

# 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”<sup>1</sup>取組みである。論文のオープンアクセス（以下、「OA」と記す）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、OECD や 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)<sup>2</sup>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。そして2018年6月に閣議決定された「統合イノベーション戦略」<sup>3</sup>の、「第2章 知の源泉」には、「(2)オープンサイエンスのためのデータ基盤の整備」が掲げられており、「③今後の方向性及び具体的に講ずる主要施策」として“研究データの管理や公開・共有に従事する研究者等の意識向上や基礎的な知識の習得のための取組や、研究者や大学・国研等における現状、取組等についての調査・分析を行い、研究者等の意識向上等に資する方策を検討する”（p. 18）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、2018年10月から11月にかけて、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ウェブ調査紙調査を実施した。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。1,516名（回答率69.1%）の回答を分析し、2016年11月から12月にかけて実施したベースライン調査<sup>4</sup>（以下、「2016年調査」と記す）と比較することによって、政策立案や研究マネジメントに資することを目指した。調査項目は2016年調査を踏襲しつつ、その後の動向、すなわち研究助成機関・団体（以下、「助成機関」と記す）によるデータマネジメントプラン（以下、「DMP」と記す）の要求やオープンサイエンスに関する議論をふまえて、新たにDMPの作成経験やデータ公開によって得られたインセンティブ等に関する質問を行った。

---

<sup>1</sup> 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.  
[https://www8.cao.go.jp/cstp/kokusaiteki/g7\\_2016/2016communique.html](https://www8.cao.go.jp/cstp/kokusaiteki/g7_2016/2016communique.html)

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

<sup>3</sup> 内閣府. 統合イノベーション戦略. 2018, 82p.  
[https://www8.cao.go.jp/cstp/tougosenryaku/tougo\\_honbun.pdf](https://www8.cao.go.jp/cstp/tougosenryaku/tougo_honbun.pdf)

<sup>4</sup> 池内有為, 林和弘, 赤池伸一. 研究データ公開と論文のオープンアクセスに関する実態調査. 文部科学省科学技術・学術政策研究所, 2017, NISTEP RESEARCH MATERIAL No.268, 108p. <https://doi.org/10.15108/rm268>

## (1) データと論文の公開状況

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」）を図 3 に示す何らかの方法で公開した経験をもつ回答者（以下、「データ公開率」）は 51.9%（787 名）、論文を OA にした経験をもつ回答者（以下、「論文の OA 率」）は 78.0%（1,182 名）であった（図 1）。

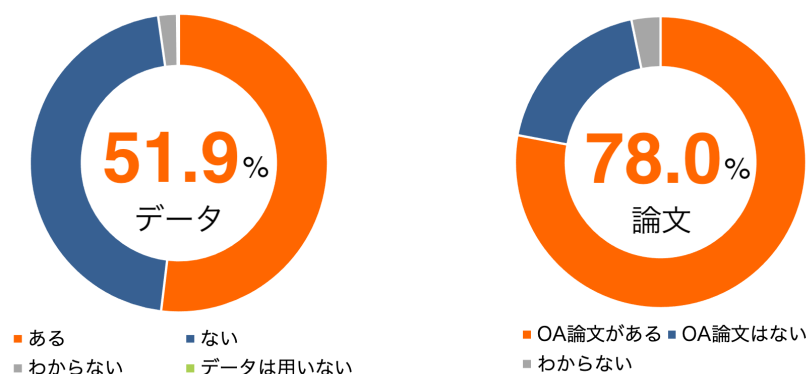


図 1 データと論文の公開経験 (いずれも n=1,516)

2016 年調査におけるデータ公開率は 51.0%であり、0.9 ポイントしか増加していなかったのに対して、論文の OA 率は 70.9%から 78.0%まで 7.1 ポイント増加していた。

図 2 に示すように、分野別のデータ公開率は、生物科学の 66.5%からの工学の 40.2%まで差がみられた。さらに、分野別の公開率の変化をみると、化学、数学、地球科学、心理学、人文学・社会科学は増加したのに対して、医学、計算機科学、物理学・天文学は減少しており、分野による増減の差も表れた。

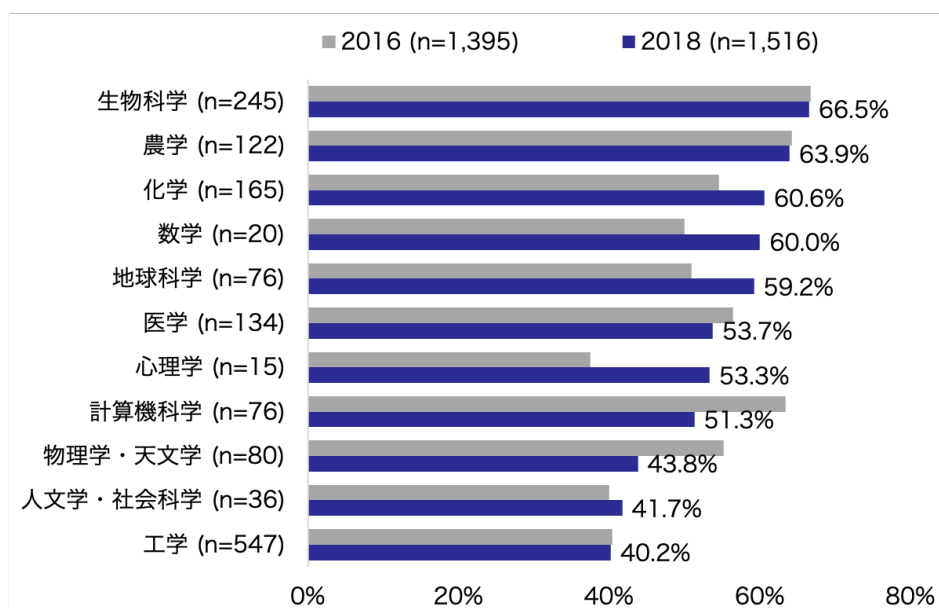


図 2 分野別データ公開率

データの公開方法は、「論文の補足資料」(47.6%)、「個人や研究室のウェブサイトへの掲載」(45.7%)の順に選択率が高く(図 3)、2016 年調査とは順位が逆転していた。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」が 27.1%、「特定分野のリポジトリ」が 18.6%にとどまっていたが、「特定分野のリポジトリ」の割合は他に比較して増加していた。「学術系 SNS」<sup>5</sup>(7.2%)、「コード共有サービス」<sup>6</sup>(5.3%)、「データ共有サービス」<sup>7</sup>(2.4%)の選択率は低かったものの、後述するように公開データの入手先としては比較的によく用いられていた(図 5)。

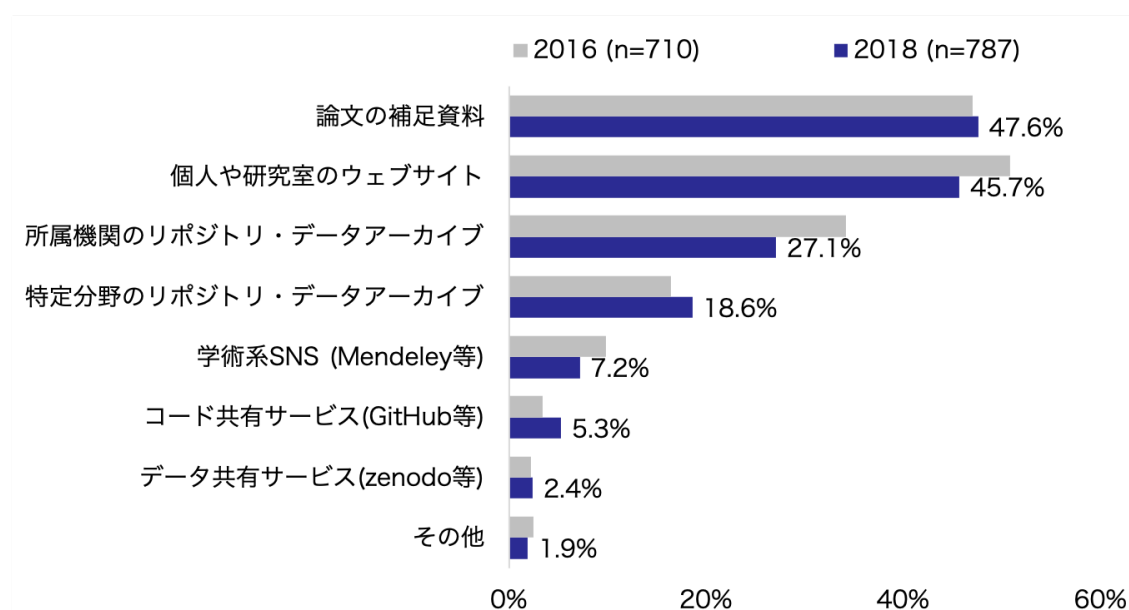


図 3 データの公開方法（複数回答）

データを公開した理由のうち選択率が高かったのは「研究成果を広く認知してもらいたいから」(54.9%)、次いで「論文を投稿した雑誌のポリシー（投稿規定）だから」(41.7%)であった。2016 年調査と選択率は異なるものの、順位に変化はなかった。論文の公開理由は「論文を投稿した雑誌がオープンアクセスだから」(78.3%)、次いで「研究成果を広く認知してもらいたいから」(52.6%)であった。

データを公開していない理由のうち選択率が高かったのは、「論文を投稿した雑誌のポリシーではないから」(31.9%)、「ニーズがないと思うから」(28.2%)であった。この 2 つが比較的多く選ばれているものの、2016 年調査と同様に突出した理由はみられなかった。一方、論文を公開していない理由は「投稿したい雑誌がオープンアクセスではないから」(55.4%)、次いで「資金がないから」(42.8%)に集中しており、これも 2016 年調査と同様の傾向であった。

<sup>5</sup> 質問紙では、学術系 SNS の例として Mendeley (Elsevier) と ResearchGate を示した。

<sup>6</sup> ソースコードを共有できる GitHub を示した。

<sup>7</sup> 無料でデータを公開できる figshare と zenodo (CERN) を示した。

データの未公開理由を尋ねた後に、その問題が解決された場合のデータの公開意思を尋ねた結果、2016年調査よりも「はい」の比率が低く（21.4%）、「いいえ」の比率が高くなっていた（30.7%）。また、「わからない」（48.0%）の比率も依然として高く、データ公開に対する慎重な姿勢がうかがえた（図 4）。

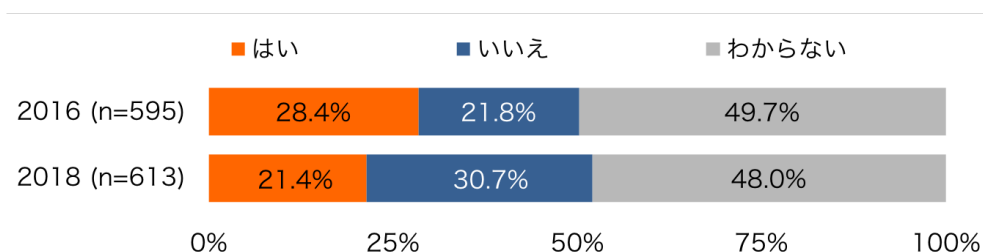


図 4 未公開理由が解決した場合のデータの公開意思

## (2) 公開データの利用状況と課題

公開データの入手経験がある回答者は 73.0%であった。公開データの入手方法のうち、最も選択率が高かったのは「個人や研究室のウェブサイト」(59.5%)、次いで「論文の補足資料」(55.7%)であった（図 5）。データ公開と同様に、入手においてもウェブサイトや論文の補足資料が中心であることがわかった。ただし、入手については「学術機関のリポジトリ」も比較的によく使われており（46.3%）、また、「特定分野のリポジトリ」や「コード共有サービス」の割合が増加していた。

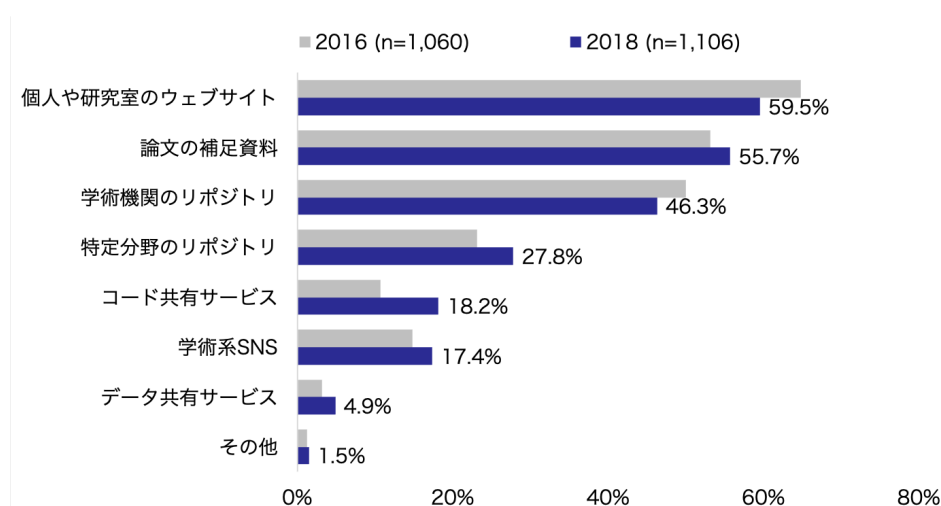


図 5 公開データの入手方法（複数回答）

公開データの入手の障壁を確認すると、1位は「データごとに品質が異なる」（37.3%）、2位は「データごとにフォーマットが異なる」（34.0%）、3位は「利用条件（営利利用が可能かどうかなど）がよくわからない」（31.5%）であった（図 6）。一方、2016年調査の1

位は「利用料金が必要」(43.1%)、2位は「利用者登録が必要」(33.3%)、3位は「利用条件がよくわからない」(33.1%)であった。2016年調査の時点では、データの入手時点で障壁を感じる回答者が多かったが、2018年調査では、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といえよう。

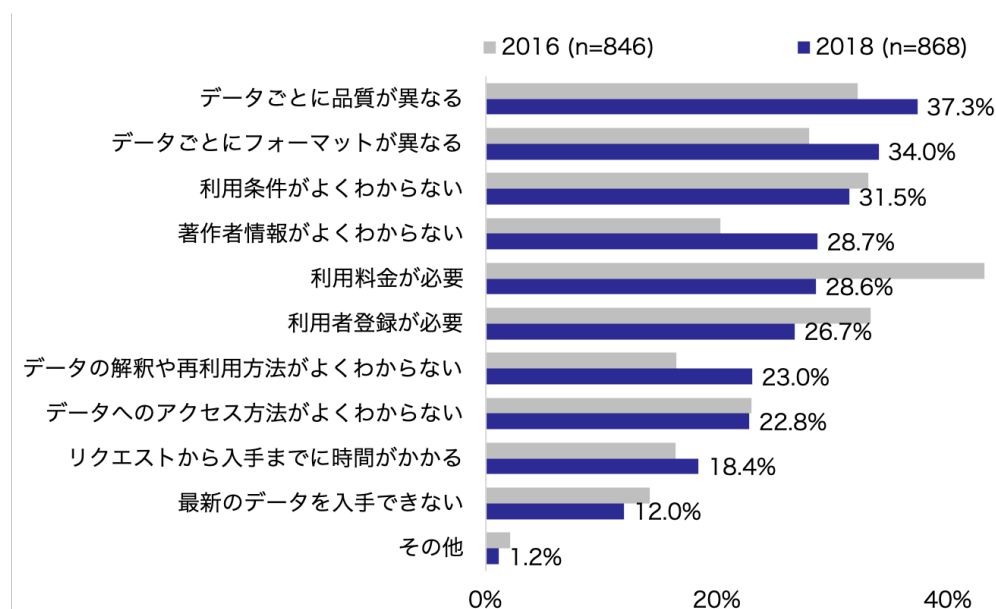


図 6 データ入手の障壁（複数回答）

### (3) データマネジメントプラン（DMP）の作成状況

DMPの作成経験をもつ回答者は18.7%、「なし」は76.1%、「わからない」は5.2%であった（図7）。Springer Nature社が日本の研究者を対象として実施した質問紙調査<sup>8</sup>では、56%がDMPの作成経験があると回答していたため、本調査の結果とは差がみられた。

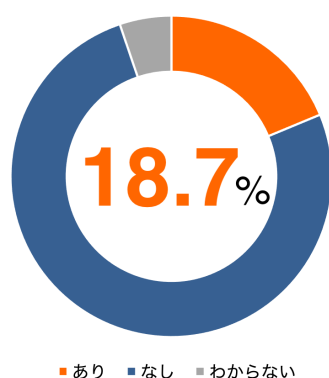


図 7 DMPの作成経験（n=1,516）

<sup>8</sup> Challenges and Opportunities for Japanese Researchers. Springer Nature, 2018.  
<https://www.springernature.com/gp/open-research/open-data/japan-survey>

作成経験がある DMP のうち、最も多かったのは、「所属機関の DMP」(46.8%)、次いで「個人や研究グループのための DMP」(38.0%)であった(図 8)。助成機関の DMP は、「科学技術振興機構 (JST)」が 26.1%,「新エネルギー・産業技術総合開発機構 (NEDO)」が 5.6%,「日本医療研究開発機構 (AMED)」が 3.5%であった。

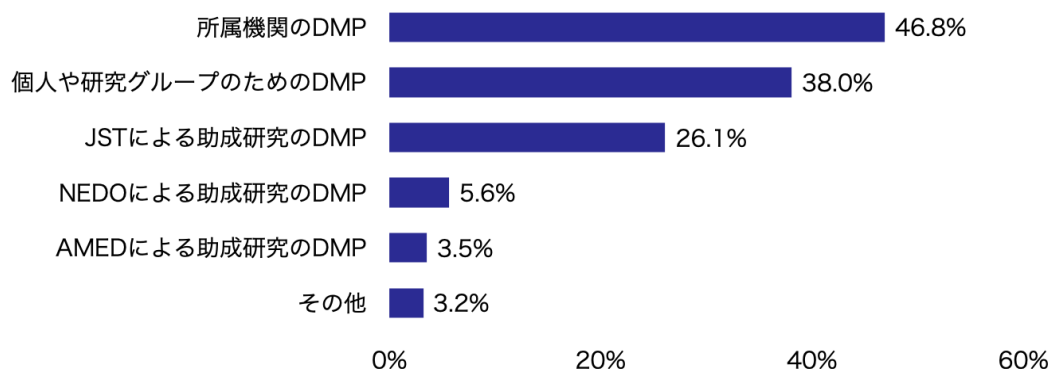


図 8 作成経験がある DMP (n=284, 複数回答)

#### (4) 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。全体的に不十分であるという認識をもつ回答者が多く、特に「データの整備・公開のための人材」(「不十分」と「やや不十分」の合計 84.6%),「データの整備・公開のための時間」(同 80.3%),「データの整備・公開のための資金」(同 78.7%)の充足度が低かった。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、「十分」と「ほぼ十分」をあわせて 47.6%であり、充足していると考えている回答者は半数に満たなかった(図 9)。ただし「公開用のリポジトリ」の充足度を「わからない」とする回答者は、2016 年調査の 27.1%から 17.2%まで 9.9 ポイント減少しており、公開用のリポジトリの認知度が向上した可能性が示唆された。

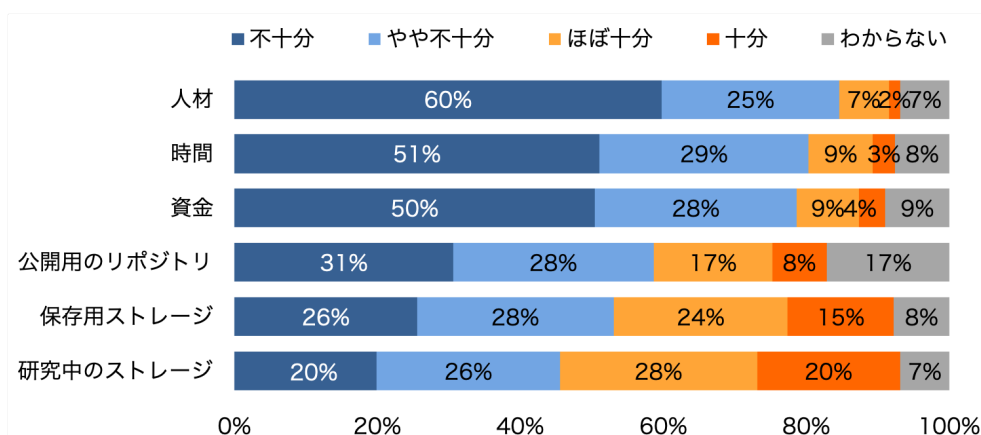


図 9 データ公開に関する資源の充足度 (n=1,513)



データを公開する場合の懸念は、「引用せずに利用される可能性」（「問題である」と「やや問題である」の合計 84.2%），次いで新規項目の「データの所有権・契約」（同 75.9%），「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」（同 69.1%），「誤解や誤用の可能性」（同 66.4%），「二次利用に関して責任が生じる可能性」（同 66.0%）の順に懸念が強かった（図 10）。

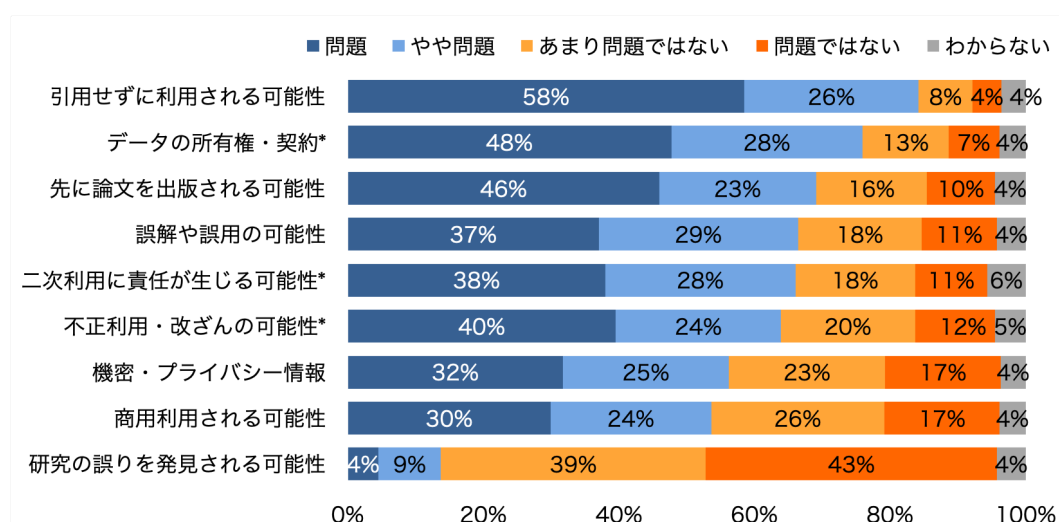


図 10 データを公開する場合の懸念の強さ（n=1,396）

2016 年調査や先行研究によってデータ公開に対する懸念が強いことが明らかにされてきた。そこで研究者が懸念しているような問題が実際に起きているのかどうかを明らかにするために、データ公開経験をもつ回答者に自由記述で尋ねた結果、223 名から回答が得られた。うち、何らかの問題が起きたと記述していたのは 24.2%であり、75.8%は特に問題が起きていない旨を記述していた。表 1 に示すように、最も多かったのは公開したデータに対する「問い合わせ等への対応」（7.6%）であった。次いで「引用せずに利用された」（6.3%）が多く、うち 3 名は盗用・剽窃行為であると指摘していた。

表 1 データ公開によって実際に起きた問題

内容	人数	比率
なし	169	75.8%
問い合わせ等への対応	17	7.6%
引用せずに利用された	14	6.3%
先取権の喪失	9	4.0%
誤用された	7	3.1%
更新のコストがかかる	2	0.9%
徒労感	2	0.9%
その他	3	1.3%
合計	223	100.0%

## (5) データ公開のインセンティブ

データ公開のインセンティブを明らかにするために、まず、データ公開経験をもつ回答者に実際に得られた良い結果を尋ねたところ、195 名から回答が得られた。表 8 に回答の内容を 7 項目に分類した結果を示す。なお、1 名の回答に複数の内容を含むコメントはそれぞれカウントしたため、合計 256 件となっている。最も多かったのは「研究上の利点」(40.6%)、次いで「研究・データ・研究者のビジビリティ向上」(25.8%)、「科学・分野の進展」(10.5%)、「人とのつながり」(10.2%)の順であった。

表 2 データ公開によって得られた良い結果 (n=195)

内容	件数	比率
研究上の利点	104	40.6%
研究・データ・研究者のビジビリティ向上	66	25.8%
科学・分野の進展	27	10.5%
人とのつながり	26	10.2%
評価	11	4.3%
個人的な利点	9	3.5%
その他	13	5.1%
合計	256	100.0%

また、研究にデータを用いている回答者を対象として、データ公開によって得られるインセンティブの重要度を尋ねた。最も重要であると考えられていたのは「データに紐づいた論文の引用」(「重要」と「やや重要」の合計 90.5%)、次いで「データの引用 (論文と同様に、参考文献リストにデータ作成者やデータ名、識別子などを記載する)」(同 88.0%)であった (図 11)。図 10 に示したように「引用されずに利用される可能性」は最も重要な懸念であったのと同時に、「論文やデータを引用されること」はデータ公開のインセンティブとして重要視されていることが明らかになった。

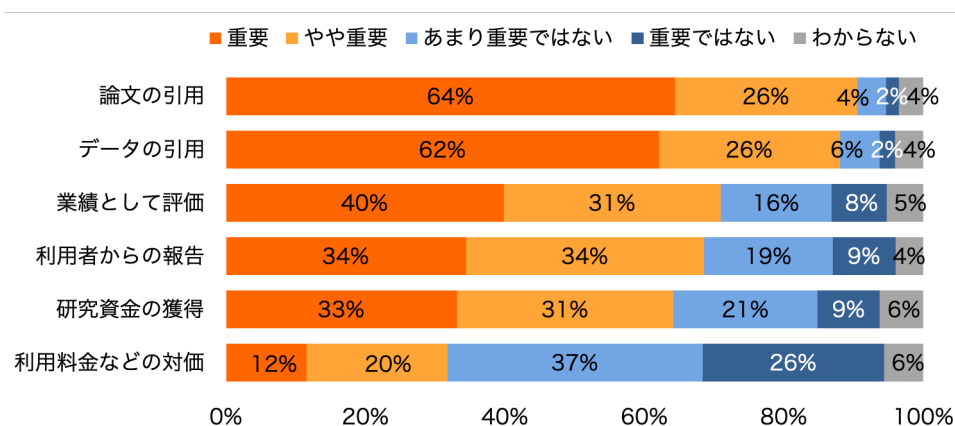


図 11 データ公開によるインセンティブの重要度 (n=1,513)

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、助成機関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、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のかを継続的に調査するとともに、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。