 | 695 |
| 誤解や誤用の可能性 | 82 | 100 | 181 | 449 | 584 |
| 引用せずに利用される可能性 | 67 | 26 | 78 | 321 | 904 |
| 研究の誤りを発見される可能性 | 90 | 526 | 576 | 122 | 82 |
| 公開したデータを使って自分より先に論文を出版される可能性 | 66 | 50 | 99 | 295 | 886 |

0=わからない, 1=問題ではない, 2=あまり問題ではない, 3=やや問題である, 4=問題である

Q30. カレントデータを公開した場合（公開したカレントデータは）、ご自身と同じ分野、すなわち Q1 で選択した分野の研究者の多くが理解できると思われますか？ *

表 47 カレントデータの理解：同じ分野

| | | |
|----------|-------|--------|
| できると思う | 838 | 60.0% |
| やや難しいと思う | 325 | 23.3% |
| 難しいと思う | 145 | 10.4% |
| わからない | 88 | 6.3% |
| 合計 | 1,396 | 100.0% |

Q31. カレントデータを公開した場合（公開したカレントデータは）、異分野の研究者の多くが理解できると思われますか？ *

表 48 カレントデータの理解：他分野

| | | |
|----------|-------|--------|
| できると思う | 184 | 13.2% |
| やや難しいと思う | 522 | 37.4% |
| 難しいと思う | 597 | 42.8% |
| わからない | 93 | 6.7% |
| 合計 | 1,396 | 100.0% |

Q32. カレントデータを公開した場合（公開したカレントデータについて）、ご関心がある項目をお選び下さい。[複数選択可]

表 49 カレントデータのメトリクスへの関心

| | | |
|------------------|-------|--------|
| チェックあり（関心あり） | 1,198 | 85.8% |
| あてはまるものはない（関心なし） | 195 | 14.0% |
| 無回答 | 3 | 0.2% |
| 合計 | 1,396 | 100.0% |

表 50 関心のあるメトリクス（n=1,198, 複数選択）

| メトリクス | 人数 | 比率 |
|------------------------|-----|-------|
| 引用数（データを利用した文献数） | 874 | 73.0% |
| 閲覧数 | 858 | 71.6% |
| ダウンロード数 | 805 | 67.2% |
| リンク数（別のウェブサイトからのリンク） | 270 | 22.5% |
| オルトメトリクス（SNS やブログの言及数） | 122 | 10.2% |
| その他 | 32 | 2.7% |

Q33. カレントデータを整備・公開する上で、より詳しく知りたいと思われる項目をお選び下さい [複数選択可]

表 51 カレントデータの管理方法への関心

| | | |
|---------------------|-------|--------|
| チェックあり（詳しく知りたい項目あり） | 1,257 | 90.0% |
| あてはまるものはない | 137 | 9.8% |
| 無回答 | 2 | 0.1% |
| 合計 | 1,396 | 100.0% |

表 52 カレントデータの整備・公開に関して詳しく知りたい項目 (n=1,257, 複数選択)

| 項目 | 人数 | 比率 |
|----------------------|-----|-------|
| 知的財産権やライセンス | 869 | 69.1% |
| データの安全な管理方法 (セキュリティ) | 836 | 66.5% |
| 適切なデータ形式 | 754 | 60.0% |
| データに含まれる機密情報の処理 | 697 | 55.4% |
| 適切なリポジトリ | 610 | 48.5% |
| 適切なメタデータ*標準 | 494 | 39.3% |
| データのバックアップ方法 | 410 | 32.6% |
| データのバージョン管理方法 | 407 | 32.4% |
| その他 | 16 | 1.3% |

*作成者やキーワードなど、データを検索するために付与するデータの要約情報(例:ジャパンリンクセンターによる[研究データのメタデータ仕様\(PDF\)](#))

Q34. カレントデータの整備や公開を、ご自身や共同研究者にかわって図書館員やデータキュレーターなどの第三者が行う場合、分野の知識や専門性が必要であると考えられる項目をお選び下さい。カレントデータが多様な場合は、もっとも難しいと考えられるデータについてお答え下さい。[複数選択可]

表 53 第三者によるカレントデータの整備や公開における専門性

| | | |
|-----------------------------|-------|--------|
| チェックあり (専門性を必要とする) | 1,302 | 93.3% |
| あてはまるものはない (いずれも専門性を必要としない) | 93 | 6.7% |
| 無回答 | 1 | 0.1% |
| 合計 | 1,396 | 100.0% |

表 54 データの整備や公開の際に専門性を必要とする項目 (n=1,302, 複数選択)

| 項目 | 人数 | 比率 |
|-------------------|-----|-------|
| 適切なデータ形式への変換 | 781 | 60.0% |
| データを再利用しやすいように整える | 769 | 59.1% |
| 適切なメタデータ標準の選択 | 656 | 50.4% |
| 適切なリポジトリの選択 | 531 | 40.8% |
| メタデータの作成 | 445 | 34.2% |
| データを異分野の研究者に紹介する | 426 | 32.7% |
| 機関のリポジトリによるデータ公開 | 347 | 26.7% |
| その他 | 31 | 2.4% |

※Q35の自由回答は別途 Web 上に電子付録として掲載する(<http://doi.org/10.15108/rm268>)。

調査資料-268

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

2017 年 12 月

文部科学省 科学技術・学術政策研究所

科学技術予測センター

池内 有為, 林 和弘, 赤池 伸一

〒100-0013

東京都千代田区霞が関 3-2-2 中央合同庁舎第 7 号館 東館 16 階

TEL : 03-3581-0605 FAX : 03-3503-3996

A Survey on Open Research Data and Open Access

December 2017

Ui IKEUCHI, Kazuhiro HAYASHI and Shinichi AKAIKE

Science and Technology Foresight Center

National Institute of Science and Technology Policy (NISTEP)

Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan

<http://doi.org/10.15108/rm268>



<http://www.nistep.go.jp>