

## 原著論文

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

### An Analysis of Open Research Data Practice, Drivers, and Barriers in Japan

池 内 有 為  
*Ui IKEUCHI*

#### *Résumé*

**【Purpose】** This study contributes to the development of open science policies and appropriate support frameworks by clarifying the status and perceptions of open research data, and analyzing drivers and barriers among researchers in Japan.

**【Methods】** From November to December 2016, 1,983 researchers from the Science and Technology Experts Network operated by the National Institute of Science and Technology Policy took part in a survey. The responses were analyzed for participants' age and disciplines and compared with open access status and previous research results. The reasons for and experience with open data were investigated; the characteristics of researchers who have open data experience were clarified using binomial regression analysis. Reasons for non-disclosure of data and degree of barriers to disclosure were investigated. The characteristics of fields with high and low open data rates compared with the strength of journal open data policies were also clarified.

**【Results】** Responses were received from 1,398 researchers (70.5%) who belonged to universities, companies, and public institutions/organizations. Of the respondents, 51.0% had experience publishing data; the main reason for disclosure was the improvement of cognition of research results and journal policies. The reasons for non-disclosure varied, including the lack of a journal policy, and only 28.4% respondents would publicize their data if the issues were resolved. Of the researchers, 75.8% had obtained published data and 97.1% exhibited a willingness to use published data. Respondents recognize that there is a lack of human resources, funds, time, and data repositories, which are required resources for data publishing. There were strong concerns, particularly among younger researchers belonging to universities, regarding career risks, such as using data

---

池内有為：筑波大学大学院図書館情報メディア研究科，文部科学省科学技術・学術政策研究所，〒305-8550 茨城県つくば市春日1-2

Ui IKEUCHI: Graduate School of Library, Information and Media Studies, University of Tsukuba; National Institute of Science and Technology Policy (NISTEP), Ministry of Education, Culture, Sports, Science and Technology (MEXT), 1-2, Kasuga, Tsukuba-shi, Ibaraki, 305-8550, Japan  
e-mail: ikeuchi.ui@gmail.com

受付日：2017年4月21日 受理日：2017年12月11日

without citation or priority loss. There were also strong concerns regarding confidentiality, commercial use, and misuse.

The resulting analysis suggests that important issues for promoting data publishing include improving the availability of open research data and providing support staff to set aside time for researchers. To solve career risk concerns and the lack of incentives, it is important to make the publishing of data a recognized research achievement.

- I. はじめに
  - A. オープンサイエンスの進展と課題
  - B. 用語の定義
- II. 先行研究と研究課題
  - A. データ公開に関する実態調査
  - B. データ公開の理由と障壁
  - C. データ公開と関連のある属性
  - D. 研究課題
- III. 調査・分析方法
  - A. 調査対象
  - B. 質問項目の設定
  - C. 分析方法
- IV. 結果
  - A. 回答率と回答者の属性
  - B. 日本の研究者によるデータ公開の状況と認識
  - C. データ公開者および分野による特徴の分析
- V. 考察
  - A. データ公開の状況と必要性
  - B. データ公開の推進要因と公開データの入手環境の整備
  - C. データ公開の阻害要因と対策
- VI. おわりに

## I. はじめに

### A. オープンサイエンスの進展と課題

研究成果への自由なアクセスを可能にするオープンカラシップ (open scholarship) は、公的資金を財源とする研究に対する社会的要求と、ICT 技術の発展を背景として急速に広がっている<sup>1)</sup>。論文のオープンアクセス (以下、「OA」と記す) に次いで研究の根拠となるデータの公開も大きな潮流となっており<sup>2)</sup>、両者をあわせてオープンサイエンスと呼ぶことが増えている<sup>3)</sup>。

2007年に経済開発協力機構(OECD)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」<sup>4)</sup>を、2013年にG8科学技術大臣会合が科学研究データ公開の基本原則を含む共同声明<sup>5)</sup>を公開したことを契機として、各国・地域の政府組織や助成機関、学術雑誌、研究機関、学会などによってデータ公開が推進され、義務化の動きも広がっている<sup>6)</sup>。その背景には、研究効率の向上、研究結果の検証による透明性と質の向上、技術革新の加速、経済への波及効果、地球規模の課題への効果的な取り組み、共同研究の推進、市民科学の促進、教育での活用といった狙いがある<sup>7)</sup>。

日本においてもオープンサイエンスに関する議論が重ねられており、2015年の内閣府「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」報告書<sup>8)</sup>(以下、内閣府報告書)、2016年の「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)<sup>9)</sup>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)<sup>10)</sup>において、国外の動向や実現に向けた課題が示されている。そして第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”<sup>11)</sup>と述べられている。

データ公開に関する政策はOA政策とあわせて検討されているものの<sup>5)</sup>、OAと比較して成熟度が低いと指摘されており<sup>7)</sup>、国内外の政策文書において分野などの状況に応じた対応が重要であると述べられている。たとえば、内閣府報告書では“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”<sup>8)</sup>[p.14]としている。また、Schmidtら<sup>12)</sup>によるデータ公開に関する国際調査を紹介した近藤<sup>13)</sup>は、地域によって差があることに言及した上で、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”であると述べている。日本のデータ公開の状況を把握するための手段として、データの引用索引であるData Citation Index (DCI)が考えられるが、2016年現在は収録レコード数が随時増減しているため<sup>14)</sup>、調査時点によって結果が大きく異なってしまう。そこで本研究は、日本の研究者を対象とした質問紙調査を実施することによって、データ公開の状況を明らかにすることを第一の目的とする。調査にあたっては、内閣府報告書<sup>8)</sup>で指摘されている“各分野の専門家・研究者、技術者の意見”を広く収集して、データ

公開に対する認識を明らかにすることを旨とする。

また、現時点では、日本の助成機関は本格的にデータ公開を義務化していないが<sup>15)</sup>、学術雑誌によるデータ公開ポリシーは投稿者の国や地域にかかわらず適用され、経年的に増加している<sup>16)</sup>。たとえば2016年12月、Springer Nature社は同社の傘下にある出版社の600以上の雑誌に4種類のデータ共有ポリシー<sup>17)</sup>のいずれかを分野横断で採用することを発表した<sup>18)</sup>。第5期科学技術基本計画では、日本の総論文数を増やすこと、および日本の論文の10%が世界の被引用回数トップ10%論文になることを目指すとしているが<sup>11)</sup>[p.30]、データ公開に適切に対応できない場合、論文投稿の障壁となる可能性がある。データ公開の要求が厳しい学術雑誌に掲載されている論文ほどデータ公開率が高い<sup>19)</sup>ことから、データ公開が要求されているにもかかわらず、公開が遅れている分野への支援は重要であると考えられる。また、データを公開している論文は、公開していない論文と比較して被引用回数が多いという調査結果も複数ある<sup>20)~23)</sup>。このように、データ公開は論文投稿においても重要な要素となりつつあるが、多くの分野の研究者にとって馴染みのない研究プロセスであり負担が大きいと予想される。さらに、日本の研究者は研究に割ける時間が減少していること<sup>24)</sup>、および研究開発費や論文の生産性が低迷していることも指摘されている<sup>25)</sup>。従って、データ公開における問題を明らかにした上で、その解決を図ることは喫緊の課題であるといえよう。そこで本研究は、データ公開の推進要因と阻害要因を分析することを第二の目的とする。研究を通じてデータ公開の状況、および推進要因と阻害要因を明らかにすることによって、オープンサイエンス政策や適切な支援体制の構築に資することを目指す。

## B. 用語の定義

ここでは用語の定義を行う。以下の定義は質問紙にも使用し、意味がより明確になるように具体例を付けた。

本研究における「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。

III章B節で述べるプレテストの結果、研究ごとに扱うデータの種類や量が異なる研究者が存在することを確認した。そこで一部の質問では「論文などの成果を发表済みの、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」と定義した上で回答を求めた。

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）ことを指す。特定の人以外はアクセスできない状態とする。

データの「公開」とは、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）」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する。いわゆるゴールドOAとグリーンOAの別を問わず、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、雑誌など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などを含む。

## II. 先行研究と研究課題

本章では、先行研究の概要と研究課題を示す。A節では、データ公開に関する国内外の実態調査について述べる。B節では、先行研究で明らかにされたデータ公開の理由と障壁を整理する。C節では、データ公開に関連する年齢などの属性についての知見を紹介する。D節では、以上の知見に基づく研究課題を示す。なお、先行研究では「共有（sharing）」という表現で本研究における「公

開」や「提供」を示している場合があった。そこで本章では、「公開」の意味を含む文献は原文通りに「共有」と記し、「提供」のみを扱う文献は取り上げないこととした。

### A. データ公開に関する実態調査

研究者を対象としたデータ公開の実態に関する質問紙調査は、2000年ごろから実施されてきた。しかし、2009～10年と2013～14年の2度にわたって調査を実施したTenopirら<sup>26)</sup>によれば、データ公開に影響を与える要因は経年的に変化している。また、近年、データ公開に関する技術やネットワーク環境は大きく変化している<sup>6)</sup>ことから、本節では主に最近の調査を中心に紹介する<sup>12), 26)～36)</sup>。

第1表に各調査の概要を示す。特定の国の調査の場合は「主な回答者」に国名を記載し、国際調査の場合は記載しなかった。「回答数」では、日本の研究者数を括弧に入れて示した。主な公開理由と障壁については次節で述べる。

データの公開率について、Tenopirら（2009/2010）<sup>27), 37)</sup>の調査で「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36%であり、ノルウェー研究評議会のための報告書<sup>32)</sup>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28%であった。Ferguson<sup>33)</sup>の報告では「データを一般に公開している」回答者は52%であり、Berghmansら<sup>36)</sup>の報告では、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<sup>38)</sup>。調査によって尋ね方が異なるものの、最近の調査ほど公開率が高い傾向にあった。データ公開が一般的な分野やデータ公開に関心のあるコミュニティを対象とした調査では、相対的に公開率が高かった。たとえば、生物多様性分野の主要な3誌の著者を対象としたHuangら<sup>28)</sup>の調査では、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%であり、California Digital LibraryのKratzら<sup>39)</sup>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

第1表 データ公開に関する質問紙調査

調査主体	調査年	主な回答者	回答数	主な公開理由	主な障壁
Tenopir ら <sup>27)</sup> /DataONE	2009 /2010	環境科学, 生態学等の研究者	1,329	[引用が重要]	時間, 資金, 権利, 公開する場所
Huang ら <sup>28)</sup>	2011	生物多様性分野の研究者	372	雑誌のデータアーカイブポリシー	出版前の公開, 公開の困難さ
Kim ら <sup>29)</sup>	2012 /2013	米国のSTEM分野の研究者	1,317	分野の規範, 科学的利他性, 雑誌の圧力, キャリアベネフィット	公開のための労力や時間についての認識
Kim ら <sup>30)</sup>	2012 /2013	米国のSTEM分野の研究者	1,298	データ共有に対する態度, データリポジトリの可用性	
Kim ら <sup>31)</sup>	2012 /2013	米国の生物学分野の研究者	608	助成機関と雑誌の圧力, データリポジトリとメタデータ標準の可用性	
Danvad 社 <sup>32)</sup>	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	[研究の進展, 研究の倫理的な義務]	時間, インフラ, インセンティブ, 将来の出版可能性
Tenopir ら <sup>26)</sup> /DataONE	2013 /2014	環境科学, 生態学等の研究者	1,015	[引用が重要]	出版前の公開, 誤解や誤用
Ferguson <sup>33)</sup> /Wiley 社	2014	多分野の研究者	2,250+ (不明)	研究コミュニティ, 研究のインパクト, 公益	知的財産権・機密, 所属機関・助成機関の要求がない
Schmidt ら <sup>12)</sup> /Belmont Forum	2014	地球科学・環境学等の研究者, 技術員等	1,253 (26)	科学研究と成果実装の加速, 研究成果の認知向上	出版前の公開, 法的制約, 信用や正当な評価の喪失
Fecher ら <sup>34)</sup>	2014	独国の自然科学, 社会科学等の研究者	1,564	評価やデータ引用による“評判経済”	他の研究者が先に出版すること
小野ら <sup>35)</sup>	2015	日本の地球科学・環境学等の研究者	38 (38)	科学の発展に貢献, 分野間連携の促進	時間, 提供作業, 自分が優先的に使う
Berghmans ら <sup>36)</sup>	2016	多分野の研究者	1,162	[分野で重視, 出版者に提供したので論文とともに公開]	[十分な訓練を受けていない, 自分分野では功績にならない]

(日本の研究者の回答数) [データ公開に対する認識]

共有経験を有していた。また、日本の研究者について、Ferguson<sup>33)</sup>の報告では「データを一般に公開している」回答者は44%であり、小野ら<sup>35)</sup>の調査ではデータを「任意のユーザーに公開している」回答者は8%であった。

データの公開方法について、各調査で回答率が高かった順に3件ずつ挙げると、Huang ら<sup>28)</sup>の調査では論文の補足資料(52%)、パブリックデータベース(38%)、機関のサイト(25%)であり、Ferguson<sup>33)</sup>の報告では論文の補足資料(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<sup>39)</sup>の調査では直接コンタクト(87%)、リポジトリ(54%)、雑誌のウェブサイト(37%)、Berghmans ら<sup>36)</sup>の報告では論文の補足資料(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。これらの結果か

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は一定数行われているものの、まだ主流とはいえないことがわかる。

## B. データ公開の理由と障壁

データ公開の推進要因や阻害要因となりうる公開理由と障壁は、インタビュー調査などの方法でも明らかにされている。そこで第1表に示した調査に加えて、研究者がデータを公開する理由と障壁を明らかにしている最近の調査、すなわちWallis ら<sup>40)</sup>やVan den Eynden ら<sup>41)</sup>の調査、Fecher ら<sup>42)</sup>のシステムティックレビュー、Nature news チーム<sup>43)</sup>のインタビュー調査、Wicherts ら<sup>44)</sup>による公開データの分析を参照して、複数の先行研究で指摘されていた項目を第2表に示す。その際、Fecher ら<sup>42)</sup>によるデータ共有の概念枠組みとKim ら<sup>29)</sup>によるモデルを参考にし

第2表 先行研究によるデータ公開の理由と障壁

		公開理由		障壁	
		制度／認識	資源	制度／認識	
外的要因	分野	学術雑誌のポリシー <sup>12)</sup> , 28), 29), 31), 40), 41) 助成機関のポリシー <sup>12)</sup> , 31), 40), 42) 規範 <sup>12)</sup> , 29), 33), 36), 41), 42)	*分野リポジトリ <sup>27)</sup> , 30), 43)		
	所属機関	所属機関のポリシー <sup>12)</sup> , 28), 42)	*機関リポジトリ <sup>27)</sup> , 30), 40), 43)	機密・プライバシー情報 <sup>33)</sup> , 42) 商用利用 <sup>12)</sup> , 42) 知的財産権 <sup>12)</sup> , 27), 33), 42)	
内的要因	研究者	業績 <sup>26)</sup> , 27), 29), 33), 34), 41) 科学の発展 <sup>12)</sup> , 32), 35), 41) 科学的利他性 <sup>12)</sup> , 29), 33), 40), 41) 研究成果の認知度向上 <sup>12)</sup> , 41) オープンデータへの貢献 <sup>12)</sup> , 41) †データの提供・被提供経験 <sup>32)</sup> , 36), 46) †公開データの利用経験 <sup>32)</sup> , 45)	*資金 <sup>27)</sup> , 46) *時間 <sup>27)</sup> , 29), 32), 35) 人材 <sup>29)</sup> , 35), 36), 41)	先取権の喪失 <sup>12)</sup> , 26), 28), 32), 34), 35), 41), 43) 誤解や誤用 <sup>12)</sup> , 26), 32), 42) 引用せずに利用される <sup>12)</sup> , 42) 研究の誤りを発見される <sup>12)</sup> , 41), 44)	

†提供を受ける・利用するだけで公開しないという結果も示されている

\*充足している場合は公開理由となる

た。以下では、まず Fecher ら<sup>42)</sup>と Kim ら<sup>29)</sup>の研究の概要を述べた後、第2表について説明する。

Fecher ら<sup>42)</sup>は、研究データ共有プロセスの概念枠組みを構築するために、98報の学術論文のシステムティックレビューと603名のデータの二次利用者を対象とした調査を行った。その結果、6つのカテゴリ、すなわちデータの提供者、研究機関、研究コミュニティ、規範、データのインフラストラクチャー、データの利用者による概念枠組みを提示し、データ共有のより良いインセンティブとなるデータ共有ポリシーが研究成果の質の向上や科学の進展に必要なであると結論づけた。

Kim ら<sup>29)</sup>は、研究者のデータ共有行動に影響を与える要因を機関による環境と個人的な動機にわけてモデル化し、それぞれの影響の強さを明らかにするために質問紙調査を行った。STEM(自然科学・技術・工学・数学)の43分野1,317名から得た回答について、マルチレベル分析を行った結果、データ共有行動と関連がみられたのは、専門分野の規範や雑誌の出版社によるプレッシャー、および個人的な動機である科学的利他性やキャリアベネフィットに関する認識であった。また、データ共有に労力と時間が必要であることを認識している研究者はデータを共有する可能性が低いことを明らかにした。

第2表では、データ公開を行う研究者の立場か

らみた主なデータ公開理由と障壁を、分野や所属機関などの外的要因と研究者の動機や懸念などの内的要因に分類した。障壁のうち、充足している場合は公開理由となる資源、たとえば分野リポジトリなどには※印を付した。

各要因について、類似の内容は統合した。たとえば、Schmidt ら<sup>12)</sup>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「誤解や誤用の懸念」とした。同様に、Tenopir ら<sup>27)</sup>による「データがコントロールできなくなる恐れ」やノルウェーの報告書<sup>32)</sup>の「研究者がデータを理解できないことへの懸念」も「誤解や誤用の懸念」とした。また、Schmidt ら<sup>12)</sup>の「信用や正当な評価を失う恐れ」や Wicherts ら<sup>44)</sup>の「再分析によってエラーや異なる結論が示される恐れ」は「研究の誤りを発見される可能性」とした。キャリアベネフィット(採用・昇進)、評価、引用などは「業績」、支援、労力などは「人材」とした。

また、個人や研究グループ内におけるデータ共有(提供・被提供)経験<sup>32)</sup>, 36)や、公開データの利用経験とデータ公開経験についての調査も行われている。Piwowar<sup>45)</sup>は遺伝子発現マイクロアレ

イデータを作成した論文の著者を対象として調査を行い、過去にデータを再利用した経験がある場合にデータを公開する傾向があることを明らかにした。ノルウェーの報告書<sup>32)</sup>では、他の研究者のデータを利用したいという要望がデータ公開の増加につながることを示唆されている。一方、Tenopirら(2009/2010)<sup>27)</sup>の調査では、「簡単にアクセスできるならば他の研究者のデータセットを使いたい」という設問に「強く同意する」または「同意する」を選んだ回答者は84%であったのに対して、データを公開している研究者は36%であった。*Science*誌<sup>46)</sup>の査読者を対象とした調査においても、回答者の76%は同僚にデータ提供を求めたことがあるにもかかわらず、80%は自身のデータを整備するための資金がないとしていた。つまり、データ利用を希望したり提供を受けるだけで自身は公開をためらう研究者が存在する可能性が示唆されている。そこで、第2表では注記(†)を付けた上で「公開理由」に掲載した。

### C. データ公開と関連のある属性

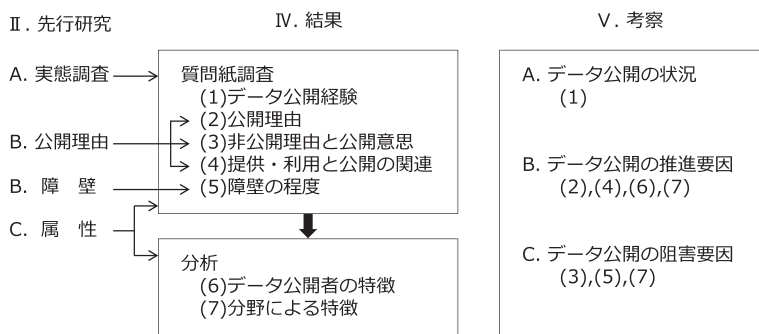
第2表に示した公開理由と障壁の程度は、年齢<sup>12), 26), 27)</sup>や分野<sup>12), 26), 27), 32), 33)</sup>などの属性によって差があることが明らかにされている。たとえば年齢について、Schmidtら<sup>12)</sup>の調査では、オープンデータを世に出す際の障壁として挙げられた10項目のうち、「データを公開する前に成果を出版したいという願望(第2表では「先取権の喪失」)」が最も重大な障壁であると認識

されており、回答者の年齢が若いほどその傾向が強かった(最も強いのは31~35歳)。Tenopirら(2013/2014)<sup>26)</sup>の調査でも、11項目の障壁のうち、「先に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。分野について、Ferguson<sup>33)</sup>は健康科学の研究者のデータ公開率は48%でありプライバシーや倫理的な懸念があるが、公益に資することが将来的なデータ共有の動機になること、その一方で人文・社会科学はデータ公開率が36%であり、研究のインパクトやビジビリティを高める場合や助成機関の要求がデータ共有の動機になることなどを報告している。

### D. 研究課題

先行研究によって明らかにされたデータ公開の状況、公開理由と障壁、および公開と関連する属性を参考として、研究目的である研究データ公開の状況とデータ公開の推進・阻害要因を明らかにするための研究課題7点を設定した。第1図に、先行研究と研究課題、研究目的の関連と論文の構成を示す。なお、ここでは研究課題の意図と先行研究の関連について述べ、具体的な質問項目や分析方法は次章に示す。

まず、(1) データ公開経験の有無を尋ね、データを公開している研究者と公開していない研究者を判別する。続いて、(2) 日本の研究者によるデータ公開の主な理由を明らかにするために、公開している研究者を対象として、先行研究の「公開理由」と「充足している場合は公開理由となる



第1図 先行研究・研究課題・研究目的の関連と論文の構成

障壁（資源）」からあてはまるものを選択してもらう。(3) 公開していない研究者には、「公開理由がないこと（たとえば、雑誌のポリシーがないこと）」を選択肢として示し、あてはまるものを選択してもらう。また、非公開理由が解決した場合にデータを公開する意思があるかどうかを尋ねることによって、先行研究で示された「公開理由」を整備・提供すればデータ公開を促すことができるかどうかを確認する。(4) 全ての回答者にデータ提供および公開データ利用経験を尋ねて、データ公開経験との関連があるかどうかを明らかにする。(5) データ公開の「障壁」は、データを公開していない研究者のみならず公開経験がある研究者にとっても問題であることが明らかにされていた。そこでデータ公開経験の有無にかかわらず、全ての回答者に「データを公開する場合」を想定して、先行研究で明らかにされたデータ公開の障壁がどの程度問題であるのかを尋ねる。

また、データ公開の状況を相対的に明らかにするために、データ公開よりも進展しているとされているOA論文<sup>7)</sup>について、公開率や理由などをデータと同様に質問して結果を比較する。先行研究で関連が示された属性（年齢、分野）による違いについても確認して、先行研究との比較や考察を行う。分野については(7)で総合的に分析する。

以上の結果によるデータ公開の推進要因と阻害要因は、それぞれ多数存在すると考えられる。そこで、(6) データを公開している研究者の特徴を分析することによって主要な推進要因の組み合わせを明らかにする。また、(7) 分野別のデータ公開率や障壁を確認した上で、特徴的な分野の推進・阻害要因を検討する。

### III. 調査・分析方法

前章で述べた研究課題を明らかにするために、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（以下、科学技術予測センター）においてWebアンケートシステムを用いた質問紙調査を実施した。プレテストは大学や企業の研究者を対象として3回実施し、それぞれ15名、8名、

12名の協力を得た。本調査の質問数は合計35問であり、内容によって7パートに分割した（付録「質問一覧」参照）。本章では、調査対象と選定理由、質問項目の設定方法、および分析方法について述べる。

#### A. 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」とした。科学技術専門家ネットワークとは、科学技術に関する動向や見解などを収集するための、産学官の研究者、技術者、マネージャーなど約2,000名の専門家集団である。メンバーは紹介によって選出されるため無作為抽出ではないが、調査時点での所属の構成比率は大学60.4%、企業22.7%、公的機関・団体16.0%、その他0.9%であり、多分野かつ幅広い年齢層の研究者が含まれることから選定した。ただし、本調査はI章B節で定義した通り、「研究の成果である論文やスライドの根拠となる」データを対象としているため、研究分野を尋ねるQ1の選択肢に「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）」を挙げて、当該回答者には自由回答のみ記入していただいた。

#### B. 質問項目の設定

本節では、まず、質問紙全体にかかわる方針について説明する。続いて、質問の構成を研究課題と対応させながら示すとともに、質問の設定方法について述べる。

##### 1. 全体の方針

データ公開の経験がない回答者や馴染みが薄い回答者が存在すると予想されたこと、データ公開の経験があっても回答者によって想定する内容が異なると予想されたことから、各セクションのはじめにI章B節の定義と具体例を示した。さらに、こうした説明部分は読み飛ばされる可能性があることから、経験の有無と方法を同一の質問で尋ねることによって、回答者ごとの認識の違いによる回答のずれを防ぐよう努めた。たとえばデー



タ公開の経験を尋ねる質問では、「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）」というように具体的なデータ公開方法を列挙して、1つ以上を選択した回答者はデータ公開経験が「ある」とみなし、あわせて排他的な選択肢として「いいえ」、「わからない」を提示してデータ公開経験の有無を判断することとした。また、プレテストで「論文もデータも公開していないと答え難い」という指摘があったことから、回答の偏りを防ぐため、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。質問の順番は、回答が比較的容易であると考えられるOA論文に関する質問から始めた。

先行研究の結果と比較を行うためには同じ質問を行うことが望ましいが、全体の質問数と回答者の疲労効果<sup>47)</sup>を抑制するために次の方針で作成した。(1) 先行研究に似たような選択肢がある場合は、第2表と同様に統合する、(2) 尺度項目の一部は複数選択方式に変更して、尺度を選択する質問を合計30項目以下とする<sup>48)</sup>。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。

プレテストにおいて、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があった。そこでデータ公開の障壁に関する質問などについては「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」と定義した上で回答を求めた。また、データ公開経験が「ない」または「わからない」を選択した回答者にも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。

分野別の分析には、調査対象者が自己申告済みのESI (Essential Science Indicators)<sup>49)</sup>の22分野を用いた。質問紙では、回答の負担を軽減するため米国科学審議会 (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、および「その他」を提示した (Q1, Q6, Q14, Q17)。

分野別の分析には、調査対象者が自己申告済みのESI (Essential Science Indicators)<sup>49)</sup>の22分野を用いた。質問紙では、回答の負担を軽減するため米国科学審議会 (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、および「その他」を提示した (Q1, Q6, Q14, Q17)。

## 2. 質問の構成と設定方法

第3表に、研究課題に対応する質問内容と選択肢の参考文献を示す。「課題」は対応する研究課題の番号、「No」は質問番号、「OA」はOA論文について尋ねた質問番号を表す。以下では、研究課題に対応する質問の目的や選択肢の設定方法、および(6)と(7)の分析の概要を述べる。なお、破線以下については本論文では言及しない。

### (1) データの公開経験

Ferguson<sup>33)</sup>、Huangら<sup>28)</sup>、Kratzら<sup>39)</sup>の調査を参考に選択肢（データ公開の方法）を作成した (Q18)。データ提供に関する選択肢（著者に直接コンタクトなど）は削除し、プレテストの結果をふまえて表現を修正した。公開データの入手経験 (Q11)、カレントデータ公開の有無 (Q27) も同じ選択肢を示した。

### (2) データの公開理由

第2表に示した公開理由を選択肢とした (Q19)。データの提供・被提供、および公開データの利用経験は公開と結びつかないとする結果も示されているため、別途(4)で尋ねた上で関連を確認する。公開理由のうち「業績になるから」について、論文の場合はOAかどうかに限らず出版すること自体が業績になる<sup>42)</sup>ため、OA論文の選択肢には含めなかった。

### (3) データの非公開理由

第2表の公開理由がないこと、および障壁のうち充足している場合は公開理由となる資源、

第3表 質問の構成

課題	No	OA	質問内容	参考文献*	資源	制度
(1)	Q18	Q2	データの公開経験	F, H, K1		
(2)	Q19	Q3	データの公開理由	第2表	○	○
(3)	Q20	Q4	データの非公開理由	第2表, T	○	○
(3)	Q21	Q5	データの公開意思			
(4)	Q9		データの提供頻度			
(4)	Q10		データの被提供頻度			
(4)	Q11		公開データの入手経験	F, H, K1		
(4)	Q12		公開データ入手における問題	S	○	
(4)	Q13		公開データの利用目的	K2, P		
(4)	Q14	Q6	利用したことがある公開データの分野	T		
(4)	Q17		今後利用してみたい公開データの分野	T		
(5)	Q24		カレントデータの所有権	第2表, O		○
(5)	Q25		カレントデータの機密情報	O	○	○
(5)	Q28		カレントデータの公開に必要な資源の状況	第2表, S, T	○	
(5)	Q29		カレントデータ公開の懸念	第2表, S, T		○
(5)	Q30		カレントデータの理解 (自分野)			
(5)	Q31		カレントデータの理解 (異分野)			
(5)	Q34		カレントデータの管理・公開の専門性	J	○	
—	Q35		自由回答			
	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	○	
	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		○	
	Q26		カレントデータの望ましい保存期間		○	○
	Q27		カレントデータ公開の有無	F, H, K1		
	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	○	

\*F=Ferguson<sup>33)</sup>, H=Huangら<sup>28)</sup>, J=Jonesら<sup>52)</sup>, K1=Kratzら<sup>39)</sup>, K2=Kratzら<sup>50)</sup>, O=OECD<sup>7)</sup>, P=Piwowar<sup>45)</sup>, S=Schmidtら<sup>12)</sup>, T=Tenopirら(2009/2010)<sup>27)</sup>およびTenopirら(2013/2014)<sup>26)</sup>

すなわち時間、資金、リポジトリを選択肢として示した(Q20)。公開理由のうち「他の研究者からのリクエストに応じて」は、Tenopirら(2009/2010)<sup>27)</sup>の選択肢を参考に「ニーズがないと思うから」とした。また、こうした非公開理由が解決された場合のデータ公開意思について尋ねた(Q21)。

#### (4) データの提供・利用状況と公開の関連

データの提供および被提供の頻度を「よくある」、「たまにある」、「ほとんどない」、「まったくない」の4件法で尋ねた(Q9, Q10)。先行研究<sup>27), 46)</sup>から、データの提供を受けるだけで自身では提供しない、あるいは公開しない回答者が存在すると仮定して、提供と被提供経験の両方を確認すること

とした。

データの入手経験は、データ公開経験と同じ選択肢で尋ねた(Q11)。また、データを入手しても研究には利用しない回答者が存在する可能性があるため、Kratzら<sup>50)</sup>の調査を参考に、入手した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」<sup>51)</sup>、「(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)。公開データ入手における問題は、Schmidtら<sup>12)</sup>のデータにアクセスする際の負担感(3件法)を参考として複数選択方式で尋ねた。Schmidtら<sup>12)</sup>の選択肢のうち、「データの収集基準が異なる」は、結果として「データごとに

品質が異なる」ことにつながると考えられること、両者の負担の度合いがほぼ同じであったことから前者のみを採用した (Q12)。公開データの利用意思是、漠然とした回答とならないように、既に研究に利用したことがある論文 (Q6) やデータ (Q14) の分野を尋ねた上で、今後、データを利用してみたい分野について複数選択方式で尋ねた (Q17)。

#### (5) 障壁の程度

データを公開する場合の障壁の程度を明らかにするために、必要な資源の状況 (Q28) と懸念 (Q29) について、それぞれの重要度を4件法で尋ねた。その際、データによって障壁の程度が異なると考えられるため、回答者がどのデータについて回答すればよいのか特定できるよう、カレントデータを対象として回答を求めた。Tenopirら<sup>26), 27)</sup>は複数選択方式で、Schmidtら<sup>12)</sup>は「重大な障壁である」、「障壁である」、「軽微な障壁である」の3件法で尋ねているが、Schmidtらの結果では、「重大な障壁である」の選択率が21%から54%まで差がみられたことから、尺度項目とすることが適していると判断した。なお、先行研究では論文の出版前にデータを公開することが想定されているが、プレテストでこうした公開の仕方を理解できないとするコメントが複数みられたため、質問文に「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)?」と記載した (Q29)。また、障壁のうち知的財産権についてはプレテストで「わからない」を選択する回答者やコメントが多数みられたため、別途カレントデータの所有権をもつ人や機関を尋ねてデータ公開経験との関連を確認した (Q24)。カレントデータの機密情報の選択肢はOECDの報告書<sup>7)</sup>を参考に「個人情報」、「企業・商業上の機密情報」、「健康情報 (遺伝情報, 医療情報)」、「その他, 守秘義務がある情報」とした (Q25)。

カレントデータの管理・公開の専門性に関する項目は、Jonesら<sup>52)</sup>による研究データ管理サービスのガイドを援用した (Q34)。また、先行研究では誤解や誤用の可能性がデータ公開における障

壁とされており、公開データを解釈可能かどうか重要であると考えられることから、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「難しいと思う」、「やや難しいと思う」、「できると思う」の3件法で尋ねた (Q30, Q31)<sup>53)</sup>。

#### (6) データ公開者の特徴

(1)~(5)の結果を用いて、データを公開している研究者の特徴を明らかにするために2項ロジスティック回帰分析を行った。

#### (7) 分野による特徴

(1)~(5)の結果を分野別に確認した上で、特徴的な分野について考察した。具体的には、先行研究において公開理由の一つとされている雑誌のポリシーの強度 (要求の厳しさ) と実際の公開状況を分野ごとに比較して、要求が高いにもかかわらずデータ公開が行われていない分野、すなわちデータ公開が論文投稿の妨げになる可能性がある分野の課題を明らかにした。また、雑誌のポリシーの強度と比較して公開が行われている、すなわち雑誌のポリシー以外の要因でデータ公開が進んでいると考えられる分野の特徴を確認した。

## C. 分析方法

集計に先立って、回答データのクリーニングを行った。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て修正した。また、その他や自由回答の記述で誤字・脱字と判断できる箇所は適宜修正した。尺度項目は、評定数値が大きいほどポジティブな回答に、小さいほどネガティブな回答に統一した。自由記述はTTM (TinyTextMiner) for Mac OSXを用いて頻出語を抽出した後、類義語の整理を行い、文脈から内容を分類してまとめた。

統計分析について、有意水準 $p$ は5%として、有意な場合は適宜 $p$ 値を示す。ソフトウェアはSPSS Version24を用いた。各質問項目について、属性 (年齢, 分野) および所属による差があるかどうかを次の手順で確認した。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。また、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。いずれも「わからない」という回答は欠損値とした。なお、カイ二乗検定の際に期待度数が5以下のセルが表全体の20%を超えることを防ぐため、合計8以下のカテゴリは除去した。

研究課題(6)の2項ロジスティック回帰分析について、目的変数はデータ公開経験の有無として、説明変数はデータ公開と関連がみられた項目から選択した。具体的な選択手順は次章で述べる。研究課題(7)の雑誌のポリシーの強度は、2014年に調査したインパクトファクター(以下、IF)の高い論文誌220誌(ESIの22分野、各10誌)の結果<sup>54)</sup>を用いた。「必須(論文の投稿・出版条件として要求する)」=4、「推奨(論文の根拠として公開することを推奨する)」=3、「受諾(データを公開したい場合は受け付ける)」=2、「なし」=1で重み付けして、分野別の公開率と比較する。両者に相関がみられた場合は強度が妥当であると判断して、強度に対して公開率が低い分野の非公開理由や障壁について考察する。

#### IV. 結果

本章では、まずA節で調査結果の概要を示した後に、B節ではデータ公開の状況と認識に関する調査結果を、C節ではデータを公開している研究者の特徴と、分野による特徴の分析結果をそれぞれ示す。

#### A. 回答率と回答者の属性

調査期間は2016年11月30日から12月9日として、11月30日に電子メールでアンケートへの協力を依頼した。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、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。調査依頼の送付数は1,983名、最終回答数は1,406名(70.9%)であった。このうち、III章A節で言及した「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名を除く1,398名(70.5%)の回答を本研究の分析対象とする。

回答者の所属と年齢層別の集計結果を第4表に示す。所属は大学(60.3%)、企業(23.2%)の比率が高く、年齢層は31~40歳(46.3%)、41~50歳(29.4%)の比率が高いが、科学技術専門家ネットワークの構成比率と比較して、特に回答率が低い所属や年齢層はなかった。

#### B. 日本の研究者によるデータ公開の状況と認識

本節では、データ公開経験、公開理由、非公開理由と解決した場合の公開意思、データの提供・利用経験、障壁の程度に関する結果を示す。その際、各質問の「その他」や「自由回答」(Q35)の記述もあわせて示す。なお、自由回答でオープンサイエンスに関する何らかの記述を行った回答者は400名であった(謝辞などを除く)。

##### 1. データ公開経験

研究データの公開経験と方法を確認するため、「これまでに、研究データを以下の方法で公開し

第4表 回答者の所属と年齢

	大学		企業		公的機関・団体		その他		合計	
	人数	比率	人数	比率	人数	比率	人数	比率	人数	比率
21-30歳	11	0.8%	15	1.1%	4	0.3%	0	0.0%	30	2.1%
31-40歳	411	29.4%	120	8.6%	114	8.2%	2	0.1%	647	46.3%
41-50歳	249	17.8%	107	7.7%	54	3.9%	1	0.1%	411	29.4%
51-60歳	103	7.4%	54	3.9%	31	2.2%	0	0.0%	188	13.4%
61-70歳	69	4.9%	28	2.0%	20	1.4%	5	0.4%	122	8.7%
合計	843	60.3%	324	23.2%	223	16.0%	8	0.6%	1,398	100.0%

第5表 データ公開の経験と方法

公開経験／公開方法	人数	比率
あり（公開方法を1つ以上選択）	713	51.0%
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%)
公開したことはない	656	46.9%
わからない	27	1.9%
研究にデジタルデータは用いない	2	0.1%
回答者数	1,398	100.0%

第6表 データ公開と OA 論文の経験

	Q18. データ公開		Q2. OA 論文	
	人数	比率	人数	比率
あり	713	51.0%	991	70.9%
なし	656	46.9%	347	24.8%
わからない	27	1.9%	60	4.3%
研究にデータは用いない*	2	0.1%		
回答者数	1,398	100.0%	1,398	100.0%

\*データのみの選択肢

たご経験はありますか？」と尋ねた（Q18）。結果を第5表に示す。質問では、7種類の公開方法を複数選択方式で示すとともに、同時に選ば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。第5表の「あり」（713名）は、公開方法を1つ以上選択した回答者数の小計である。各公開方法に付した数字は選択した人数を、括弧内の数字は「あり」を選んだ713名に対する比率を示している（以降の表でも同様に示す）。

データの公開経験がある回答者は全体の51.0%であった。所属別では大学が56.9%、公的機関・団体が55.6%、企業が32.4%であり、差がみられた（ $p < 0.001$ ）。公開方法は個人や研究室のウェブサイトや論文の補足資料が多く、所属機関や特定分野のリポジトリは公開経験がある回答者の34.2%、16.4%にとどまった。「その他」には、「プロジェクト研究のWebサイト」などの記述

があった。

データ公開と OA 論文の経験を比較すると、OA 論文が「あり」とする回答者は991名（70.9%）、「なし」は374名（24.8%）、「わからない」は60名（4.3%）であり、OA 論文の方がデータ公開よりも実践されている<sup>7)</sup>ことを数値で裏付ける結果となった（第6表）。データ公開と OA 論文両方の経験がある回答者は568名（40.6%）、いずれの経験もない回答者は214名（15.3%）であった。データの公開経験の有無と OA 論文経験の有無には関連がみられ（ $p < 0.001$ ）、データを公開した経験のある研究者は OA 論文の経験もある傾向が明らかになった。

## 2. データを公開した理由

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。結果を第7表に示す。横線の下「あてはまるものはな

第7表 データ公開とOA論文の理由（複数回答）

	Q19. データ公開			Q3. OA論文		
	順位	人数	比率	順位	人数	比率
研究成果を広く認知してもらいたいから	1	417	58.5%	2	465	46.9%
論文を投稿した雑誌のポリシーだから <sup>†</sup>	2	312	43.8%	1	810	81.7%
科学研究や成果実装を推進したいから	3	186	26.1%	3	144	14.5%
他の研究者からのリクエストに応じて	4	181	25.4%	4	101	10.2%
所属機関のポリシーだから	5	169	23.7%	6	82	8.3%
オープンデータ/OA に貢献したいから	6	75	10.5%	5	94	9.5%
分野・コミュニティの規範だから	7	68	9.5%	7	29	2.9%
業績になる場合があるから <sup>*</sup>	8	66	9.3%	—	—	—
助成機関のポリシー（助成条件）だから	9	45	6.3%	8	20	2.0%
その他		14	2.0%		24	2.4%
あてはまるものはない		5	0.7%		1	0.1%
回答者数		713			991	

<sup>†</sup>論文を投稿した雑誌のポリシー（投稿規定）だから/論文を投稿した雑誌がオープンアクセスだから

<sup>\*</sup>データのための選択肢

い」は、排他的選択肢である（以降の表でも同様に示す）。

データを公開した理由の1位は「研究成果の認知向上」（58.5%）であり、「その他」にも「次のポジションのための成果・技術アピール」といった記述がみられるなど、研究成果を発信しようとする積極的な姿勢がうかがえた。2位は「雑誌のポリシー」（43.8%）であり、より公開率が高いOA論文では1位（81.7%）であった。「その他」では、研究の信頼性の担保（2件）や結果の検証を可能にする（2件），“新たな研究者とのつながりの構築”，“データを公開すると実験設備の利用料が安くなる”，査読者や編集者に勧められたなどの記述がみられた。科学的利他性<sup>29)</sup>に関する項目，すなわち，「科学研究や成果実装の推進」は3位（26.1%），「他の研究者からのリクエストに応じて」は4位（25.4%），「オープンデータへの貢献」は6位（10.5%）であった。

7位の「分野・コミュニティの規範」（9.5%）は，OA論文の2.9%よりは選択率が高いものの1割未満にとどまり，選択率が高い分野であっても3割未満（数学28.6%，植物学・動物学25.0%，分子生物学22.6%）であった。8位の「業績になる場合があるから」（9.3%）の分野別の回答率を

算出して「分野・コミュニティの規範」との関連を確認したところ，規範であると認識されている分野では業績として評価される傾向にあった（Pearsonの相関係数は $r=0.60$ ,  $p<0.01$ ）。なお，経済学，精神医学・心理学，宇宙科学は，規範と業績ともに選択率が0%であった。一方，分野・コミュニティの規範と科学的利他性に関する項目には関連がみられず，規範であると認識されている分野であれば科学的利他性が高いとはいえなかった。

### 3. データを公開していない理由と公開意思

データの公開経験およびOA論文が「なし」または「わからない」を選択した回答者に，それぞれの理由を複数選択方式で尋ねた（Q20）。結果を第8表に示す。

いずれも1位は「雑誌のポリシーではないから」であったが，データ公開は26.4%，OA論文は60.4%と比率が異なっていた。また，公開理由の順位はほぼ同一であったが，非公開理由の順位や選択率は異なっていた。データ公開は1位から5位までの選択率が20%台であり，回答者の多くが選択した理由がなかった一方で，OA論文は1位の「雑誌のポリシーではないから」（60.4%）

第8表 データと論文の非公開理由（複数回答）

	Q20. データ			Q4. 論文		
	順位	人数	比率	順位	人数	比率
論文を投稿した雑誌のポリシーではないから <sup>†</sup>	1	180	26.4%	1	246	60.4%
時間が必要だから	2	177	25.9%	5	10	2.5%
所属機関にデータ公開/OA 方針がないから	3	153	22.4%	3	47	11.5%
ニーズがないと思うから	4	151	22.1%	6	9	2.2%
業績にならないから <sup>*</sup>	5	148	21.7%	—	—	—
リポジトリなどの公開手段がないから	6	61	8.9%	6	9	2.2%
資金が必要だから	7	52	7.6%	2	161	39.6%
助成機関のポリシー（助成条件）ではないから	8	42	6.1%	8	4	1.0%
分野・コミュニティで推奨されていないから	9	29	4.2%	4	13	3.2%
その他		61	8.9%		24	5.9%
あてはまるものはない		88	12.9%		28	6.9%
回答者数		683			407	

<sup>†</sup>論文を投稿した雑誌のポリシー（投稿条件）ではないから／投稿したい雑誌がオープンアクセスではないから

<sup>\*</sup>データのみの選択肢

と2位の「資金が必要だから」（39.6％）に集中していた。

5位の「業績にならないから」（21.7％）については、自由回答で“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”など、所属機関や助成機関による業績化や評価（5件）、引用（3件）が保証されるならばデータ公開の可能性が高まることが示唆された。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”とする意見もみられた。

「その他」には、(a) 不要（9件）、(b) 公開のインセンティブの欠如（8件）、(c) 公開するデータがない（4件）といった趣旨の記述がみられた。(a) データ公開は不要であるとする意見は、「論文や補足資料で充分である」、「必要ならば著者に直接問い合わせれば良い」、「必要性が感じられない」<sup>55)</sup>という3点に集約される。(b) 公開のインセンティブの欠如について、自由回答においても同様の指摘が18件あり、“公開した人は公開の労力が必要なの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”といった記述がみられた。インセンティブの

第9表 データ公開とOA論文の意思

	Q21. データ公開		Q5. OA 論文	
	人数	比率	人数	比率
はい	169	28.4%	296	78.1%
いいえ	130	21.8%	18	4.7%
わからない	296	49.7%	65	17.2%
回答者数	595	100.0%	379	100.0%

具体例としては、先に述べた業績や経済的利益が示された。(c) 公開するデータがないことについて、自由回答では3件の指摘があり“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”といった記述がみられた。

続いて、非公開理由を1つ以上選択した回答者に「Q20（Q4）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？（論文をオープンアクセスにしたいと思われますか？）」

第10表 データの提供・被提供経験とデータ公開率

	Q9. 提供			Q10. 被提供		
	人数	比率	公開率	人数	比率	公開率
よくある	104	7.4%	75.5%	75	5.4%	67.6%
たまにある	511	36.6%	65.4%	466	33.3%	65.6%
ほとんどない	469	33.5%	45.1%	490	35.1%	48.7%
まったくない	303	21.7%	32.1%	359	25.7%	35.7%
わからない	10	0.7%		7	0.5%	
無回答	1	0.1%		1	0.1%	
回答者数	1,398	100.0%		1,398	100.0%	

と尋ねた (Q21)。結果を第9表に示す。

OA論文は「はい」の比率が78.1%と高いが、データ公開は28.4%と低く、「わからない」が約半数(49.7%)を占めた(第9表)。そこでデータ公開について「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、偏りはみられなかった。つまり、データの非公開理由は雑誌や機関のポリシーがない、時間不足、ニーズがない、業績にならないなど多岐にわたり、かつ、解決されれば公開につながるような特定の理由がないことが明らかになった。

#### 4. 公開データの利用状況

##### a. データ提供と被提供経験

個人的なデータの提供経験がデータ公開経験と関連があるかどうかを確認するために、共同研究者を除く他の研究者へのデータの提供および被提供経験を尋ねた(Q9, Q10)。第10表にそれぞれの人数と比率、データ公開率を示す。

提供経験の方が被提供経験の頻度よりもやや高く、データの提供、被提供経験の頻度が高い回答者ほどデータ公開を行っていることが明らかになった(いずれも $p < 0.001$ )。データを専ら提供する、あるいは提供される研究者がいることも想定したが、それぞれ1名ずつしか存在しなかった。また、データの提供・被提供経験の相関は高く(Spearmanの相関係数 $r = 0.77, p < 0.001$ )、データを提供している回答者は、提供を受けてもいるという傾向が明らかになった。以上の結果から、個人的なデータのやりとりをよく行う研究者はデータを公開する傾向にあるといえる。

##### b. 公開データの入手経験と目的

公開データを入手して再利用している研究者は、自身もデータを公開しているという先行研究<sup>45)</sup>の結果を検証するために、入手経験と利用目的を尋ねた。まず、公開データの入手経験を確認するため、「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?」と複数選択方式で尋ねた(Q11)。入手先を1つ以上選択した回答者は1,060名(75.8%)であり、回答者の約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(第11表)。データの入手経験がある回答者のデータ公開率は59.1%、入手経験がない回答者の公開率は25.8%であり、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( $p < 0.001$ )。

続いて、入手した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、Kratzら<sup>50)</sup>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(第2図)。また、Piwowar<sup>45)</sup>がデータの再利用経験とデータ公開経験に関連があることを示していたのと同様に、本研究においてもデータの再利用経験、すなわち「研究の参考」、「研究に再利用」、「研究を再現」は、全てデータ公開との関連がみられた(それぞれ $p < 0.01, p < 0.01, p < 0.05$ )。

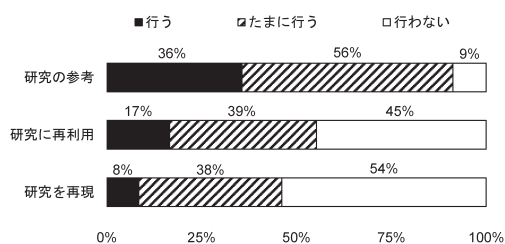
以上の結果から、公開データを入手している研究者は自身もデータを公開している傾向があること、公開データを研究に活用したり研究を再現し



第11表 公開データの入手経験とデータ公開率

入手経験/入手先	人数	比率	公開率
あり（入手先を1つ以上選択）	1,060	75.8%	59.1%
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	530	(50.0%)	
特定分野のリポジトリ・データアーカイブ	245	(23.1%)	
学術系 SNS	157	(14.8%)	
コード共有サービス	113	(10.7%)	
データ共有サービス	34	(3.2%)	
その他	13	(1.2%)	
なし	318	22.7%	25.8%
わからない	20	1.4%	
回答者数	1,398	100.0%	

\*リポジトリ等の例示は省略（第5表と同じ）



第2図 公開データの利用目的と頻度（回答者数 1,060名）

たりしている研究者はデータを公開している傾向があることがわかった。

#### c. 利用したことがある公開データの分野と利用してみたい分野

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、13分野を示して複数選択方式で尋ねた。具体的には、利用経験がある回答者（1,060名）に「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」と尋ね（Q14）、全ての回答者（1,398名）に「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」と尋ねた（Q17）。比較のため、全ての回答者に論文の利用状況も尋ねた（Q6）。

今後、利用してみたい公開データの分野について、1つ以上の分野を選択した回答者は97.1%にのぼった。また、回答者一人あたりの平均選択

数は2.79であり、データを利用したことがある分野（2.03）や論文（2.67）よりも有意に多かった。先行研究<sup>27)</sup>と同様に、研究者の多くに公開データを利用したいという意思がみられた。

#### d. 公開データ入手における問題

公開データの入手における問題点を複数選択方式で尋ねたところ（Q12）、1つ以上選択した回答者は約8割（79.8%）にのぼった（第12表）。利用料金や利用者登録、利用条件、データの品質などが主な問題であり、「その他」では「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件）についての記述がみられた。

#### 5. データ公開の障壁の程度

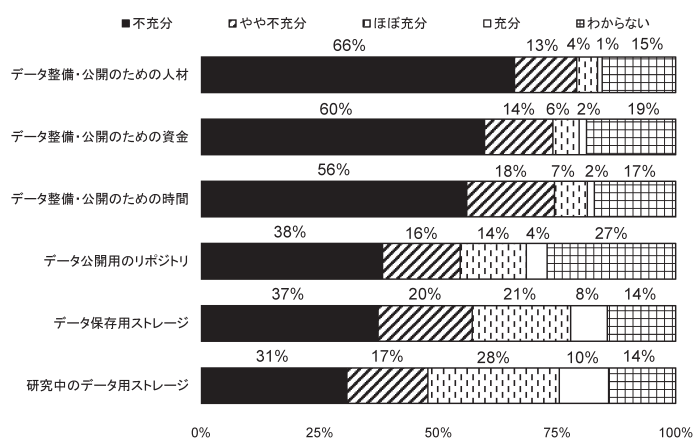
データ公開の障壁の程度を明らかにするため、研究にデータを用いる回答者を対象として資源の充足状況（Q28）や懸念（Q29）について尋ね、データ公開経験の有無との関連を確認した。なお、Q28とQ29は尺度項目がそれぞれ6問ずつ続くため、Cronbachのアルファ係数 $\alpha$ を算出したところ、Q28は $\alpha = 0.86$ 、Q29は $\alpha = 0.73$ であり、いずれの回答も内の一貫性を有していることが確認された。

##### a. データを公開する場合の資源の状況

カレントデータを整備し、公開する場合の資源の充足状況を「不十分」、「やや不十分」、「ほぼ充

第12表 公開データを入手する際の問題

問題	人数	比率
あり（問題を1つ以上選択）	846	79.8%
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%)
最新のデータを入手できない	120	(14.2%)
その他	18	(2.1%)
問題を感じたことはない	214	20.2%
回答者数	1,060	100.0%



第3図 カレントデータの管理・公開資源の充足度（回答者数1,396名）

分]、「充分」の4件法で尋ねた（Q28）。結果を「不十分」の選択率が高い順に示す（第3図）。なお、「データ保存用ストレージ」は、質問紙では「研究終了後のデータ保存用ストレージ」と記した。

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」についても、「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%と高く、

「充分」と「ほぼ充分」をあわせても18.3%にとどまった。

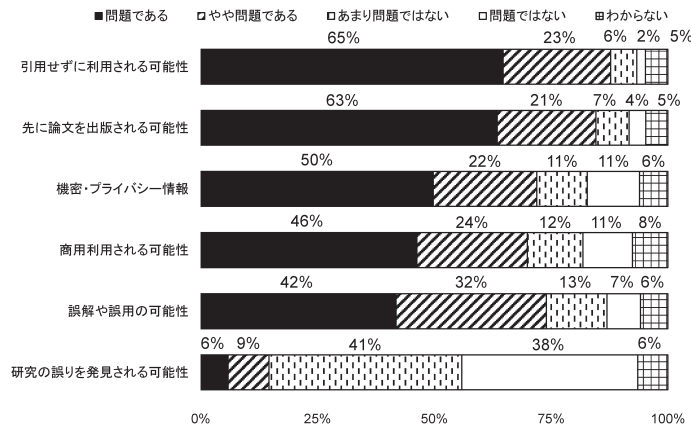
データ公開経験の有無との関連は、人材、資金、時間、公開用のリポジトリにみられた（すべて $p < 0.001$ ）。つまりデータを公開している研究者は、これらの資源の充足度が比較的高い傾向にあることがわかった。

#### b. データの整備や公開における専門性

人的支援の可能性を探るため、「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

第13表 データの整備や公開の際の専門性

専門性	人数	比率
あり (1つ以上選択)	1,302	93.3%
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%)
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%
回答者数	1,396	100.0%



第4図 カレントデータを公開する場合の懸念 (回答者数 1,396名)

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 (Q34)。回答者の93.3%が1つ以上の項目を選択していたが、選択率が低い項目、すなわち「機関のリポジトリによるデータ公開」(26.7%)や「データを異分野の研究者に紹介する」(32.7%)、「メタデータの作成」(34.2%)は、専門性の必要度がそれほど高くないという点において、相対的に支援や人材育成に取り組みやすいことが示唆された (第13表)。

「その他」には“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”などの記述があり、自由回答では、“サポートスタッ

フがないと継続的に公開することは不可能と思われる”という記述もみられた。一方で、第三者による支援は不可能・困難 (11件)、“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。

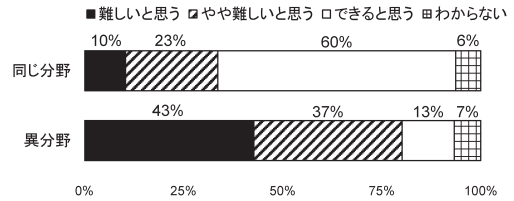
#### c. データを公開する場合の懸念

カレントデータを公開する場合の懸念を「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 (Q29)。結果を「問題である」の選択率が高い順に示す (第4図)。なお、「先に論文を出版される可能性」は、質問紙では「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」と記した。全体的に「問題である」と認識している回答者が多いが、「研究の誤りを発見される可能性」だけは「問題である」と「やや問題である」をあわせても 14.6%であった。

最も懸念が強い「引用せずに利用される可能性」および「先に論文を出版される可能性」は、年齢による差がみられた。前者について「問題である」の選択率が最も高いのは 36~40 歳 (68.7%) であり ( $p<0.05$ )、後者について「問題である」の選択率が最も高いのは 31~35 歳 (70.3%)、次いで 36~40 歳 (68.1%) であった ( $p<0.01$ )。また、所属別では大学の研究者が「問題である」を選ぶ比率が有意に高く (それぞれ 70.9%, 69.2%, いずれも  $p<0.05$ )、キャリア形成期の大学の研究者による懸念が強いことが明らかになった。自由回答では、引用せずに利用される可能性について“盗用”, “剽窃”, “無断利用”, “フリーライダー”といった表現による懸念が示された。論文の出版前の公開については、特許 (17 件), および先取権, 業績評価, 採用・昇進などのキャリアとの関連から否定的な見解が示され (10 件), “絶対にありえない”, “自殺行為である”など強い表現もみられた。論文出版後の公開であっても, “論文に掲載されたデータは非常に再公開しづらいため, (中略) 厳しい就職・プロモーション環境の中ではマイナスの要因が強い”と抵抗を示す記述もみられる (2 件)。対策として, 自由回答ではタイムスタンプなど公開日情報の管理 (2 件) や“オリジナリティの担保”が必要であるとの記述がみられた。

第 4 図の結果を所属別に確認すると, 「機密・プライバシー情報」と「商用利用される可能性」は, 企業の回答者の懸念が強く, 「問題である」, 「やや問題である」の合計は, それぞれ 84.9%と 86.0%であり, 自由回答では特許のため公開は困難であるという意見が多数みられた。「商用利用される可能性」に対しては, 大学や公的機関・団体の回答者の懸念も強い (それぞれ 74.2%, 67.1%)。「誤解や誤用の可能性」に関連する自由回答では, 単なる解釈の違いではなく“悪意の



第 5 図 カレントデータの理解: 同じ分野と異分野 (回答者数 1,396 名)

ある引用”, “揚げ足取り”, “あらさがし”, “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”など利用者の態度や不適切な批判を懸念する記述がみられた。

データ公開経験の有無については, 「機密・プライバシー情報」, 「商用利用される可能性」と関連がみられた (いずれも  $p<0.01$ )。つまりデータを公開している研究者は, 機密・プライバシー情報や商用利用される可能性に関する懸念が弱い傾向にある。

また, 先行研究では知的財産権が懸念となっていたことから, カレントデータの所有権が回答者以外にあるかどうかと, その種類を複数選択方式で尋ねた (Q24)。回答者 1,396 名のうち, 83.5%が所有権をもつ人・組織を 1 つ以上選択しており<sup>56)</sup>, 「なし (自分のみ)」は 8.2%, 「わからない」も 8.2%, 無回答は 0.1%であった。なお, データ公開経験との関連はみられなかった。

#### d. データの理解の難しさ

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 3 件法で尋ねた (Q30, Q31)。その結果, 同じ分野の研究者は 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 (第 5 図)。また, 異分野の研究者による理解の難しさとデータ公開経験の有無には関連がみられ, 難しいと考えている研究者ほどデータを公開していない傾向にあった ( $p<0.05$ )。

第14表 データ公開経験の有無と関連のある項目

説明変数（候補）
① OA 論文
② 公開データの入手経験
②' 公開データの利用目的: ②'-1 研究の参考, ②'-2 研究に再利用, ②'-3 研究を再現
③ データ提供経験
④ データを整備・公開するための時間
⑤ 商用利用される可能性
⑥ 異分野の研究者による理解
⑦ 所属 (3=大学, 2=公的機関・団体, 1=企業) ※データ公開経験ありの回答率が高い順序
除外した項目
・ データ被提供経験 (③と相関あり $r=0.77, p<0.01$ )
・ データ公開用のリポジトリ (④と相関あり $r=0.49, p<0.01$ )
・ データを整備・公開するための人材 (④と相関あり $r=0.71, p<0.01$ )
・ データを整備・公開するための資金 (④と相関あり $r=0.56, p<0.01$ )
・ 機密・プライバシー情報 (⑤と相関あり $r=0.60, p<0.01$ )

### C. データ公開者および分野による特徴の分析

#### 1. データを公開している研究者の特徴

前節の結果により、データの公開経験の有無に関連のある15項目が明らかになった。このうち、どのような要因の組み合わせをもつ研究者がデータを公開しているのかを明らかにするために、経験の有無を目的変数として2項ロジスティック回帰分析を行った。

まず、15項目から説明変数の候補を選択した(第14表)。多重共線性が生じないよう項目同士の相関分析を行い、0.4以上の相関がみられる項目はデータ公開経験との関連がより高い方の項目を残して除外した(5項目)。公開データの利用目的は、公開データの入手経験がある回答者のみ尋ねているため、片方ずつ回帰式に投入して、あてはまりの良い方を採用することとした。また、線形回帰分析によってVIF値が高い項目がないことを確認した。完全分離の状態にある項目もなかった。

こうして得られた7項目(②を含む)と9項目(②'を含む)を説明変数として、変数増加法による分析を行った。順序尺度は最小の値(「まったくない」などネガティブな選択肢)を参照カテゴリとした。ステップワイズにおける有意確率を、投入=0.05、除去=0.10として実施した結果、「OA論文」、「公開データの入手経験」、「デー

タ提供経験」、「データを整備・公開するための時間」、「所属」の5変数による回帰式が得られた。しかし、各変数の欠損値によって全体の欠損値が大きくなるため、改めて5変数を用いた分析を行った。第15表に5変数を用いた結果を示す(データ公開経験「あり」=577、「なし」=498)。

以上の結果から、研究課題(6)データ公開経験をもつ研究者の特徴は、OA論文、公開データの入手経験、データの提供経験があり、データを整備・公開するための時間が比較的充分であり、大学や公的機関・団体に所属していると予測できる。回帰式によるデータ公開経験の有無の予測正解率は、あり=75.6%、なし=64.1%、全体=70.2%であった。

#### 2. 分野による特徴

第16表に分野別の回答者数と主な結果を示す。排列はデータ公開率が高い順とした。表に示した質問項目のうち、データ公開、OA論文、データ入手経験、機密ありの比率は分野による差がみられた(それぞれ $p<0.01$ ,  $p<0.001$ ,  $p<0.01$ ,  $p<0.001$ )。データ公開意思是回答者数が少ないため検定できなかった。

III章C節に示した方法で算出した分野別の雑誌のポリシーの強度( $x$ )と、本調査でデータ公開経験がある回答者の比率( $y_1$ )との相関分析

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

第15表 データ公開に影響を与える要因

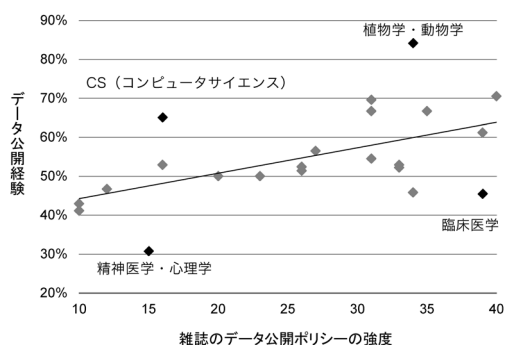
	偏回帰係数	Wald $\chi^2$	有意確率
OA論文(あり)	0.605	13.422	0.000
公開データの入手経験(あり)	1.227	46.404	0.000
データ提供経験(まったくない)		58.696	0.000
データ提供経験(ほとんどない)	0.431	4.767	0.029
データ提供経験(たまにある)	1.258	41.351	0.000
データ提供経験(よくある)	1.572	26.381	0.000
時間(不十分)		54.572	0.000
時間(やや不十分)	0.873	24.685	0.000
時間(ほぼ充分)	1.710	33.524	0.000
時間(充分)	1.494	6.341	0.012
所属(企業)		27.762	0.000
所属(公的機関・団体)	0.824	12.23	0.000
所属(大学)	0.998	27.65	0.000
定数	-3.136	118.644	0.000

第16表 分野別結果

分野	回答数	データ公開	OA論文	データ公開意思	雑誌ポリシー	データ入手経験	機密あり
植物学・動物学	19	84.2%	84.2%	33.3%	34	89.5%	5.3%
分子生物学・遺伝学	44	70.5%	77.3%	30.0%	40	97.7%	43.2%
環境科学・生態学	46	69.6%	84.8%	25.0%	31	73.9%	37.8%
微生物学	48	66.7%	72.9%	15.4%	35	85.4%	50.0%
免疫学	12	66.7%	75.0%	0.0%	31	75.0%	75.0%
CS	63	65.1%	74.6%	35.3%	16	85.7%	47.5%
生物学・生化学	85	61.2%	70.6%	19.4%	39	81.2%	45.9%
農学	46	56.5%	78.3%	15.8%	27	80.4%	32.6%
複合領域	66	54.5%	77.3%	14.8%	31	75.8%	47.0%
化学	121	52.9%	57.0%	28.3%	33	76.9%	48.3%
物理学	51	52.9%	88.2%	27.3%	16	74.5%	47.1%
地球科学	42	52.4%	85.7%	47.1%	26	90.5%	36.6%
神経科学	46	52.2%	78.3%	35.0%	33	82.6%	34.8%
薬理学・毒物学	35	51.4%	80.0%	13.3%	26	71.4%	58.8%
社会科学	10	50.0%	80.0%	0.0%	20	80.0%	30.0%
宇宙科学	8	50.0%	75.0%	66.7%	23	87.5%	25.0%
数学	15	46.7%	73.3%	0.0%	12	73.3%	20.0%
材料科学	118	45.8%	69.5%	33.9%	34	77.1%	60.2%
臨床医学	22	45.5%	72.7%	33.3%	39	50.0%	68.2%
経済学	7	42.9%	85.7%	0.0%	10	71.4%	42.9%
工学	467	41.1%	63.8%	32.2%	10	70.0%	55.0%
精神医学・心理学	13	30.8%	100.0%	33.3%	15	46.2%	69.2%
全体		51.2%	70.9%	28.7%		76.0%	49.3%
回答者数	1,384	1,384	1,384	586		1,384	1,369

を行ったところ、正の相関がみられた (Pearsonの相関係数  $r=0.53$ ,  $p<0.01$ )。つまり雑誌によるデータ公開ポリシーが強い分野の研究者ほど、データを公開している傾向が確認された。また、

$x$ が「雑誌のポリシーの強度」を示す説明変数として妥当であるかどうかを確認するために、データ公開経験の理由として雑誌のポリシーを選択した回答者の比率 ( $y_2$ ) やデータ公開の方法として



第6図 分野別データ公開経験とポリシーの強度

論文の補足資料を選択した回答者の比率 ( $y_3$ ) との相関分析を行ったところ、いずれも正の相関がみられた ( $y_2$  は  $r=0.67$ ,  $p<0.01$ ,  $y_3$  は  $r=0.74$ ,  $p<0.01$ )。

続いて、ポリシーの強度と比較してデータ公開経験がある回答者の比率が低い分野、すなわち回帰式  $y_1=0.0065x+0.3772$  の外れ値を確認すると、臨床医学と精神医学・心理学であった(第6図)。たとえば、臨床医学でIFが高い10誌のうち、9誌はデータ公開を出版の条件として要求しており、1誌はデータ公開を推奨しているのに対して、データ公開経験がある回答者は45.5%にとどまっている。逆にポリシーの強度と比較してデータ公開経験がある回答者の比率が高い分野は、植物学・動物学とコンピュータサイエンス(以下、CS)であった。以下では、これらの4分野に着目しながら分野別の結果を示す。

#### a. データの公開・非公開理由

分野別に非公開理由の選択率をそれぞれ確認して比較したところ、精神医学・心理学は「公開手段がない」の選択率が全分野で最も高かった(44.4%, 平均は8.9%)。他には臨床医学、精神医学・心理学ともに特に選択率が高い項目はなかったが、非公開理由が解決された場合に公開したいと考える回答者はいずれも33.3%であり、平均(28.7%)よりも高かった。

公開理由について、植物学・動物学とCSは「研究成果を広く認知してもらいたいから(植物=75.0%), (CS=82.9%,  $p<0.01$ )」, 「科学研究や成

果実装を推進したいから(植物=43.8%), (CS=46.3%,  $p<0.01$ )」, 「他の研究者からのリクエストに応じて(植物=37.5%), (CS=34.1%)」, 「オープンデータに貢献したいから(植物=37.5%), (CS=22.0%)」で順位が一致しており、いずれも全体より選択率が高かった。また、全体では「他の研究者からのリクエスト」に次いで「所属機関のポリシー」を選ぶ回答者が多かったが、両分野では「オープンデータに貢献したいから」が多く選ばれていた。「その他」では、「それが当然のことと思っているから(植物)」という回答もみられた。つまり、両分野では科学的利他性に基づく公開理由が他分野と比較してよく選ばれていた。

#### b. データの提供・利用状況

データの提供・被提供経験は分野による差がみられた(いずれも  $p<0.001$ )。まったくない=1, ほとんどない=2, たまにある=3, よくある=4と重み付けした場合の順位を確認すると、精神医学・心理学は提供、被提供ともに22位(最下位)であり、臨床医学は提供が19位、被提供が18位であった。一方、植物学・動物学はそれぞれ5位と2位、CSは8位と4位であり、比較的良好にデータの提供を受けていた。

公開データの入手経験の順位を確認すると、精神医学・心理学が22位(46.2%), 臨床医学が21位(50.0%)と低く、植物学・動物学は3位(89.5%), CSは5位(85.7%)であった(第16表)。「研究に再利用」する頻度も分野による差があり( $p<0.001$ )、「行わない」比率が低い順に並べると、精神医学・心理学が22位(100.0%), 臨床医学が21位(90.9%)であり、CSは3位(24.1%), 植物学・動物学は5位(29.4%)であった(平均44.8%)。つまり、精神医学・心理学と臨床医学は、公開データの入手も再利用も行われにくい傾向にあり、植物学・動物学とCSは入手も再利用も行われる傾向にあることが明らかになった。

#### c. データ公開の障壁

カレントデータの管理・公開資源の充足度と分野の関連を調べた結果、資金、時間、データ公開用リポジトリ、研究中のストレージに差がみられ

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

第17表 分野別カレントデータの管理・公開資源の充足度

資源の充足度	低		高	
データ整備・公開のための資金 ( $p<0.05$ )	臨床医学*	1.12	植物学・動物学*	1.75
	宇宙科学	1.14	数学	1.75
	精神医学・心理学*	1.17	分子生物学・遺伝学	1.53
データ整備・公開のための時間 ( $p<0.01$ )	精神医学・心理学*	1.17	植物学・動物学*	2.06
	臨床医学*	1.18	数学	1.86
	免疫学	1.27	分子生物学・遺伝学	1.59
データ公開用のリポジトリ ( $p<0.01$ )	精神医学・心理学*	1.08	数学	2.57
	社会科学	1.20	植物学・動物学*	2.56
	薬理学・毒物学	1.35	CS*	2.22
研究中のデータ用ストレージ ( $p<0.05$ )	免疫学	1.55	CS*	2.69
	精神医学・心理学*	1.58	植物学・動物学*	2.47
	農学	1.79	物理学	2.47

第18表 分野別カレントデータを公開する場合の懸念の強さ

懸念	強		弱	
引用せずに利用される可能性 ( $p<0.01$ )	材料科学	1.22	数学	1.91
	化学	1.24	宇宙科学	1.88
	農学	1.24	CS*	1.70
先に論文を出版される可能性 ( $p<0.01$ )	免疫学	1.00	数学	2.00
	化学	1.29	CS*	1.89
	農学	1.31	宇宙科学	1.75
機密・プライバシー情報 ( $p<0.01$ )	臨床医学*	1.35	宇宙科学	2.63
	免疫学	1.58	植物学・動物学*	2.59
	分子生物学・遺伝学	1.60	数学	2.38
商用利用される可能性 ( $p<0.01$ )	免疫学	1.25	植物学・動物学*	3.00
	臨床医学*	1.42	宇宙科学	2.75
	薬理学・毒物学	1.66	環境科学・生態学	2.37
誤解や誤用の可能性 ( $p<0.05$ )	精神医学・心理学*	1.54	宇宙科学	2.38
	臨床医学*	1.58	数学	2.27
	農学	1.63	CS*	2.26

た。充足度を、不十分=1、やや不十分=2、ほぼ充分=3、充分=4と重み付けして比較した場合に、充足度が低い3分野と高い3分野を第17表に示す。ポリシーの強度との乖離がみられる臨床医学などの4分野には、\*印を付した(第18表でも同様に示す)。精神医学・心理学分野は非公開理由のうち「公開手段がない」の選択率が高かったが、データ公開用リポジトリの充足度も低かった。

カレントデータを公開する場合の懸念と分野については、「研究の誤りを発見される可能性」以外の項目に差がみられた。問題である=1、やや

問題である=2、あまり問題ではない=3、問題ではない=4と重み付けして比較した場合に、懸念が強い3分野と弱い3分野を第18表に示す。臨床医学は機密・プライバシー情報と商用利用される可能性について懸念が強く、カレントデータに含まれる機密情報の分野別順位は3位(68.2%)であった(第16表)。

## V. 考察

前章の結果を総合して、本研究の目的であるデータ公開の状況、データ公開の推進要因、および阻害要因について考察を行うとともに、今後の



課題について述べる。研究課題ごとの結果と考察の関係は第1図に示した通りであるが、データ公開の推進要因と阻害要因は相互に関連しているため、結果の一部を別の考察に用いた箇所もある。

## A. データ公開の状況と必要性

日本の研究者によるデータ公開経験の結果について、先行研究との比較を行う。続いて、そもそもデータ公開は不要であるとする意見や公開するデータがないとする意見について検討する。

### 1. データ公開の状況

データ公開の経験がある回答者は51.0%であった。II章で触れた先行研究のうち、分野横断で実施された調査の結果、すなわち Tenopirら(2009/2010)<sup>27)</sup>の36%、ノルウェーの報告書(2013)<sup>32)</sup>の28%、Ferguson(2014)<sup>33)</sup>の52%よりも高いか同程度であった。また、Ferguson<sup>33)</sup>の報告では世界の研究者52%に対して日本の研究者は44%であったこと、本調査と同時期に実施して1,162名の回答を得た Berghmansら(2016)<sup>36)</sup>の報告では世界の研究者が66%であったことから、本調査の結果はおおむね妥当であると考えられる。データ公開経験がある日本の研究者は、経年的に増加している可能性も示唆された。

### 2. データ公開の必要性

ここでは、IV章B節3項「データを公開していない理由と公開意思」に示した「その他」の回答と自由回答のうち、(a) 不要、(c) 公開するデータがない、とする記述から、データ公開の必要性について考察する。(b) 公開のインセンティブの欠如については、C節で検討する。

(a) データ公開は不要であるとする意見のうち、「論文や補足資料で充分である」については、補足資料の有用性を示す調査結果<sup>57), 58)</sup>があるものの、再利用には適さないこと<sup>59)</sup>や、リンク切れ<sup>60)</sup>、質の問題<sup>61)</sup>などが指摘されていることから、充分とは言い難いと考えられる。また、「必要ならば著者に直接問い合わせれば良い」という意見については、植物学と動物学を対象とした調

査によって、出版後20年が経過した論文の約8割はデータが入手できないことが明らかにされている<sup>62)</sup>。以上のことから、基本的にはデータを公開して長期にわたる再利用に供することが望ましく、論文や補足資料で十分な場合については例外として検討する方が良いのではないかと考えられる。「必要性が感じられない」については、研究者の立場から、公開の意義や再利用の可能性、保存の必要性が感じられないといった見解が示されていた。一方、オープンサイエンス政策<sup>7), 8), 10)</sup>では、現時点では想定されていないようなイノベーションの創出や分野を超えたデータの利活用などが期待されている。このギャップを埋めるためには、データの再利用によるベストプラクティスや経済的な効果などの具体的な将来像を示す必要があると考えられる。

(c) 公開するデータがないことについては、純粋数学など分野による事情についての記述がみられた。こうした分野については別途調査と検討を行った上で、一律の公開を求めないといった対応が必要であると考えられる。この点は今後の課題としたい。

## B. データ公開の推進要因と公開データの入手環境の整備

データ公開の推進要因として、日本の研究者によるデータの公開理由を先行研究と比較しながら検討する。また、データ公開の経験がある研究者は公開データを活用していることから、データの利用や公開を促進するリポジトリの整備について考察する。

### 1. データの公開理由

データの公開理由の1位は「研究成果の認知向上」(58.5%)であり、雑誌のポリシーの強度と比較してデータ公開率が高い植物学・動物学(75.0%)とCS(82.9%)の選択率は特に高かった。そこで、第3項では研究成果の認知向上を推進するためのリポジトリについて考察する。2位は「雑誌のポリシー」(43.8%)であったが、より公開率が高いOA論文は1位であった。デー

タ公開を求める雑誌が増加することによって、公開率が上がると考えられる。

Kim ら<sup>29)</sup>の調査では、科学的利他性が主なデータ公開理由の一つであり、関連する「科学研究や成果実装の推進」は、Schmidt ら<sup>12)</sup>、ノルウェーの報告書<sup>32)</sup>、小野ら<sup>35)</sup>の調査でも主な公開理由となっていたが、本調査では3位(26.1%)にとどまった。また、「他の研究者からのリクエストに応じて」は4位(25.4%)、「オープンデータへの貢献」は6位(10.5%)であり、選択率はOA論文よりも高かったものの主要な理由とまでは言い難かった。ただし、植物学・動物学とCSではこれらの選択比率が高く、雑誌のポリシーなどの強制力によらずデータ公開が進んでいる分野では、科学的利他性が高いことが示唆された。

「所属機関のポリシー」は5位(23.7%)、「助成機関のポリシー」は9位(6.3%)であり、同じく主要な理由ではなかった。なお、いずれもOA論文よりも選択率が高かったが、OA論文のポリシーを制定している機関の方が多いと予想されることから、これらの回答が実態に即しているかどうか、質問が適切であったかどうかの検証を今後の課題としたい。また、「分野・コミュニティの規範」は7位(9.5%)であったが、分野別の比率は8位の「業績になる場合があるから」(9.3%)との関連がみられた。データ公開と業績については、C節で考察する。

## 2. 公開データの利用

Piowar の調査<sup>45)</sup>と同様に、研究データを公開している研究者の特徴として、公開データの入手経験があることがわかった。また、回答者は複数分野(平均2.7分野)の公開データを利用したいと希望していることも確認された。ノルウェーの報告書<sup>32)</sup>で示唆されていたように、公開データの利用が進むことによってデータ公開につながる可能性があることから、公開データの入手環境を整備することが重要であると考えられる。

しかし、公開データの入手先(第11表)やデータの公開先(第5表)で最も多かったのは、いずれも個人や研究室のウェブサイト、次いで論

文の補足資料であった。先行研究<sup>28), 33), 36), 39)</sup>と同様に、所属機関や特定分野のリポジトリは十分に活用されていないこと、さらに、データ公開用リポジトリの充足度や認知度が低いことも確認された(第3図)。個人や研究室のウェブサイトは、機関や分野のリポジトリと比較すると、永続的に公開することや発見可能性を高めることなどが困難であると考えられる。また、公開データを手入手する際に、何らかの問題があるとする回答者は約8割にのぼった(第12表)。そこで、次項では公開データの入手先という観点からもリポジトリについて考察する。

## 3. リポジトリの整備

植物学・動物学やCSはデータを公開するためのリポジトリの充足度が高く、公開データの入手や研究利用も盛んである。こうした環境によって自身もデータを公開して研究成果の認知向上をはかろうとする意識や科学的利他性が醸成されていると考えられる。両分野のように、データの利用と公開のサイクルを推進するためには、研究者が容易にデータを手入手し、活用できるとともに、容易にデータを公開できるリポジトリを整備することが重要であると考えられる。

よく再利用されるデータは、適切に整備されて永続性のあるリポジトリで公開されていることが明らかにされている<sup>63)~65)</sup>。具体的には、FAIRデータ原則<sup>66)</sup>に準じてデータを発見可能(Findable)、アクセス可能(Accessible)、相互運用可能(Interoperable)、再利用可能(Re-usable)な状態で登録することが望ましい<sup>67)</sup>と考えられる。また、第12表に示した利用料金や利用者登録といった入手における問題への対応も必要であろう。さらに、データ公開における誤解や誤用などの懸念を低減するために、そのデータを用いた論文や著者の情報を参照できるよう、DOI(Digital Object Identifier)などの永続的なリンクを備える必要があると考えられる。データを整備するための資源については、次節で考察する。

### C. データ公開の阻害要因と対策

第1項ではデータの非公開理由について、第2項ではデータ公開の重要な阻害要因である、研究者の時間不足と人的支援の可能性について検討する。第3項では、キャリアリスクの懸念やインセンティブの欠如への対策としてデータ公開の業績化と評価について考察する。第4項では、公開対象とはならない機密やプライバシー情報などを含むセンシティブデータについて、将来的な公開と活用の可能性について述べる。

#### 1. データの非公開理由

データの公開経験がないことについて、3割以上の回答者が選択した理由はなかった。2割以上の回答者が選択した理由は、雑誌や所属機関にポリシーがない、時間がない、ニーズがないと思う、業績にならないなど多岐にわたった。OA論文の経験がない理由については、雑誌のポリシーがないこと（約6割）と資金（約4割）に集中しており、それ（ら）が解決すれば8割近くの回答者が実践しても良いと考えていた。一方、データ公開の経験がない理由については、解決したとしても実践しようとする回答者は3割未満にとどまり、解決すればデータ公開につながる理由も特定できなかった。データ公開経験がない研究者はデータ公開に対して慎重であることや、A節2項に示した「公開するデータがない」分野、雑誌によるデータ公開の要求も公開率も低い精神医学・心理学分野の特性などをふまえた上で、さらなる調査と検討を行いたい。

#### 2. データ公開のための資源と人的支援

データを整備し、公開しようとする場合の資源は、特に人材、資金、時間が不足していると認識されていた。また、雑誌のポリシーと比較してデータ公開率が低い臨床医学と精神医学・心理学の研究者は、資金と時間が不足しているという認識が強かった。一方、データを公開している研究者は公開のための時間の充足度が高い傾向にあった。

資金については、データ公開の理由として記述

されていたインフラの利用料金の割引や、内閣府報告書に示されている“高品質なデータを提供した研究者に適切な報酬（助成金や昇進機会）を与える”<sup>8)</sup>[p.22]といった経済的支援の方法をとることができるならば、公開のインセンティブにもつながると考えられる。時間については、I章で示した通り、データ公開に限らない日本の研究界における問題となっている。日本の研究者が、データ公開のための時間やデータ公開のための技術や知識を身につけるための時間を新たに確保することは困難であろう。「学術情報のオープン化の推進について（審議まとめ）」<sup>9)</sup>では、データ公開について“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ともに、必要に応じて複数の大学等が共同して、データキュレーター等を育成するシステムを検討し、推進する”と述べられている。そこで、時間不足を補うための人的支援について検討する。

データ公開のための人的支援の例として、大学・研究図書館による研究データ管理（research data management）、あるいは研究データサービス（research data services）があり、米国<sup>68), 69)</sup>や欧州<sup>70), 71)</sup>の主要大学<sup>72)</sup>のみならず、小規模大学<sup>73), 74)</sup>やその他の地域<sup>75)~77)</sup>においても多数の事例が報告されている。また、支援サービスを担うデータライブラリアンの職務については、精査や体系化が行われており<sup>78), 79)</sup>、機関の研究支援部署やIT部署、外部組織などと連携して支援体制を構築することが望ましいと指摘されている。日本においても、データ管理や公開の実践に向けた資料として、ジャパンリンクセンター（JaLC）による「研究データへのDOI登録ガイドライン」<sup>80)</sup>や機関リポジトリ推進委員会（IRPC）<sup>81)</sup>によるNISOの入門書『研究データ管理』の翻訳版<sup>82)</sup>、デジタルリポジトリ連合（DRF）<sup>83)</sup>による研究者へのインタビュー調査結果<sup>84), 85)</sup>などが公開されている。

それでもなお、データを扱うことはフォーマットの決まった出版物などに比べて難易度が高いと考えられるが、機関リポジトリによるデータ公開

など、比較的支援や人材育成に取り組みやすいと考えられる項目が明らかになった（第13表）。一方、約6割の回答者が「専門性が必要である」と考えていた、適切なデータ形式への変換やデータを再利用しやすいように整えることなどについては、人材を育成するよりも、専門知識を有する人材をデータキュレーターといった新たな専門職として雇用することや、そのキャリアパスを検討する方が効率的であるかもしれない。また、第三者による支援には肯定的な意見も否定的な意見もみられたことから、研究者のニーズを明らかにすることを今後の課題としたい。

### 3. キャリアリスクの懸念とデータ公開の業績化

「引用せずに利用される可能性」および「先に論文を出版される可能性」は、データ公開経験の有無にかかわらず、最も問題視する回答者が多かった。また、先行研究<sup>12), 26)</sup>と同様に、キャリア形成期である若年層の懸念が強かった。そこで、この2点をキャリアリスクの問題として考察した上で、業績化や評価による対策について検討する。

#### a. 論文の引用とデータ引用

データを再利用する際の引用方法には、データを用いた論文を引用する方法とデータを直接引用する方法がある。論文の引用は、自由回答にもみられたように公開のインセンティブになる<sup>20)~23)</sup>と考えられている。データを用いた論文の引用が徹底されれば、データ公開の懸念の1位であった「引用せずに利用される可能性」が低減されるとともに、データ公開理由の1位であった研究成果の認知向上にもつながると考えられる。

データ引用については、科学的知識の共有を推進する国際イニシアティブであるFORCE11によるデータ引用の原則<sup>86)</sup>とガイドラインが公開されており、2016年にはElsevier社が1,800を超える雑誌に採択した<sup>87)</sup>。データ引用は、論文の引用と同様に研究成果の認知向上に効果があると指摘されている<sup>88)</sup>。データを直接引用することによって、出版物と同様の研究成果である<sup>89)</sup>という認識も浸透するのではないだろうか。また、

引用によってデータの参照が保証されれば、間接的に誤解や誤用に対する懸念も低減されるかもしれない。しかし、データ引用はまだ充分に行われていないとする調査結果<sup>90), 91)</sup>もあることから、出版者や学協会などのステークホルダーによるアドボカシー活動や、データの利用規約によるデータ引用の浸透が重要であると考えられる。

#### b. 先取権

論文の出版前にデータ公開を求めることについては、本研究のみならず、II章に示した先行研究<sup>12), 26), 28)</sup>や環境学分野の調査<sup>92)</sup>でもデータ公開の障壁となっていることが明らかにされている。生命科学分野など、1990年代ごろから論文出版前にデータを公開するよう求めてきた分野もあるが<sup>93)</sup>、同分野のように公開によって利点が生じる場合や、データジャーナル<sup>94)</sup>のように業績や引用、査読による質保証といったインセンティブが十分に提供される場合を除いて、論文出版前のデータ公開は受け入れ難いと考えられる。一方で、データ引用が徹底されることによって先取権が確保されるならば、また、データ公開そのものが業績として認められるならば、「先に論文を出版される可能性」への懸念は低減されると考えられる。そこで、データ公開の業績化について検討する。

#### c. データ公開の業績化

データ公開を業績化することは、キャリアリスクの懸念への対策であると同時に、自由回答にみられたようにデータ公開のインセンティブになると考えられ、国内外の政策文書<sup>7), 8), 10), 95)</sup>でも検討が行われている。また、データ公開の業績化は、データ公開が規範ではない分野への啓発につながると考えられる。ただし、業績化に疑問を呈する意見もみられたことから、データの取得、整備、公開のうち、どこまでを業績とするかについては、慎重な議論が必要であると考えられる。

データ公開を業績とするためには、公開データの評価が必要であろう。評価の副次的な効果として、公開データの利用におけるデータの品質の問題（第12表）や、自由回答にみられた悪意のある批判への懸念が低減されると考えられる。しかし、

分野によっては公開データの質が調査され<sup>96),97)</sup>、評価の枠組み<sup>98)</sup>や標準<sup>99)</sup>、リポジトリによる基準<sup>100)</sup>が検討されているものの、多くの分野では評価のためのプロトコルが確立されていない<sup>101)</sup>。それでは、どのような方法でデータの評価を行えばよいのだろうか。

最も厳正な方法として、学術雑誌が専門家の査読によって掲載論文の質を保証しているように、データに対しても査読を行うことが提案されている<sup>102)~104)</sup>。データの査読を経てデータ論文 (data paper や data description) を掲載するデータジャーナル<sup>105)</sup>は、データの信頼性を保証する仕組みとして有効であると考えられる。また、一部の雑誌では補足資料やデータの査読を行っている<sup>106)</sup>。しかし、研究者数や論文数の増大によって査読制度が限界を迎えつつあることが指摘されている<sup>107)</sup>。現状において、さらに論文の根拠となるデータまで査読することは、極めて困難であろう。

データを評価するための次善の策として、そのデータを用いた論文が確認できる DOI などの永続的な識別子を付与する仕組みや、データの引用数といったメトリクスをリポジトリに備えることが望ましいと考えられる。また、*Journal of Visualized Experiments* (<https://www.jove.com>) のように実験やデータの生成過程を動画によって確認可能にすること、データ利用者からのフィードバックを掲載して他の研究者による評価を確認可能にすること、ORCID (<https://orcid.org>) のような研究者・研究業績データベースへのリンクを充実させてデータ作成者の情報を確認可能にすることなども業績化のための基盤として有効であると考えられる。

#### 4. 機密・プライバシー情報と商用利用への懸念

企業の研究者は機密・プライバシーや商用利用への懸念が強く、特許のため公開は困難であるという意見も多数みられた。分野別では、臨床医学分野の研究者もこれらの懸念が強かった。第5期科学技術基本計画では、国家安全保障等に係るデータ、商業目的で収集されたデータを公開適用

対象外とし、個人のプライバシーや財産的価値があるデータについては制限事項を設けるとしている<sup>11)</sup> [p.32] ことから、こうしたセンシティブ情報を含むデータの公開を行う必要はないと考えられる。

一方、オープンサイエンスは、“企業や個人においては、科学的成果を活用・再利用して新しい製品や新しいサービス（市場）を生み出すこと”<sup>8)</sup> [p.1] を期待している。大学や公的機関・団体の研究者も商用利用への懸念が強かったが、研究成果が企業のイノベーションに一定の役割を果たしている<sup>108)</sup> ことから、商用利用に関する議論とコンセンサスの形成が必要であると考えられる。また、センシティブ情報を保護しつつ、データを流通させようとする取り組みも行われている。たとえば、内閣府の知的財産戦略本部は、民間投資によるデータと人工知能 (AI) の利活用を目指して、“営業秘密”に関するデータを保護しつつ、公開によって価値を生み得るデータ<sup>109)</sup>を流通させる仕組みを検討している<sup>110)</sup>。また、Project Data Sphere (<https://www.projectdatasphere.org/>) のように、がん治療を目的として産学連携で個人情報や商業的な機密に配慮しながら臨床データと解析ツールを公開する取り組み<sup>111)</sup>もみられる。センシティブ情報は原則的に非公開としつつも、将来的には新たな手法を取り入れ、成果を生むことができれば、産学官の互恵的なデータ公開の機運が高まるのではないだろうか。

## VI. おわりに

本研究の結果、データ公開を推進するための重要な課題は、データの流通を促進するためのリポジトリの整備、研究者の時間不足を補うための人的支援、およびデータ公開の業績化であった。データ公開の推進要因や阻害要因は相互に関連していることから、1つの課題に対応することで複数の課題の解決や低減につながる場合があること、一方で、複数の課題に同時に対応しなければ、データ公開の推進は困難であることも示唆された。

本調査の限界として、データ公開に関連するデータの管理や保存については十分に明らかにできなかったことを挙げる。今後はより広く、研究のライフサイクル全体からデータ公開に関する状況と課題を検討したい。また、データ公開に対する自由回答は多岐にわたり、研究環境や倫理など学術界全体におよぶ問題に関する指摘もみられた。インタビュー調査などによって、より丁寧に研究者の意向を汲み取ることが重要であると考えられる。

1665年に *Philosophical Transactions of the Royal Society* や *Journal des Sçavans* が刊行されてから、研究成果を論文として公開することの価値が研究者に認められ、受け入れられ、定着するまでには相当の年月を要した。オープンスカラシップの進展はめざましいが、データを第三者が理解できるように公開し、それによって新たな成果を生むサイクルが多く分野で定着するまでには、さまざまな問題を解決することや一定の時間が必要であろう。本研究で明らかにした課題に取り組むことによって、科学研究の進展に寄与したい。

## 謝 辞

本研究は、2014年度三田図書館・情報学会研究助成、および2016年度総合地球環境学研究所コアプロジェクトFS(14200075)「社会課題解決型研究のアクター間における知識情報ギャップの可視化と克服」の経費を受けて実施致しました。

質問紙調査は、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で客員研究官として実施致しました。調査票と結果は報告書として公開致します。また、筑波大学図書館情報メディア系研究倫理審査の承認を受けて実施致しました(第28-58号、2016年9月6日)。

本調査およびプレテストにご協力を賜りました皆様、調査実施にご助力を賜りました文部科学省科学技術・学術政策研究所の林和弘上席研究官、林宣子様、質問紙作成にご助言を賜りました筑波大学の逸村裕教授、小野寺夏生名誉教授、溝上智恵子教授、緑川信之教授に、心よりお礼申し上げます。

また、査読者の皆様から多数の貴重なご意見を賜りました。厚くお礼申し上げます。

## 注・引用文献

- 1) Borgman, Christine L. Big Data, Little Data, No Data: Scholarship in the Networked World. The MIT Press, 2015. 416p.
- 2) 倉田敬子. e-Science とは. 情報の科学と技術. 2013, vol. 63, no. 9, p. 352-357. <http://doi.org/10.18919/jkg.63.9.352>, (accessed 2017-3-30).
- 3) 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 2017-3-30).
- 4) OECD Principles and Guidelines for Access to Research Data from Public Funding. OECD Publications, 2007, 22p. <http://www.oecd.org/science/sci-tech/38500813.pdf>, (accessed 2017-3-30).
- 5) Foreign & Commonwealth Office. G8 Science Ministers Statement. GOV.UK. 2013-6-13. <https://www.gov.uk/government/news/g8-science-ministers-statement>, (accessed 2017-3-30).
- 6) The Royal Society. Science as an Open Enterprise. The Royal Society Science Policy Centre, 2012, 104p. <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/>, (accessed 2016-3-30).
- 7) OECD. Making Open Science a Reality. OECD Publishing, 2015. OECD Science, Technology and Industry Policy Papers, no. 25, 112p. <http://doi.org/10.1787/5jrs2f963zsl-en>, (accessed 2017-3-30).
- 8) 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 2017-3-30).
- 9) 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/b\\_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm), (accessed 2017-3-30).
- 10) 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 2017-3-30).
- 11) 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 2017-3-30).
- 12) 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 2017-3-30).
  - 13) 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 2017-3-30).
  - 14) たとえば, 2015年から2016年の間に化学は10万件以上レコード数が増えている一方で, 人口統計学は41,854件から5,458件まで減少している。
  - 15) 2016年2月に科学技術振興機構戦略研究推進部・研究プロジェクト推進部は「戦略的創造研究推進事業におけるデータマネジメント実施方針」を公表した。国外では, 一般に助成金の申請時にデータマネジメント計画 (Data Management Plan, DMP) の提出を義務付けているのに対して, この方針は一部の助成金に採択後のDMP提出を求めるものであり, データは非公開とすることも可能である。科学技術振興機構戦略研究推進部・研究プロジェクト推進部, 戦略的創造研究推進事業におけるデータマネジメント実施方針. 2016, 3p. [http://www.senryaku.jst.go.jp/teian/koubo/data\\_houshin.pdf](http://www.senryaku.jst.go.jp/teian/koubo/data_houshin.pdf), (accessed 2017-3-30).
  - 16) Stodden, Victoria; Guo, Peixuan; Ma, Zhaokun. Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLOS ONE*. 2013, vol. 8, no. 6, e67111. <http://doi.org/10.1371/journal.pone.0067111>, (accessed 2017-3-30).
  - 17) "Research data policies". Springer Nature. <http://www.springernature.com/gp/group/data-policy/>, (accessed 2017-3-30).
  - 18) "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 2017-3-30).
  - 19) Vines, Timothy H. et al. Mandated data archiving greatly improves access to research data. *The FASEB Journal*. 2013, vol. 27, no. 4, p. 1304-1308. <http://doi.org/10.1096/fj.12-218164>, (accessed 2017-3-30).
  - 20) Piwowar, Heather A.; Day, Roger S.; Fridsma, Douglas B. Sharing detailed research data is associated with increased citation rate. *PLOS ONE*. 2007, vol. 2, no. 3, e308. <http://doi.org/10.1371/journal.pone.0000308>, (accessed 2017-3-30).
  - 21) Vandewalle, Patrick. Code sharing is associated with research impact in image processing. *Computing in Science & Engineering*. 2012, vol. 14, no. 4, p. 42-47. <http://doi.org/10.1109/MCSE.2012.63>, (accessed 2017-3-30).
  - 22) Henneken, Edwin. A.; Accomazzi, Alberto. "Linking to data: Effect on citation rates in astronomy". *Astronomical Data Analysis Software and Systems XXI*. Paris, 2011-11-6/10, Astronomical Society. 2012, p. 763-766. <http://www.aspbooks.org/publications/461/763.pdf>, (accessed 2017-3-30).
  - 23) Dorch, Bertil F.; Drachen, Thea M.; Ellegaard, Ole. The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union*. 2015, vol. 11, no. A29A, p. 172-175. <https://doi.org/10.1017/S1743921316002696>, (accessed 2017-3-30).
  - 24) 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 2017-3-30).
  - 25) Phillips, Nicky. Striving for a research renaissance. *Nature*. 2017, vol. 543, no. 7646, S7 (Nature Index). <http://doi.org/10.1038/543S7a>, (accessed 2017-3-30).
  - 26) 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 2017-3-30).
  - 27) 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 2017-3-30).
  - 28) Huang, Xiaolei et al. Willing or unwilling to share primary biodiversity data: Results and implications of an international survey. *Conservation Letters*. 2012, vol. 5, no. 5, p. 399-406.
  - 29) 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 2017-3-30).
  - 30) Kim, Youngseek; Zhang, Ping. Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data reposi-

- tories. *Library & Information Science Research*. 2015, vol. 37, no. 3, p.189-200. <http://doi.org/10.1016/j.lisr.2015.04.006>, (accessed 2017-3-30).
- 31) 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 2017-3-30).
- 32) Sharing and Archiving of Publicly Funded Research Data: Report to the Research Council of Norway. Danvad, 2014, 74p.
- 33) 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>, (accessed 2017-3-30).
- 34) 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>, (accessed 2017-3-30).
- 35) 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 2017-3-30).
- 36) Berghmans, Stephane et al. Open Data: The Researcher Perspective. *Centre for Science and Technology Studies (CWTS)*, 2017, 47p. <https://www.cwts.nl/download/f-53w2.pdf>, (accessed 2017-10-30).
- 37) 以下, 括弧内の数字は調査年を示す。
- 38) 出版物の付録や補足資料, データジャーナルでのデータ出版, データリポジトリ, その他, いずれの方法でもデータを公開していない回答者が34%であった。
- 39) 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 2017-3-30).
- 40) 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 2017-3-30).
- 41) Van den Eynden, Veerle; Bishop, Libby. Sowing the Seed: Incentives and Motivations for Sharing Research Data, a Researcher's Perspective. *Knowledge Exchange*, 2014, 46p. [http://repository.jisc.ac.uk/5662/1/KE\\_report-incentives-for-sharing-researchdata.pdf](http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf), (accessed 2017-3-30).
- 42) 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 2017-3-30).
- 43) 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 2017-3-30).
- 44) Wicherts, Jelte M.; Bakker, Marjan; Molenaar, Dylan. Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLOS ONE*. 2011, vol. 6, no. 11, e26828. <http://doi.org/10.1371/journal.pone.0026828>, (accessed 2017-3-30).
- 45) Piwowar, Heather A. Who shares? Who doesn't?: Factors associated with openly archiving raw research data. *PLOS ONE*. 2011, vol. 6, no. 7, e0018657. <http://doi.org/10.1371/journal.pone.0018657>, (accessed 2017-3-30).
- 46) Challenges and opportunities. *Science*. 2011, vol. 331, no. 6018, p. 692-693. <http://doi.org/10.1126/science.331.6018.692>, (accessed 2017-3-30).
- 47) “質問の後半にみられるいいかげんな回答や間違い”を指す。鈴木淳子. “選択回答法”. *質問紙デザインの技法*. 第2版, ナカニシヤ出版, 2016, p. 196.
- 48) Dörnyei, Zoltán. *外国語教育学のための質問紙調査入門: 作成・実施・データ処理*. 八島智子, 竹内理監訳. 松柏社, 2006, 196p.
- 49) Clarivate Analytics 社の研究動向・統計データベース. <http://ip-science.thomsonreuters.jp/products/esi/>, (accessed 2017-3-30).
- 50) Kratz, John E.; Strasser, Carly. Making data count. *Scientific Data*. 2015, vol. 2, 150039 (2015). <http://doi.org/10.1038/sdata.2015.39>, (accessed 2017-3-30).
- 51) Kratzらは“論文の主な結論を支持する”と“論文の主な結論に到達する”という質問を行っているが, プレテストで両者の違いが分かりにくいとの指摘が複数あり, 詳しい説明を行うと煩雑になるため統合した。
- 52) Jones, Sarah; Pryor, Graham; Whyte, Angus. *How to Develop Research Data Management Services: A Guide for HEIs*. Digital Curation Centre, 2013, 22p. <http://www.dcc.ac.uk/sites/>



- default/files/documents/publications/How-to-develop-RDM-services\_finalMay2013rev.pdf, (accessed 2017-3-30).
- 53) 最終的な尺度項目数は、合計 19 (7 問) である。
- 54) 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](http://doi.org/10.20651/jslis.62.1_20), (accessed 2017-10-30).
- 55) 現時点で、研究者の立場からは再利用の可能性や保存の必要性が感じられないといった見解や、“データをパターン化されたシステムで公開する意義を感じない”という記述がみられた。
- 56) 最も多かったのは「共同研究者」(82.1%) であり、以下、「所属機関」(58.2%), 「所属機関以外の研究機関」(9.5%), 「所属機関以外の企業」(6.4%), 「研究助成機関」(5.7%) の順であった。
- 57) Kafkas, Senay et al. Database citation in supplementary data linked to Europe PubMed Central full text biomedical articles. *Journal of Biomedical Semantics*. 2015, vol. 6, 1. <http://doi.org/10.1186/2041-1480-6-1>, (accessed 2017-3-30).
- 58) Anagnostou, Paolo et al. When data sharing gets close to 100%: What human paleogenetics can teach the open science movement. *PLOS ONE*. 2015, vol. 10, no. 3, e0121409. <http://doi.org/10.1371/journal.pone.0121409>, (accessed 2017-3-30).
- 59) Kenyon, Jeremy; Sprague, Nancy; Flathers, Edward. The journal article as a means to share data: A content analysis of supplementary materials from two disciplines. *Journal of Librarianship and Scholarly Communication*. 2016, vol. 4, eP2112. <http://doi.org/10.7710/2162-3309.2112>, (accessed 2017-3-30).
- 60) Anderson, Nicholas R.; Tarczy-Hornoch, Peter; Bumgarner, Roger E. On the persistence of supplementary resources in biomedical publications. *BMC Bioinformatics*. 2006, vol. 7, 260. <http://doi.org/10.1186/1471-2105-7-260>, (accessed 2017-3-30).
- 61) Jubb, Michael. “Introduction: Scholarly communications: Disruptions in a complex ecology”. *The Future of Scholarly Communication*. Shorley, Deborah; Jubb, Michael eds. Facet, 2013, p. XIII-XXXVI.
- 62) Vines, Timothy H. et al. The availability of research data declines rapidly with article age. *Current Biology*. 2014, vol. 24, no. 1, p. 94-97. <http://doi.org/10.1016/j.cub.2013.11.014>, (accessed 2017-3-30).
- 63) Horton, Laurence; Katsanidou, Alexia. Purposing your survey: Archives as a market regulator, or how can archives connect supply and demand?. *IASSIST Quarterly*. 2011, vol. 35, no. 4, p. 18-23. [http://www.iassistdata.org/sites/default/files/iqvol35\\_horton.pdf](http://www.iassistdata.org/sites/default/files/iqvol35_horton.pdf), (accessed 2017-3-30).
- 64) Faniel, Ixchel M.; Kriesberg, Adam; Yakel, Elizabeth. Social scientists’ satisfaction with data reuse. *Journal of the Association for Information Science and Technology*. 2016, vol. 67, no. 6, p. 1404-1416. <http://doi.org/10.1002/asi.23480>, (accessed 2017-3-30).
- 65) Peters, Isabella et al. Research data explored: An extended analysis of citations and altmetrics. *Scientometrics*. 2016, vol. 107, no. 2, p. 723-744. <http://doi.org/10.1007/s11192-016-1887-4>, (accessed 2017-3-30).
- 66) FORCE11. The FAIR Data Principles. <https://www.force11.org/group/fairgroup/fairprinciples>, (accessed 2017-3-30).
- 67) Wilkinson, Mark D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016, vol. 3, 160018 (2016). <http://doi.org/10.1038/sdata.2016.18>, (accessed 2017-3-30).
- 68) Fearon, David Jr. et al. SPEC Kit 334: Research Data Management Services. *Association of Research Libraries*, 2013, 220p.
- 69) Tenopir, Carol et al. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*. 2014, vol. 36, no. 2, p. 84-90. <http://doi.org/10.1016/j.lisr.2013.11.003>, (accessed 2017-3-30).
- 70) Tenopir, Carol et al. Research data services in European academic research libraries. *LIBER Quarterly*. 2017, vol. 27, no. 1, p. 23-44. <http://doi.org/10.18352/lq.10180>, (accessed 2017-3-30).
- 71) Cox, Andrew M.; Pinfield, Stephen. Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Sciences*. 2014, vol. 46, no. 4, p. 299-316. <http://doi.org/10.1177/0961000613492542>, (accessed 2017-3-30).
- 72) Si, Li et al. Investigation and analysis of research data services in university libraries. *The Electronic Library*. 2015, vol. 33, no. 3, p. 417-449. <http://doi.org/10.1108/EL-07-2013-0130>, (accessed 2017-3-30).
- 73) Clement, Ryan et al. Team-based data management instruction at small liberal arts colleges. *IFLA Journal*. 2017, vol. 43, no. 1, p. 105-118. <http://doi.org/10.1177/0340035216678239>, (accessed 2017-3-30).
- 74) Shorish, Yasmeen. Data curation is for every-

- one! The case for master's and baccalaureate institutional engagement with data curation. *Journal of Web Librarianship*. 2012, vol. 6, no. 4, p. 263-273. <http://doi.org/10.1080/19322909.2012.729394>, (accessed 2017-3-30).
- 75) Renwick, Shamin; Winter, Marsha; Gill, Michelle. Managing research data at an academic library in a developing country. *IFLA Journal*. 2017, vol. 43, no. 1, p. 51-64. <http://doi.org/10.1177/0340035216688703>, (accessed 2017-3-30).
- 76) Primary Research Group. *International Survey of Academic Library Data Curation Practice*. Primary Research Group, 2013, 62p.
- 77) Kahn, Michelle et al. Research data management in South Africa: How we shape up. *Australian Academic & Research Libraries*. 2014, vol. 45, no. 4, p. 296-308. <http://doi.org/10.1080/0048623.2014.951910>, (accessed 2017-3-30).
- 78) Rice, Robin; Southall, John. *The Data Librarian's Handbook*. Facet Publishing, 2016, 192p.
- 79) Davidson, Joy. "5 Supporting early-career researchers in data management and curation". *Mastering Digital Librarianship: Strategy, Networking and Discovery in Academic Libraries*. Mackenzie, Alison; Martin, Lindsey eds. Facet Publishing, 2013, p. 83-102.
- 80) ジャパンリンクセンター運営委員会. 研究データへのDOI登録ガイドライン. 2015, 26p. [http://doi.org/10.11502/rd\\_guideline\\_ja](http://doi.org/10.11502/rd_guideline_ja), (accessed 2017-3-30).
- 81) 2017年度よりオープンアクセス推進協会(JP-COAR)に移行した。
- 82) Strasser, Carly. 研究データ管理. 機関リポジトリ推進委員会訳. 2016, 28p. <http://id.nii.ac.jp/1280/00000195/>, (accessed 2017-3-30).
- 83) 2016年度末に解散した。
- 84) 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/drif/index.php?onlineworkshop2015>, (accessed 2017-3-30).
- 85) 平成28年度DRFオンラインワークショップ「第2回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/drif/index.php?onlineworkshop2016>, (accessed 2017-3-30).
- 86) Joint Declaration of Data Citation Principles. FORCE11. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>, (accessed 2017-3-30).
- 87) Data citation. Elsevier. <https://www.elsevier.com/about/open-science/research-data/data-citation>, (accessed 2017-3-30).
- 88) Starr, Joan et al. Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*. 2015, e1. <http://doi.org/10.7717/peerj.cs.1>, (accessed 2017-3-30).
- 89) Martone, Maryann 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 2017-3-30).
- 90) Mooney, Hailey. Citing data sources in the social sciences: Do authors do it?. *Learned Publishing*. 2011, vol. 24, no. 2, p. 99-108. <http://doi.org/10.1087/20110204>, (accessed 2017-3-30).
- 91) Mayo, Christine; Vision, Todd J.; Hull, Elizabeth A. The location of the citation: Changing practices in how publications cite original data in the Dryad Digital Repository. *International Journal of Digital Curation*. 2016, vol. 11, no. 1, p. 150-155. <http://doi.org/10.2218/ijdc.v11i1.400>, (accessed 2017-3-30).
- 92) Soranno, Patricia A. et al. It's good to share: Why environmental scientists' ethics are out of date. *BioScience*. 2015, vol. 65, no. 1, p. 69-73. <https://doi.org/10.1093/biosci/biu169>, (accessed 2017-3-30).
- 93) McCain, Katherine W. Mandating sharing: Journal policies in the natural sciences. *Science Communication*. 1995, vol. 16, no. 4, p. 403-431.
- 94) 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*. 2015, vol. 148, p. 4-9. <http://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-STT148J-4.pdf>, (accessed 2017-3-30).
- 95) Expert Advisory Group on Data Access; Medical Research Council; Cancer Research UK; Wellcome Trust; Economic and Social Research Council. *Establishing Incentives and Changing Cultures to Support Data Access*. 2014, 28p. <http://wellcomelibrary.org/item/b20928129#>, (accessed 2017-3-30).
- 96) Roche, Dominique G. et al. Public data archiving in ecology and evolution: How well are we doing?. *PLOS Biology*. 2015, vol. 13, no. 11, e1002295. <http://doi.org/10.1371/journal.pbio.1002295>, (accessed 2017-3-30).
- 97) Peer, Limor; Green, Ann; Stephenson, Elizabeth. Committing to data quality review. *International Journal of Digital Curation*. 2014, vol. 9, no. 1, p. 263-291. <http://doi.org/10.2218/ijdc.v9i1.317>, (accessed 2017-3-30).
- 98) Wuest, Thorsten; Mak-Dadanski, Jakub; Thoben,

- Klaus-Dieter. "Data quality in materials science: A quality management manual approach". *Advances in Production Management Systems: Innovative and Knowledge-Based Production Management in a Global-Local World*. Ajaccio, France, 2014-09-20/24, APMS. Springer, 2014, vol. 438, p.42-49. [http://doi.org/10.1007/978-3-662-44739-0\\_6](http://doi.org/10.1007/978-3-662-44739-0_6), (accessed 2017-3-30).
- 99) Zimmerman, Ann S. New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, & Human Values*. 2008, vol. 33, no. 5, p. 631-652. <http://doi.org/10.1177/0162243907306704>, (accessed 2017-3-30).
- 100) たとえば以下のものなど。"Quality assurance". UK Data Archive. <http://www.data-archive.ac.uk/create-manage/format/quality>, (accessed 2017-3-30).
- 101) Brase, Jan et al. Approach for a joint global registration agency for research data. *Information Services & Use*. 2009, vol. 29, no. 1, p. 13-27.
- 102) Lawrence, Bryan et al. Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*. 2011, vol. 6, no. 2, p. 4-37. <http://doi.org/10.2218/ijdc.v6i2.205>, (accessed 2017-3-30).
- 103) Costello, Mark J. et al. Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*. 2013, vol. 28, no. 8, p. 454-461. <http://doi.org/10.1016/j.tree.2013.05.002>, (accessed 2017-3-30).
- 104) Parsons, Mark A.; Duerr, Ruth; Minster, Jean-Bernard. Data citation and peer review. *Eos, Transactions American Geophysical Union*. 2010, vol. 91, no. 34, p. 297-298.
- 105) 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 2017-3-30).
- 106) たとえば, 王立化学会の査読者用ガイドラインでは, 全ての電子補足資料 (Electronic supplementary information) を確認すること, 通常は完全なチェックを求められない X 線結晶データについても十分な注意を払う必要があること, さらに専門家に確認を依頼することもできるとしている。"Reviewer responsibilities: Reviewer procedure and policies". Royal Society of Chemistry. <http://www.rsc.org/journals-books-databases/journal-authors-reviewers/reviewer-responsibilities/>, (accessed 2017-3-30).
- 107) 佐藤翔. 査読の抱える問題とその対応策. *情報の科学と技術*. 2016, vol. 66, no. 3, p.115-121. [https://doi.org/10.18919/jkg.66.3\\_115](https://doi.org/10.18919/jkg.66.3_115), (accessed 2017-3-30).
- 108) 池内健太, 元橋一之, 田村龍一, 塚田尚稔. 科学・技術・産業データの接続と産業の科学集約度の測定. *科学技術・学術政策研究所*, 2017, DP-142, 45p. <http://doi.org/10.15108/dp142>, (accessed 2017-3-30).
- 109) 現状ではデータは知的財産権の保護対象ではないが, "収集・蓄積・保管等するために一定の投資又は労力を投じることが必要であり, 営業(事業)活動上の利益として保護すべきもの"を「価値あるデータ」と定義している。
- 110) 新たな情報財検討委員会. 新たな情報財検討委員会報告書: データ・人工知能 (AI) の利活用促進による産業競争力強化の基盤となる知財システムの構築に向けて. 知的財産戦略本部検証・評価・企画委員会, 2017, 47p. [http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho\\_hyoka\\_kikaku/2017/johozai/houkokusho.pdf](http://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kensho_hyoka_kikaku/2017/johozai/houkokusho.pdf), (accessed 2017-3-30).
- 111) Bertagnolli, Monica M. et al. Advantages of a truly open-access data-sharing model. *New England Journal of Medicine*. 2017, vol. 376, no. 12, p. 1178-1181. <http://doi.org/10.1056/NEJMsb1702054>, (accessed 2017-03-30).

## 要 旨

**【目的】** 日本の研究者による研究データ公開の状況と認識を明らかにすること、および推進要因と阻害要因を分析することによって、オープンサイエンス政策や適切な支援体制の構築に資することを目的とする。

**【方法】** 2016年11月から12月にかけて、科学技術・学術政策研究所が運営する科学技術専門家ネットワークに登録されている1,983名の研究者を対象とした質問紙調査を実施した。回答は、年齢や分野との関連を分析し、論文のオープンアクセス状況や先行研究の結果と比較した。推進要因として公開理由や公開データの利用経験を調査し、2項ロジスティック回帰分析によってデータ公開経験がある研究者の特徴を明らかにした。阻害要因として非公開理由や障壁の程度を調査した。また、雑誌のデータ公開ポリシーの強度に対してデータ公開率が高い分野と低い分野の特徴を明らかにした。

**【結果】** 大学、企業、公的機関・団体等の研究者1,398名（回答率70.5%）から回答を得た。データの公開経験がある回答者は51.0%であり、主な公開理由は研究成果の認知向上と雑誌のポリシーであった。非公開理由は雑誌のポリシーがないことなど多岐にわたり、それらが解決されても公開するという回答者は28.4%にとどまった。公開データの入手経験は75.8%、今後の利用意思は97.1%にみられた。データ公開のための資源は、人材、資金、時間、公開用リポジトリの充足度が低いと認識されている。また、データを引用せずに利用されることや先取権の喪失などキャリアリスクの懸念が強く、特に大学の若手研究者に顕著であった。機密、商用利用、誤用への懸念も強い。推進要因と阻害要因の分析から、データ公開を推進するための重要な課題は公開データの入手環境の整備と研究者の時間を確保するための人的支援であることが示唆された。また、キャリアリスクの懸念やインセンティブの欠如への対策としてデータ公開の業績化が重要であることを示した。

## 付録 質問一覧

質問	必須	方式
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。	*	
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。	*	
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。	*	m
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。	*	m
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せんか？	*	
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。	*	m
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。		m
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。		m
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？		4
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？		4
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？	*	m
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。	*	m
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。	*	3
Q14. 研究に利用したことがある（参考にしたリ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。	*	m
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。		m
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。		m
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。	*	m
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？	*	m
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。	*	m
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。	*	m
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？	*	
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？		
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。		s
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。		m
Q25. カレントデータには、以下の機密情報が含まれていますか？		m
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せんか？ 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）		
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？	*	
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）？	*	4
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）？	*	4
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せんか？	*	3
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せんか？	*	3
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。		m
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。		m
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。		m
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。		

複数選択法 =m, スケール (3 件法 =3, 4 件法 =4, データサイズ =s)