

研究データ公開と論文のオープンアクセスに関する 実態調査 2018

2020 年 5 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

【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

【Authors】

IKEUCHI, Ui	Affiliated Fellow Science and Technology Foresight Center, National Institute of Science and Technology Policy (NISTEP), MEXT
HAYASHI, Kazuhiro	Senior Research Fellow 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「研究データ公開と論文のオープンアクセスに関する実態調査2018」，NISTEP RESEARCH MATERIAL, No. 289, 文部科学省科学技術・学術政策研究所.

DOI: <https://doi.org/10.15108/rm289>

IKEUCHI, Ui and HAYASHI, Kazuhiro “A Survey on Open Research Data and Open Access 2018,” NISTEP RESEARCH MATERIAL, No. 289, National Institute of Science and Technology Policy, Tokyo.

DOI: <https://doi.org/10.15108/rm289>

研究データ公開と論文のオープンアクセスに関する実態調査 2018

池内 有為, 林 和弘

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可能な限り社会に広く公開するオープンサイエンスによって、科学の発展やイノベーションの創出が期待されている。オープンサイエンス政策は世界的に推進されており、統合イノベーション戦略（2018年）にその実態把握が明記された。

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月にかけて科学技術専門家ネットワークを活用したウェブ質問紙調査を実施した。結果、回答者1,516名のうち51.8%がデータの、78.0%が論文の公開経験を有していた。2016年調査の結果と比較すると、全体としてはデータの公開が進んでいなかったが、分野による差が表れた。また、助成機関等が要求するデータマネジメントプラン（DMP）の作成経験をもつ回答者は18.7%にとどまった。データ公開経験をもつ回答者によって研究上のインセンティブがあることが示されたものの、研究者のデータ公開に対する懸念は依然として強く、84.2%が引用せずに利用される可能性を、75.9%がデータの所有権・契約を、69.1%が先に論文を出版される可能性を、「問題」または「やや問題」として認識していた。また、データを公開しようとする場合の資源の不足感も強く、84.6%は人材が、80.3%は時間が、78.7%は資金が、それぞれ「不足」または「やや不足」として認識していた。

A Survey on Open Research Data and Open Access 2018

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

IKEUCHI, Ui and HAYASHI, Kazuhiro

ABSTRACT

Open science, in which the results of publicly funded research and data published as open as possible to society, is expected to advance science and create innovation. Open science policies are being promoted globally, and their realization and monitoring has been stipulated in the Integrated Innovation Strategy (2018).

The National Institute of Science and Technology Policy (NISTEP) Science and Technology Foresight Center conducted a web questionnaire survey from October to November 2018 using its Science and Technology Experts Network. As a result, of the 1,516 respondents, 51.8% of the respondents had data, and 78.0% had experience in publishing papers. Compared to the results of the 2016 survey, data disclosure has not so progressed as a whole, but there are differences between research fields. Only 18.7% of respondents have experience in creating a data management plan (DMP)

required by a funding organization. Although respondents with experience in data disclosure have shown incentives for research, concerns about data disclosure by researchers remain strong, they recognized that it was a "problem" or a "slight problem" to the possibility of being used without citations (84.2%), data ownership and contracts (75.9%), and the possibility of publishing a paper earlier (69.1%). There is also a strong sense of lack of resources when trying to release data, with 84.6% recognizing that human resources, 80.3% have time, and 78.7% have funds, which are "insufficient" or "slightly insufficient", respectively.

目次

概要	i
(1) データと論文の公開状況	ii
(2) 公開データの利用状況と課題	iv
(3) データマネジメントプラン（DMP）の作成状況	v
(4) データ公開の障壁	vi
(5) データ公開のインセンティブ	viii
1. 調査の概要	1
1.1 オープンサイエンスの動向と本調査の目的	1
1.2 主要な概念の定義	2
1.3 調査項目	4
(1) 新規質問項目	6
(2) 修正・追加した項目	7
(3) 削除した項目	7
1.4 調査対象	8
1.5 調査方法と期間	9
1.6 回答率と回答者の属性	9
(1) 回答者の所属	10
(2) 回答者の年齢層	10
(3) 回答者の専門分野	10
(4) カレントデータの量	11
2. 調査結果	13
2.1 データと論文の公開状況	13
(1) データと論文の公開経緯	13
(2) データの公開方法	18
(3) データの公開理由	19
(4) 論文の公開理由	20
(5) データの未公開理由	21
(6) 論文の未公開理由	23
(7) データと論文の公開意思	24
2.2 データの提供（共有）状況	25
(1) データの提供頻度	25
(2) データの提供方法【新規】	27
2.3 公開データの利用状況と課題	28

(1) 公開データの入手経験	28
(2) 公開データの入手方法	29
(3) 公開データ入手の障壁	30
(4) 公開データの利用目的	31
(5) 公開データの探索方法	33
(6) 論文の探索方法	34
2.4 データマネジメントプラン（DMP）の作成状況【新規】	35
(1) DMP の作成経験	35
(2) 作成経験がある DMP の種類	36
(3) DMP の作成理由	37
(4) DMP の未作成理由	38
2.5 データ公開の障壁	38
(1) カレントデータの公開に必要な資源の状況	38
(2) カレントデータの公開に対する懸念【項目追加】	40
(3) 論文の OA に対する懸念【新規】	41
(4) データ公開のネガティブな影響【新規】	42
(5) 他の研究者によるカレントデータの理解	44
2.6 データ公開のインセンティブ【新規】	44
(1) データ公開のポジティブな影響	44
(2) データ公開によるインセンティブの重要度	46
2.7 データの整備・公開に対する認識	47
(1) データの整備・公開に関するリテラシーへの関心	47
(2) データを整備・公開する上でより詳しく知りたい項目	47
(3) データ整備・公開の専門性	48
3. まとめ	51
謝辞	54
参考文献	55
資料	61
(1) 調査依頼文・質問票	61
(2) 単純集計表	77
(3) クロス集計表	96

図表目次

図 1	データと論文の公開経験（いずれも n=1,516）	ii
図 2	分野別データ公開率	ii
図 3	データの公開方法（複数回答）	iii
図 4	未公開理由が解決した場合のデータの公開意思	iv
図 5	公開データの入手方法（複数回答）	iv
図 6	データ入手の障壁（複数回答）	v
図 7	DMP の作成経験（n=1,516）	v
図 8	作成経験がある DMP（n=284, 複数回答）	vi
図 9	データ公開に関する資源の充足度（n=1,513）	vi
図 10	データを公開する場合の懸念の強さ（n=1,396）	vii
図 11	データ公開によるインセンティブの重要度（n=1,513）	viii
図 12	回答者の所属（n=1,516）	10
図 13	回答者の専門分野（n=1,516）	11
図 14	カレントデータの量（n=1,513）	11
図 15	データと論文の公開経験（n=1,516）	13
図 16	データ公開経験の経年変化	14
図 17	論文の公開経験の経年変化	14
図 18	分野別データ公開率	15
図 19	分野別論文の公開率	16
図 20	分野別データと論文の公開率	17
図 21	データの公開方法（複数回答）	18
図 22	データの公開理由（複数回答）	19
図 23	論文の公開理由	20
図 24	データの未公開理由（複数回答）	21
図 25	論文の未公開理由	23
図 26	データの公開意思	24
図 27	論文の公開意思	24
図 28	データの提供経験	25
図 29	分野別データ提供経験	26
図 30	データの提供方法（n=1,027, 複数回答）	27
図 31	公開データの入手経験の有無	28
図 32	分野別公開データの入手経験	28
図 33	公開データの入手方法	29
図 34	データ入手の障壁の有無	30
図 35	データ入手の障壁（複数回答）	31
図 36	公開データの利用目的（n=1,106, 複数回答）	32
図 37	2016 年調査：公開データの利用目的（n=1,060）	32
図 38	公開データの探索方法	33

図 39	論文の探索方法	34
図 40	DMP の作成経験 (n=1,516)	35
図 41	分野別 DMP の作成経験 (n=1,516)	36
図 42	作成経験がある DMP (n=284, 複数回答)	36
図 43	DMP の作成理由 (n=284, 複数回答)	37
図 44	DMP の未作成理由 (n=1,153, 複数回答)	38
図 45	データ公開に関する資源の充足度 (n=1,513)	39
図 46	2016 年調査：データ公開に関する資源の充足度 (n=1,396)	39
図 47	カレントデータの公開に関する懸念 (n=1,513)	40
図 48	2016 年調査：カレントデータの公開に関する懸念 (n=1,396)	41
図 49	論文の OA に対する懸念 (n=1,516)	41
図 50	データ公開によって問題が起きた経験 (n=223)	42
図 51	他の研究者によるデータの理解 (n=1,513)	44
図 52	2016 年調査：他の研究者によるデータの理解 (n=1,396)	44
図 53	データ公開によるインセンティブの重要度 (n=1,513)	46
図 54	データリテラシーへの関心の有無 (n=1,513)	47
図 55	データを整備・公開する上でより詳しく知りたい項目	48
図 56	データの整備や公開における専門性の有無 (n=1,513)	49
図 57	専門性を必要とする項目 (n=1,426)	49
表 1	データ公開によって実際に起きた問題	vii
表 2	データ公開によって得られた良い結果 (n=195)	viii
表 3	調査目的と調査項目	4
表 4	質問項目一覧	5
表 5	回答者の年齢層	10
表 6	データの未公開理由：「その他」の回答 (n=73)	22
表 7	データ公開によるデメリット	43
表 8	データ公開によって得られた良い結果 (n=195)	45
表 9	データ公開によって得られた良い結果の詳細 (n=195)	45
表 10	所属機関	77
表 11	年齢層	77
表 12	回答者の研究分野	78
表 13	オープンアクセスの状況	78
表 14	論文を OA にした理由 (n=1,182, 複数回答)	79
表 15	論文が OA ではない理由 (n=285, 複数回答)	79
表 16	オープンアクセスの意思	80
表 17	論文やプレプリントを OA にする際の障壁 (n=1,516)	80
表 18	論文の探索ツール (n=1,516, 複数回答)	81
表 19	データ提供経験	82
表 20	データ提供方法 (n=1,027, 複数回答)	82

表 21	公開データの入手経験	83
表 22	公開データの入手方法 (n=1,106, 複数回答)	83
表 23	公開データ入手における問題の有無	84
表 24	公開データを入手する際の問題 (n=868, 複数回答)	84
表 25	公開データの利用目的 (n=1,106, 複数回答)	84
表 26	公開データの探索ツール (n=1,103, 複数回答)	85
表 27	DMP 作成経験の有無	86
表 28	作成経験がある DMP (n=284, 複数回答)	86
表 29	DMP を作成した理由 (n=284, 複数回答)	86
表 30	DMP を作成していない理由 (n=1,232, 複数回答)	87
表 31	データ公開経験の有無	88
表 32	データを公開した方法 (n=787, 複数回答)	88
表 33	データを公開した理由 (n=713, 複数回答)	89
表 34	データを公開していない理由 (n=695, 複数回答)	90
表 35	データ公開意思	90
表 36	カレントデータの量	91
表 37	カレントデータ公開の有無	92
表 38	カレントデータの管理資源の充実度 (n=1,513)	92
表 39	カレントデータを公開する際の障壁 (n=1,513)	93
表 40	カレントデータの理解：同じ分野	93
表 41	カレントデータの理解：他分野	94
表 42	データの整備・公開方法への関心の有無	94
表 43	データの整備・公開について詳しく知りたい項目 (n=1,400, 複数回答) .	94
表 44	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	95
表 45	データの整備や公開に専門性を必要とする項目 (n=1,426, 複数回答)	95
表 46	分野別集計結果	96

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)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。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」³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⁴（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。

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.
https://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.
<https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>

³ 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

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

(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）。

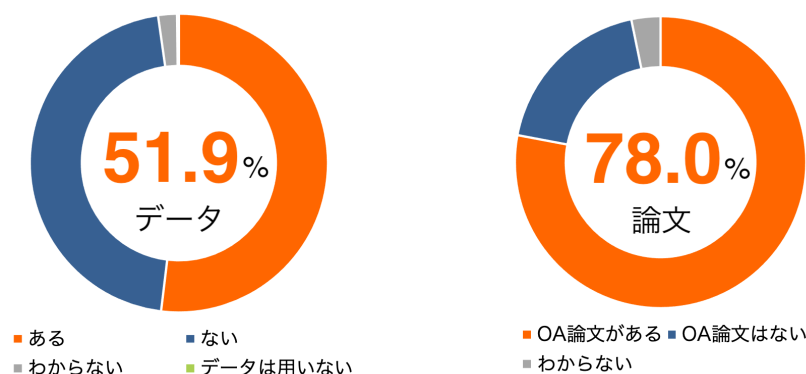


図 1 データと論文の公開経験 (いずれも n=1,516)

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。

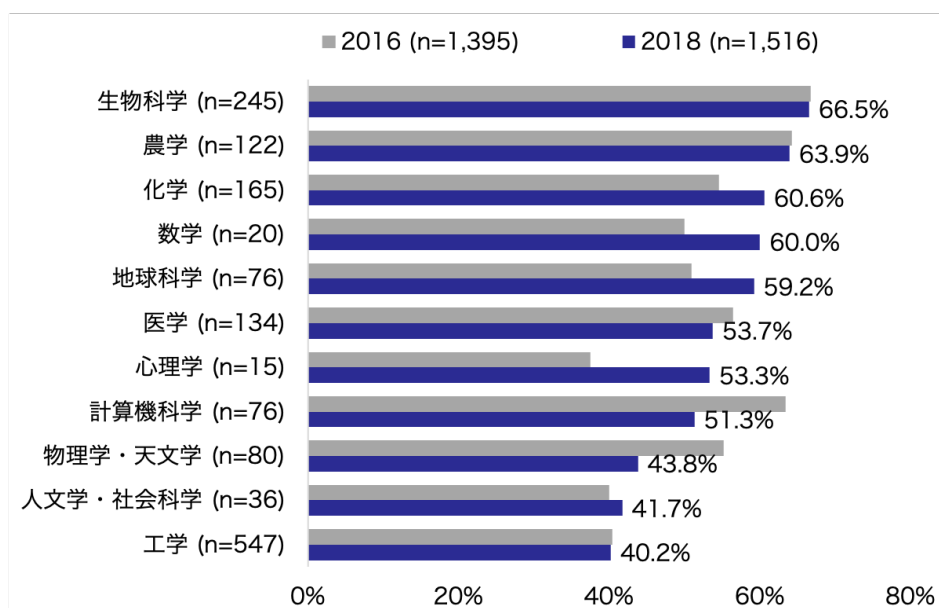


図 2 分野別データ公開率

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」⁵（7.2%）、「コード共有サービス」⁶（5.3%）、「データ共有サービス」⁷（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好く用いられていた（図 5）。

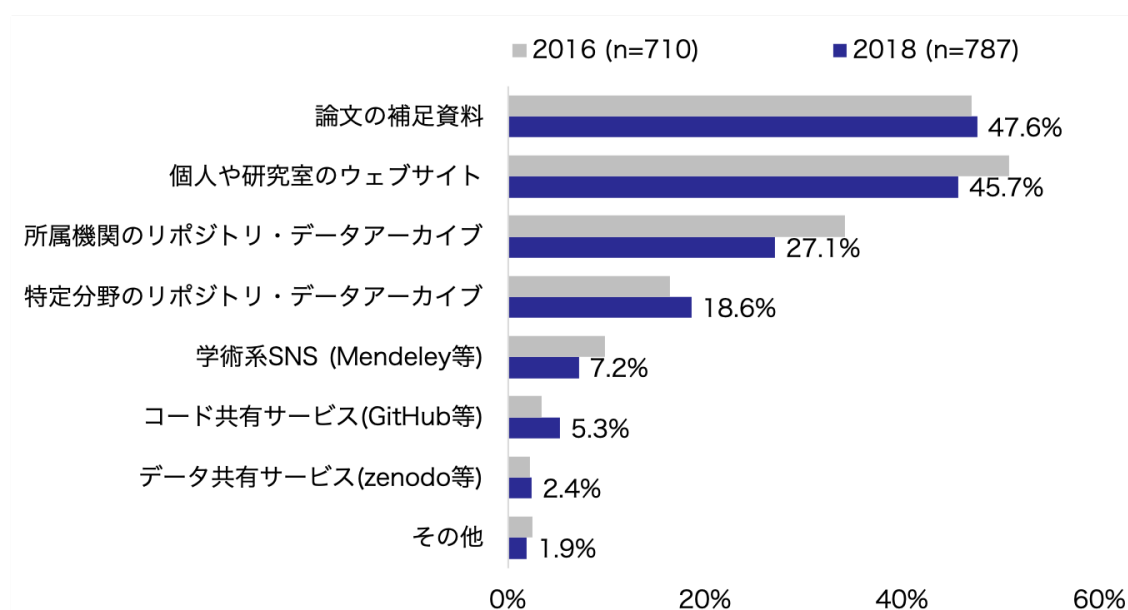


図 3 データの公開方法（複数回答）

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。

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

⁶ ソースコードを共有できる GitHub を示した。

⁷ 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）。

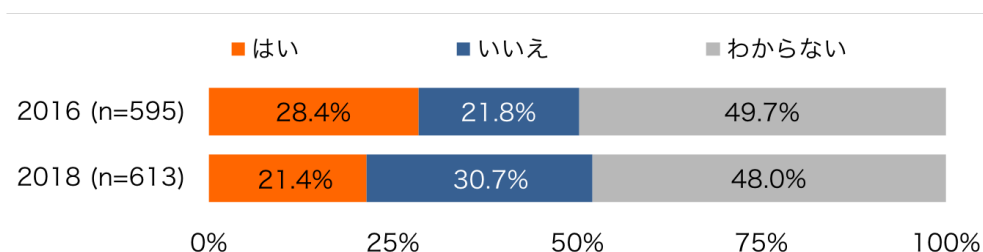


図 4 未公開理由が解決した場合のデータの公開意思

(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。

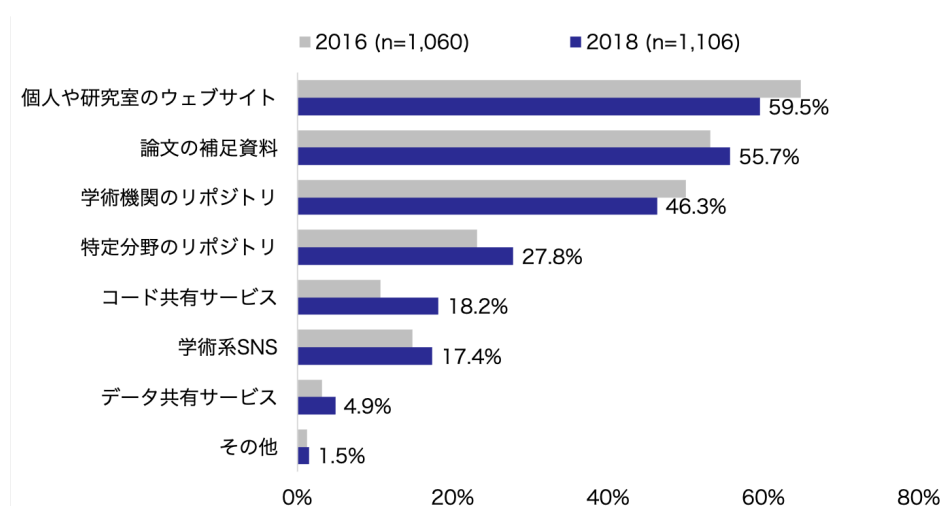


図 5 公開データの入手方法（複数回答）

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。

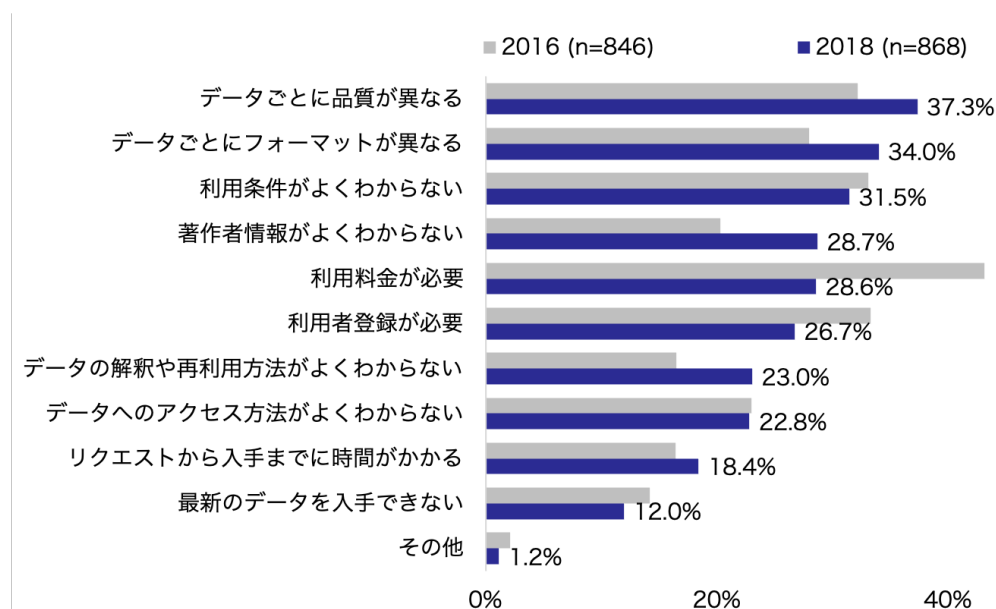


図 6 データ入手の障壁（複数回答）

(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⁸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。

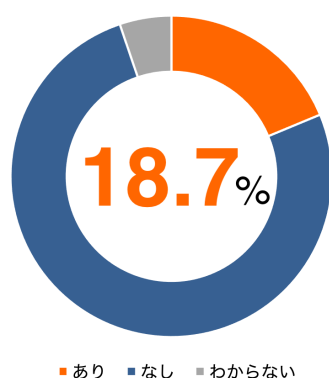


図 7 DMP の作成経験（n=1,516）

⁸ 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。

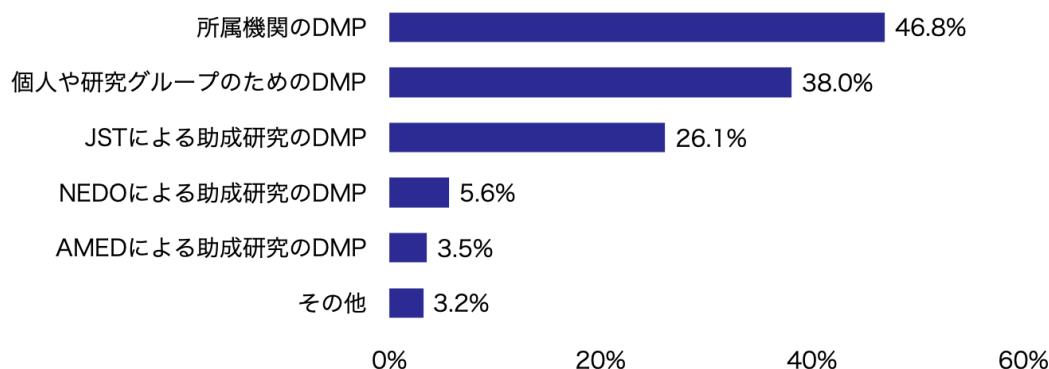


図 8 作成経験がある DMP (n=284, 複数回答)

(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。

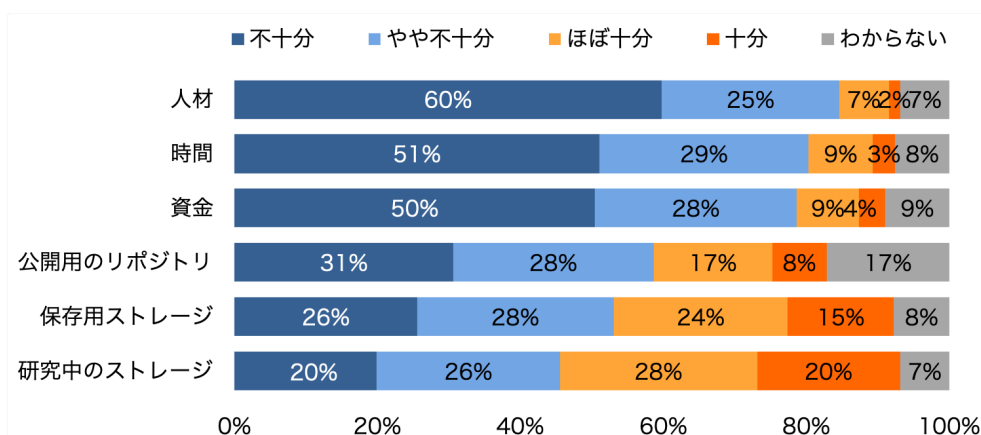


図 9 データ公開に関する資源の充足度 (n=1,513)

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）。

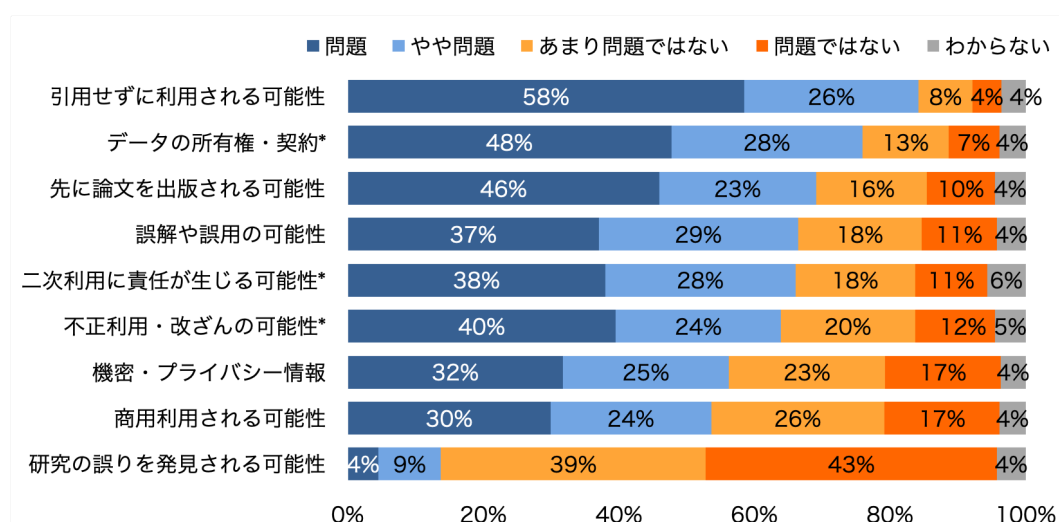


図 10 データを公開する場合の懸念の強さ（n=1,396）

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。

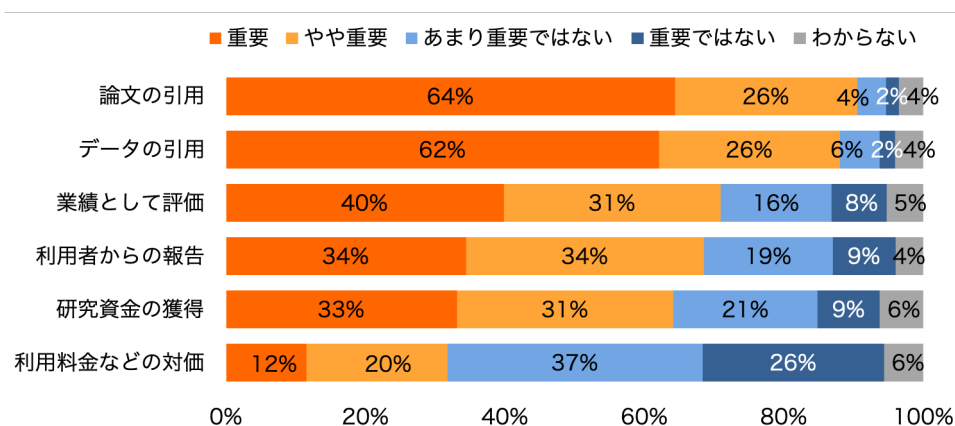


図 11 データ公開によるインセンティブの重要度 (n=1,513)

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。

本編

1. 調査の概要

1.1 オープンサイエンスの動向と本調査の目的

科学研究の成果として論文の根拠となる研究データを公開し、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）が公開した『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や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、各国・地域の政府組織や学術機関による対応が加速している⁴。研究データの共有や公開によって、研究の効率化⁵や研究不正への対策⁶、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⁷、イノベーションの創出などが期待されている。2020年1月にはGoogle Dataset Search^{ix}の正式版がリリースされるなど、データ公開とその活用は産学官を超えた大きなうねりとなっている。また、研究データの公開と論文のオープンアクセス（以下、「OA」と記す）は「オープンサイエンス」と呼ばれるようになり⁸、オープンサイエンス政策の策定やオープンサイエンスのための基盤の開発が進められている⁹。

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きた^{10 11 12}。一方、日本の研究者を対象とした調査は十分に行われていなかったため、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てウェブ質問紙調査^{13 14 15}を実施した（以下、「2016年調査」と記す）。本調査は、この2016年調査をベースライン調査と位置づけて、オープンサイエンスの実践状況及び課題の経年変化や新たな動向の影響について明らかにするために、2018年10月から11月にかけて実施されたフォローアップ調査である。ここでは、2016年調査以降の日本の主な動きとして、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（Data Management Plan, 以下「DMP」と記す）の要求と「統合イノベーション戦略」（2018年）について述べる。

各国・地域における研究データ公開を促す方策の一つとして、助成機関が研究資金を申請する際にデータの整備・公開・保存の方法などを記したDMP（書類）の提出を義務化する場合が増えている。DMPは、2003年に米国衛生研究所（NIH）が義務化したのを先駆けとして、英国、欧州、オーストラリア、カナダ、南アフリカなどの助成機関が義務化を進めており、これによってデータ公開が推進されているとの指摘がある。日本においては、2017年から科学技術振興機構（JST）が、2018年からは新エネルギー・産業技術総合開発機構（NEDO）と日本医療研究開発機構（AMED）がDMPの提出を求めるようになった¹⁶。そこで本調査では、日本の研究者のDMPの作成状況についても明らかにすることとした。

^{ix} <https://datasetsearch.research.google.com>

「第5期科学技術基本計画（2016～2020年度）」¹⁷において、“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”（p.32）と述べられている。そのフォローアップとして2018年6月に閣議決定された「統合イノベーション戦略」¹⁸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」が挙げられており、“公的資金による研究成果としての研究データについては、データインフラを通して機械判読可能化を促進する”（p.16-19）ことが主要施策の1つとされている。そして、「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）と述べられており、2016年調査は“先行的な調査”（p.19、注記）として言及されている。

以上の動向をふまえた本調査の目的は、日本の研究者によるデータと論文の公開状況や認識を、2016年調査や先行研究と比較することによって経年的に明らかにすることである。また、新たにDMPの作成状況やデータ公開のインセンティブについて明らかにすることとした。具体的には、(1)データと論文の公開状況、(2)データの提供（共有）状況、(3)公開データの利用状況と課題、(4)DMPの作成状況、(5)データ公開の障壁、(6)データ公開のインセンティブ、(7)データの整備・公開に対する認識を調査し、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を目指した。また、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、図書館職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されることを期待している。

1.2 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を以下に示す。2016年調査においても同様の定義を用いた。

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。2016 年調査のプレテストの結果、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、回答者がどのデータについて回答すればよいかを明確にするために定義した。

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（雑誌自体が OA）、グリーン OA（購読費モデルの雑誌論文の著者最終原稿をリポジトリ等で公開）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。

研究分野

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 分野とした。具体的には、(1)工学、(2)天文学、(3)化学、(4)物理学、(5)地球科学、(6)数学、(7)計算機科学、(8)農学、(9)生物科学、(10)医学、(11)心理学、(12)社会科学、(13)人文学である。なお、質問紙では(7)計算機科学は“コンピュータサイエンス”と記した。また、選択肢として「その他」も提示した。

1.3 調査項目

調査項目は、原則として 2016 年調査と同様とした。調査項目の設定にあたって参照した先行研究や項目の設定意図の詳細は、2016 年調査の報告書 (RM268) ¹³ をご確認ください。表 3 に調査目的と、それらを明らかにするための調査項目を示した。本調査の新規項目 (*印) や 2016 年調査からの修正点については後述する。

表 3 調査目的と調査項目

調査目的	調査項目
1. データと論文の公開状況	データと論文の公開状況 データと論文の公開理由と未公開理由 未公開理由が解決した場合の公開意思
2. データの提供（共有）状況	データの提供（共有）経験 データの提供（共有）方法
3. 公開データの利用状況と課題	公開データの利用状況と目的 公開データの入手における障壁 公開データの探索ツール
4. データマネジメントプラン(DMP)の作成状況	DMP の作成状況* DMP の作成理由と未作成理由*
5. データ公開の障壁	データを公開する場合の資源の充足度 データや論文を公開する場合の懸念 データを公開した際に実際に起きた問題*
6. データ公開のインセンティブ	データを公開した際に得られた良い経験* データ公開によるインセンティブの重要度*
7. データの整備・公開に対する認識	データの整備・公開について関心がある項目 データの整備・公開の専門性

*新規調査項目

作成した質問項目は、回答しやすいように表 4 の順番に配列した。まず、事実として「1. 研究分野」から「6. 研究データの公開」について尋ねた上で、「7. 最近の研究のためのデータ」を想定して頂いた上で、データを公開・整備する場合の状況や認識について尋ねた。実際の質問票は資料として掲載する。

表 4 質問項目一覧

No.	質問	必須	方式	新規
1. 研究分野について				
Q1	研究分野	○		
2. 学術論文について				
Q2	論文のオープンアクセス (OA) 経験	○		
Q3	論文を OA にした理由	○	m	
Q4	論文を OA にしていない理由	○	m	
Q5	論文の OA 意思	○		
Q6	OA への懸念	○	4	新規
Q7	論文の探索ツール		m	
3. 研究データの提供について				
Q8	データの提供 (共有) 経験		4	
Q9	データの提供 (共有) 方法		m	新規
4. 公開データの利用について				
Q10	公開データの入手経験	○	m	
Q11	公開データ入手における問題	○	m	
Q12	入手した公開データの利用目的		m	修正
Q13	公開データの探索ツール		m	
5. データマネジメントプラン (DMP) について				
Q14	DMP の作成経験	○	m	新規
Q15	DMP を作成した理由	○	m	新規
Q16	DMP を作成していない理由	○	m	新規
6. 研究データの公開について				
Q17	研究データの公開経験	○	m	
Q18	研究データを公開した理由	○	m	
Q19	研究データの公開によって良い結果が得られた経験		記述	新規
Q20	研究データの公開によって問題が起きた経験		記述	新規
Q21	研究データを公開していない理由	○	m	
Q22	研究データの公開意思	○		
Q23	研究データ公開のインセンティブの重要度	○	4	新規
7. 最近の研究のためのデータについて				
Q24	カレントデータの総量		s	
Q25	カレントデータの公開経験	○		
Q26	カレントデータの整備・公開に必要な資源の充足度	○	4	修正
Q27	カレントデータを公開しようとする場合の懸念	○	4	追加
Q28	カレントデータの理解の容易さ (同分野)		3	
Q29	カレントデータの理解の容易さ (異分野)		3	
Q30	カレントデータの整備・公開に関する知識・関心		m	
Q31	カレントデータの整備や公開の専門性		m	
8. 自由回答				
Q32	オープンサイエンスや調査に対する自由回答		記述	

※方式：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)

以下では表 4 に示した質問項目のうち、(1)新規に設定した項目、(2)修正を加えた項目、(3)削除した項目について述べる。なお、2016 年調査と同様に、本調査においても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、先行研究を参照しつつ似たような選択肢は統合すること、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抑制するために、「わからない」という選択肢を加えることとした。

(1) 新規質問項目

2016 年調査の結果から導いた仮説、及び 2016 年調査の後に出版された先行研究や政策の状況をふまえて、8 件の質問を新たに設定した。

- a) データ公開の懸念について深掘り検証するための質問（データ公開への懸念が強いが論文の公開に対しても同様の懸念を抱くのではないかと、また、研究者が懸念していることは実際に起きているのか）

Q6. 論文やプレプリントをオープンアクセスにしようとする場合、次の点はどの程度問題となりますか？

Q20. 研究データの公開によって問題が起きたご経験がありましたら、差し支えない範囲でお書き下さい。

- b) データ公開のインセンティブを検討するための質問（実際にどのようなインセンティブが得られているのか、研究者が重要視するインセンティブは何か）

Q19. 研究データの公開によって良い結果が得られたご経験がありましたら、差し支えない範囲でお書き下さい。

Q23. 研究データ公開のインセンティブとして、次の点はどの程度重要だと思われますか？

- c) データの共有方法を尋ね、適切な共有基盤の検討に資するための質問（選択肢は先行研究¹⁹を参照した）

Q9. 共同研究者を除く他の研究者に研究データを提供した方法をお選び下さい。

- d) データマネジメントプランに関する質問（選択肢は先行研究¹⁹を参照した）

Q14. これまでに、DMP を作成したご経験はありますか？

Q15. DMP を作成した理由として、あてはまるものをお選び下さい。

Q16. DMP を作成していない理由として、あてはまるものをお選び下さい。

(2) 修正・追加した項目

回答者の負担軽減、及び2016年調査の結果をふまえた項目の修正や追加を行った。

a) 回答の負担を低減するための修正

- ① 頻度を3件法で尋ねる→複数選択方式〔Q12. 公開データの利用目的〕
- ② 回答必須項目を任意に変更〔Q28. Q29. カレントデータの理解の容易さ〕

b) 2016年調査の結果をふまえた項目の追加と修正

Q27. カレントデータを公開しようとする場合、次の点はどの程度問題となりますか（ありましたか）？あてはまるものをそれぞれお選び下さい。

- ① 2016年調査の自由回答から、「不正利用、改ざんの可能性」^x、「二次利用に関して責任が生じる可能性」、「データの所有権・契約」を追加
- ② 「論文の出版前にデータを公開する」という前提を削除

c) その他、表現の修正

DMPと区別するために、“カレントデータを管理・公開しようとする場合”という表現を“カレントデータを整備・公開しようとする場合”に変更するなど、軽微な修正を行った。

(3) 削除した項目

全体の質問数を抑制するために、11件の質問を削除した。質問の目的と削除した意図、2016年調査の質問項目を示す。

a) 異分野でのデータ利活用を前提とした質問（現状では異分野データの活用はほとんど行われていないことから、時期尚早であると考えた）

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。

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。

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。

b) リポジトリやメタデータに備えるべき事項を検討するための質問（経年変化はわずかなのではないかと考えた）

^x 自由回答において、“理解不足による誤用”と“盗用や悪意のある改ざん”の両方に関する記述がみられたことから、既存の選択肢「誤解や誤用の可能性」と区別するために追加した。

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。
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。
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。

- c) データの被提供頻度を尋ねる質問（先行研究から、データの提供を受けるだけで自らは提供しない研究者がいることを想定したが、提供頻度と被提供頻度はほぼ同じであったため）

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？

- d) データの保存期間に関する質問（研究不正への対応のため、所属機関の規則で10年保存と定められている場合が多いと予想される）

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？
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）

- e) データ公開の懸念と関連する質問（データの状況と懸念を比較するために設定したが、懸念に集約することとした）

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。
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？

1.4 調査対象

調査対象は2016年調査と同様、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」²⁰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、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。ただし、科学技術専門家ネットワークの構成員は毎年一部入替えがあるため、パネル調査ではないことに御留意いただきたい。

なお、本調査は1.2に示したように、「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「論文出版や学会等での口頭発表は行っていない」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。

1.5 調査方法と期間

調査方法は、ウェブアンケートシステムを用いた質問紙調査とした。具体的には、Questant 社のアンケートシステムを用いて質問の設定と回答の収集を行った。プレテストは 2018 年 8 月 23 日に実施して、質問の順序やワーディング等の修正を行った。

本調査の期間は、2018 年 10 月 19 日から 11 月 2 日とした。調査への協力依頼は、10 月 19 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 10 月 26 日と 31 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。

11 月 2 日以降も回答入力があったため、最終的に 11 月 9 日の回答までを結果に含めた。アンケートシステムの不具合等があった場合は別の方法でも受け付けることとしていたが、すべての回答がアンケートシステムで提出された。

なお、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 37 名が「その他」を選択していたが、該当する選択肢に修正した。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。

1.6 回答率と回答者の属性

調査依頼の送付数は 2,195 名、最終回答数は 1,548 名（回答率 70.5%）であった。このうち、1.4 で述べたように口頭発表や論文出版を行っている研究者 1,516 名（69.1%）の回答を分析対象とした。以下では、回答者の構成比率を所属、年齢、分野別に示す。また、回答の前提として、回答者が扱っているデータの量を示す。

(1) 回答者の所属

回答者の所属は、大学が 987 名(全体の 65.1%)で最も多く、次いで企業 270 名(17.8%), 公的機関・団体 251 名(16.6%), その他 8 名(0.5%) の順であった(図 12)。

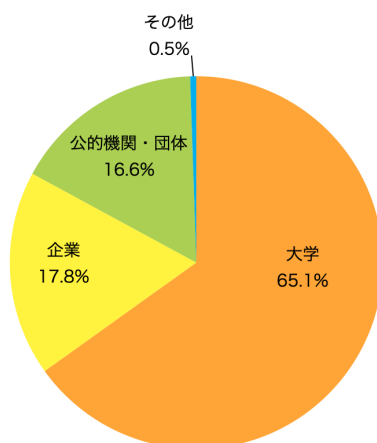


図 12 回答者の所属 (n=1,516)

(2) 回答者の年齢層

回答者の年齢層は、40 代が 753 名(全体の 49.7%)で最も多く、次いで 30 代が 440 名(29.0%), 50 代が 218 名(14.4%) の順であった(表 5)。

表 5 回答者の年齢層

年代	回答者数	比率
20	13	0.9%
30	440	29.0%
40	753	49.7%
50	218	14.4%
60	84	5.5%
70	7	0.5%
非開示	1	0.1%
合計	1,516	100.0%

(3) 回答者の専門分野

Q1 では回答者の専門分野を尋ねた。分野別の回答率を図 13 に示す。工学が 547 名(36.1%)で最も多く、次いで生物科学が 245 名(16.2%), 化学が 165 名(10.9%)であった。回答者数が 10 名以下であった天文学(6 名)は物理学と、人文学(5 名)は社会科学

学とあわせて「物理学・天文学」（合計 80 名，5.3%），「人文学・社会科学」（合計 36 名，2.4%）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。

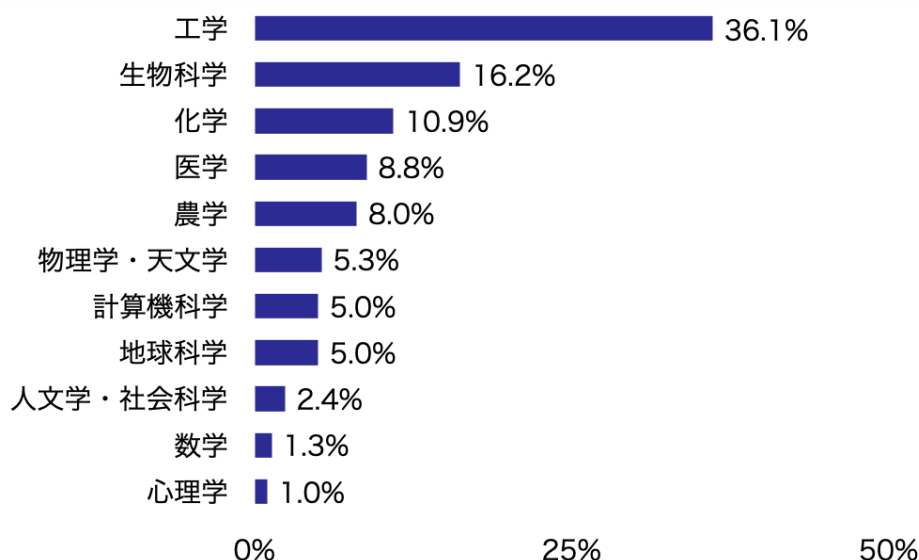


図 13 回答者の専門分野 (n=1,516)

(4) カレントデータの量

回答者が研究に用いているデータの量を確認するために，研究にデータを用いる回答者を対象として「Q24. 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 14 に結果を示す。

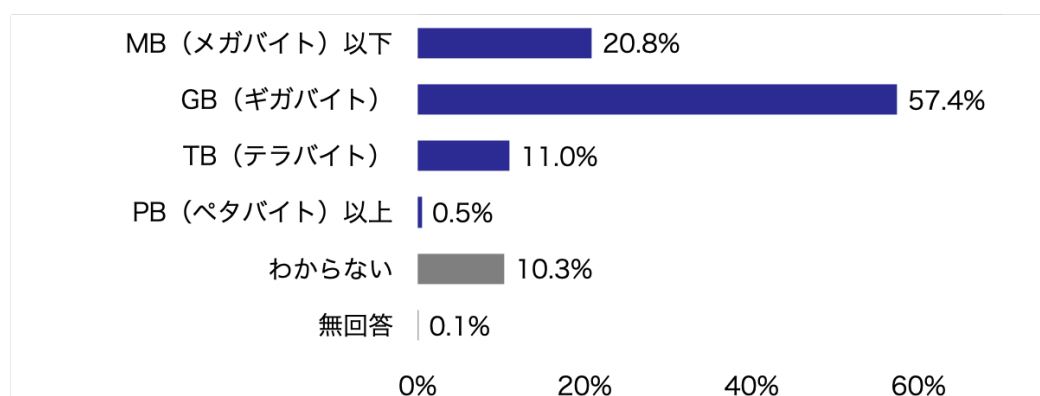


図 14 カレントデータの量 (n=1,513)

最も選択率が高かったのは，「GB (ギガバイト)」(57.4%)，次いで「MG (メガバイト) 以下」(20.8%)，「TB (テラバイト)」(11.0%) の順であった。「PB (ペタバイト) 以上」

は 0.5% (7 名) であった。データ量を「わからない」とする回答者は 10.3% (156 名), 無回答は 0.1% (1 名) であった。

2. 調査結果

調査結果を 2016 年調査の結果と比較しながら、調査目的（表 3）のグループごとに示す。すなわち、「1. データと論文の公開状況」、「2. データの提供（共有）状況」、「3. 公開データの利用状況と課題」、「4. データマネジメントプラン（DMP）の作成状況【新規】」、「5. データ公開の障壁」、「6. データ公開のインセンティブ【新規】」、「7. データの整備・公開に対する認識」の順に示す。

なお、Q1 から Q31 までの単純集計表は資料として報告書の末尾に掲載し、Q32 の自由回答は別途 Web 上に電子付録として掲載する。

2.1 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に、「Q17. 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。データ及び論文の公開方法を 1 つ以上選択した回答者は、データ及び論文の公開経験が「ある」とみなして集計した。図 15 にデータと論文の公開経験の有無についての集計結果を示す。

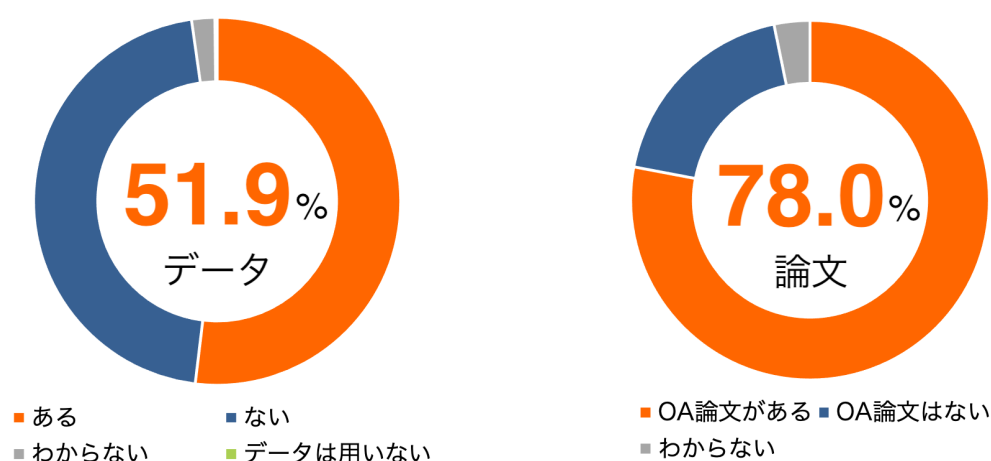


図 15 データと論文の公開経験（n=1,516）

データの公開経験をもつ回答者（以下、「データ公開率」）は 51.9%、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%であり、論文の OA 率の方が高かった。

データ公開率の経年変化を確認するために、2016 年調査と 2018 年調査の結果を比較した。図 16 にそれぞれの結果を示す。

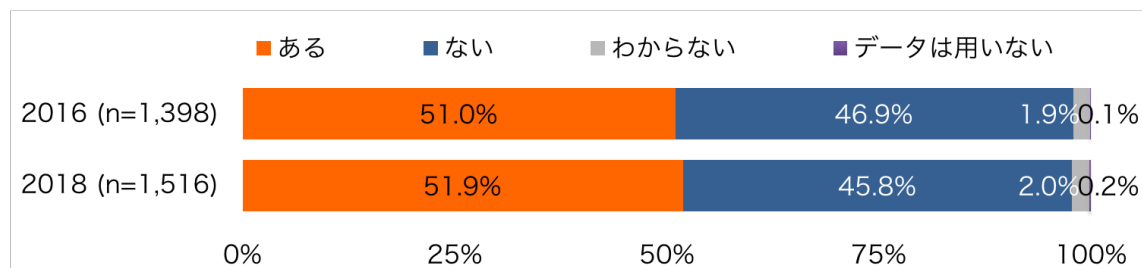


図 16 データ公開経験の経年変化

2016 年調査のデータ公開率は 51.0%，2018 年調査は 51.9%であり，0.9 ポイントしか増加していなかった。公開経験がない回答者は 46.9%から 45.8%まで 1.1 ポイント減少していた。

論文の OA 率についても経年変化を確認するため、2016 年調査と 2018 年調査の結果を比較した。図 17 にそれぞれの結果を示す。

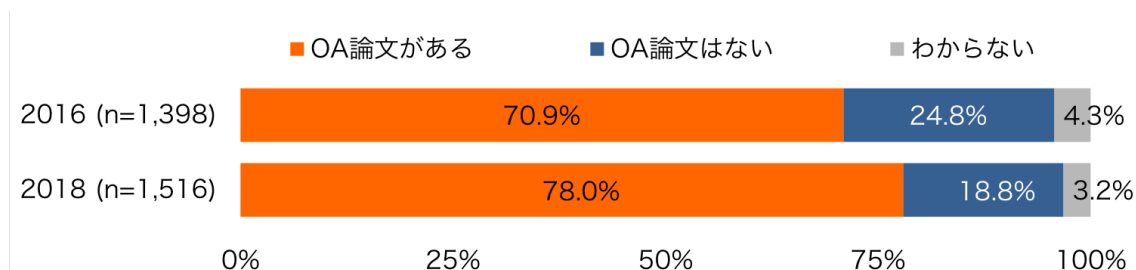


図 17 論文の公開経験の経年変化

論文の OA 率は 70.9%から 78.0%まで 7.1 ポイント増加しており，公開経験がない回答者は 24.8%から 18.8%まで 6.0 ポイント減少していた。

図 18 に、2016 年調査と 2018 年調査の分野別データ公開率を示す。配列は、2018 年調査において公開率が高かった分野の順とした。分野に付した「n」、及びグラフの右に付した比率（％）は 2018 年調査の値である。2016 年調査の値については、2016 年調査の報告書（RM268）¹³ をご参照いただきたい（以下、同様）。また、分野によって回答者数が異なり、特に心理学（n=15）や数学（n=20）は少数であることにご留意いただきたい。

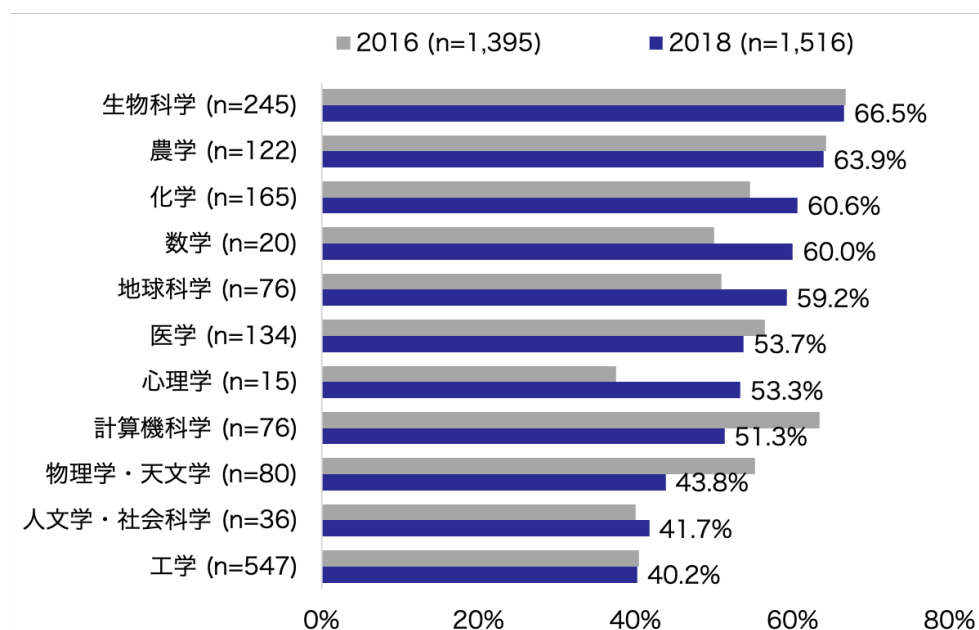


図 18 分野別データ公開率

データの公開率が最も高かったのは生物科学(66.5%), 次いで農学(63.9%), 化学(60.6%)の順であり、低かったのは工学 (40.2%), 人文学・社会科学 (41.7%), 物理学・天文学 (43.8%) であり、半数に満たなかった。2016 年調査と 2018 年調査を比較して 10 ポイント以上の差がみられたのは、心理学 (15.8 ポイント増加), 数学 (10.0 ポイント増加), 計算機科学 (12.1 ポイント減少), 物理学・天文学 (11.4 ポイント減少) の 4 分野であった。

図 19 に、2016 年調査と 2018 年調査の分野別論文の公開率を示す。配列は、2018 年調査において公開率が高かった分野の順とした。分野に付した「n」、及びグラフの右に付した比率（%）は 2018 年調査の値である。

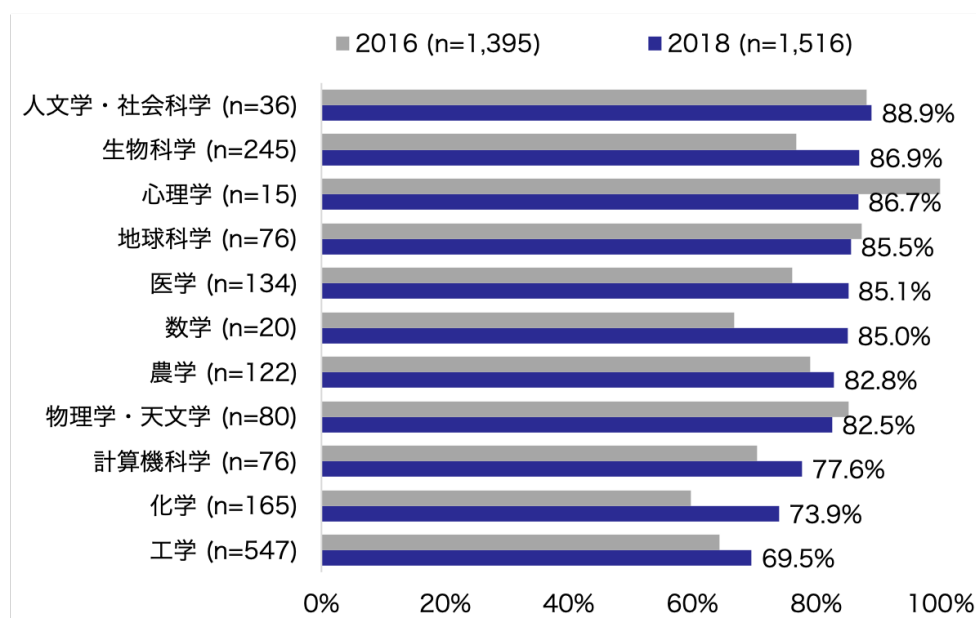


図 19 分野別論文の公開率

論文の公開率が最も高かったのは人文学・社会科学(88.9%), 次いで生物科学(86.9%), 心理学(86.7%)の順であり, 低かったのは工学(69.5%), 化学(73.9%), 計算機科学(77.6%)であった。2016 年調査と 2018 年調査を比較して 10 ポイント以上の差がみられたのは, 数学 (18.3 ポイント増加), 化学 (14.2 ポイント増加), 生物科学 (10.2 ポイント増加), 心理学 (13.3 ポイント減少) の 4 分野であった。

分野別のデータ公開率と論文の OA 率に相関があるかどうかを確認するために Pearson の相関分析を行ったが, 有意な相関はみられなかった。すなわち, データ公開率が高い(低い)分野であれば論文の OA 率が高い(低い)とはいえなかった。

図 20 に、分野別のデータ公開率と論文の OA 率の散布図を示す。なお、分野別クロス集計結果は資料として表 46 に掲載した。

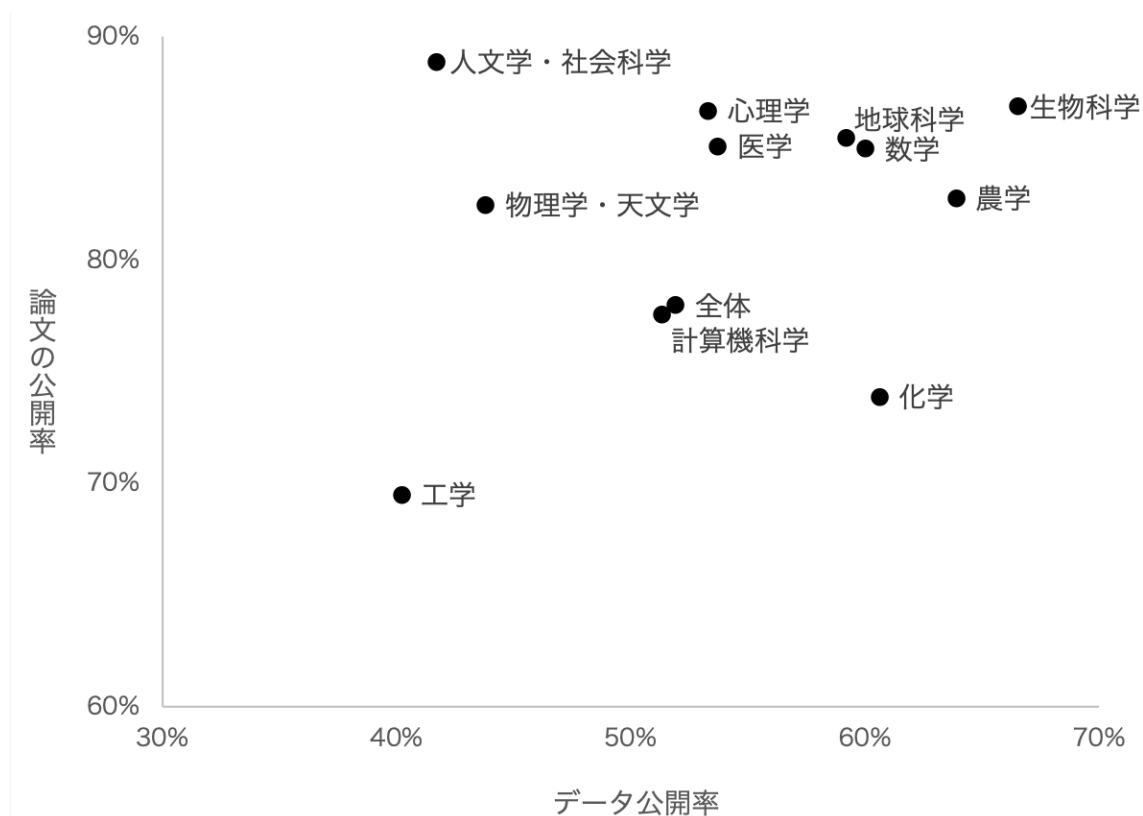


図 20 分野別データと論文の公開率

データと論文の公開率がいずれも高かったのは生物科学，いずれも低かったのは工学であった。論文の公開率と比較してデータ公開率が高かったのは化学であり，逆にデータの公開率と比較して論文の公開率が高かったのは人文学・社会科学等であった。

(2) データの公開方法

図 21 に、2016 年調査と 2018 年調査のデータ公開方法を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率（％）は、2018 年調査の値である。

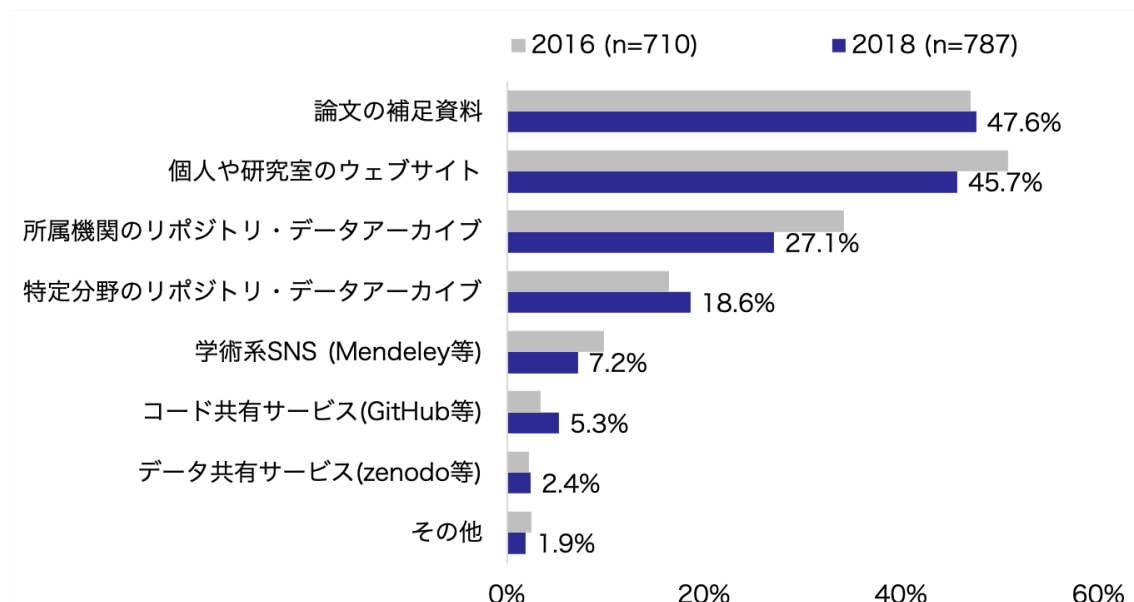


図 21 データの公開方法（複数回答）

データの公開方法として選択率が最も高かったのは、「論文の補足資料（supplementary materials）」（47.6%）、次いで「個人や研究室のウェブサイトへの掲載」（45.7%）であった。順位でみると、2016 年調査では 1 位が「個人や研究室のウェブサイト」、2 位が「論文の補足資料」であり、1 位と 2 位が逆転していた。「所属機関のリポジトリ・データアーカイブ」は 27.1%、「特定分野のリポジトリ・データアーカイブ（DDBJ^{xi}や ICPSR^{xii}など）」は 18.6%であった。リポジトリやアーカイブの選択率がそれほど高くないという傾向は、2016 年調査と同様であった。「その他」には、データ論文^{xiii} 21（2 名）などの記述があった。

2017 年に実施された Springer Nature の調査¹⁹においても、1 位は「論文の補足資料」（51%）、2 位は「研究室または個人のウェブサイト」（27%）、3 位「分野特有のリポジトリまたはデータアーカイブ」（25%）であり、ほぼ同様の結果であった。現状では、データの主な公開先は、論文の補足資料やウェブサイトであると考えられる。

^{xi} <https://www.ddbj.nig.ac.jp/index.html>

^{xii} <https://www.icpsr.umich.edu/icpsrweb/>

^{xiii} data paper（データ論文）：公開データの情報や来歴、ライセンスなどを記述した文書で、主としてデータジャーナルに掲載される。

(3) データの公開理由

データ公開経験がある回答者を対象として、公開理由を複数選択方式で尋ねた。図 22 に 2016 年調査と 2018 年調査の結果を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率（％）は、2018 年調査の値である。

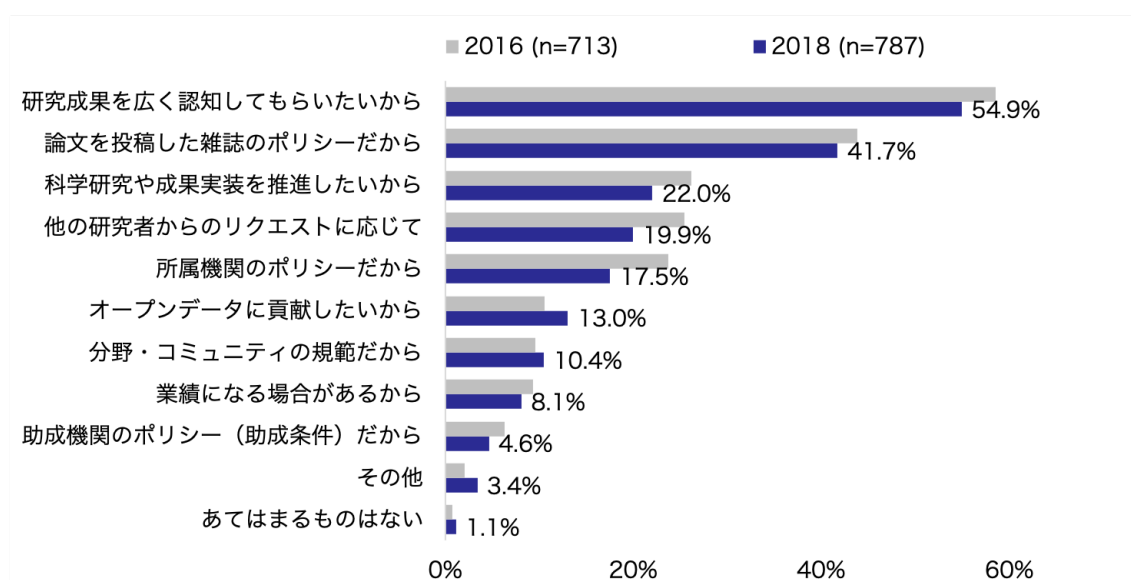


図 22 データの公開理由（複数回答）

データの公開理由として選択率が最も高かったのは、「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。2016 年調査と比較して選択率が増加していたのは、「オープンデータに貢献したいから」（2.5 ポイント）と「分野・コミュニティの規範だから」（0.9 ポイント）であった。

Springer Nature の調査¹⁹におけるデータ共有の動機は、1 位が「分野の研究の進展のため」（50%）、2 位が「透明性とデータの再利用のため」（42%）であった。本調査とは選択肢が異なるものの、データ共有への意識が高まっている可能性が示唆された。

(4) 論文の公開理由

論文の公開経験がある回答者を対象として、公開理由を複数選択方式で尋ねた。図 23 に 2016 年調査と 2018 年調査の結果を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率（％）は、2018 年調査の値である。

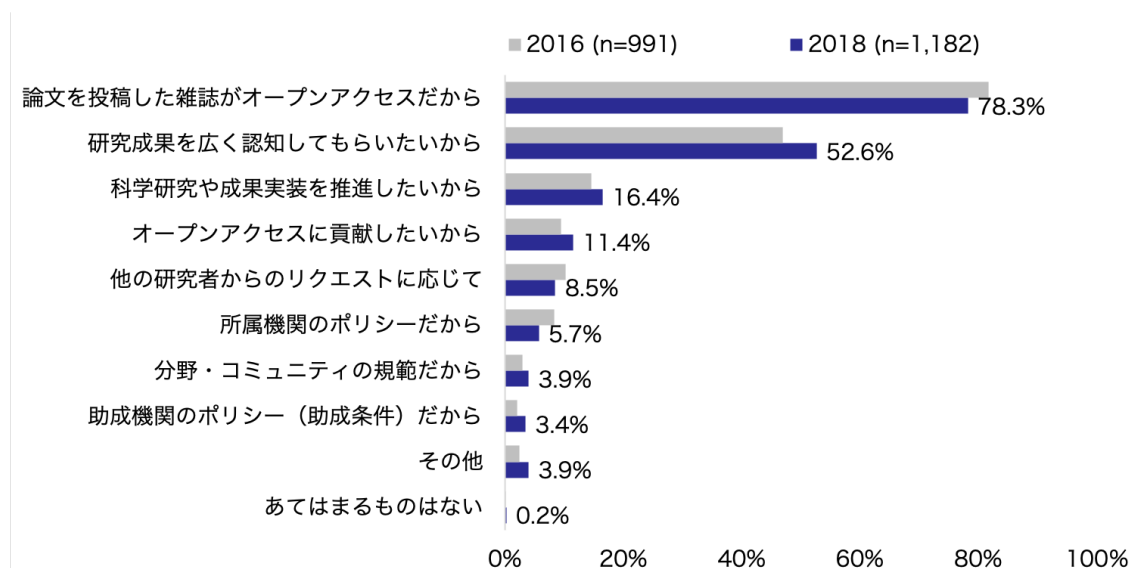


図 23 論文の公開理由

論文の公開理由として選択率が最も高かったのは、「論文を投稿した雑誌がオープンアクセスだから」(78.3%)であり、次いで「研究成果を広く認知してもらいたいから」(52.6%)、「科学研究や成果実装を推進したいから」(16.4%)と続く。2016 年調査と 2018 年調査の順位を比較すると、4 位の「オープンアクセスに貢献したいから」(11.4%)と 5 位の「他の研究者からのリクエストに応じて」(8.5%)が入れ替わっていた。選択率に最も差がみられたのは「研究成果を広く認知してもらいたいから」(52.6%)であり、2016 年調査の 46.9%から 5.7 ポイント増加していた。

(5) データの未公開理由

データ公開の経験が「なし」、「わからない」を選択した回答者を対象として、未公開理由を複数選択方式で尋ねた。図 24 に、2016 年調査と 2018 年調査の結果を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率(%)は、2018 年調査の値である。

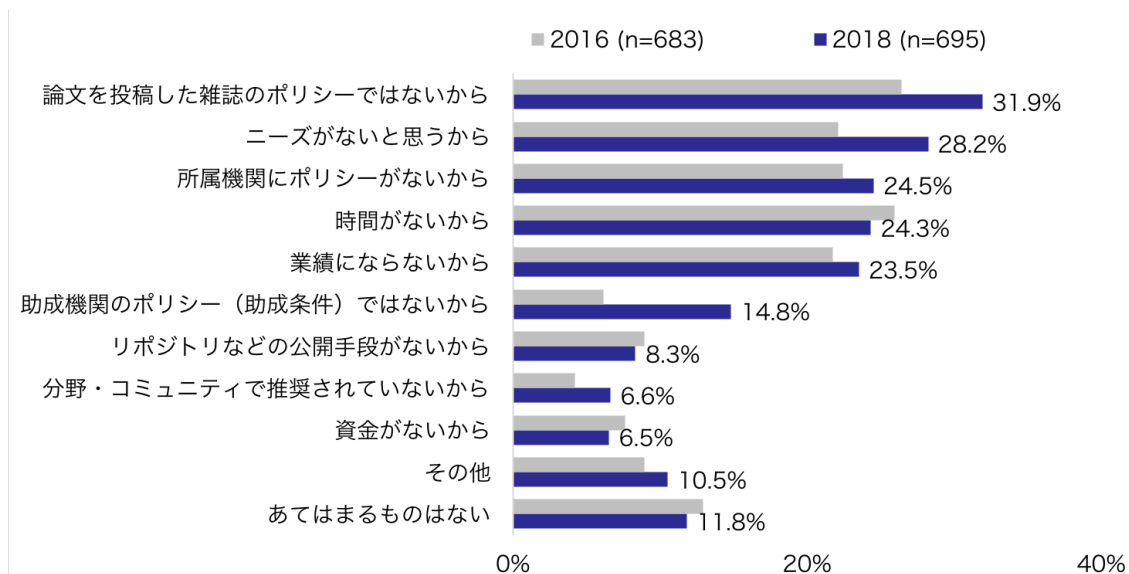


図 24 データの未公開理由（複数回答）

2018 年調査において選択率が最も高かったのは「論文を投稿した雑誌のポリシーではないから」(31.9%)、次いで「ニーズがないと思うから」(28.2%)、「所属機関にポリシーがないから」(24.5%)、「時間がないから」(24.3%)、「業績にならないから」(23.5%)の順であった。2016 年調査では、これらの 5 項目が 20%～25%を占めており、突出した理由がみられなかったが、2018 年調査では「雑誌のポリシーではないから」と「ニーズがないと思うから」がやや増加していた。選択率が最も上がっていたのは「助成機関のポリシー（助成条件）ではないから」であり、6.1%から 14.8%まで 8.7 ポイント増加していた。2016 年の時点では日本の助成機関は DMP を要求していなかったことが影響していると考えられる。

「その他」には 10.5% (73 名) による記述があった。表 6 に回答をタグ付けして集計し、6 項目に分類した結果を示す。なお、()内の数字は人数を示しているが、1 名の回答に 2 項目の内容が含まれている場合があったため、合計は 78 件である。

表 6 データの未公開理由：「その他」の回答 (n=73)

内容と分類
1. データの性質上公開できない (26) 機密 (10), 知的財産権 (5), 公開するべきではない (4), 規約・契約上不可 (3), 特許 (2), 機関のポリシー (2)
2. データ公開は必要ない (14) 論文で十分 (7), 不要・意味がない (7)
3. データ公開によるリスクがある (12) 無断利用などのリスク (10), 論文と二重投稿になる可能性 (2)
4. データ公開の慣習がない (12) 公開によるメリットがない (6), 公開の慣習や意識がない (3 名), 直接提供している (3)
5. データ公開のための環境が不十分 (7) 公開の許諾を得るためのコスト (2), 雑誌にデータを掲載するための料金が高額 (2), 共通フォーマットがない (1), シェアするための仕組みがない (1), データ量が大きい (1)
6. その他 (7) データを公開するべきではない (4), 公開するようなデータがない (3)

()内の数字は人数を示す。

未公開理由のうち最も多かったのは、データの性質上公開できない (合計 26 名) という意見であり、機密が含まれる (10 名), 知的財産権・著作権の問題がある (5 名), 公開するべきではない (4 名) といった記述がみられた。次いでデータ公開は必要ない (合計 14 名) という意見が多く, 論文への記載で十分である (7 名), データ公開は不要・意味がない (7 名) といった記述がみられた。3 番目に多かったのは, データ公開によるリスクがある (合計 12 名) という意見であり, 無断利用やデータが独り歩きするリスク (10 名) や, 論文と二重投稿になる可能性 (2 名) を危惧する記述がみられた。4 番目は分野等にデータ公開の慣習がない (合計 12 名) という意見であり, 公開によるメリットがない・感じられない (6 名), 公開の慣習や意識がない (3 名), 希望者には直接提供している (3 名) といった記述がみられた。5 番目はデータ公開のための環境が不十分である (合計 6 名) という意見であり, 公開の許諾を得るためのコストが高い (2 名), “他人が説明なしで系統立てて理解できるような共通フォーマットを思いつかないから”, “特定の資金を投入して取得したデータであり, そのデータを公平にシェアする仕組みがないから” といった記述がみられた。その他には, “共同研究者の内部情報になるため, 公開すべきではないデータであるから”, “公開できるほどのデータセットをまだもっていない” 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 11.9%であったのに対して、次項に示すように論文は 5.6%であった。データの未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。

(6) 論文の未公開理由

論文の公開経験が「なし」、「わからない」を選択した回答者を対象として、未公開理由を複数選択方式で尋ねた。図 25 に、2016 年調査と 2018 年調査の結果を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率(%)は、2018 年調査の値である。

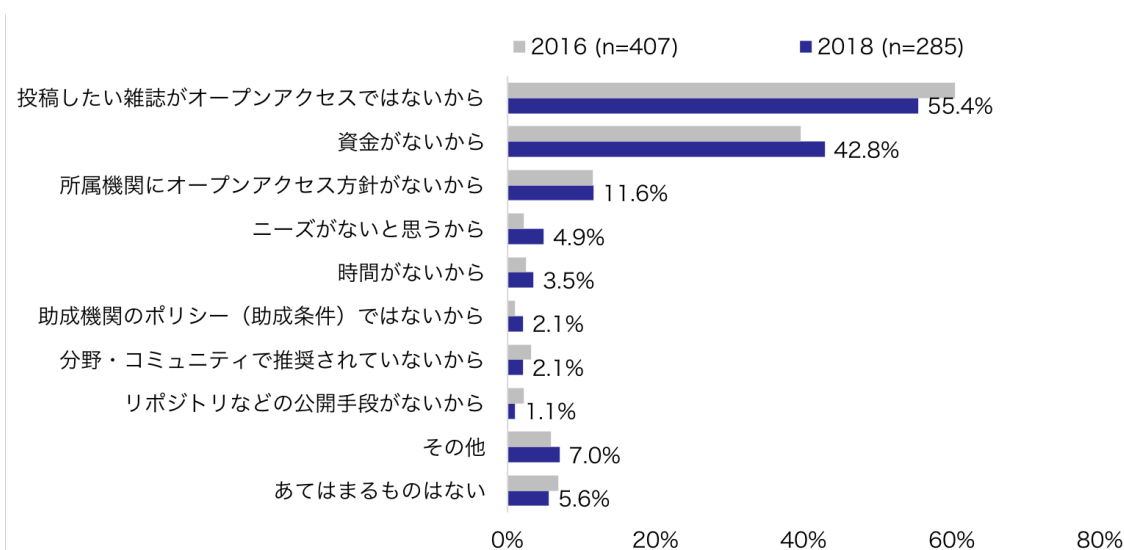


図 25 論文の未公開理由

2018 年調査において選択率が最も高かったのは「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%)であった。2016 年調査と同様に、これらの 2 項目の選択率が高かった。「所属機関にオープンアクセス方針がないから」(11.6%)はやや高いものの、その他「ニーズがないと思うから」(4.9%)など 5 項目の選択率は 5%未満であった。

「その他」には 7.0% (20 名) による記述があった。最も多かったのは、OA ジャーナルの信頼性が低いとする記述 (7 名) であり、ハゲタカジャーナル (predatory journal)^{xiv} 22 の存在を指摘する記述 (うち 3 名) や、“オープンアクセスジャーナルは栄枯盛衰があり、氾濫したため品質も IF^{xv} 23 も低下してしまった印象があるため”、“まともな査読をせずに

^{xiv} predatory journal (ハゲタカジャーナル)：査読付き OA 誌を装いつつ、APC (論文出版手数料) の徴収を目的とした粗悪雑誌。

^{xv} Impact Factor (インパクトファクター)：特定のジャーナル (学術雑誌) に掲載された論文が特定の年または期間内にどれくらい頻繁に引用されたかを平均値で示す尺度。毎年、Clarivate Analytics 社が Journal Citation Reports で発表している。

論文を掲載するような出版社・ジャーナルが存在し、信頼性の疑わしい論文と思われる可能性があるから」といった記述がみられた。次いで、OA について考えたことがない、今後は検討するといった記述がみられた（3 名）。この他、“メリットがない”、“インパクトが不明瞭”、“数学論文はほとんどプレプリントの段階で公表されているのでわざわざ資金を投入する必要がない”といった記述がみられた。

(7) データと論文の公開意思

データの非公開理由を 1 つ以上選択した回答者に、「Q22. Q21 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?」と尋ねた。図 26 に、2016 年調査と 2018 年調査の結果を示す。

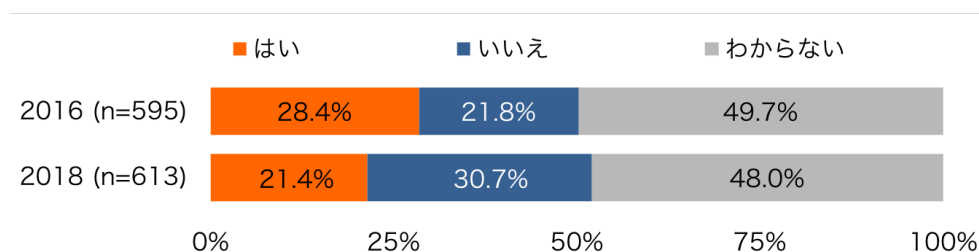


図 26 データの公開意思

データの 2016 年調査と 2018 年調査の結果を比較すると、「はい」が 28.4%から 21.4%まで 7.0 ポイント減少していた。「いいえ」は 21.8%から 30.7%まで 8.8 ポイント増加、「わからない」は 49.7%から 48.0%まで 1.8 ポイント減少していた。「わからない」とする回答者の比率は減少しているものの、「はい」と「いいえ」の比率が逆転しており、データ公開に対してより慎重になっている傾向がみられた。

また、論文の非公開理由を 1 つ以上選択した回答者に、「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?」と尋ねた。図 27 に、2016 年調査と 2018 年調査の結果を示す。

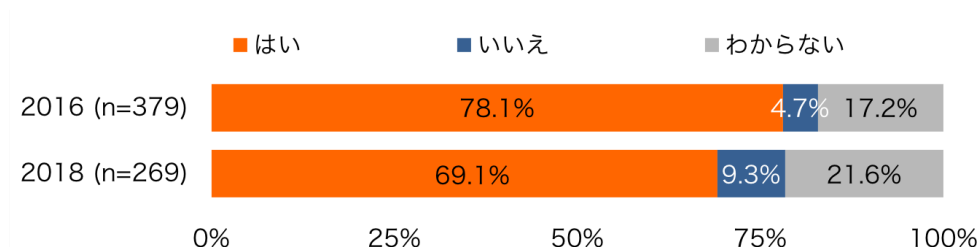


図 27 論文の公開意思

論文の 2016 年調査と 2018 年調査の結果を比較すると、「はい」が 78.1%から 69.1%まで 9.0 ポイント減少していた。「いいえ」は 4.7%から 9.3%まで 4.5 ポイント増加、「わからない」は 17.2%から 21.6%まで 4.4 ポイント増加していた。つまり、「はい」の選択率が

減少し、「いいえ」と「わからない」増加しており、論文の公開に対してより慎重になっている傾向がみられた。ただし、(1)に示したように OA 経験をもつ回答者の比率は 70.9% から 78.0%まで増加していたため、論文を公開しようと考えている研究者は既に実践しているとも推測できる。

データと論文の結果を比較すると、論文の方が公開意思をもつ回答者の比率が高いという傾向は変わっていなかった。また、公開意思をもつ回答者の比率が 2016 年調査よりも低下しているという傾向はデータと論文に共通していた。

2.2 データの提供（共有）状況

(1) データの提供頻度

データの提供（共有）状況を確認するために、「Q8. 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？」と尋ねた。「よくある」から「まったくない」までの 4 件法で尋ね、「わからない」という選択肢も示した。図 28 に 2016 年調査と 2018 年調査の結果を示す。

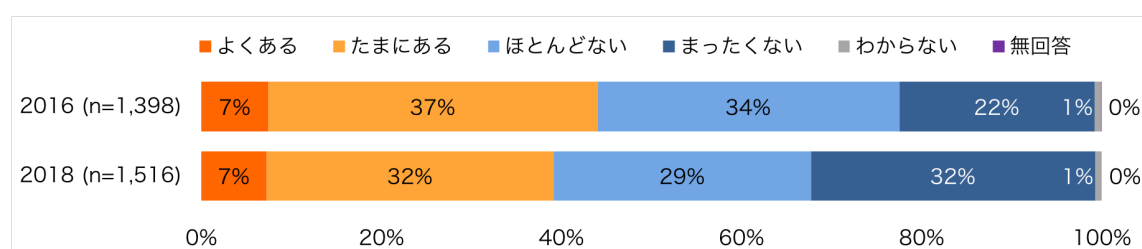


図 28 データの提供経験

2016 年調査と 2018 年調査の結果を比較すると、2018 年調査の方がデータ提供の頻度が低い傾向がみられた。「よくある」は 7.4%と 7.3%でほぼ変わらないが、「たまにある」は 36.6%から 31.9%まで 4.7 ポイント減少、「ほとんどない」は 33.5%から 28.6%まで 4.9 ポイント減少しており、「まったくない」は 21.7%から 31.5%まで 9.9 ポイント増加していた^{xvi}。

^{xvi} 2016 年調査では、回答者が他の研究者にデータを提供した経験に加えて、他の研究者から提供を受けた経験（被提供経験）も尋ねた。その結果、被提供経験よりも提供経験の方がやや高かった。本調査では質問文は変更しなかったが、提供経験のみを尋ねたことが結果に影響を与えた可能性もあると考えられる。

図 29 に 2016 年調査と 2018 年調査による分野別データ提供経験をもつ回答者の比率、すなわち「よくある」、「たまにある」、「ほとんどない」を選択した回答者の比率を示す。分野に付した「n」、及びグラフの右に付した比率（％）は 2018 年調査の値である。なお、全体では 2016 年調査が 78.1%，2018 年調査は 67.7%であった。

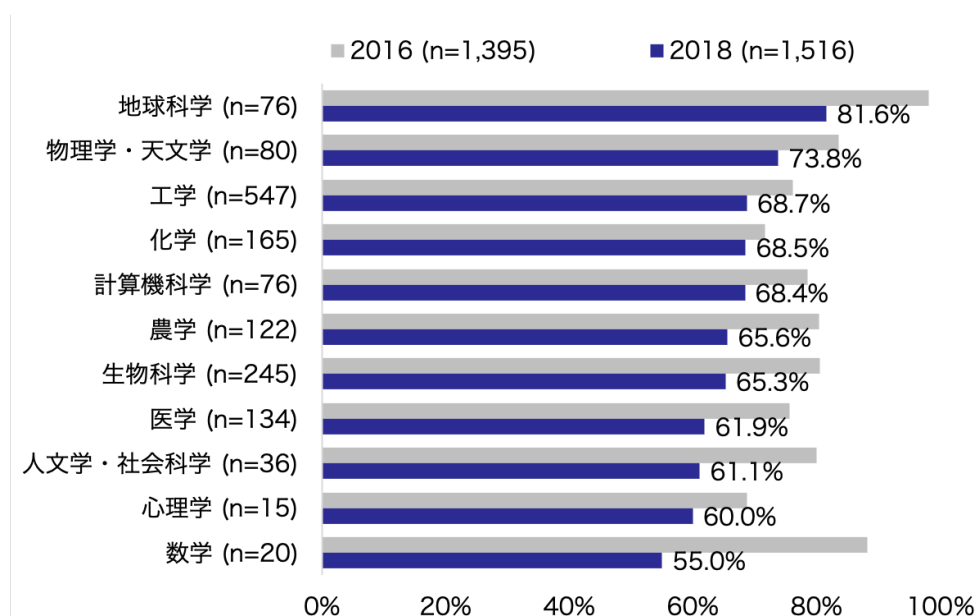


図 29 分野別データ提供経験

データ提供経験をもつ回答者の比率が最も高かったのは地球科学（81.6%）、次いで物理学・天文学（73.8%）、工学（68.7%）の順であり、低かったのは数学（55.0%）、心理学（60.0%）、人文学・社会科学（61.1%）であった。分野別のデータ提供経験とデータ公開率に相関があるかどうかを確認するために Pearson の相関分析を行ったが、有意な相関はみられなかった。つまり、データ提供経験がある回答者の比率が高い（低い）分野はデータ公開率が高い（低い）とはいえなかった。たとえば、工学は 11 分野のうちデータ公開率が最も低かったが、データの提供率は 3 位であった。ただし、前述の通り回答者数が少ない分野があるため、分野別の分析に十分な回答者数を確保することを今後の課題としたい。

(2) データの提供方法【新規】

データの提供（共有）状況を確認するために、データ提供経験を有する回答者を対象として、「Q9. 共同研究者を除く他の研究者に研究データを提供した方法をお選びください」と尋ねた。図 30 に、4 項目を示して複数選択方式で尋ねた結果を示す。

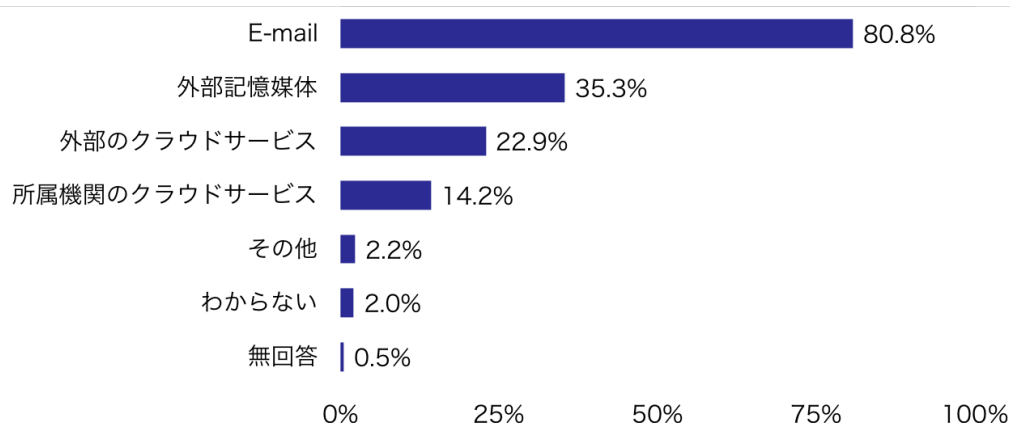


図 30 データの提供方法 (n=1,027, 複数回答)

データの提供方法のうち選択率が最も高かったのは「E-mail」(80.8%)であり、次いで「外部記憶媒体 (USB フラッシュメモリ, 記録型 CD/DVD など)」(35.3%), 「外部のクラウドサービス (Dropbox, Google Drive など)」(22.9%), 「所属機関のクラウドサービス」(14.2%) の順であった。「その他」(2.2%) には, 印刷して渡す (10 件), 直接渡す (5 件), SNS 経由 (3 件), 口頭で伝える (2 件), 郵便 (2 件), “特定の研究に関する研究者のメーリングリスト”などの記述がみられた。

Springer Nature の調査¹⁹においても, プライベートなデータの共有に用いられている方法の 1 位は「email」(65%), 2 位は「USB またはフラッシュドライブ」(41%), 3 位は「ファイルシェアリングサービス」(39%) であった。本調査と同様の傾向であったといえる。

2.3 公開データの利用状況と課題

(1) 公開データの入手経験

公開データの入手経験を確認するために、「Q10. 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。図 31 に 2016 年調査と 2018 年調査の結果を示す。

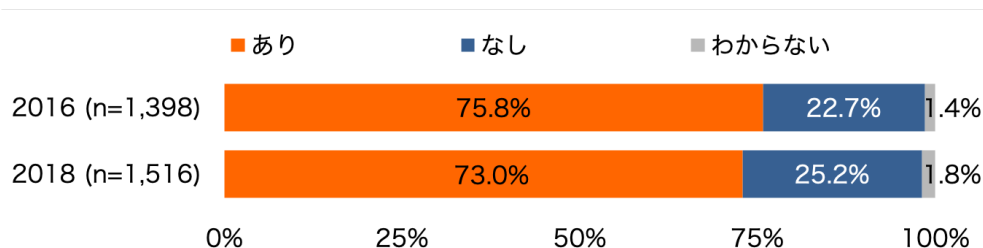


図 31 公開データの入手経験の有無

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 73.0%, 排他的選択肢である「ない」は 25.2%, 「わからない」は 1.8% 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。2016 年調査では, 公開データの入手経験がある回答者は 75.8% であったため, 2.8 ポイント減少していた。

図 32 に 2016 年調査と 2018 年調査の分野別データ入手経験を示す。分野に付した「n」, 及びグラフの右に付した比率 (%) は 2018 年調査の値である。

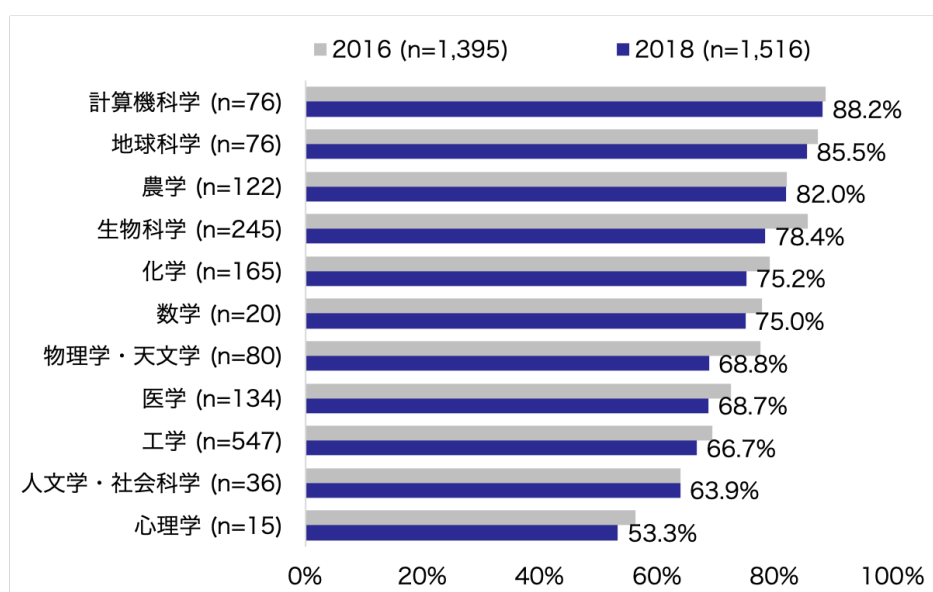


図 32 分野別公開データの入手経験

公開データの入手経験をもつ回答者の比率が最も高かったのは計算機科学 (88.2%)、次いで地球科学 (85.5%)、農学 (82.0%) の順であり、低かったのは心理学 (53.3%)、人文科学・社会科学 (63.9%)、工学 (66.7%) であった。分野別の公開データの入手経験とデータ公開率に相関があるかどうかを確認するために Pearson の相関分析を行ったが、有意な相関はみられなかった。すなわち、公開データの入手経験がある回答者の比率が高い（低い）分野はデータ公開率が高い（低い）とはいえなかった。

(2) 公開データの入手方法

図 33 に、2016 年調査と 2018 年調査における公開データの入手方法を示す。配列は、2018 年調査で選択率が高かった入手方法の順とした。グラフの右に付した比率 (%) は、2018 年調査の値である。

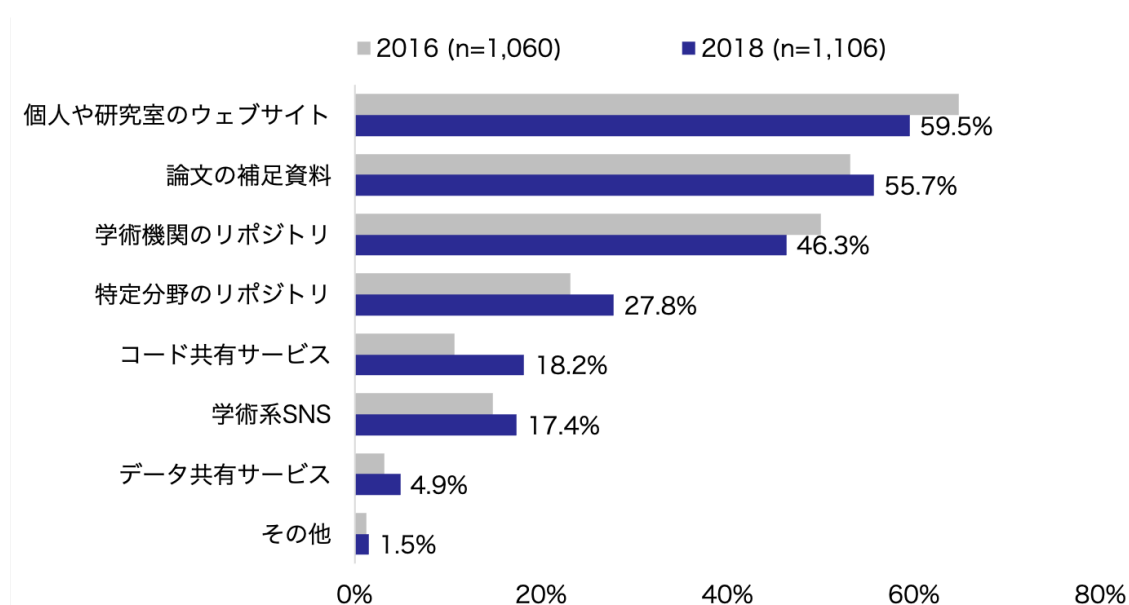


図 33 公開データの入手方法

公開データの入手方法のうち、選択率が最も高かったのは「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%)、「学術機関のリポジトリ (大学や NASA のリポジトリなど)」(46.3%) の順であった。データの公開先は 1 位が「論文の補足資料」、2 位が「個人や研究室のウェブサイト」であったが、入手先の順位は逆であった。ただし、「個人や研究室のウェブサイト」の選択率は 5.3 ポイント減少し、「論文の補足資料」の選択率は 2.6 ポイント増加していたため、両者の差は縮まっている。

「学術機関のリポジトリ」(46.3%) は公開先としては 27.1%にとどまったが、データの入手先としては比較的よく活用されていた。「特定分野のリポジトリ」(27.8%)、「コード共有サービス (GitHub など)」(18.2%)、「学術系 SNS (Mendeley, ResearchGate など)」(17.4%)、「無料のデータ共有サービス (figshare, zenodo など)」(4.9%) は、いずれも選択率が低かった。しかし、2016 年調査と比較するとすべて選択率が上昇しており、特に

「コード共有サービス」は 7.5 ポイント増加していた。

「その他」には、官公庁や自治体が公開しているオープンデータ（13 名），“国立国会図書館，国立公文書館”，“図書館”，“学術団体の Web サイト”などが挙げられていた。また，Sci-Hub^{xvii} 24 という回答もみられた。

(3) 公開データ入手の障壁

公開データの入手における障壁を明らかにするために，データ入手経験がある回答者（1,106 名）を対象として，「Q11. 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。図 34 に 2016 年調査と 2018 年調査の結果を示す。

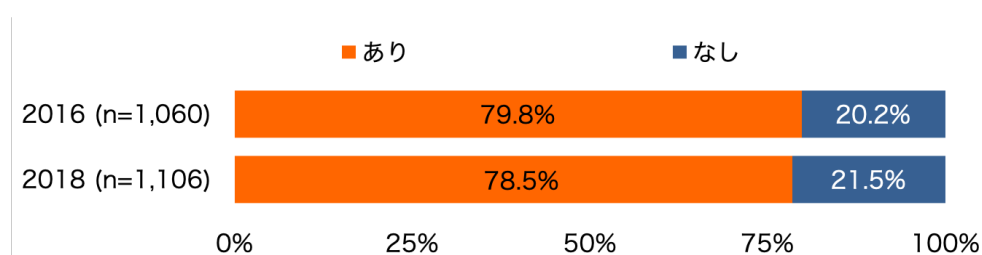


図 34 データ入手の障壁の有無

2016 年調査では，1 つ以上の問題を選択した回答者，すなわち障壁があると考えられる回答者は 79.8%であった。2018 年調査では 78.5%であったため，1.3 ポイント減少していた。

^{xvii} Sci-Hub：有料論文を含む学術論文のフルテキストをダウンロードできる違法サイト。

図 35 に、2016 年調査と 2018 年調査によるデータ入手の障壁の選択率を示す。配列は、2018 年調査において選択率が高かった項目の順とした。グラフの右に付した比率(%)は、2018 年調査の値である。

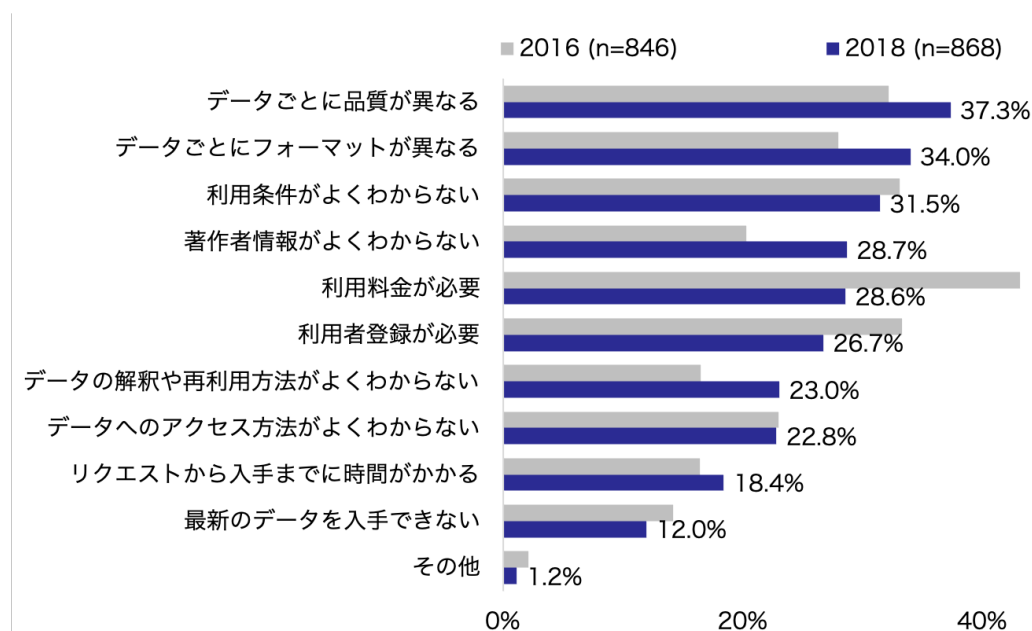


図 35 データ入手の障壁（複数回答）

2016 年調査と 2018 年調査では、問題であると考えられている項目の選択率に変化がみられた。2018 年調査の 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。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、データ入手の時点で障壁を感じる回答者が多かったが、2018 年調査では先行研究¹⁰と同様に、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。なお、選択率が最も上昇したのは「著作者情報がよくわからない」(8.4 ポイント増加)であり、低下したのは「利用料金が必要」(14.5 ポイント減少)であった。

「その他」では、リンク切れ (3 名)、データの信頼性の判断が困難 (3 名)、メタデータ不足 (2 名)、“ダウンロード方法が適切でない場合。例えば、大量のデータをダウンロードしなければならないのに、GUI での作業が伴うため、数ヶ月その作業だけに費やさねばならず、諦めたことがあった”といった記述がみられた。

(4) 公開データの利用目的

公開データの利用目的として、「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」の 3 点を挙げて、「Q12.

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。」と複数選択方式で尋ねた。図 36 に結果を示す。

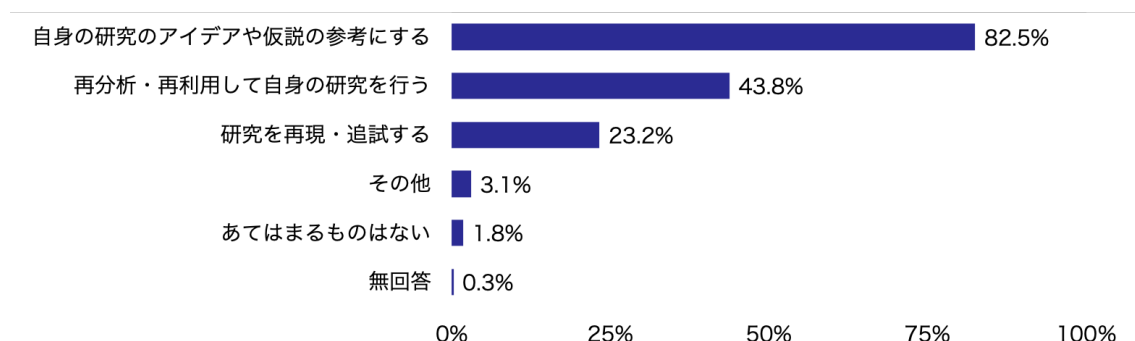


図 36 公開データの利用目的 (n=1,106, 複数回答)

公開データの利用目的のうち、選択率が最も高かったのは「自身の研究のアイデアや仮説の参考にする」(82.5%), 次いで「再分析・再利用して自身の研究を行う」(43.8%), 「研究を再現・追試する」(23.2%) の順であった。「その他」には、テストデータとして用いる (7 名), データを用いて行われた研究を分析する, 理解する (7 名), 授業等の教材として用いる (6 名), 引用する (5 名), 自身の結果と比較する (4 名) といった記述がみられた。

2016 年調査では、それぞれの頻度を「行う」、「たまに行う」、「行わない」の 3 件法で尋ねた。結果を図 37 に示す。

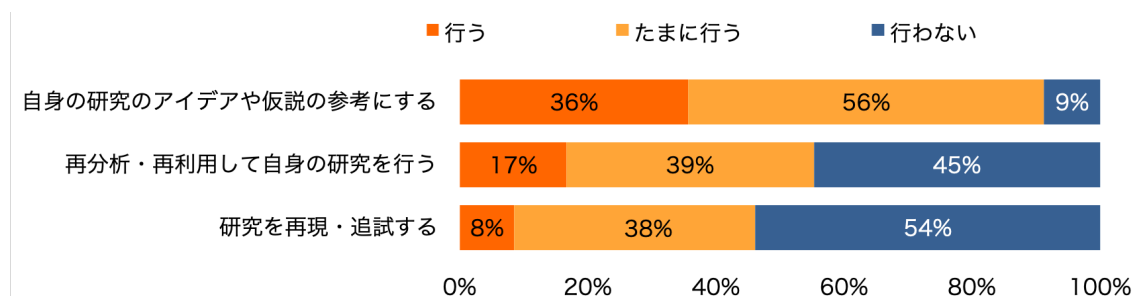


図 37 2016 年調査：公開データの利用目的 (n=1,060)

2016 年調査の結果は Kratz らによる先行研究²⁵と同様であり、「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」の順に頻度が高かった。2018 年調査と比較するために「行う」と「たまに行う」を合計すると、「自身の研究のアイデアや仮説の参考にする」は 91.2%, 「再分析・再利用して自身の研究を行う」は 55.3%, 「研究を再現・追試する」は 46.1%であり、いずれも 2018 年調査よりも選択率が高かった。

(5) 公開データの探索方法

公開データの探索方法を明らかにするために、「Q13. 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。図 38 に 2016 年調査と 2018 年調査の項目別選択率を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率（％）は、2018 年調査の値である。

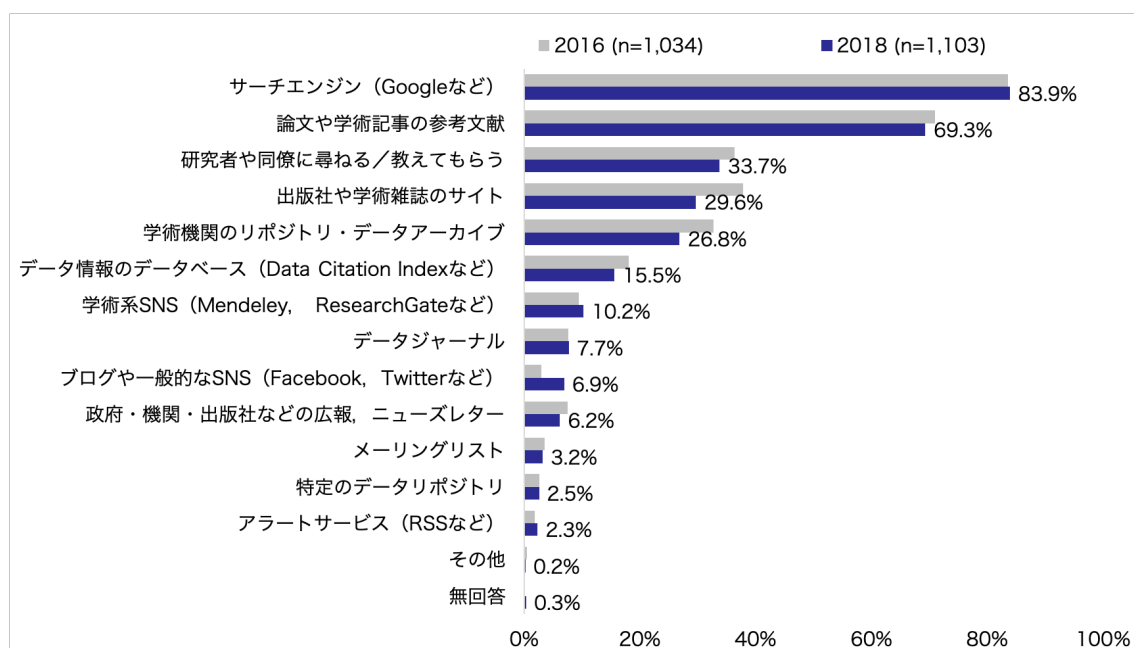


図 38 公開データの探索方法

公開データの探索方法のうち、最もよく利用されていたのは、「サーチエンジン (Google, Google Dataset Search など)」(83.9%)、次いで「論文や学術記事の参考文献」(69.3%)であった。2016 年調査も同じく「サーチエンジン」、「参考文献」の順であり、他の選択肢と比較して突出していた。

続いて、「研究者や同僚に尋ねる／教えてもらう」(33.7%)、「出版社や学術雑誌のサイト (Elsevier, Wiley など)」(29.6%)、「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)」(26.8%) が比較的によく用いられていた。「特定のデータリポジトリ」(2.5%) については具体的な名称の記述を求めたところ、Gene Expression Omnibus (GEO)^{xviii} (2 名)、Cambridge Crystallographic Data Centre (CCDC)^{xix} (2 名)、National Center for Biotechnology Information (NCBI)^{xx} (以下、各 1 名)、日本 DNA データバンク (DDBJ)、タンパク質構造データベース (PDB)^{xxi}、cBioPortal^{xxii}などが挙げられ

^{xviii} <https://www.ncbi.nlm.nih.gov/geo/>

^{xix} <https://www.ccdc.cam.ac.uk>

^{xx} <https://www.ncbi.nlm.nih.gov>

^{xxi} <https://www.rcsb.org>

^{xxii} <https://www.cbioportal.org>

ていた。「その他」では、“基本的にデータが公開である業界にいるので、新しいデータについては、その都度情報が入ってくる”，“GitHub など定期的に更新されるサイトを定期的に見ているので新たに探すことはない”という記述がみられた。

(6) 論文の探索方法

論文の探索方法を明らかにするために、「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」と複数選択方式で尋ねた。図 39 に 2016 年調査と 2018 年調査の項目別選択率を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率(%)は、2018 年調査の値である。なお無回答はなかった。

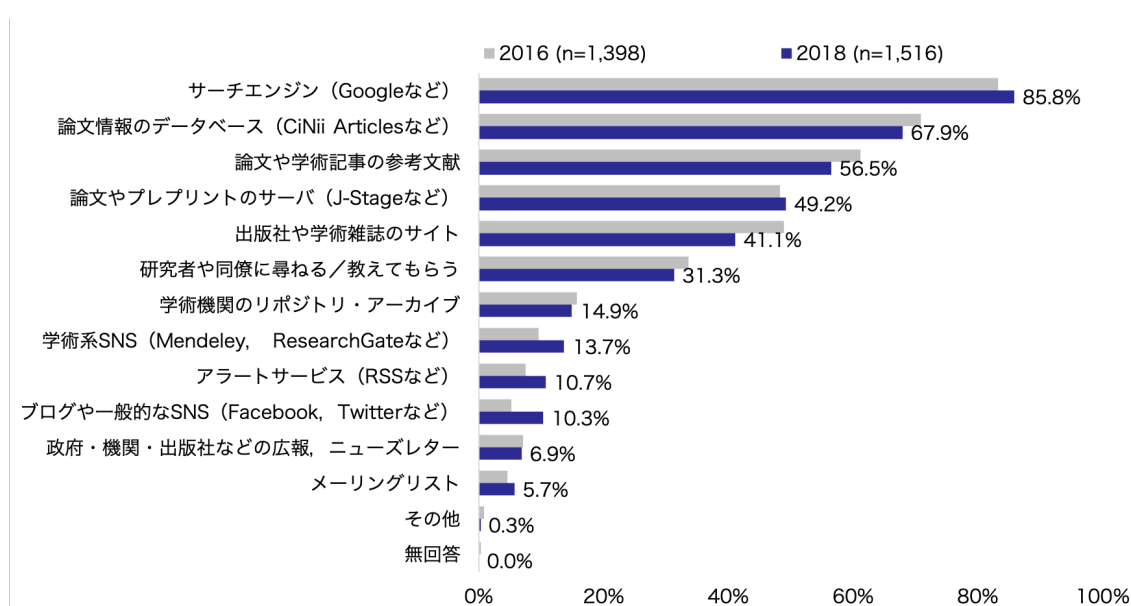


図 39 論文の探索方法

論文の探索方法のうち、最もよく利用されていたのは、「サーチエンジン (Google, Google Scholar など)」(85.8%)、次いで「論文情報のデータベース (Web of Science, Scopus, CiNii Articles など)」(67.9%)、「論文や学術記事の参考文献」(56.5%)であった。2016 年調査も同じく「サーチエンジン」、「データベース」、「参考文献」の順であり、いずれも回答者の半数以上が選択していた。「その他」では、“権威的学会のページを探す”，“同分野の研究者の Web ページ”，“オンラインデータベース上の参考文献”，“国際会議のサイト”といった記述がみられた。

データ、論文ともに最もよく利用されていたのは 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「出版社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た（データは 15.5%）。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（49.2%）が選択しているのに対して、「特定のデータリポジトリ」は2.5%にとどまった。この状況は 2016 年調査とほぼ同様であり、ほとんど変化はみられなかった。

2.4 データマネジメントプラン（DMP）の作成状況【新規】

(1) DMP の作成経験

データマネジメントプラン（DMP）の作成経験と種類を確認するために、「Q14. これまでに、DMP を作成したご経験はありますか？」と尋ねた。選択肢として 5 種類の DMP 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「作成したことはない」、「わからない」を示した。DMP を 1 つ以上選択した回答者は、DMP 作成経験が「ある」とみなして集計した。図 40 に集計結果を示す。

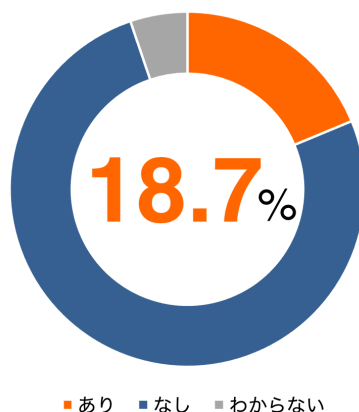


図 40 DMP の作成経験 (n=1,516)

DMP の作成経験をもつ回答者は 18.7%（284 名）、「なし」は 76.1%（1,153 名）、「わからない」は 5.2%（79 名）であった。Springer Nature の調査¹⁹では、日本の研究者の 56% が DMP の作成経験があると回答しており、本調査の結果とは差がみられた。

図 41 に分野別の DMP の作成経験をもつ回答者の比率を示す。

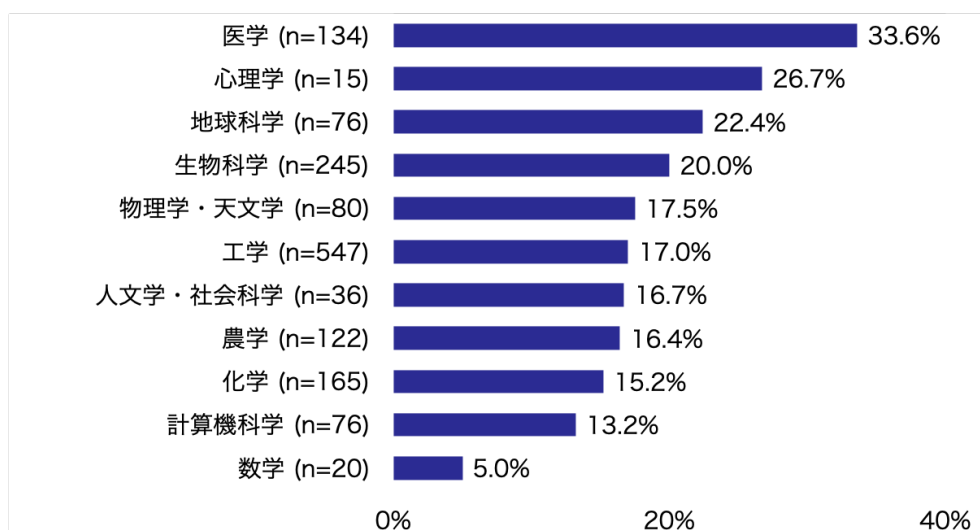


図 41 分野別 DMP の作成経験 (n=1,516)

DMP の作成経験をもつ回答者の比率が最も高かったのは医学 (33.6%), 次いで心理学 (26.7%), 地球科学 (22.4%) の順であり, 低かったのは数学 (5.0%), 計算機科学 (13.2%), 化学 (15.2%) であった。分野別の DMP 作成経験とデータ公開率に相関があるかどうかを確認するために Pearson の相関分析を行ったが, 有意な相関はみられなかった。すなわち, DMP の作成経験をもつ回答者の比率が高い (低い) 分野はデータ公開率が高い (低い) とはいえなかった。

(2) 作成経験がある DMP の種類

図 42 に, DMP の種類ごとの集計結果を示す。

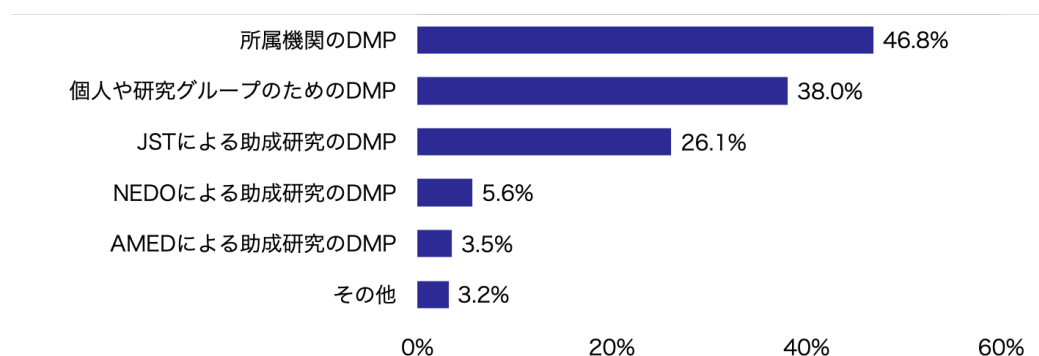


図 42 作成経験がある DMP (n=284, 複数回答)

例示した DMP のうち選択率が最も高かったのは, 「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。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。「その他」には、国外の助成機関（2 名）や、機器の利用申請のために提出したといった記述がみられた。

(3) DMP の作成理由

DMP の作成経験がある回答者を対象として、作成理由を複数選択方式で尋ねた。図 43 に結果を示す。

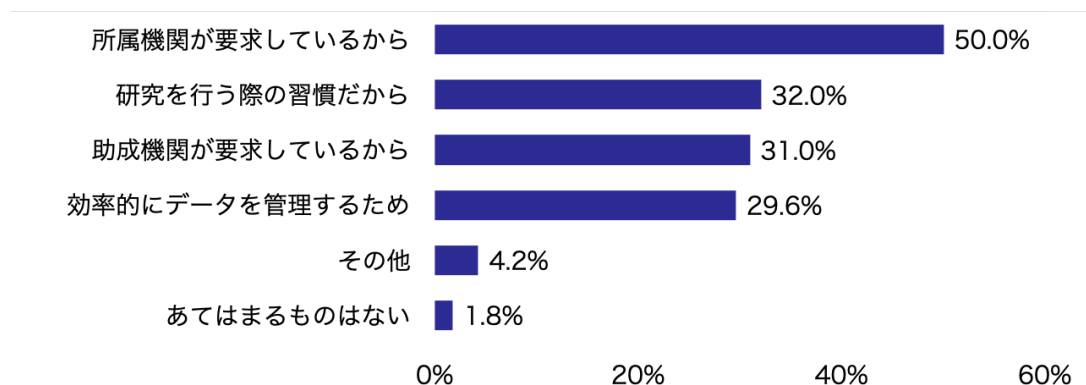


図 43 DMP の作成理由（n=284，複数回答）

DMP の作成理由のうち、選択率が最も高かったのは、「所属機関が要求しているから」（50.0%）、次いで「研究を行う際の習慣だから」（32.0%）、「助成機関が要求しているから」（31.0%）、「効率的にデータを管理するため」（29.6%）の順であった。すべての回答者（1,516 名）に対する比率をみると、DMP の作成が研究を行う際の習慣となっている回答者は 6.0%、効率的なデータ管理に資すると考えている回答者は 5.5%であった。「その他」には、共同研究のため（5 名）、研究倫理を遵守するための対応（3 名），“投稿先の雑誌が要求しているから”といった記述がみられた。

(4) DMP の未作成理由

DMP の作成経験がない回答者を対象として、未作成理由を複数選択方式で尋ねた。図 44 に結果を示す。

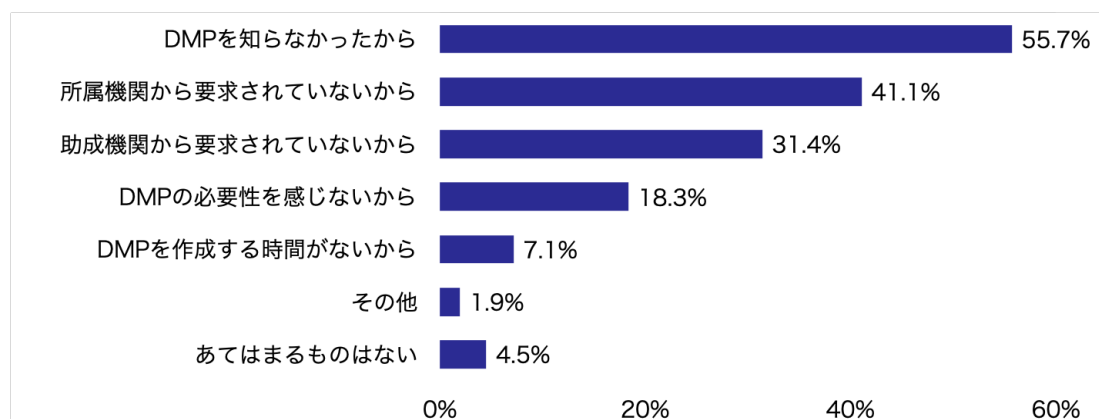


図 44 DMP の未作成理由 (n=1,153, 複数回答)

DMP の未作成理由のうち、選択率が最も高かったのは、「DMP を知らなかったから」(55.7%)、次いで「所属機関から要求されていないから」(41.1%)、「助成機関から要求されていないから」(31.4%)、「DMP の必要性を感じないから」(18.3%)、「DMP を作成する時間がないから」(7.1%) の順であった。「その他」には、別の方法で管理しているから (7 名)、別の人 (PI など) が作成しているから (2 名)、別の規則があるから (2 名)、分野に慣習がないから (2 名)、“ジャーナルから要求されていなかったから”、“どのようなデータが得られるか、研究開始前に予測することが困難だから”といった記述がみられた。

なお、Springer Nature の調査¹⁹によれば、回答者 763 名のうち 34%は自身の主要な助成機関が DMP を要求しているかどうかを認識していなかった。また、JST を主要な助成機関とみなしている回答者 115 名のうち、JST の要求について正しく回答していたのは 12%であった。このことから、「助成機関から要求されていないから」を選択した回答者の中には、助成機関による要求を正しく認識していない者が含まれている可能性もあると考えられる。

2.5 データ公開の障壁

(1) カレントデータの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に、6 項目を挙げて「Q26. 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)? あてはまるものをそれぞれお選び下さい。」と尋ねた。選択肢は、「不十分」から「十分」までの 4 件法として、「わからない」も設けた。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 45)。

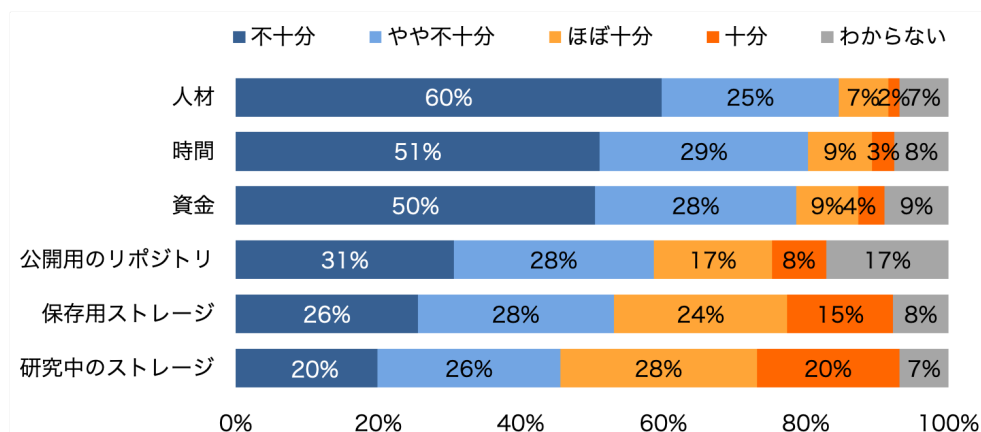


図 45 データ公開に関する資源の充足度 (n=1,513)

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。充足度が最も高い「研究中のデータ用ストレージ」であっても、「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。

比較のため、図 46 に 2016 年調査の結果を示す。配列は 2018 年調査と同様に、「不十分」と「やや不十分」の合計選択率が高い順とした。

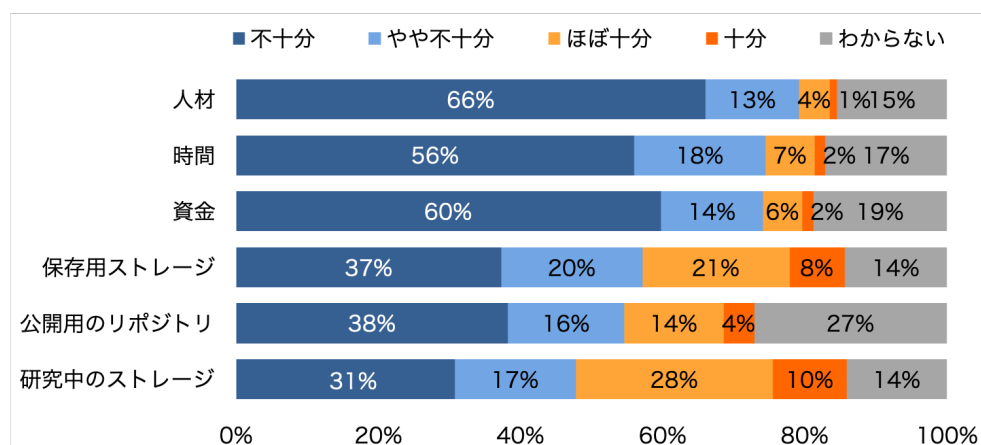


図 46 2016 年調査：データ公開に関する資源の充足度 (n=1,396)

2016 年調査と 2018 年調査の結果を比較すると、全体的に充足度は低いものの、やや改善されていた。また、2016 年調査ではいずれの項目も「わからない」とする回答が 14%を超えていたが、2018 年調査では低減していた。特に「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 ポイント減少していた。このことから、「公開用のリポジトリ」の認知度がやや向上した可能性が示唆された。

(2) カレントデータの公開に対する懸念【項目追加】

データ公開に対する懸念の強さを確認するために、9項目を挙げて「Q27. カレントデータを公開しようとする場合、次の点はどの程度問題となりますか（なりましたか）？あてはまるものをそれぞれお選び下さい。」と尋ねた。選択肢は、「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件法として、「わからない」も設けた。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 47）。なお、2018 年調査で新たに追加した項目には*印を付した。

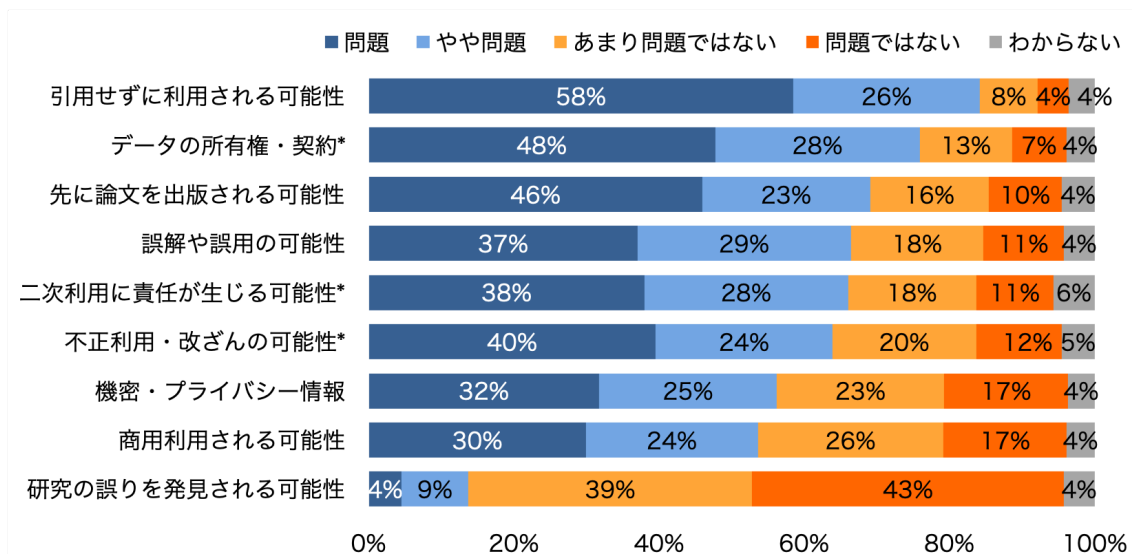


図 47 カレントデータの公開に関する懸念（n=1,513）

カレントデータの公開に関する問題のうち、最も懸念が強かった項目は、「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%）であった。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。

比較のため、図 48 に 2016 年調査の結果を示す。配列は 2018 年調査と同様に、「不十分」と「やや不十分」の合計選択率が高い順とした。

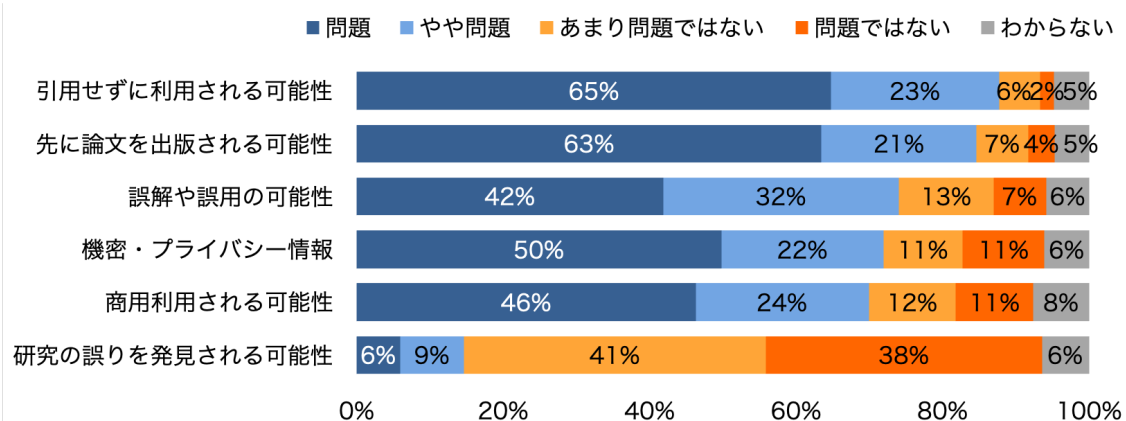


図 48 2016 年調査：カレントデータの公開に関する懸念 (n=1,396)

2016 年調査と 2018 年調査の結果を比較すると、「研究の誤りを発見される可能性」以外は懸念が強いものの、全体的にはやや低減していた。また、2016 年調査と 2018 年調査で共通の項目の順位は変化していなかった。

(3) 論文の OA に対する懸念【新規】

データ公開に対する懸念と OA に対する懸念の強さを比較するために、データ (Q27) と同様の 9 項目を挙げて「Q6. 論文やプレプリントをオープンアクセスにしようとする場合、次の点はどの程度問題となりますか？あてはまるものをそれぞれお選び下さい。」と尋ねた。選択肢は、「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 件法として、「わからない」も設けた。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 49)。

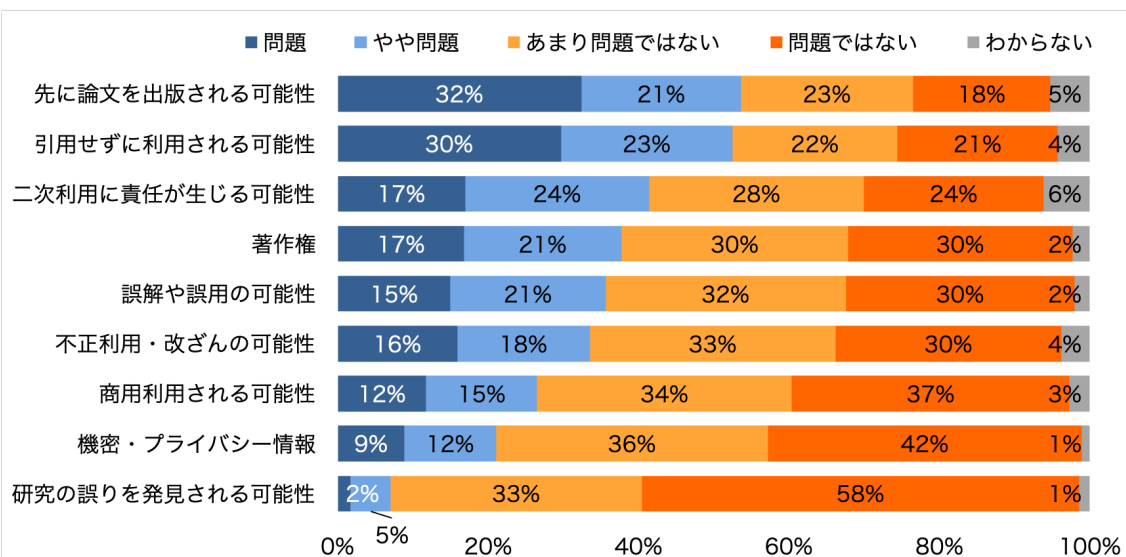


図 49 論文の OA に対する懸念 (n=1,516)

論文の OA に関する問題のうち、最も懸念が強かった項目は「公開したプレプリント（＝学術雑誌などで出版される前の原稿）を使って自分より先に論文を出版される可能性」（「問題である」と「やや問題である」の合計 53.6%）、次いで「引用せずに利用される可能性」（同 52.5%）であった。その他の項目は「問題である」と「やや問題である」の合計が 50% 未満であり、全体的に懸念が弱く、特に「研究の誤りを発見される可能性」（同 7.1%）への懸念が弱かった。

データ公開に対する懸念と順位を比較すると、多少の違いがみられた。OA の 1 位は「先に論文を出版される可能性」であった。3 位の「二次利用に責任が生じる可能性」は、データ公開（5 位）よりも高順位であった。4 位の「著作権」に相当する項目は 2 位の「データの所有権・契約」であり、データよりも低順位であった。論文の著作権は明確にされる場合が多いが、データの場合はそもそも著作権等が認められるかどうかの判断も困難であるため、懸念が強いのではないかと考えられる。

(4) データ公開のネガティブな影響【新規】

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なった。実際に、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、「Q20. 研究データの公開によって問題が起きたご経験がありましたら、差し支えない範囲でお書き下さい。」と自由記述方式で尋ねたところ、データ公開経験をもつ回答者のうち 234 名から回答が得られた。うち、4 名は論文の OA に関するコメント^{xxiii}、7 名はデータ公開によって生じた問題以外についてのコメントを記載していたため、これらを除く 223 名の回答を分類して集計した。その結果、何らかの問題が起きたと記述していたのは 24.2%（54 名）であり、75.8%（169 名）は特に問題が起きていない旨を記述していた（図 50）。

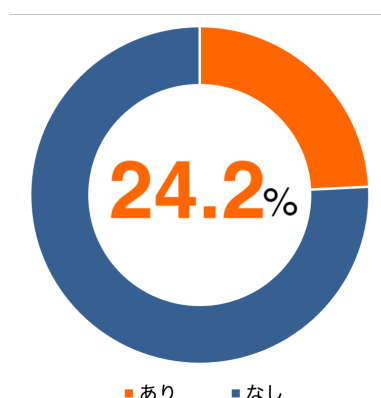


図 50 データ公開によって問題が起きた経験（n=223）

^{xxiii} OA 論文を盗用された（2 名）、OA のための費用が高額（1 名）、質の悪い論文とともに掲載される場合がある（1 名）というものであった。

表 7 に回答の内容をタグ付けして集計し、「なし」及び7項目に分類した結果を示す。

表 7 データ公開によるデメリット

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%

問題が起きたとする回答のうち、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%)であり、研究者、企業、マスコミ、学会、一般の方からの問い合わせに時間を取られる、または“対応できない”といった記述がみられた。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。3位は「先取権の喪失」(4.0%)であり、論文出版や特許出願の際の制約となったとする記述や、“類似の研究テーマが別の国の研究者によって立ち上げられてしまった”といった記述がみられた。4位は「誤用された」(3.1%)であり、“データの使用条件をあまり理解しないで使って、誤った結論を導くことがあった”との記述がみられた。5位の「更新のコストがかかる」(0.9%)には保守のための手間や予算が必要であること、「徒労感」(0.9%)には公開の手間をかけても利用されない、やりがいを感じられないという記述がみられた。「その他」(1.3%)には、商用利用された、研究者以外に利用された、“公開サイトの管理が甘く、ハッキングによってデータを消されたことがある”との記述がみられた。

(5) 他の研究者によるカレントデータの理解

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ことが重要であり、先行研究では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「できると思う」、「やや難しいと思う」、「難しいと思う」の3件法で尋ねた（Q28, Q29）。図 51 に結果を示す。

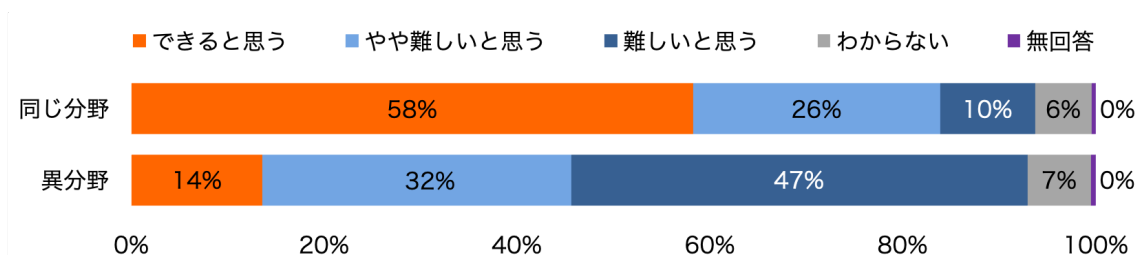


図 51 他の研究者によるデータの理解（n=1,513）

同じ分野の研究者については58.3%が理解「できると思う」を選択していたのに対して、異分野の研究者については13.6%にとどまった。また、「難しいと思う」同じ分野の研究者が9.8%、異分野の研究者は47.3%であった。

比較のため、図 52 に 2016 年調査の結果を示す。

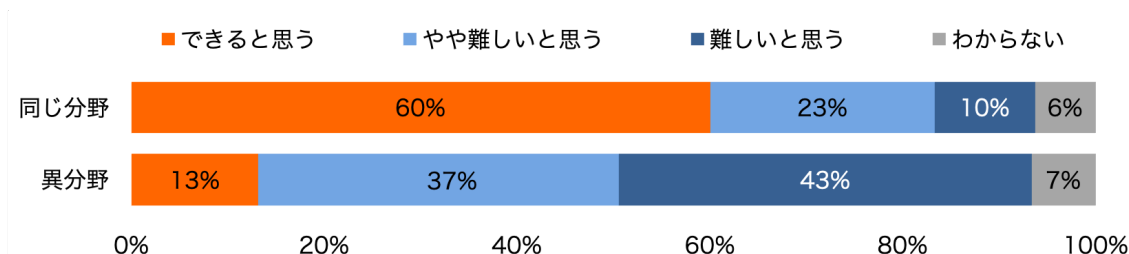


図 52 2016 年調査：他の研究者によるデータの理解（n=1,396）

2016 年調査と 2018 年調査はおおむね同様の傾向がみられた。2016 年調査では、同じ分野の研究者であれば、60.0%が理解「できると思う」を選択していたのに対して、異分野の研究者については13.2%であった。

2.6 データ公開のインセンティブ【新規】

(1) データ公開のポジティブな影響

データ公開のメリットやインセンティブとなりうる項目を明らかにするために、「Q19. 研究データの公開によって良い結果が得られたご経験がありましたら、差し支えない範囲でお書き下さい」と自由記述方式で尋ねたところ、データ公開経験をもつ回答者のうち 195

名から回答が得られた。表 8 に回答の内容をタグ付けして集計し、7 項目に分類した結果を示す。なお、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。

表 8 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%

データ公開によって得られた良い結果のうち、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。具体的な内容を表 9 に示す。

表 9 データ公開によって得られた良い結果の詳細 (n=195)

内容と詳細
1. 研究上の利点 共同研究の契機 (82), 研究の進展 (10), 研究の信頼性 (7), データ管理の実践 (3), 受託研究 (2)
2. 研究・データ・研究者のビジビリティ向上 認知度向上 (36), 引用 (20), 学生の獲得 (6), 広報 (4)
3. 科学・分野の進展 科学の進展 (15), 再利用 (9), 社会貢献 (3)
4. 人とのつながり 人脈 (14), コミュニケーションの契機 (12)
5. 評価 研究・研究者・組織の評価, 信頼 (11)
6. 個人的な利点 研究の効率化 (4), 商業的利益 (3), 実感が得られた (2)
7. その他 教育 (3), その他 (10)

()内の数字は人数を示す。

「その他」には，“Web へのアクセス解析によって，関心のある団体・企業を知ることができる”，“論文の本文を端的に表現できるようになり，読みやすくなっているように感じる”，“論文査読者が著者の情報を知りたい際に，おそらく著者をネットで検索しホームページを見る。その際に，査読者が著者の研究者レベルを判断する基準として使用しているようである”といった記述がみられた。

(2) データ公開によるインセンティブの重要度

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明らかにするために，6 項目を挙げて「Q23. 研究データ公開のインセンティブとして，次の点はどの程度重要だと思いますか？」と尋ねた。選択肢は「重要」から「重要ではない」までの 4 件法として，「わからない」も設けた。図 53 に結果を示す。

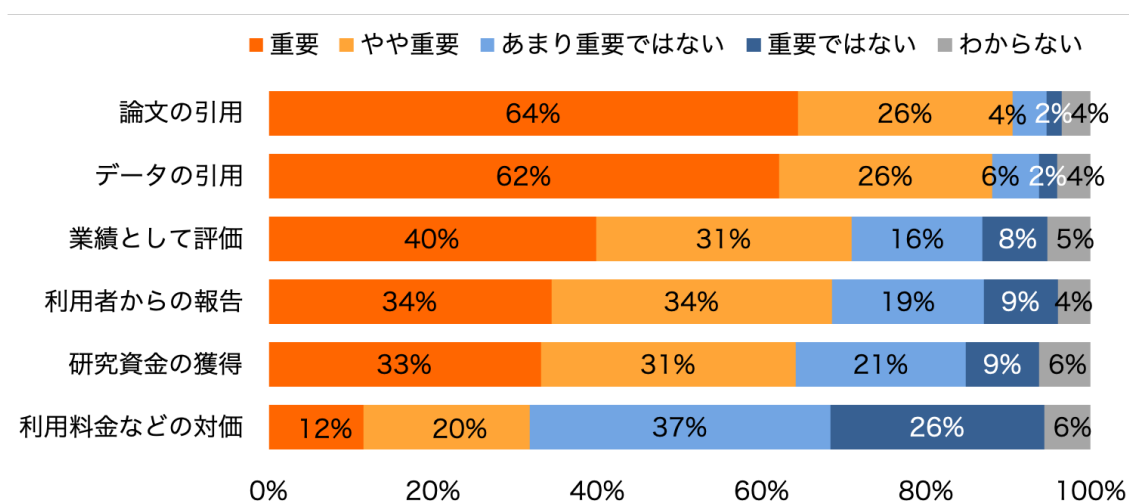


図 53 データ公開によるインセンティブの重要度 (n=1,513)

データ公開によるインセンティブとして，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。比較的重要度が高いとみなされていたのは「業績として論文と同様に評価されること」（同 70.9%），「利用者からの報告」（同 68.5%），「研究資金の獲得」（同 64.2%）であった。一方，「利用料金などの対価」（同 31.8%）が重要であると考えている回答者の比率は，相対的に低かった。

論文やデータの引用はデータ公開のインセンティブとして重要視されていた。また，データを公開する場合に最も懸念されることは「引用せずに利用される可能性」であった（図 47）。つまり，研究者は引用を重要視していると推測されることから，データ公開の推進においては引用の徹底が重要であると考えられる。

2.7 データの整備・公開に対する認識

(1) データの整備・公開に関するリテラシーへの関心

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に、「Q30. 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。」と尋ねた。選択肢として8種類の項目と「その他」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「あてはまるものはない」を示した。項目を1つ以上選択した回答者は、データリテラシー教育に関心が「ある」とみなして集計した。図 54 に集計結果を示す。

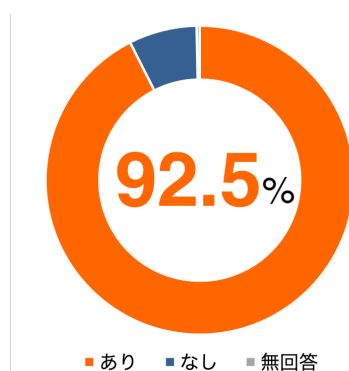


図 54 データリテラシーへの関心の有無 (n=1,513)

データの整備・公開に関する項目に関心をもつ回答者の比率は 92.5%, 「なし」は 7.1%, 無回答は 0.3%であった。2016 年調査では、関心をもつ回答者の比率は 90.0%であったため、2.5 ポイント増加していた。

(2) データを整備・公開する上でより詳しく知りたい項目

図 55 に、2016 年調査と 2018 年調査におけるデータを整備・公開する上でより詳しく知りたい項目の集計結果を示す。配列は、2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率 (%) は、2018 年調査の値である。

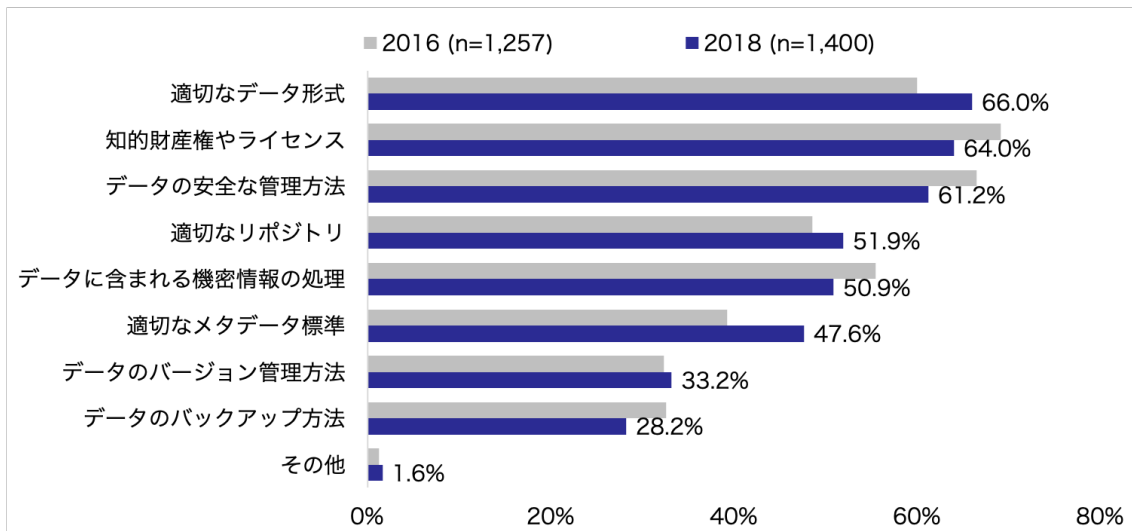


図 55 データを整備・公開する上でより詳しく知りたい項目

データを整備・公開する上でより詳しく知りたい項目のうち、選択率が最も高かったのは、「適切なデータ形式」(66.0%)、次いで「知的財産権やライセンス」(64.0%)、「データの安全な管理方法」(61.2%)、「適切なりポジトリ」(51.9%)の順であった。2016年調査と比較すると、やや順位に変動があるものの、関心が高い項目の傾向はおおむね同じであった。関心が高まっていたのは、「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」で39.3%から47.6%まで8.3ポイント増加していた。また、「適切なデータ形式」も60.0%から66.0%まで6.0ポイント増加していた。「その他」では、“幅広い研究者が理解できる内容にするための労力を削減するためのフォーマットや自動化の方法”など省力化の方法（2名），“データ公開・整備のための予算獲得や運用，人材獲得方法”など資金，人材といったリソースの獲得方法や情報（2名），“参加者の個人情報の保護等倫理的配慮とのバランス”といった記述がみられた。また，そもそもデータ整備・公開は不要である，公開したくない，整備・公開のコストを誰が負担するのか，といったコメントもみられた。

(3) データ整備・公開の専門性

データの管理や公開の支援人材に必要な専門性を明らかにするために、「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」と尋ねた。選択肢として7種類の項目と「その他」を複数選択方式で示すとともに，これらとは同時に選べない排他的選択肢として「あてはまるものはない（いずれも専門性を必要としない）」を示した。項目を1つ以上選択した場合は，データの整備・公開に専門性が「ある」と考えているとみなして集計した。図 56 に集計結果を示す。

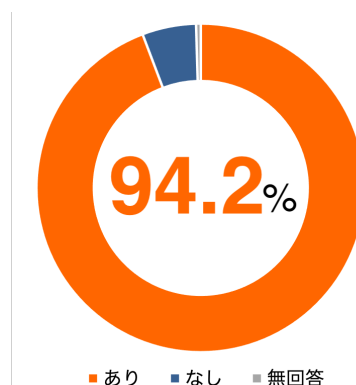


図 56 データの整備や公開における専門性の有無 (n=1,513)

データの整備・公開に何らかの専門性が必要であると考えている回答者は 94.2%，なしは 5.3%，無回答は 0.5%であった。2016 年調査 (n=1,396) では。専門性が必要であると考えていた回答者は 93.3%であったため，0.9 ポイント増加した。

図 57 に，2016 年調査と 2018 年調査の項目別選択率を示す。配列は，2018 年調査において選択率が高かった公開方法の順とした。グラフの右に付した比率 (%) は，2018 年調査の値である。

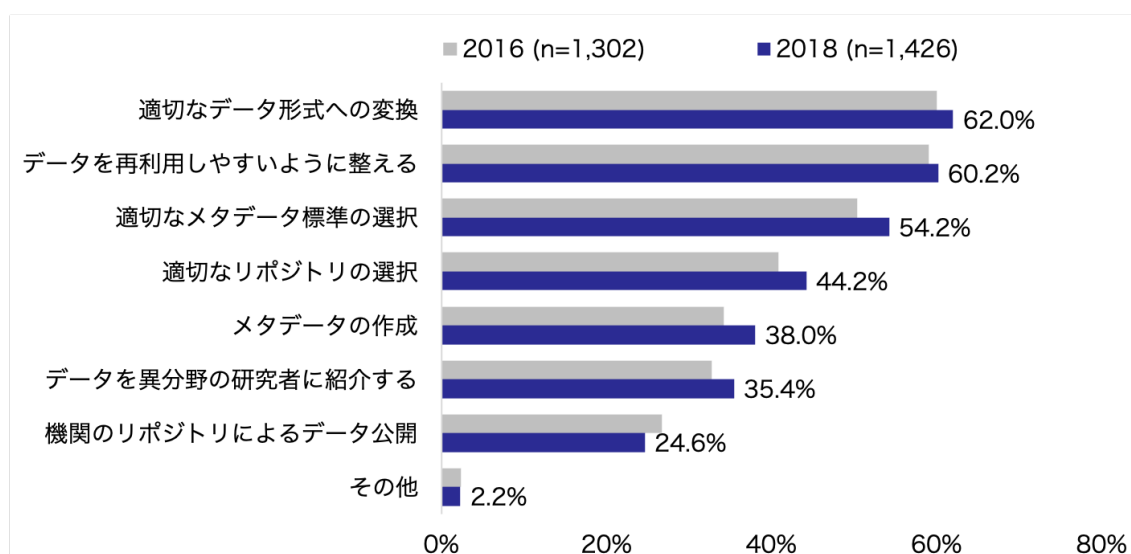


図 57 専門性を必要とする項目 (n=1,426)

2016 年調査と 2018 年調査を比較すると，順位に変化はなかった。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」(62.0%)，「データを再利用しやすいように整える」(60.2%)，「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」(54.2%) であった。比較的，選択率が低かったのは「機関のリポジトリによるデータ公開」(24.6%) や「データを異分野の研究者に紹介する」(35.4%)，「メタデータの作成」(38.0%) であった。

「その他」では、“（ソフトウェアのソースコードについて）適切なライセンスの選択”，“データを利用することによって可能になること。その範囲の宣言。このデータを用いることによって，どういった研究に使うことができるか，何を知ることができるか”といった記述がみられた。また，そもそも第三者による支援は困難・不可能であるとする意見が 12 件あった。たとえば，“研究の内容が理解できないとほとんど無理。テクニックで済むものではない”，“試験条件，試験方法等，そのデータがどのように取得されたものかが分からないと意味がない”，“他者が行うことはほぼ無理でしょう。活発な研究室ならば，ラボに 1 人は専門のスタッフがいれば可能かもしれません”といった記述がみられた。同様の指摘は 2016 年調査においても 11 件あったことから，人的支援の可能性については，より詳細な調査と議論が必要であると考えられる。

3. まとめ

日本の研究者によるデータと論文の公開状況や認識を明らかにすることを目的として、2018年に科学技術専門家ネットワークを活用したウェブ質問紙調査を実施した。さらに、大学、企業、公的機関・団体に所属する研究者1,516名（回答率69.1%）による回答を分析した結果を2016年調査や先行研究の結果と比較することによって、オープンサイエンスの課題や支援の可能性を検討することを目指した。

データ公開率（51.9%）は2016年調査から0.9ポイントしか増加しておらず、論文のOA率（78.0%）が7.1ポイント増加しているのと比較すると、ほとんど変化していないことが明らかになった。分野別のデータ公開率を確認すると、生物科学の66.5%から工学の40.2%まで差がみられた。論文の公開率も人文・社会科学（88.9%）から工学（69.5%）まで差がみられたが、両者の関連はみられなかった。つまり、データ公開率が高い（低い）分野は論文の公開率が高い（低い）とはいえなかった。

データの公開理由は1位が「研究成果を広く認知してもらいたいから」、2位が「論文を投稿した雑誌のポリシーだから」、論文の公開理由は1位が「論文を投稿した雑誌のポリシーだから」、2位が「研究成果を広く認知してもらいたいから」であり、2016年調査と同様であった。データの未公開理由は2016年調査では突出した項目がなかったが、本調査では1位の「論文を投稿した雑誌のポリシーではないから」と2位の「ニーズがないと思うから」の選択率がやや高かった。論文の未公開理由は2016年調査と同様に、1位の「論文を投稿した雑誌のポリシーではないから」と2位の「資金がないから」に集中していた。また、ハゲタカジャーナルの存在やOA誌の質を疑問視する記述もみられた。

データの提供（共有）経験は67.7%が有していた。分野別にみると、地球科学（81.6%）から数学（55.0%）まで差がみられた。データの提供方法のうち最も多かったのはE-mail、次いで外部記憶媒体（USBフラッシュメモリなど）、外部のクラウドサービス（Dropboxなど）の順であった。

公開データの入手経験は73.0%が有していた。分野別にみると、計算機科学（88.2%）から心理学（53.3%）まで差がみられた。公開データの入手において問題があるとしていた回答者は78.5%であり、1位は「データごとに品質が異なる」、2位は「データごとにフォーマットが異なる」、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」であった。2016年調査ではデータの入手時点における問題（利用料金や利用者登録が必要であること）の選択率が高かったが、本調査ではデータの利活用における問題の選択率が高かった。

公開データや論文の探索ツールはGoogleなどのサーチエンジンが最もよく使われていた。論文は論文情報のデータベース（CiNii Articlesなど）が2番目によく使われているのに対して、データ情報のデータベース（Data Citation Indexなど）は6位にとどまり、2位は論文や学術記事の参考文献であった。今後、データ情報のデータベースが拡充することによって、公開データの入手が容易になることが期待される。

DMP の作成経験をもつ回答者は 18.7%であった。分野別にみると、医学 (33.6%) から数学 (5.0%) まで差がみられた。作成経験がある DMP の 1 位は所属機関の DMP、2 位は個人や研究グループのための DMP であった。また、DMP の作成理由の 1 位は「所属機関が要求しているから」であった。DMP を作成していない理由の 1 位は「DMP を知らなかったから」であり、現状では DMP の認知度がそれほど高くないことが示唆された。

データ公開に必要な資源の充足度は 2016 年調査と比較するとやや改善されていたものの全体的に低く、特に人材、時間、資金が不足していると考えられていた。データ公開に必要な人材像について、第三者がデータの整備や公開する場合に専門性が必要であると考えた回答者は 94.2%であった。専門性が必要であると考えられている項目の 1 位は「適切なデータ形式への変換」、2 位は「データを再利用しやすいように整える」、3 位は「適切なメタデータ標準の選択」であった。これらの知識や技術を習得した支援人材の育成が課題であると考えられる。一方、選択率が最も低かったのは「機関のリポジトリによるデータ公開」(n=1,426, 24.6%) であり、機関リポジトリを管理する図書館等による支援の可能性が示唆された。

データ公開に対する懸念も依然として強いことが明らかになった。一方で、実際にデータを公開している回答者の記述から、実際に問題が起きるケースはそれほど多くないことが示唆された。また、データ公開によって、研究上の利点が得られたことや、研究・データ・研究者のビジビリティが向上したことなどが明らかになった。このようなデータ公開によって起きたことは、これまで十分に明らかにされてこなかった。こうした事例を共有することによって、データ公開のインセンティブやデータを公開する上で注意すべき点が整理され、データ公開を後押しできるのではないかと考えられる。

データ公開に対する懸念のうち、最も強かったのは引用されずに利用されることであった。また、データ公開のインセンティブのうち、最も重視されていたのは論文やデータの引用であった。研究者が安心してデータを公開できる環境づくりの第一歩として、データ引用を学術界の慣習とすることが重要であるといえよう。そのためには、研究機関、学術出版社、学協会、助成機関といったステークホルダーの協力が必要であると考えられる。

データの公開率はほとんど変化がみられなかったものの、データの公開理由や公開データの利用状況などの変化から、オープンサイエンスに対する認識が少しずつ高まっている可能性が示唆された。たとえば、データ公開理由のうち 2016 年調査と比較して選択率が増加していたのは、「オープンデータに貢献したいから」と「分野・コミュニティの規範だから」であった。前述の通り、2016 年調査ではデータを入手する時点で障壁を感じる回答者が多かったが、2018 年調査では先行研究と同様に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。データの公開先や公開データの入手先は依然として「論文の補足資料」や「個人や研究室のウェブサイト」が多いものの、「特定分野のリポジトリ・データアーカイブ」の選択率が増えていた。また、データ公開に必要な資源のうち、「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 27.1%から 17.2%まで減少していた。

オープンサイエンスを取り巻く環境は常に変化している。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いくと考えられる。研究のための資金や時間など限られ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することによって、適切な学術政策に資することを今後の課題としたい。

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>
- 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>
- ³ Foreign & Commonwealth Office. G8 Science Ministers Statement. GOV.UK. 2013-6-13, <https://www.gov.uk/government/news/g8-science-ministers-statement>
- 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. <https://www.nistep.go.jp/wp/wp-content/uploads/NISTEP-STT146J-12.pdf>
- ⁵ 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
- ⁶ 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. https://doi.org/10.18919/jkg.66.3_98
- ⁷ 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>
- ⁸ 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>
- ⁹ OECD. Making Open Science a Reality. OECD Publishing, 2015, OECD Science, Technology and Industry Policy Papers, no. 25, 112p. <https://doi.org/10.1787/5jrs2f963zs1-en>
- ¹⁰ Tenopir, Carol et al. Data sharing by scientists: Practices and perceptions. PLOS ONE. 2011, vol. 6, no. 6, e21101. <https://doi.org/10.1371/journal.pone.0021101>

- ¹¹ 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.
<https://doi.org/10.1371/journal.pone.0134826>
- 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. <https://doi.org/10.1371/journal.pone.0146695>
- ¹³ 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>
- ¹⁴ 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—オープンサイエンスの課題と展望—. STI Horizon. 2017, vol. 3, no. 4, p. 27-32.
<https://doi.org/10.15108/stih.00106>
- ¹⁵ 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, no. 79, p. 21-57.
http://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00003152-00000079-0021
- ¹⁶ 池内有為. データマネジメントプラン (DMP) —FAIR 原則の実現に向けた新たな展開 (連載: オープンサイエンスのいま). 情報の科学と技術. 2018, vol. 68, no. 12, p. 613-615. https://doi.org/10.18919/jkg.68.12_613
- ¹⁷ 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>
- ¹⁸ 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
- ¹⁹ Challenges and Opportunities for Japanese Researchers. Springer Nature, 2018.
<https://www.springernature.com/gp/open-research/open-data/japan-survey>
- ²⁰ <http://www.nistep.go.jp/activities/st-experts-network>
- ²¹ 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. <https://doi.org/10.11501/9497651>
- ²² 千葉浩之. ハゲタカジャーナル問題: 大学図書館員の視点から. カレントアウェアネス

ス. 2019, no. 341, p. 12-14. <https://doi.org/10.11501/11359093>

²³ インパクトファクター（文献引用影響率）とは. クラリベイト・アナリティクス.

<https://clarivate.jp/products/journal-citation-reports/impact-factor/>

²⁴ 大谷周平, 坂東慶太. 論文海賊サイト Sci-Hub を巡る動向と日本における利用実態. 情報の科学と技術. 2018, vol. 68, no. 10, p. 513-519.

https://doi.org/10.18919/jkg.68.10_513

²⁵ Kratz, John E.; Strasser, Carly. Making data count. Scientific Data. 2015, vol. 2, 150039 (2015). <https://doi.org/10.1038/sdata.2015.39>

資料

資料

(1) 調査依頼文・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。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](#)の「オープンサイエンスのためのデータ基盤の整備」では、主要施策として“研究分野の特性等を踏まえたオープン・アンド・クローズ戦略を考慮したデータポリシーやデータマネジメントプランの策定を促進し、これらに基づく研究データの管理・公開等を促進するとともに、公的資金による研究成果としての研究データについては、データインフラを通して機械判読可能化を促進する”とされています。

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などが期待されており、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考えられます。

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、分野や所属機関による特性、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。なお、本調査は2016年に実施した[ベースライン調査](#)の後続調査という位置づけです。ベースライン調査の結果は、オープンサイエンスに関する内閣府や文部科学省の議論で活用されております。**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大変参考になっております**ので、ぜひ研究者の皆様からの率直なご意見をお聞かせ下さい。

調査結果は、NISTEPの刊行物や学術出版物として公表する予定です。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。

2018年10月19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 池内為有

<回答要領>

- (1) 2018 年 11 月 2 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 32 問で、回答には 30 分程度を要します。必須項目（*印）は 19 問です。
なお、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) あらかじめ質問の内容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 ブラウザーの「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。
- ・ 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がないため、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。

以上の留意事項をふまえて、ご協力をよろしくお願い致します。

<アンケート担当>

科学技術予測センター

担当者名（問い合わせアドレス）

本調査は、オープンサイエンス（学術論文や、その根拠となる研究データを公開すること）についておうかがいします。

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)論文出版や学会等での口頭発表は行っていない【→Q32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うかがいします。

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
- ・ セルフアーカイブ：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【→Q6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。[複数選択可] *

- (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【※黄色の回答は他の選択肢とは同時に選べません】

【すべて→Q6へ】

Q4. 論文をオープンアクセスと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] *

- (1) 資金がないから
- (2) 時間がないから
- (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】

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. 論文やプレプリントをオープンアクセスにしようとする場合、次の点はどの程度問題となりますか？ あてはまるものをそれぞれお選び下さい。*

	問題ではない	あまり問題ではない	やや問題である	問題である	わからない
商用利用される可能性	1	2	3	4	0
機密・プライバシー情報	1	2	3	4	0
著作権	1	2	3	4	0
誤解や誤用の可能性	1	2	3	4	0
不正利用・改ざんの可能性	1	2	3	4	0
二次利用に関して責任が生じる可能性	1	2	3	4	0
引用せずに利用される可能性	1	2	3	4	0
研究の誤りを発見される可能性	1	2	3	4	0
公開したプレプリント*を使って自分より先に論文を出版される可能性	1	2	3	4	0

* 学術雑誌などで出版される前の原稿

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]

- (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:]

3. 研究データの提供について

本調査における「**研究データ**」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。

Q8. 共同研究者を除く他の研究者に研究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 【→Q10 へ】
- (5) わからない 【→Q10 へ】

Q9. 共同研究者を除く他の研究者に研究データを提供した方法をお選びください。[複数回答可]

- (1) E-mail
- (2) 外部記憶媒体（USB フラッシュメモリ、記録型 CD/DVD など）
- (3) 所属機関のクラウドサービス
- (4) 外部のクラウドサービス（Dropbox, Google Drive など）
- (5) その他 [具体的に：]
- (6) 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。

Q10. 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〕 ＊

- (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
- (7) 学術系 SNS（Mendeley, ResearchGate など）
- (8) その他〔具体的に： 〕

(9) ない【→Q14 へ】

(10) わからない【→Q14 へ】

Q11. 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〕 ＊

- (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（営利利用が可能かどうかなど）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

Q12. 入手した公開データの利用目的について、あてはまるものをお選び下さい。[複数選択可]

- (1) 自身の研究のアイデアや仮説の参考にする
- (2) 再分析・再利用して自身の研究を行う
- (3) 研究を再現・追試する
- (4) その他 [具体的に：]
- (5) あてはまるものはない

Q13. 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]

- (1) サーチエンジン (Google, Google Dataset Search 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：]

5. データマネジメントプラン（DMP）について

データマネジメントプラン（DMP）についておうかがいします。

本調査における「DMP」とは、研究データを適切に管理・公開・保存することを目的として、研究開始から研究終了後までのデータの取り扱いについて記載する文書を指します。助成機関や研究機関が研究費獲得の際に提出を求める場合があります。

【DMP の記載内容の例】

- ・ データの概要（データの種類、量など）
- ・ データの公開・非公開の方針（一般公開／共同研究者のみ／非公開、公開時期など）
- ・ データの保存場所（研究中・研究後）

Q14. これまでに、DMP を作成したご経験はありますか？ [複数選択可] *

- (1) 科学技術振興機構（JST）による助成研究の DMP
- (2) 日本医療研究開発機構（AMED）による助成研究の DMP
- (3) 新エネルギー・産業技術総合開発機構（NEDO）による助成研究の DMP
- (4) 所属機関の DMP
- (5) 個人や研究グループのための DMP
- (6) その他 [具体的に：]
- (7) ない 【→Q16 〜】
- (8) わからない 【→Q16 〜】

Q15. DMP を作成した理由として、あてはまるものをお選び下さい。 [複数選択可] *

- (1) 研究を行う際の習慣だから
- (2) 効率的にデータを管理するため
- (3) 助成機関が要求しているから
- (4) 所属機関が要求しているから
- (5) その他 [具体的に：]
- (6) あてはまるものはない
【すべて→Q17 〜】

Q16. DMP を作成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〕 *

- (1) DMP を知らなかったから
- (2) DMP の必要性を感じないから
- (3) DMP を作成する時間がないから
- (4) 助成機関から要求されていないから
- (5) 所属機関から要求されていないから
- (6) その他〔具体的に： 〕
- (7) あてはまるものはない

6. 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うかがいします。

(再掲) 本調査における「研究データ」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。

Q17. 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〕 *

- (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) 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
- (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【→Q21 〜】
- (10) わからない【→Q23 〜】
- (11) 研究にデジタルデータは用いない【→Q32 〜】

Q18. 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] *

- (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

Q19. 研究データの公開によって良い結果が得られたご経験がありましたら、差し支えない範囲でお書き下さい。（最大 500 文字） [例：共同研究につながった]

Q20. 研究データの公開によって問題が起きたご経験がありましたら、差し支えない範囲でお書き下さい。（最大 500 文字）

【すべて→Q23 へ】

Q21. 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] *

- (1) 資金がないから
- (2) 時間がないから
- (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【→Q23へ】

Q22. 0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

Q23. 研究データ公開のインセンティブとして、次の点はどの程度重要だと思われますか？ あてはまるものをそれぞれお選び下さい。 *

	重要ではない	あまり重要ではない	やや重要である	重要である	わからない
利用者からの報告	1	2	3	4	0
利用料金などの対価	1	2	3	4	0
データに紐づいた論文の引用	1	2	3	4	0
データの引用*	1	2	3	4	0
研究資金の獲得	1	2	3	4	0
業績として論文と同様に評価されること	1	2	3	4	0

* 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

7. 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“についてお聞きいたします（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。

Q24. 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。

- (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

Q25. 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：以降の質問は、（カッコ内）をお読みの上でお答え下さい
- (2) いいえ：以降の質問は、公開することを想定してお答え下さい
- (3) わからない：以降の質問は、公開することを想定してお答え下さい

Q26. 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）？ あてはまるものをそれぞれお選び下さい。*

	不十分	やや 不十分	ほぼ 十分	十分	わから ない
研究中のデータ用ストレージ	1	2	3	4	0
データ公開用のリポジトリ	1	2	3	4	0
研究終了後のデータ保存用ストレージ	1	2	3	4	0
データの整備・公開のための時間	1	2	3	4	0
データの整備・公開のための人材	1	2	3	4	0
データの整備・公開のための資金	1	2	3	4	0

Q27. カレントデータを公開しようとする場合、次の点はどの程度問題となりますか（なりましたか）？ あてはまるものをそれぞれお選び下さい。*

	問題で はない	あまり 問題で はない	やや 問題で ある	問題で ある	わから ない
商用利用される可能性	1	2	3	4	0
機密・プライバシー情報	1	2	3	4	0
データの所有権・契約	1	2	3	4	0
誤解や誤用の可能性	1	2	3	4	0
不正利用・改ざんの可能性	1	2	3	4	0
二次利用に関して責任が生じる可能性	1	2	3	4	0
引用せずに利用される可能性	1	2	3	4	0
研究の誤りを発見される可能性	1	2	3	4	0
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	1	2	3	4	0

- Q28. 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
- Q29. 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
- Q30. 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選び下さい [複数選択可]
- (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

Q31. 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]

- (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）

8. 自由回答

Q32. 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）

ご協力ありがとうございました。

(2) 単純集計表

回答者の属性

表 10 所属機関

機関	人数	比率
大学	987	65.1%
企業	270	17.8%
公的機関・団体	251	16.6%
その他	8	0.5%
合計	1,516	100.0%

表 11 年齢層

年代	人数	比率
20	13	0.9%
30	440	29.0%
40	753	49.7%
50	218	14.4%
60	84	5.5%
70	7	0.5%
非開示	1	0.1%
合計	1,516	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 12 回答者の研究分野

研究分野	人数	比率
工学	547	36.1%
生物科学	245	16.2%
化学	165	10.9%
医学	134	8.8%
農学	122	8.0%
地球科学	76	5.0%
コンピュータサイエンス (CS)	76	5.0%
物理学	74	4.9%
社会科学	31	2.0%
数学	20	1.3%
心理学	15	1.0%
天文学	6	0.4%
人文学	5	0.3%
合計	1,516	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 13 オープンアクセスの状況

OA 経験	人数	比率
OA の論文がある	1,182	78.0%
OA の論文はない	285	18.8%
わからない	49	3.2%
合計	1,516	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 14 論文を OA にした理由 (n=1,182, 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	925	78.3%
研究成果を広く認知してもらいたいから	622	52.6%
科学研究や成果実装を推進したいから	194	16.4%
オープンアクセスに貢献したいから	135	11.4%
他の研究者からのリクエストに応じて	100	8.5%
所属機関のポリシーだから	67	5.7%
分野・コミュニティの規範だから	46	3.9%
助成機関のポリシー（助成条件）だから	40	3.4%
その他	46	3.9%
あてはまるものはない	2	0.2%

- 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 15 論文が OA ではない理由 (n=285, 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	158	55.4%
資金がないから	122	42.8%
所属機関にオープンアクセス方針がないから	33	11.6%
ニーズがないと思うから	14	4.9%
時間がないから	10	3.5%
助成機関のポリシー（助成条件）ではないから	6	2.1%
分野・コミュニティで推奨されていないから	6	2.1%
リポジトリなどの公開手段がないから	3	1.1%
その他	20	7.0%
あてはまるものはない	16	5.6%

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 16 オープンアクセスの意思

OA 意思	人数	比率
はい	186	69.1%
いいえ	25	9.3%
わからない	58	21.6%
合計	269	100.0%

Q6. 論文やプレプリントをオープンアクセスにしようとする場合、次の点はどの程度問題となりますか？ あてはまるものをそれぞれお選び下さい。 *

表 17 論文やプレプリントを OA にする際の障壁 (n=1,516)

	0	1	2	3	4
商用利用される可能性	2.7%	36.9%	33.8%	14.8%	11.7%
機密・プライバシー情報	1.1%	41.7%	36.1%	12.2%	8.8%
著作権	2.3%	29.9%	30.1%	20.9%	16.8%
誤解や誤用の可能性	2.0%	30.4%	31.9%	20.6%	15.0%
不正利用・改ざんの可能性	3.8%	29.9%	32.7%	17.6%	15.9%
二次利用に関して責任が生じる可能性	6.1%	23.9%	28.5%	24.5%	17.0%
引用せずに利用される可能性	4.3%	21.3%	21.9%	22.8%	29.7%
研究の誤りを発見される可能性	1.5%	58.1%	33.4%	5.3%	1.7%
公開したプレプリント (= 学術雑誌などで出版される前の原稿) を使って自分より先に論文を出版される可能性	5.3%	18.2%	22.9%	21.2%	32.4%

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

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 18 論文の探索ツール (n=1,516, 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)	1,301	85.8%
論文情報のデータベース (Web of Science, Scopus, CiNii Articles など)	1,029	67.9%
論文や学術記事の参考文献	856	56.5%
論文やプレプリントのサーバ (PubMed Central, arXiv, J-Stage など)	746	49.2%
出版社や学術雑誌のサイト (Elsevier, Wiley など)	623	41.1%
研究者や同僚に尋ねる／教えてもらう	474	31.3%
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)	226	14.9%
学術系 SNS (Mendeley, ResearchGate など)	207	13.7%
アラートサービス (RSS など)	162	10.7%
ブログや一般的な SNS (Facebook, Twitter など)	156	10.3%
政府・機関・出版社などの広報, ニュースレター	104	6.9%
メーリングリスト	87	5.7%
その他	4	0.3%
無回答	0	0.0%

3. 研究データの提供について

Q8. 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

頻度	人数	比率
よくある	110	7.3%
たまにある	483	31.9%
ほとんどない	434	28.6%
まったくない	478	31.5%
わからない	11	0.7%
無回答	0	0.0%
合計	1,516	100.0%

Q9. 共同研究者を除く他の研究者に研究データを提供した方法をお選びください。[複数選択可]

表 20 データ提供方法 (n=1,027, 複数回答)

方法	人数	比率
E-mail	830	80.8%
外部記憶媒体 (USB フラッシュメモリ, 記録型 CD/DVD など)	363	35.3%
外部のクラウドサービス (Dropbox, Google Drive など)	235	22.9%
所属機関のクラウドサービス	146	14.2%
その他	23	2.2%
わからない	21	2.0%
無回答	5	0.5%

4. 公開データの利用について

Q10. 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

入手経験	人数	比率
あり（チェックあり）	1,106	73.0%
なし	382	25.2%
わからない	28	1.8%
合計	1,516	100.0%

表 22 公開データの入手方法（n=1,106，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	658	59.5%
論文の補足資料	616	55.7%
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）	512	46.3%
特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）	307	27.8%
コード共有サービス（GitHub など）	201	18.2%
学術系 SNS（Mendeley, ResearchGate など）	192	17.4%
データ共有サービス（figshare, zenodo など）	54	4.9%
その他	17	1.5%

Q11. 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

問題	人数	比率
あり（チェックあり）	868	78.5%
問題を感じたことはない	238	21.5%
合計	1,106	100.0%

表 24 公開データを入手する際の問題（n=868, 複数回答）

問題	人数	比率
データごとに品質が異なる	324	37.3%
データごとにフォーマットが異なる	295	34.0%
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	273	31.5%
著作者情報がよくわからない	249	28.7%
利用料金が必要	248	28.6%
利用者登録が必要	232	26.7%
データの解釈や再利用方法がよくわからない	200	23.0%
利用したいデータへのアクセス方法がよくわからない	198	22.8%
リクエストから入手までに時間がかかる	160	18.4%
最新のデータを入手できない	104	12.0%
その他	10	1.2%

Q12. 入手した公開データの利用目的について、あてはまるものをお選び下さい。

表 25 公開データの利用目的（n=1,106, 複数回答）

利用目的	人数	比率
自身の研究のアイデアや仮説の参考にする	912	82.5%
再分析・再利用して自身の研究を行う	484	43.8%
研究を再現・追試する	257	23.2%
その他	34	3.1%
あてはまるものはない	20	1.8%
無回答	3	0.3%

Q13. 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 26 公開データの探索ツール (n=1,103, 複数回答)

公開データの探索ツール	人数	比率
サーチエンジン (Google, Google Dataset Search など)	925	83.9%
論文や学術記事の参考文献	764	69.3%
研究者や同僚に尋ねる／教えてもらう	372	33.7%
出版社や学術雑誌のサイト (Elsevier, Wiley など)	327	29.6%
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)	296	26.8%
データ情報のデータベース (Data Citation Index など)	171	15.5%
学術系 SNS (Mendeley, ResearchGate など)	112	10.2%
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)	85	7.7%
ブログや一般的な SNS (Facebook, Twitter など)	76	6.9%
政府・機関・出版社などの広報, ニュースレター	68	6.2%
メーリングリスト	35	3.2%
特定のデータリポジトリ	28	2.5%
アラートサービス (RSS など)	25	2.3%
その他	2	0.2%
無回答	3	0.3%

5. データマネジメントプラン（DMP）について

Q14. これまでに、DMP を作成したご経験はありますか？〔複数選択可〕 *

表 27 DMP 作成経験の有無

DMP 作成経験	人数	比率
あり（チェックあり）	284	18.7%
なし	1,153	76.1%
わからない	79	5.2%
合計	1,516	100.0%

表 28 作成経験がある DMP（n=284, 複数回答）

作成した DMP	人数	比率
所属機関の DMP	133	46.8%
個人や研究グループのための DMP	108	38.0%
科学技術振興機構（JST）による助成研究の DMP	74	26.1%
新エネルギー・産業技術総合開発機構（NEDO）による助成研究の DMP	16	5.6%
日本医療研究開発機構（AMED）による助成研究の DMP	10	3.5%
その他	9	3.2%

Q15. DMP を作成した理由として、あてはまるものをお選び下さい。〔複数選択可〕 *

表 29 DMP を作成した理由（n=284, 複数回答）

理由	人数	比率
所属機関が要求しているから	142	50.0%
研究を行う際の習慣だから	91	32.0%
助成機関が要求しているから	88	31.0%
効率的にデータを管理するため	84	29.6%
その他	12	4.2%
あてはまるものはない	5	1.8%

Q16. DMP を作成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 30 DMP を作成していない理由 (n=1,232, 複数回答)

理由	人数	比率
DMP を知らなかったから	686	55.7%
所属機関から要求されていないから	506	41.1%
助成機関から要求されていないから	387	31.4%
DMP の必要性を感じないから	226	18.3%
DMP を作成する時間がないから	88	7.1%
その他	24	1.9%
あてはまるものはない	55	4.5%

6. 研究データの公開について

Q17. 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 31 データ公開経験の有無

データ公開経験	人数	比率
あり（チェックあり）	787	51.9%
なし	695	45.8%
わからない	31	2.0%
データは用いない	3	0.2%
合計	1,516	100.0%

表 32 データを公開した方法（n=787，複数回答）

公開方法	人数	比率
論文の補足資料	375	47.6%
個人や研究室のウェブサイト	360	45.7%
所属機関のリポジトリ・データアーカイブ	213	27.1%
特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）	146	18.6%
学術系 SNS（Mendeley, ResearchGate など）	57	7.2%
コード共有サービス（GitHub など）	42	5.3%
データ共有サービス（figshare, zenodo など）	19	2.4%
その他	15	1.9%

Q18. 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 33 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%

Q19. 研究データの公開によって良い結果が得られたご経験がありましたら、差し支えない範囲でお書き下さい。

Q20. 研究データの公開によって問題が起きたご経験がありましたら、差し支えない範囲でお書き下さい。

Q21. 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 34 データを公開していない理由 (n=695, 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	222	31.9%
ニーズがないと思うから	196	28.2%
所属機関にポリシーがないから	170	24.5%
時間がないから	169	24.3%
業績にならないから	163	23.5%
助成機関のポリシー（助成条件）ではないから	103	14.8%
リポジトリなどの公開手段がないから	58	8.3%
分野・コミュニティで推奨されていないから	46	6.6%
資金がないから	45	6.5%
その他	73	10.5%
あてはまるものはない	82	11.8%

Q22. Q21 の理由が解決された場合、研究データを公開したいと思いますか？ *

表 35 データ公開意思

データ公開意思	人数	比率
はい	131	21.4%
いいえ	188	30.7%
わからない	294	48.0%
合計	613	100.0%

※Q17.でデータを公開した経験がない回答者（695 名）から、Q21.データを公開していない理由を「あてはまるものはない」とした回答者 82 名を除いた。

Q23. 研究データ公開のインセンティブとして、次の点はどの程度重要だと思われますか？ あてはまるものをそれぞれお選び下さい。 *

	0	1	2	3	4
データに紐づいた論文の引用	3.5%	1.9%	4.1%	26.1%	64.4%
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）	4.0%	2.2%	5.7%	25.9%	62.1%
業績として論文と同様に評価されること	5.3%	7.9%	15.9%	31.1%	39.9%
利用者からの報告	4.0%	9.0%	18.5%	34.1%	34.4%
研究資金の獲得	6.2%	8.9%	20.7%	31.1%	33.1%
利用料金などの対価	5.6%	26.0%	36.5%	20.3%	11.5%

0=わからない, 1=重要ではない, 2=あまり重要ではない, 3=やや重要, 4=重要

7. 最近の研究のためのデータについて

Q24. 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 36 カレントデータの量

データ量	人数	比率
MB（メガバイト）以下	315	20.8%
GB（ギガバイト）	868	57.4%
TB（テラバイト）	166	11.0%
PB（ペタバイト）以上	7	0.5%
わからない	156	10.3%
無回答	1	0.1%
合計	1,513	100.0%

Q25. 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 37 カレントデータ公開の有無

カレントデータ公開	人数	比率
はい	255	16.9%
いいえ	1,163	76.9%
わからない	95	6.3%
合計	1,513	100.0%

Q26. 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）？あてはまるものをそれぞれお選び下さい。 *

表 38 カレントデータの管理資源の充実度（n=1,513）

	0	1	2	3	4
データの整備・公開のための人材	6.9%	59.7%	24.9%	6.9%	1.6%
データの整備・公開のための時間	7.6%	51.0%	29.3%	9.0%	3.1%
データの整備・公開のための資金	9.0%	50.4%	28.2%	8.7%	3.6%
データ公開用のリポジトリ	17.2%	30.6%	28.0%	16.6%	7.6%
研究中のデータ用ストレージ	7.8%	25.5%	27.6%	24.3%	14.8%
研究中のデータ用ストレージ	6.9%	19.9%	25.6%	27.6%	20.0%

0=わからない, 1=不十分, 2=やや不十分, 3=ほぼ十分, 4=十分

Q27. カレントデータを公開しようとする場合、次の点はどの程度問題となりますか
（なりましたか）？あてはまるものをそれぞれお選び下さい。*

表 39 カレントデータを公開する際の障壁（n=1,513）

	0	1	2	3	4
引用せずに利用される可能性	3.6%	4.2%	8.0%	25.7%	58.5%
データの所有権・契約	3.9%	7.5%	12.7%	28.2%	47.7%
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	4.5%	10.1%	16.3%	23.1%	46.0%
誤解や誤用の可能性	4.2%	11.1%	18.2%	29.3%	37.1%
二次利用に関して責任が生じる可能性	5.7%	10.6%	17.6%	28.0%	38.0%
不正利用・改ざんの可能性	4.6%	11.8%	19.8%	24.4%	39.5%
機密・プライバシー情報	3.6%	17.1%	23.0%	24.5%	31.7%
商用利用される可能性	3.8%	17.0%	25.6%	23.7%	29.9%
研究の誤りを発見される可能性	4.2%	43.0%	39.1%	9.3%	4.5%

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

Q28. 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 40 カレントデータの理解：同じ分野

理解（同分野）	人数	比率
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	882	58.3%
やや難しいと思う	387	25.6%
難しいと思う	149	9.8%
わからない	89	5.9%
無回答	6	0.4%
合計	1,513	100.0%

Q29. 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 41 カレントデータの理解：他分野

理解（同分野）	人数	比率
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	206	13.6%
やや難しいと思う	484	32.0%
難しいと思う	716	47.3%
わからない	100	6.6%
無回答	7	0.5%
合計	1,513	100.0%

Q30. 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選び下さい〔複数選択可〕

表 42 データの整備・公開方法への関心の有無

関心の有無	人数	比率
チェックあり（詳しく知りたい項目あり）	1,400	92.5%
あてはまるものはない	108	7.1%
無回答	5	0.3%
合計	1,513	100.0%

表 43 データの整備・公開について詳しく知りたい項目（n=1,400、複数回答）

項目	人数	比率
適切なデータ形式	924	66.0%
知的財産権やライセンス	896	64.0%
データの安全な管理方法（セキュリティ）	857	61.2%
適切なリポジトリ	727	51.9%
データに含まれる機密情報の処理	712	50.9%
適切なメタデータ*	667	47.6%
データのバージョン管理方法	465	33.2%
データのバックアップ方法	395	28.2%
その他	23	1.6%

*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））

- Q31. 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 44 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

専門性	人数	比率
チェックあり（専門性を必要とする）	1,426	94.2%
あてはまるものはない（いずれも専門性を必要としない）	80	5.3%
無回答	7	0.5%
合計	1,513	100.0%

表 45 データの整備や公開に専門性を必要とする項目（n=1,426，複数回答）

項目	人数	比率
適切なデータ形式への変換	884	62.0%
データを再利用しやすいように整える	858	60.2%
適切なメタデータ*標準の選択	773	54.2%
適切なリポジトリの選択	630	44.2%
メタデータ*の作成	542	38.0%
データを異分野の研究者に紹介する	505	35.4%
機関のリポジトリによるデータ公開	351	24.6%
その他	32	2.2%

*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

- Q32. 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。

※Q35 の自由回答は別途 Web 上に電子付録として掲載する。

(3) クロス集計表

表 46 分野別集計結果

分野	n	比率	データ	論文	入手	提供	DMP
生物科学	245	16.2%	66.5%	86.9%	78.4%	65.3%	20.0%
農学	122	8.0%	63.9%	82.8%	82.0%	65.6%	16.4%
化学	165	10.9%	60.6%	73.9%	75.2%	68.5%	15.2%
数学	20	1.3%	60.0%	85.0%	75.0%	55.0%	5.0%
地球科学	76	5.0%	59.2%	85.5%	85.5%	81.6%	22.4%
医学	134	8.8%	53.7%	85.1%	68.7%	61.9%	33.6%
心理学	15	1.0%	53.3%	86.7%	53.3%	60.0%	26.7%
計算機科学	76	5.0%	51.3%	77.6%	88.2%	68.4%	13.2%
物理学・天文学	80	5.3%	43.8%	82.5%	68.8%	73.8%	17.5%
人文学・社会科学	36	2.4%	41.7%	88.9%	63.9%	61.1%	16.7%
工学	547	36.1%	40.2%	69.5%	66.7%	68.7%	17.0%
全体	1,516	100.0%	51.9%	78.0%	73.0%	67.7%	18.7%

データ＝データ公開経験，論文＝論文の公開経験，入手＝データ入手経験，提供＝データ提供（共有）経験，DMP＝DMP 作成経験

調査資料-289

研究データ公開と論文のオープンアクセスに関する実態調査 2018

2020 年 5 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

〒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 2018

May 2020

IKEUCHI, Ui and HAYASHI, Kazuhiro

Science and Technology Foresight Center

National Institute of Science and Technology Policy (NISTEP)

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



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