

研究データ公開と論文のオープンアクセスに関する  
実態調査 2020

2021年11月

文部科学省 科学技術・学術政策研究所

データ解析政策研究室

池内 有為 林 和弘

【調査研究体制】

池内 有為

データ解析政策研究室  
客員研究官

[調査設計、調査実施、データ分析、報告書執筆]

林 和弘

データ解析政策研究室  
室長

[調査設計、調査実施、報告書執筆、報告書確認]

【Authors】

IKEUCHI, Ui

Affiliated Fellow

Research Unit for Data Application, National Institute of  
Science and Technology Policy (NISTEP), MEXT

HAYASHI, Kazuhiro

Director

Research Unit for Data Application, 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2020」, NISTEP RESEARCH MATERIAL, No. 316, 文部科学省科学技術・学術政策研究所.

DOI: <https://doi.org/10.15108/rm316>

IKEUCHI, Ui and HAYASHI, Kazuhiro “A Survey on Open Research Data and Open Access 2020,” NISTEP RESEARCH MATERIAL, No. 316, National Institute of Science and Technology Policy, Tokyo.

DOI: <https://doi.org/10.15108/rm316>

## 研究データ公開と論文のオープンアクセスに関する実態調査 2020

池内 有為、林 和弘

文部科学省科学技術・学術政策研究所 データ解析政策研究室

### 要旨

科学技術・学術政策研究所 (NISTEP) 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明らかにするために、2016年、2018年に続き、2020年にウェブ質問紙調査を実施した。調査対象は科学技術専門家ネットワークであり、大学、企業、公的機関・団体に所属する研究者 1,349名から回答を得た (回答率 70.5%)。

現在、研究活動を行っている 1,268名のうち、データは 44.7%、論文は 80.1%が公開経験を有していた。データの公開率は、地球科学 (70.2%) から工学 (27.7%) まで分野による差が大きかった。データの提供経験は 71.2%が、公開データの入手経験は 69.7%が有しており、分野別のデータの公開・共有・入手経験には有意な相関がみられた。助成機関等が要求するデータマネジメントプラン (DMP) の作成経験がある回答者は 20.8%であった。データ公開に必要な資源の充足度はやや改善されていたものの全体的に低く、人材は 78.5%が、時間は 72.8%が、資金は 71.8%が、それぞれ「不足」または「やや不足」として認識していた。データ公開に対する懸念も依然として強く、引用せずに利用される可能性を 89.8%が、先に論文を出版される可能性を 80.9%が「問題」または「やや問題」として認識していた。データ公開を「高く評価している」と「やや評価している」の合計は、コミュニティや機関が 33.0%、回答者自身が 45.2%であった。データ形式の変換等、研究データ管理 (RDM) を図書館員やデータキュレータに依頼したいと考える回答者は 41.1%であった。

## A Survey on Open Research Data and Open Access 2020

Research Unit for Data Application, National Institute of Science and Technology Policy (NISTEP), MEXT

IKEUCHI, Ui and HAYASHI, Kazuhiro

### ABSTRACT

The National Institute of Science and Technology Policy (NISTEP) conducted a web-based questionnaire survey in 2020, following on from 2016 and 2018, to identify the status and perceptions of data and publication by Japanese researchers. The survey was conducted by the Network of Science and Technology Professionals and received responses from 1,349 researchers affiliated with universities, companies, and public institutions and organisations (response rate 70.5%).

Of the 1,268 respondents who were currently engaged in research activities, 44.7% had published data and 80.1% had published papers. The rate of data publication varied widely between disciplines, from earth sciences (70.2%) to engineering (27.7%).

71.2% of the respondents had experience in providing data and 69.7% had experience

in obtaining public data, and there was a significant correlation between the experience of publishing, sharing and obtaining data by discipline. 20.8% of the respondents had experience in developing a Data Management Plan (DMP) as required by the funding agency. Overall, the level of resources required to publish data was low, although improving slightly, with 78.5% of respondents perceiving a 'shortage' or 'slight shortage' of human resources, 72.8% of respondents perceiving a 'shortage' of time and 71.8% of respondents perceiving a 'shortage' of funding. Concerns about data publication also remained strong, with 89.8% perceiving the possibility of data being used without citation as "problematic" or "somewhat problematic" and 80.9% perceiving the possibility of papers being published first as "problematic" or "somewhat problematic". The total of "highly" and "somewhat" valued data publication was 33.0% for communities and institutions and 45.2% for respondents themselves. 41.1% of respondents would be willing to ask a librarian or data curator to manage their research data management (RDM), such as converting data formats.

# 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 .....	vi
(4) データ公開の障壁 .....	vii
(5) データ公開のインセンティブ .....	viii
(6) 研究データ管理（RDM）に対する認識 .....	x
1. 調査の概要 .....	3
1.1 オープンサイエンスの動向と本調査の目的 .....	3
1.2 主要な概念の定義 .....	4
1.3 質問紙調査システムの変更について .....	6
1.4 調査項目 .....	6
(1) 新規質問項目 .....	9
(2) アンケートシステムの変更に伴う修正：経験等を尋ねる質問 .....	9
(3) アンケートシステムの変更に伴う修正：理由を尋ねる質問 .....	12
(4) その他の修正 .....	12
(5) 削除した項目 .....	12
1.5 調査対象 .....	14
1.6 調査期間 .....	14
1.7 回答率と回答者の属性 .....	14
(1) 回答者の所属 .....	15
(2) 回答者の年齢層 .....	15
(3) 回答者の専門分野 .....	15
2. 調査結果 .....	17
2.1 データと論文の公開状況 .....	17
(1) データと論文の公開経験 .....	17
(2) データと論文の公開方法 .....	22
(3) データの公開理由 .....	24
(4) 論文の公開理由 .....	25
(5) データの未公開理由 .....	26
(6) 論文の未公開理由 .....	28
(7) データの公開意思 .....	29
(8) 論文の公開意思 .....	29

2.2 データの共有（提供）状況.....	30
2.3 公開データの利用状況と課題.....	32
(1) 公開データの入手経験.....	32
(2) 公開データの入手方法.....	33
(3) 公開データ入手の障壁.....	35
2.4 データマネジメントプラン（DMP）の作成状況.....	38
(1) DMP の作成経験.....	38
(2) 作成経験がある DMP の種類.....	39
(3) DMP の作成理由.....	40
(4) DMP の未作成理由.....	42
2.5 データの整備や公開の障壁.....	43
(1) カレントデータの量.....	43
(2) カレントデータの整備・公開に必要な資源の状況.....	43
(3) カレントデータの公開に対する懸念.....	45
(4) データ公開のネガティブな影響.....	46
2.6 データ公開のインセンティブ.....	49
(1) データ公開のポジティブな影響.....	49
(2) データ公開によるインセンティブの重要度.....	51
(3) コミュニティや研究機関によるデータ公開に対する評価【新規】.....	52
(4) 回答者自身によるデータ公開に対する評価【新規】.....	53
2.7 研究データ管理（RDM）に対する認識.....	54
(1) RDM に関するリテラシーへの関心.....	54
(2) RDM に関してより詳しく知りたい項目.....	55
(3) RDM プロセスの依頼.....	56
3. まとめ.....	59
謝辞.....	62
参考文献.....	63
資料.....	69
(1) 調査依頼文・質問票.....	69
(2) 単純集計表.....	84
(3) クロス集計表.....	100

## 図目次

図 1	データと論文の公開経験 (n=1,268)	ii
図 2	分野別データ公開率 (2016/18/20 年)	ii
図 3	データの公開方法 (2016/18/20 年：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 (2016/18/20 年)	iv
図 5	分野別公開データの入手経験 (2016/18/20 年)	iv
図 6	公開データの入手方法 (2016/18/20 年：複数回答)	v
図 7	DMP 作成経験の経年変化 (2018/20 年)	vi
図 8	作成経験がある DMP (2018/20 年：複数回答)	vi
図 9	データの整備や公開に関する資源の充足度 (n=1,188)	vii
図 10	カレントデータの公開に関する懸念 (n=1,188)	viii
図 11	データ公開によるインセンティブの重要度 (n=1,180)	ix
図 12	データ公開に対するコミュニティ・機関と回答者自身の評価	x
図 13	研究データ管理 (RDM) に関して知りたい項目 (2016/18/20 年：複数回答)	x
図 14	研究データ管理 (RDM) に関して依頼したい項目 (n=488：複数回答)	xi
図 15	2016 年調査：アンケート入力画面	10
図 16	2020 年調査：アンケート入力画面	11
図 17	回答者の所属 (n=1,349)	15
図 18	回答者の専門分野 (n=1,349)	16
図 19	データと論文の公開経験 (n=1,268)	17
図 20	データ公開経験の経年変化 (2016/18/20 年)	18
図 21	論文の公開経験の経年変化 (2016/18/20 年)	19
図 22	分野別データ公開率 (2016/18/20 年)	19
図 23	分野別論文の公開率 (2016/18/20 年)	20
図 24	分野別データと論文の公開率 (n=1,268)	21
図 25	データの公開方法 (2016/18/20 年：複数回答)	22
図 26	論文の公開方法 (n=1,016：複数回答)	23
図 27	データの公開理由 (2016/18/20 年：複数回答)	24
図 28	論文の公開理由 (2016/18/20 年：複数回答)	25
図 29	データの未公開理由 (2016/18/20 年：複数回答)	26
図 30	論文の未公開理由 (2016/18/20 年：複数回答)	28
図 31	データの公開意思 (2016/18/20 年)	29
図 32	論文の公開意思 (2016/18/20 年)	30
図 33	データの共有経験 (2016/18/20 年)	30
図 34	分野別データ共有経験 (2016/18/20 年)	31
図 35	公開データの入手経験の有無 (2016/18/20 年)	32
図 36	分野別公開データの入手経験 (2016/18/20 年)	33
図 37	公開データの入手方法 (2016/18/20 年：複数回答)	34

図 38	データ入手の障壁の有無 (2016/18/20 年)	35
図 39	データ入手の障壁 (2016/18/20 年: 複数回答)	36
図 40	DMP の作成経験 (n=1,268)	38
図 41	DMP 作成経験の経年変化 (2018/20 年)	38
図 42	分野別 DMP の作成経験 (2018/20 年)	39
図 43	作成経験がある DMP (2018/20 年: 複数回答)	40
図 44	DMP の作成理由 (2018/20 年: 複数回答)	41
図 45	DMP の未作成理由 (2018/20 年: 複数回答)	42
図 46	カレントデータの量 (2016/18/20 年)	43
図 47	データの整備や公開に関する資源の充足度 (n=1,188)	44
図 48	データ公開資源の不足度 (2016/18/20 年)	44
図 49	カレントデータの公開に関する懸念 (n=1,188)	45
図 50	データ公開に対する懸念 (2016/18/20 年)	46
図 51	データ公開によって問題が起きた経験の有無 (n=567)	47
図 52	データ公開によって良い結果が得られた経験の有無 (n=567)	49
図 53	データ公開によるインセンティブの重要度 (n=1,180)	51
図 54	データ公開によるインセンティブの重要度 (2018 年: n=1,513)	52
図 55	コミュニティや機関によるデータ公開に対する評価 (n=1,182)	52
図 56	回答者自身によるデータ公開に対する評価 (n=1,180)	53
図 57	データ公開に対するコミュニティ・機関と回答者自身の評価	54
図 58	研究データ管理 (RDM) に関するリテラシーへの関心の有無	55
図 59	研究データ管理 (RDM) に関して知りたい項目 (2016/18/20 年: 複数回答)	55
図 60	研究データ管理 (RDM) プロセスの依頼意思 (n=1,188)	56
図 61	研究データ管理 (RDM) に関して依頼したい項目 (n=488: 複数回答)	57
図 62	データ整備・公開に専門性を必要とする項目 (2016/18 年: 複数回答)	58



## 表目次

表 1	データ入手の障壁の経年変化 (2016/18/20 年)	v
表 2	データ公開によるデメリット (2018/20 年)	viii
表 3	データ公開によって得られた良い結果 (2018/20 年)	ix
表 4	2020 年調査の目的と調査項目の概要	7
表 5	質問項目一覧	8
表 6	回答者の年齢層	15
表 7	データの未公開理由:「その他」の回答 (n=32)	27
表 8	データの公開方法と入手方法	34
表 9	データ入手の障壁の経年変化 (2016/18/20 年)	37
表 10	データ公開によるデメリット	47
表 11	データ公開によるデメリット (2018/20 年)	48
表 12	データ公開によって得られた良い結果 (2018/20 年)	50
表 13	データ公開によって得られた良い結果の詳細 (n=152)	50
表 14	回答者の所属機関	84
表 15	回答者の年齢層	84
表 16	回答者の研究分野	85
表 17	研究活動の状況	85
表 18	オープンアクセスの状況	86
表 19	論文の公開方法 (n=1,016:複数回答)	86
表 20	論文を OA にした理由 (n=1,016:複数回答)	87
表 21	論文が OA ではない理由 (n=231:複数回答)	87
表 22	オープンアクセスの意思	88
表 23	データ提供経験	88
表 24	公開データの入手経験	89
表 25	公開データの入手方法 (n=884:複数回答)	89
表 26	公開データ入手における問題の有無	90
表 27	公開データを入手する際の問題 (n=816:複数回答)	90
表 28	DMP 作成経験の有無	91
表 29	作成経験がある DMP (n=264:複数回答)	91
表 30	DMP を作成した理由 (n=264:複数回答)	91
表 31	DMP を作成していない理由 (n=891:複数回答)	92
表 32	データ公開経験の有無	92
表 33	データを公開した方法 (n=567:複数回答)	93
表 34	データを公開した理由 (n=567:複数回答)	93
表 35	データを公開していない理由 (n=593:複数回答)	94
表 36	データ公開意思	94
表 37	データ公開のインセンティブの重要度 (n=1,180)	95
表 38	コミュニティや機関によるデータ公開の評価	95

表 39	回答者自身によるデータ公開の評価 .....	96
表 40	カレントデータの量.....	96
表 41	カレントデータの管理資源の充実度 (n=1,188) .....	97
表 42	カレントデータを公開する際の障壁 (n=1,513) .....	97
表 43	研究データ管理 (RDM) に関して知りたい項目 (n=1,039 : 複数回答) ...	98
表 44	研究データ管理 (RDM) プロセスの依頼意思 .....	98
表 45	研究データ管理 (RDM) に関して依頼したい項目 (n=488 : 複数回答) ..	99
表 46	分野別集計結果.....	100

# 概要



## 概要

---

科学研究の成果であるデータや論文を公開して、学術関係者のみならず民間企業を含む一般市民が自由にアクセスして利用できるようにするオープンサイエンス政策は、世界的な広がりを見せている。

日本でも、2021年3月26日に閣議決定された『第6期科学技術・イノベーション基本計画』<sup>1</sup>では、第2章 Society 5.0の実現に向けた科学技術・イノベーション政策の「2. 知のフロンティアを開拓し価値創造の源泉となる研究力の強化」において、オープンサイエンスとデータ駆動型研究等の推進のために新たな研究システムを構築することが明記されている (p. 58～)。こうした政策を適切かつ効率的に実現していくためには、まず、現状を把握した上で、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が誘発されることが望ましいと考えられる。特に研究者が論文に加えて研究データの共有・公開することによって、データ駆動型研究が促進されるが、その実態は十分に明らかになっていない。

そこで科学技術・学術政策研究所 (NISTEP) 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明らかにするために、2016年<sup>2</sup>と2018年<sup>3</sup>に科学技術専門家ネットワークを対象としてウェブ質問紙調査を実施した。3回目となる本調査では、大学、企業、公的機関・団体に所属する研究者 1,349名 (回答率 70.5%) による回答を分析した結果を2016/2018年調査の結果と比較した。調査項目は2016/2018年調査を踏襲しつつ、新たにデータ公開に対する評価に関する質問を行った。以下では概要として、(1)データと論文の公開状況、(2)公開データの利用状況と課題、(3)データマネジメントプラン (DMP) の作成状況、(4)データ公開の障壁、(5)データ公開のインセンティブ、(6)研究データ管理 (RDM) に対する認識についての結果を示す。

---

<sup>1</sup> 内閣府. 第6期科学技術・イノベーション基本計画. 2021.

<https://www8.cao.go.jp/cstp/kihonkeikaku/index6.html>, (accessed 2021-10-18).

<sup>2</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>, (accessed 2021-10-18).

<sup>3</sup> 池内有為, 林和弘. 研究データ公開と論文のオープンアクセスに関する実態調査 2018. 文部科学省科学技術・学術政策研究所, 2020, NISTEP RESEARCH MATERIAL No.289, 96p. <https://doi.org/10.15108/rm289>, (accessed 2021-10-18).

## (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」)を公開した経験を有する回答者(以下、「データ公開率」)は44.7%、論文をOAにした経験を有する回答者(以下、「論文のOA率」)は80.1%であった(図1)。

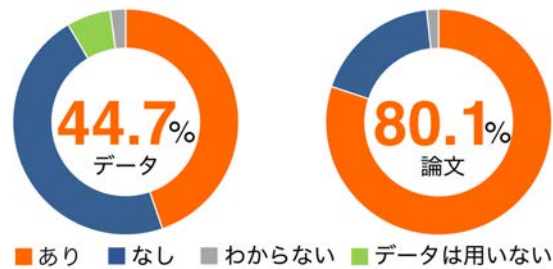


図1 データと論文の公開経験 (n=1,268)

2016年調査におけるデータ公開率は51.0%、2018年調査は51.9%であり、2年前と比較して7.2ポイント減少していた。論文のOA率は2016年調査が70.9%、2018年調査は78.0%であり、2年前と比較して2.1ポイント増加した。

図2に示すように、分野別のデータ公開率は、地球科学(70.2%)から工学(27.7%)まで42.5ポイントの差がみられた。全体の公開率は低下していたものの、地球科学と数学は増加していた。

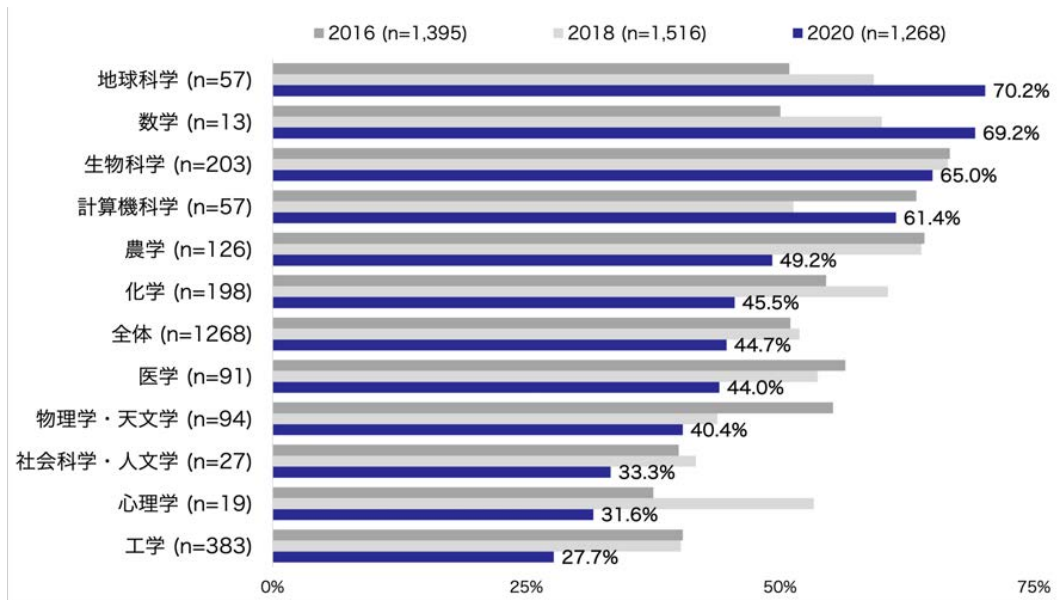


図2 分野別データ公開率 (2016/18/20年)

データの公開方法は、「論文の補足資料」(54.1%)、「個人や研究室のウェブサイトへの掲載」(31.4%)の順に選択率が高かった(図3)。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「特定分野のリポジトリ」(29.1%)が2018年度から10.5ポイント増加していた一方で、「学術機関のリポジトリ」(25.4%)は2018年度から1.7ポイント減少していた。「学術系SNS」<sup>4</sup>(10.1%)、「コード共有サービス」<sup>5</sup>(9.7%)、「データ共有サービス」<sup>6</sup>(5.3%)の選択率はそれほど高くないものの、2016/2018年調査と比較すると増加していた(図6)。

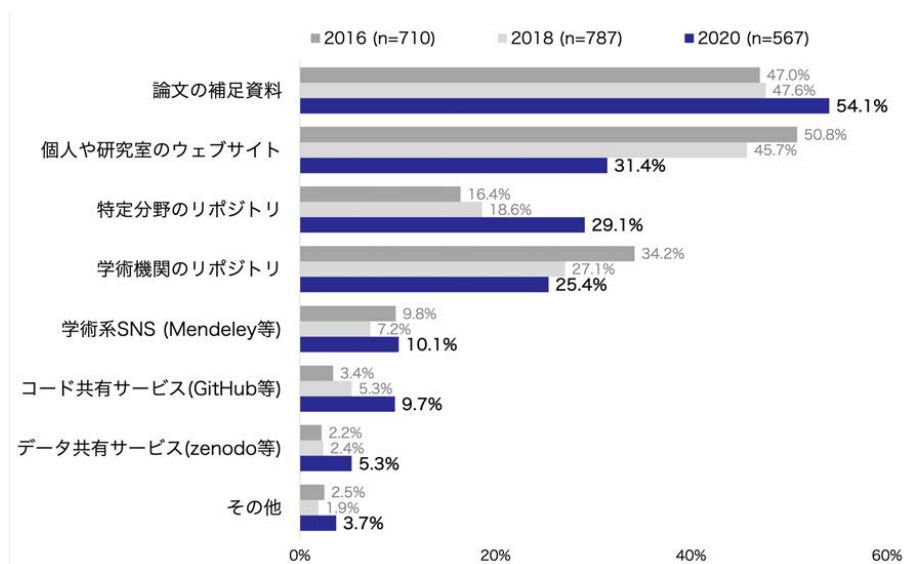


図3 データの公開方法 (2016/18/20年：複数回答)

データの公開理由は1位が「論文を投稿した雑誌のポリシーだから」(54.1%)、2位が「研究成果を広く認知してもらいたいから」(52.6%)であり、2016/2018年調査とは1位と2位の順位が逆転していた。論文の公開理由は1位が「論文を投稿した雑誌がOAだから」(75.8%)、2位が「研究成果を広く認知してもらいたいから」(57.6%)であり、2016/2018年調査と同様であった。データ、論文のいずれも雑誌のポリシーが主たる理由であった。

データの未公開理由は1位が「論文を投稿した雑誌のポリシーではないから」(38.3%)、2位が「業績にならないから」(26.6%)であった。論文の未公開理由は1位が「資金がないから」(57.6%)、2位が「論文を投稿したい雑誌がOAではないから」(40.3%)であり、2016/2018年調査とは1位と2位の順位が逆転していたものの、この2つの理由が突出しているという点は同様であった。

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、「はい」は29.4%、「いいえ」は29.2%、「わからない」は41.4%であった(図4)。全体的にデータ公開に対して慎重な姿勢は継続しているものの、やや「はい」の比率が増え、「わからない」の比率が減少している傾向がみられた。

<sup>4</sup> 質問紙では、学術系SNSの例としてMendeley (Elsevier) とResearchGateを示した。

<sup>5</sup> 同様に、ソースコードを共有できるGitHubを示した。

<sup>6</sup> 同様に、無料でデータを公開できるfigshareとzenodo (CERN)を示した。

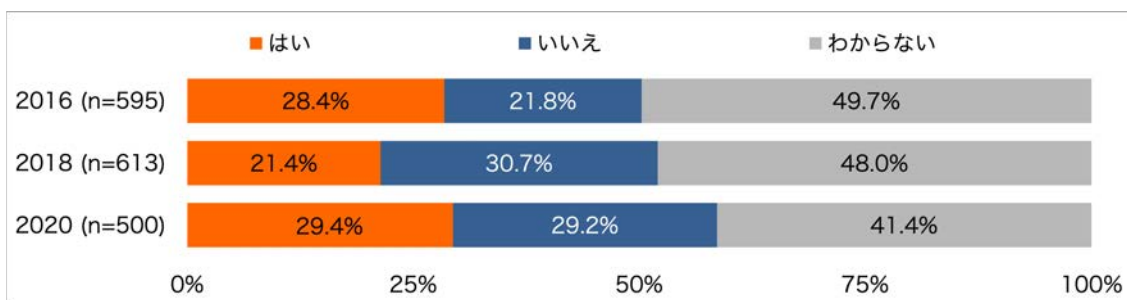


図 4 未公開理由が解決した場合のデータの公開意思 (2016/18/20 年)

## (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 69.7%であった。分野別にみると、計算機科学 (91.2%) から心理学 (47.4%) まで 43.8 ポイントの差がみられた (図 5)。分野別のデータ入手経験はデータ公開経験及びデータ共有 (提供) 経験と正の相関がみられた。すなわち、よくデータを入手している分野は、よくデータを公開している傾向や、よくデータを共有している傾向がみられた。

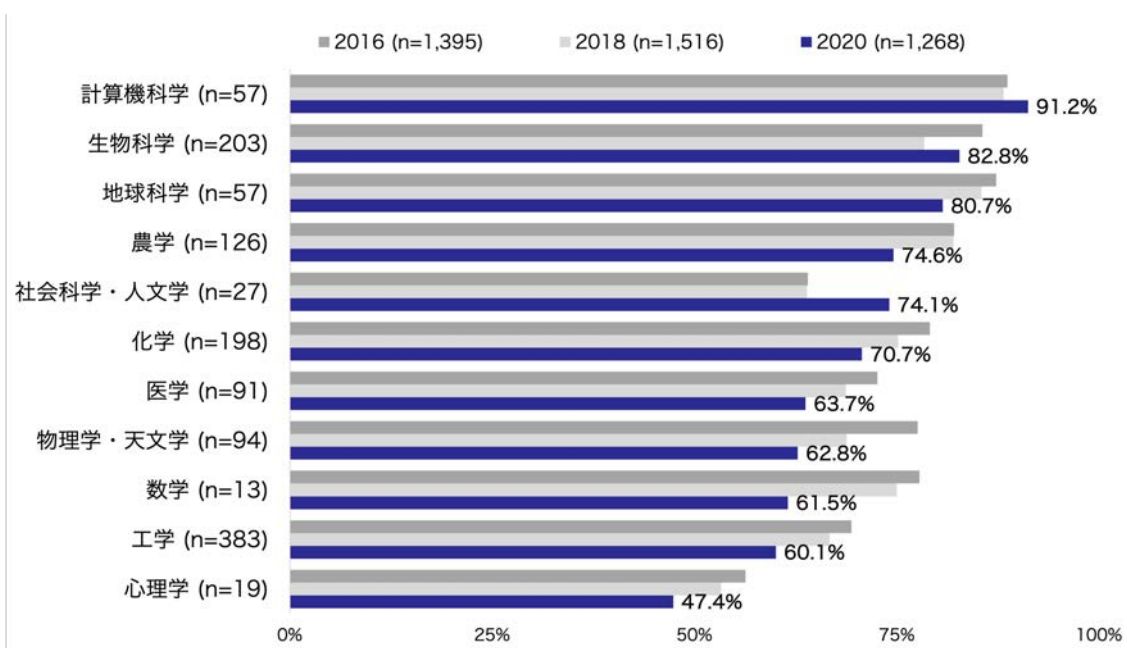


図 5 分野別公開データの入手経験 (2016/18/20 年)

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「論文の補足資料」(59.4%)、次いで「学術機関のリポジトリ」(57.6%)、「個人や研究室のウェブサイト」(38.9%)であった (図 6)。2016/2018 年調査においては、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」、3 位が「学術機関のリポジトリ」であったため変化がみられた。また「特定分野のリポジトリ」、「学術系 SNS (Mendeley 等)」、「コード共有サービス (GitHub 等)」の割合が増加していた。



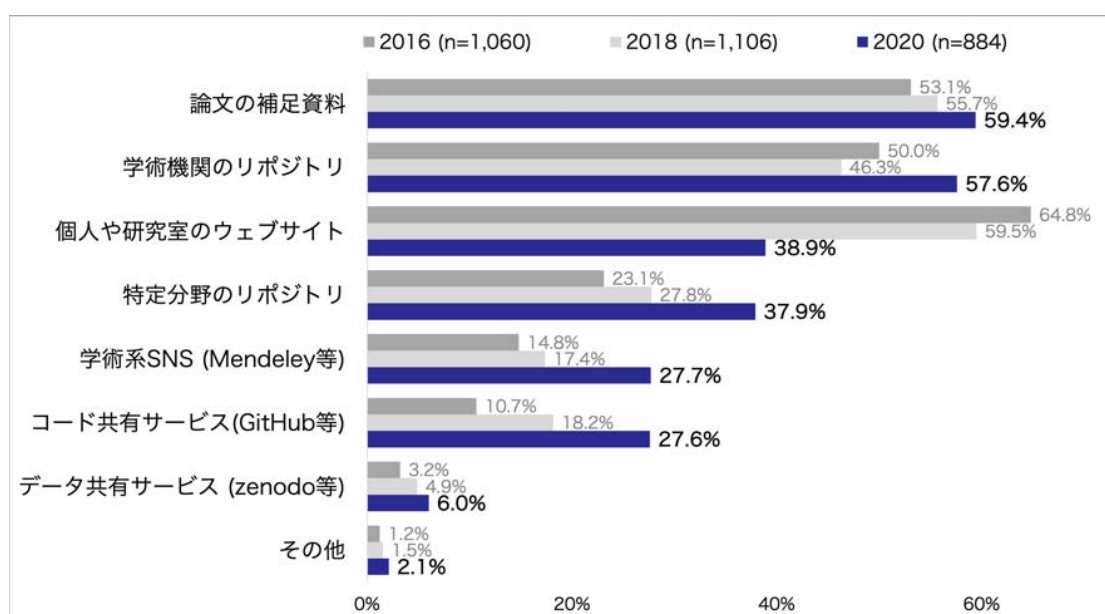


図 6 公開データの入手方法（2016/18/20 年：複数回答）

公開データの入手経験がある回答者 884 名のうち、32.0%が何らかの障壁があると回答していた。具体的な項目を確認すると、1 位は「データごとに品質が異なる」（47.7%）、2 位は「データごとにフォーマットが異なる」（45.9%）、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」（31.8%）であった（表 9）。

表 1 データ入手の障壁の経年変化（2016/18/20 年）

順位	2016 年	2018 年	2020 年
1 位	利用料金	データの品質が異なる	データの品質が異なる
2 位	利用者登録	データのフォーマット	データのフォーマット
3 位	利用条件がわからない	利用条件がわからない	利用条件がわからない
4 位	データの品質が異なる	著作者情報がわからない	著作者情報がわからない
5 位	データのフォーマット	利用料金	データの解釈・再利用
6 位	アクセス方法	利用者登録	アクセス方法
7 位	著作者情報がわからない	データの解釈・再利用	利用料金
8 位	データの解釈・再利用	アクセス方法	利用者登録
9 位	入手までの時間	入手までの時間	入手までの時間
10 位	最新データの入手	最新データの入手	最新データの入手

2016/2018 年調査と比較すると、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、2018 年調査、本調査ではこれらの順位が徐々に低下しており、データの再利用における問題の順位が徐々に高まっていた。

### (3) データマネジメントプラン（DMP）の作成状況

データマネジメントプラン（Data Management Plan, DMP）の作成経験を有する回答者は 20.8%であり、2018 年調査から 2.1 ポイント増加していた。また、「わからない」を選択した回答者は 8.9%であり、3.7 ポイント増加していた。

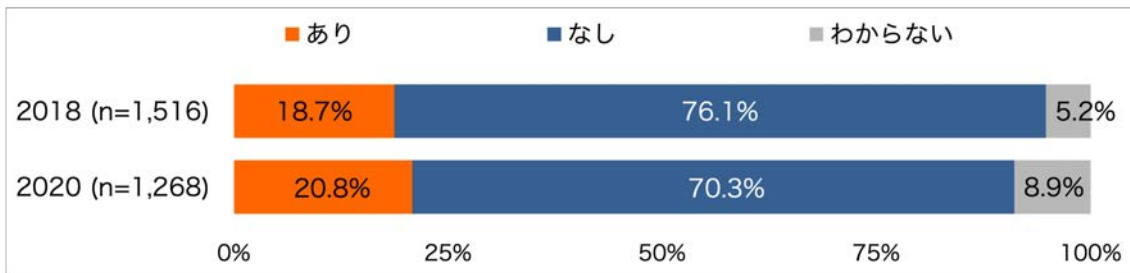


図 7 DMP 作成経験の経年変化（2018/20 年）

例示した DMP のうち選択率が最も高かったのは「科学技術振興機構（JST）」（35.6%）、2 位は「所属機関の DMP」（33.3%）、3 位は「個人や研究グループのための DMP」（30.3%）（図 8）。作成理由のうち、選択率が最も高かったのは「助成機関が要求しているから」であり、JST は 9.5 ポイント、日本医療研究開発機構（AMED）は 6.7 ポイント、新エネルギー・産業技術総合開発機構（NEDO）は 4.2 ポイント増加していた。

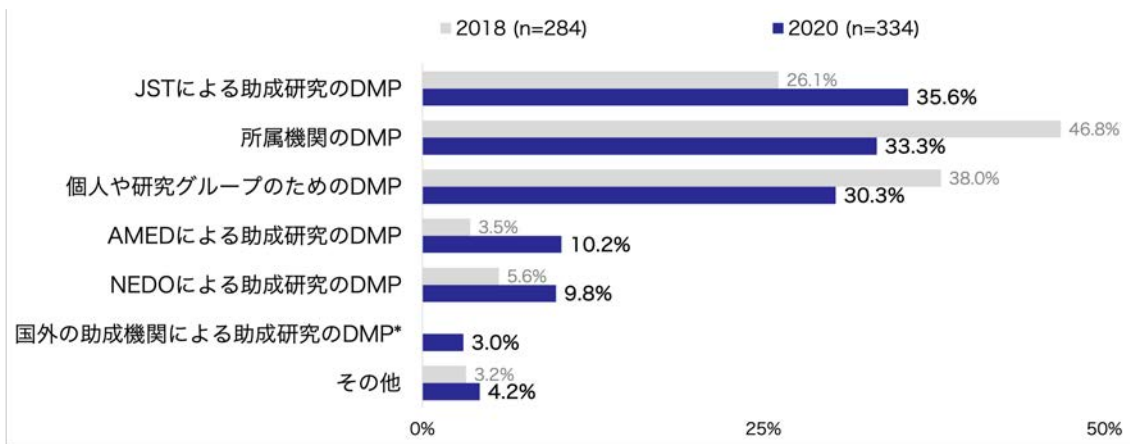


図 8 作成経験がある DMP（2018/20 年：複数回答）

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、2018 年調査と同様であった。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ものの、前回調査と比べると 3.7 ポイント減少していた。2 位は「所属機関から要求されていないから」（39.3%）、3 位は「助成機関から要求されていないから」（33.1%）であった。

#### (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、人材、時間、資金については「不十分」とする回答者が過半数であった（図 9）。「わからない」とする回答者の比率も高く、人材は 12.0%、時間は 13.1%、資金は 15.0%であった。保存用ストレージ、公開用のリポジトリ、研究中のストレージについては、「不十分」と「やや不十分」の合計選択率が 5 割を下回っており、2016/2018 年調査と比較して、やや不足感が低減している傾向がみられた。公開用のリポジトリについては「わからない」とした回答者が 27.0%であり、認知度が低いことがうかがえた。

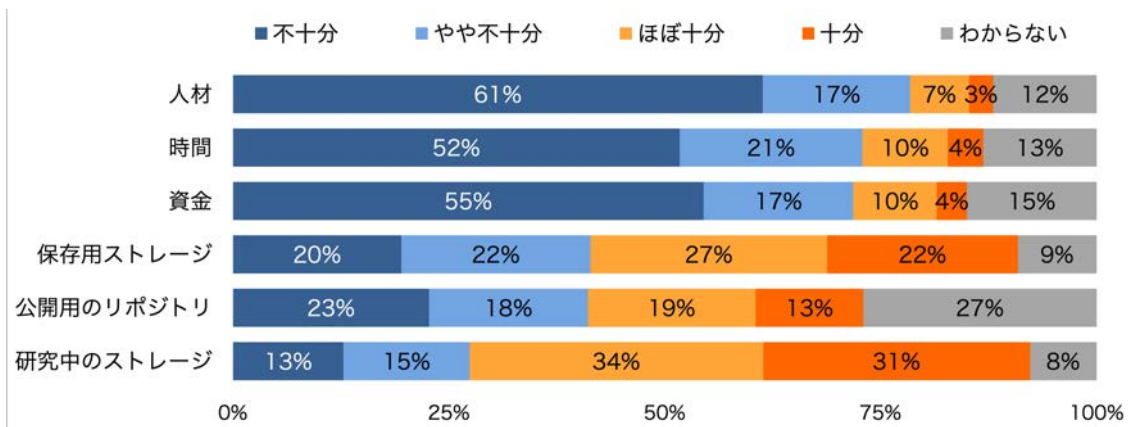


図 9 データの整備や公開に関する資源の充足度 (n=1,188)

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 89.8%）であった。次いで「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 80.9%）、「二次利用に関して責任が生じる可能性」（同 78.2%）、「不正利用・改ざんの可能性」（同 78.0%）、「データの利用権限や契約」（同 76.6%）の順に懸念が強かった。「研究の誤りを発見する可能性」（同 19.4%）は懸念をもつ回答者が比較的少なく、「問題ではない」（35.3%）と「あまり問題ではない」（37.5%）を選択する回答者の比率が高かった（図 10）。

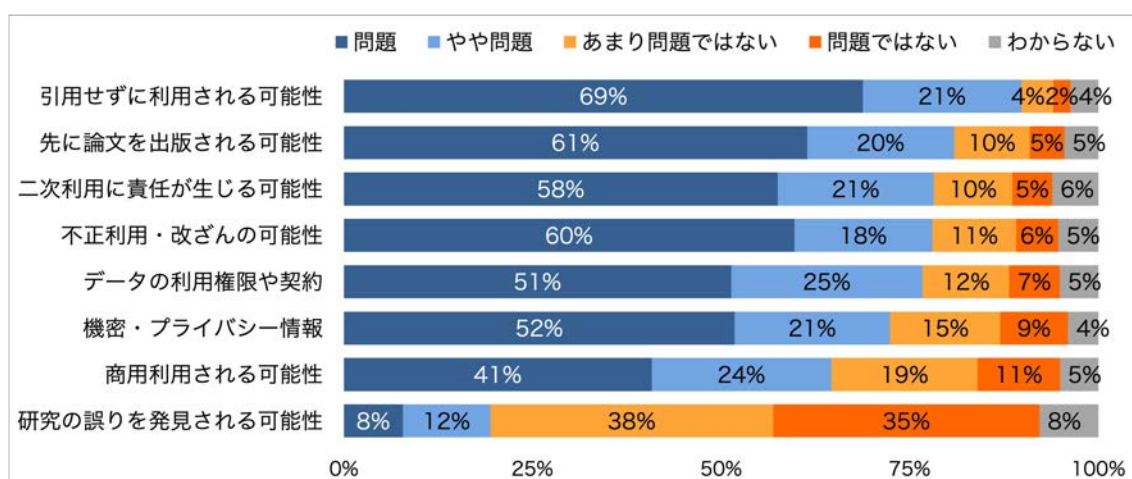


図 10 カレントデータの公開に関する懸念 (n=1,188)

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、2018 年調査からデータ公開経験を有する回答者に自由記述で尋ねることとした。その結果、データ公開経験を有する回答者 567 名のうち 33 名 (5.8%) が具体的な問題について記述していた。表 2 に示すように、最も多かったのは 2018 年調査と同様、公開したデータに対する「問い合わせ等への対応」であった。次いで「誤用された」、「引用せずに利用された」、「データの権利に関する問題」の順であった。

表 2 データ公開によるデメリット (2018/20 年)

2018 年調査	人数	2020 年調査	人数
問い合わせ等への対応	17	問い合わせ等への対応	6
引用せずに利用された	14	誤用された	5
先取権の喪失	9	引用せずに利用された	4
誤用された	7	データの権利に関する問題	4
更新のコストがかかる	2	懸念が生じた	3
徒労感	2	ミスを発見した	2
不正アクセス	1	公開に手間がかかる	2
商用利用された	1	先取権の喪失	2
研究者以外に利用された	1	長期公開・保存のコストがかかる	2
		不正アクセス	2
		情報流出	1
合計	54	合計	33

## (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、2018 年調査からデータ公開経験を

有する回答者に自由記述で尋ねることとした。その結果、データ公開経験を有する回答者 567 名のうち 130 名 (22.9%) が具体的な事柄について記述していた。表 12 に 7 項目に分類した集計結果を 2018 年調査の結果とともに示す。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 152 件となった。最も多かったのは 2018 年調査と同様に「研究上の利点」(35.5%) であり、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」(17.8%) であった。

表 3 データ公開によって得られた良い結果 (2018/20 年)

内容	2018 年		2020 年	
	件数	比率	件数	比率
研究上の利点	104	40.6%	54	35.5%
研究・データ・研究者のビジビリティ向上	66	25.8%	41	27.0%
科学・分野の進展	27	10.5%	27	17.8%
人とのつながり	26	10.2%	14	9.2%
評価	11	4.3%	7	4.6%
個人的な利点	9	3.5%	6	3.9%
その他	13	5.1%	3	2.0%
合計	256	100.0%	152	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 94.5%)、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 93.0%) 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た。これは 2018 年調査でも同様の結果であった。

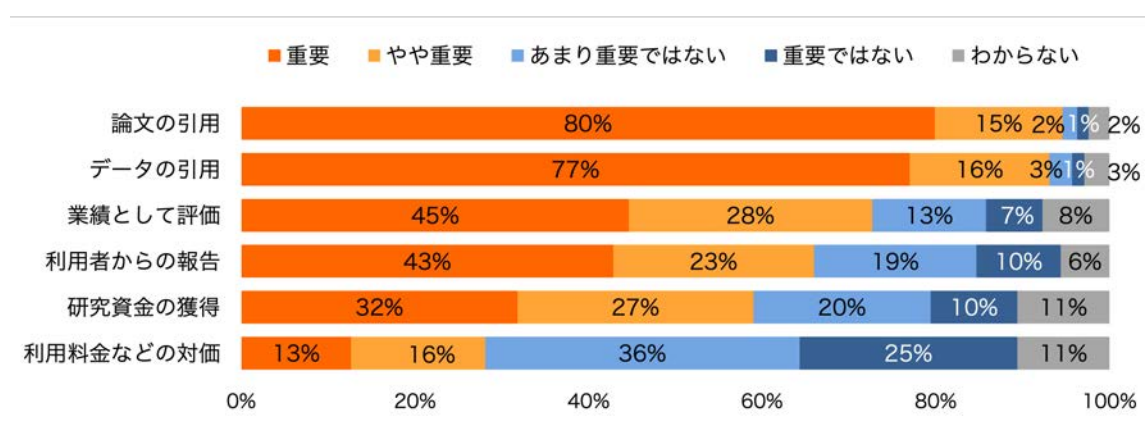


図 11 データ公開によるインセンティブの重要度 (n=1,180)

データ公開に対する評価の状況を明らかにするために、回答者自身が所属するコミュニティや機関、及び回答者自身が評価しているかどうかを尋ねた。両者を比較するために、回答から「その他」と「指導する研究者がいない」を除いて再集計した結果を示す(図 12)。

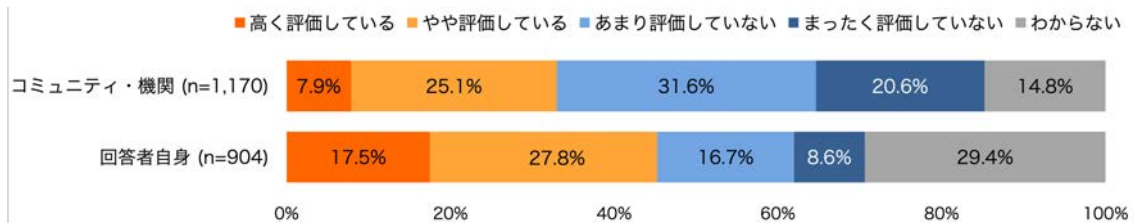


図 12 データ公開に対するコミュニティ・機関と回答者自身の評価

コミュニティや機関よりも回答者自身の方が、評価している回答者の比率が高く、評価していない回答者の比率が低かった。「わからない」とする回答者の比率も高く、回答者自身については29.4%であった。データを公開することが評価につながるようになるのかどうか、今後も継続的に調査していきたい。

## (6) 研究データ管理 (RDM) に対する認識

研究データ管理 (Research Data Management, RDM) に関するリテラシーへの関心を把握するために、データの整備や公開に関する項目を挙げて複数選択方式で尋ねたところ、回答者の87.5%が何らかの項目を選択していた。2016/2018年調査から引き続き、「適切なデータ形式」、「知的財産権やライセンス」、「データの安全な管理方法」の選択率が高かった(図 13)。

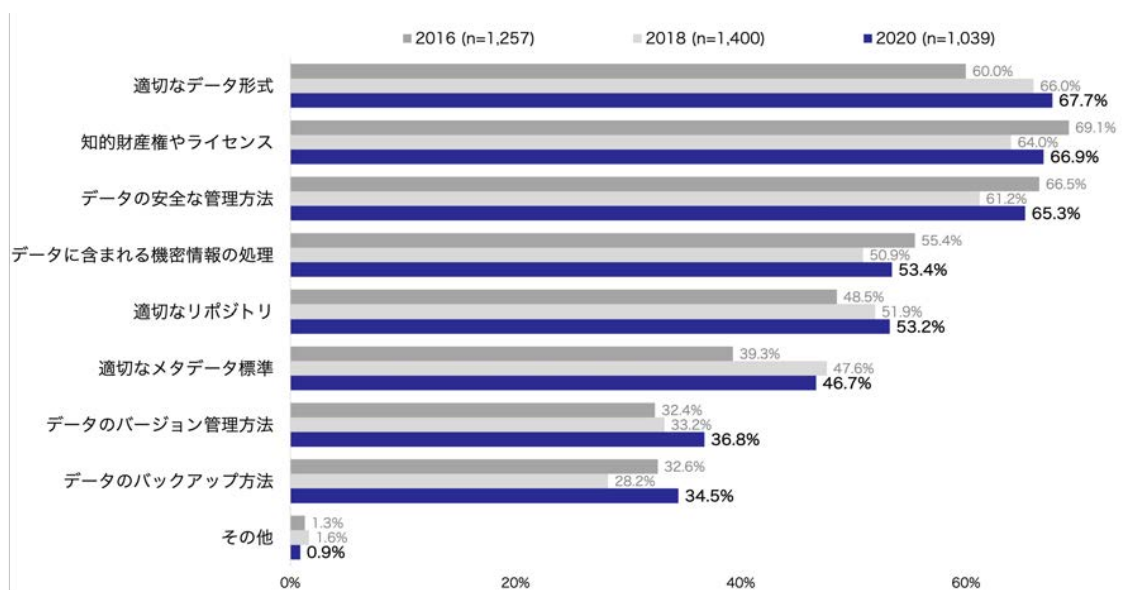


図 13 研究データ管理 (RDM) に関して知りたい項目 (2016/18/20年：複数回答)

RDM プロセスを、回答者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に依頼したいと考えるかどうかを尋ねたところ、回答者の 41.1%が依頼したいと考えていた。具体的に依頼したいプロセスとしては、「適切なデータ形式への変換」、「適切なリポジトリの選択」、「適切なライセンスの選択」の選択率が高かった（図 14）。

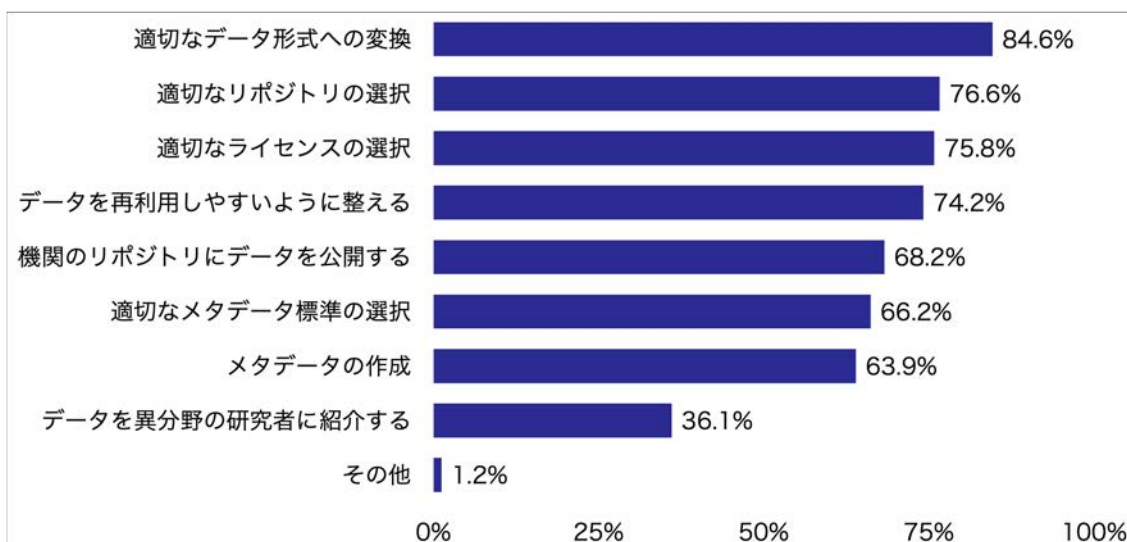


図 14 研究データ管理（RDM）に関して依頼したい項目（n=488：複数回答）

本調査によって、日本の研究者によるオープンサイエンスの実践状況や認識、課題とその経年的な変化を明らかにした。今後も継続的な調査を実施することによって、日本のオープンサイエンスがどのように変化していくのか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

科学研究の成果である論文やデータを公開し、国や分野を超えた活用を促進するオープンサイエンス政策が加速している。オープンサイエンス政策は、ICTの進展によるデジタル化とネットワーク化の特性を活かし、主に公的資金を利用した研究成果のさらなる活用・再利用によって、イノベーションの創出と科学や社会の変容を加速する研究基盤（インフラ）づくりを目指している。そしてCOVID-19によって、図らずもその重要性が幅広く認知され、ないしは再認識され、科学と社会、および科学と社会の変容を促している。そして、研究成果の迅速かつできるだけオープンな共有に向けた取り組みが加速している。

日本でも、2021年3月に閣議決定された『第6期科学技術・イノベーション基本計画』<sup>1</sup>では、「第2章 Society 5.0の実現に向けた科学技術・イノベーション政策」の「2. 知のフロンティアを開拓し価値創造の源泉となる研究力の強化」にオープンサイエンスに関連する方向性や具体的な取組が示されており、研究データの利活用に着目している。たとえば、「(1)多様で卓越した研究を生み出す環境の再構築」では、“目標”として“人文・社会科学の厚みのある研究が進み、多様な知が創出されるとともに、国内外や地域の抱える複雑化する諸問題の解決に向けて、自然科学の知と融合した「総合知」を創出・活用することが定着する”とあり（p. 51）、“具体的な取組”として“人文・社会科学の研究データの共有・利活用を促進するデータプラットフォームについて、2022年度までに我が国における人文・社会科学分野の研究データを一元的に検索できるシステム等の基盤を整備することや、“人文・社会科学の研究データの共有・利活用を促進するデータプラットフォームの基盤整備・強化や、研究データの管理・利活用機能などにおける図書館のデジタル転換等を通じた支援機能の強化”が盛り込まれている（p. 56）。また、「(2) 新たな研究システムの構築（オープンサイエンスとデータ駆動型研究等の推進）」では、“データの共有・利活用については、研究の現場において、高品質な研究データが取得され、これら研究データの横断的検索を可能にするプラットフォームの下で、自由な研究と多様性を尊重しつつ、オープン・アンド・クローズ戦略に基づいた研究データの管理・利活用を進める環境を整備する。特にデータの信頼性が確保される仕組みが不可欠となる”と述べられ（p. 59）、“【科学技術・イノベーション政策において目指す主要な数値目標】（主要指標）”として、“機関リポジトリを有する全ての大学・大学共同利用機関法人・国立研究開発法人において、2025年までに、データポリシーの策定率が100%になる／公募型の研究資金の新規公募分において、2023年度までに、データマネジメントプラン（DMP）、および、これと連動したメタデータの付与を行う仕組みの導入率が100%になる”とされている（p. 59-60）。また、“具体的な取組”として、“公的資金により得られた研究データの機関における管理・利活用を図るため、大学、大学共同利用機関法人、国立研究開発法人等の研究開発を行う機関は、データポリシーの策定を行うとともに、機関リポジトリへの研究データの収載を進める。あわせて、研究データ基盤システム上で検索可能とするため、研究データへのメタデータ

の付与を進める”ことや、“研究者の研究データ管理・利活用を促進するため、例えば、データ・キュレーター、図書館職員、URA、研究の第一線から退いたシニア人材、企業等において研究関連業務に携わってきた人材、自らの研究活動に資する場合にはポストク等の企画や、図書館のデジタル転換等の取組について、2022年度までにその方向性を定める”ことが挙げられている（p.60）。

こうした方針に従って研究成果の共有と活用を実現するためのインフラの整備、データポリシーやDMPの策定、支援人材の育成などを進めていく上では、まず現状を把握し、適切かつ効率的な計画立案が重要であると考えられる。科学技術・学術政策研究所（NISTEP）は、2020年11月に大学や公的機関、企業に所属する日本の研究者を対象として、オープンサイエンスに関するウェブ質問紙調査を実施した。これは2016年と2018年に実施した調査<sup>2 3 4 5 6</sup>（以下、それぞれ「2016年調査」、「2018年調査」、両者をまとめて「2016/2018年調査」と記す）の後続調査である。基本的には同一の質問を行っているが、新たな動向の影響を明らかにするために、一部の質問は入れ替えや修正を行っている。たとえば、2018年調査からはデータマネジメントプラン（DMP）の作成状況に関する質問を追加した。

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、2016/2018年調査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、新たにデータ公開に対する評価についても調査することとした。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た。これらの結果が、根拠に基づく政策立案（EBPM）、あるいは学術機関や企業のマネジメント層、研究支援を行う技術職員やURA、図書館職員などによるオープンサイエンスに関する議論に基礎資料を提供することを目指した。

## 1.2 主要な概念の定義

本報告書で使用した概念や用語の定義を以下に示す。2016/2018年調査における定義と基本的には同一であるが、表現を一部修正した。

### 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

論文などの成果を公表済みの、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。特定の人以上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 (Open Data)

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。本調査においては、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 (Open Access, 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

オープンサイエンスには、さまざまな要素が含まれる<sup>7</sup>が、本調査においては、データ公開 (Open Data) とオープンアクセス (Open Access) 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 質問紙調査システムの変更について

2016/2018年調査には、Questant社のウェブ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。2020年調査では、ユミルリンク社のウェブアンケートシステム（Cuenote）を用いることとなった。

Questant社のシステムでは、複数選択式の質問において排他的選択肢（他の項目とは同時に選べない選択肢）を設定することが可能であったが、調査時点においてCuenoteではこうした設定は行えなかった。そこで次節に示すように、質問項目の修正や削除といった変更を行った。質問項目の変更が回答に影響する可能性は否定できない。2016/2018年調査と本調査の結果を比較する際は、この点に留意いただきたい。

### 1.4 調査項目

調査項目は、基本的には2016/2018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、各調査の報告書（RM268<sup>2</sup>、RM289<sup>5</sup>）をご確認いただきたい。

表4に2020年調査の目的と、それらを明らかにするための調査項目を示す。本調査では、研究データ公開に対する評価に関する項目を新たに設けた（表4の\*印）。なお、前節で述べた通り、本調査に用いたシステムでは排他的選択肢を設けられなかったため、質問方法を変更せざるを得なかった。また、結果的に質問数が増えることとなったため、質問項目の削除を行った。

表 4 2020 年調査の目的と調査項目の概要

調査目的	調査項目
1. 研究分野と研究活動状況	研究分野 調査時点での研究活動状況
2. データと論文の公開状況	データと論文の公開状況と公開方法 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
3. データの提供（共有）状況	データの提供（共有）経験
4. 公開データの利用状況と課題	公開データの利用状況 公開データの入手における障壁
5. 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
6. 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
7. 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 データ公開に対する評価の状況*
8. データの整備・公開・保存に対する認識	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・公開の依頼意思

\*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 5 の順番に配列した。まず、事実として「1. 研究分野」から「5. データマネジメントプラン (DMP) の作成状況」について尋ねた上で、研究データを公開しようと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 5 質問項目一覧

質問	必須	方式	変更等
1. 研究分野について			
Q1 研究分野	○		修正
Q2 現在の研究状況	○		修正
2. 論文の公開について			
Q3 論文のオープンアクセス (OA) 経験の有無	○		修正
Q4 論文の OA 方法	○	m	修正
Q5 論文を OA にした理由	○	m	
Q6 論文を OA にしていない理由	○	m	
Q7 論文の OA 意思	○		
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	
4. 公開データの利用について			
Q9 公開データの入手経験の有無	○		修正
Q10 公開データの入手先	○	m	修正
Q11 公開データ入手における問題の有無	○		修正
Q12 公開データ入手における問題	○	m	修正
5. データマネジメントプラン (DMP) について			
Q13 DMP 作成経験の有無	○		
Q14 作成した DMP の種類	○	m	
Q15 DMP を作成した理由	○	m	
Q16 DMP を作成していない理由	○	m	
6. 研究データの公開について			
Q17 研究データの公開経験の有無	○		修正
Q18 研究データの公開方法	○	m	修正
Q19 研究データを公開した理由	○	m	
Q20 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
Q21 研究データの公開によって問題が起きた経験			記述
Q22 研究データを公開していない理由	○	m	
Q23 研究データの公開意思	○		
Q24 研究データ公開のインセンティブの重要度		4	
Q25 コミュニティや機関における研究データ公開の評価		4	新規
Q26 自身が研究データ公開を評価しているかどうか		4	新規
7. 研究データの整備・公開・保存について			
Q27 カレントデータの総量		s	
Q28 カレントデータの整備・公開・保存に必要な資源の充足度	○	4	
Q29 カレントデータを公開しようとする場合の懸念	○	4	修正
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・公開の依頼意思の有無			修正
Q32 カレントデータの整備・公開を依頼したい項目		m	修正
8. 自由回答			
Q33 オープンサイエンスや調査に対する自由回答		m	

※方式：単一選択法=無印、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 5 に示した質問項目のうち、(1)新規に設定した項目、(2)アンケートシステムの変更に伴い修正を加えた項目、及び(3)削除した項目について述べる。なお、2016/2018 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

研究データ公開に対する評価について、回答者の認識を尋ねる質問を 2 問追加した。Q25 では回答者が所属するコミュニティや機関での評価の状況について、Q.26 では回答者が指導する研究者が研究データを公開した場合に評価するかどうかを、それぞれ 4 件法で尋ねた。

Q25. 研究データを公開することは、ご自身が所属するコミュニティや機関で評価されていると思われますか？

Q26. ご自身が指導する研究者（大学院生やポスドクなど）が研究データを公開していることを、評価していらっしゃいますか？

### (2) アンケートシステムの変更に伴う修正：経験等を尋ねる質問

研究データの入手経験や公開経験を尋ねる質問について、2016/2018 年調査では、図 15 に示すように(1)から(8)の公開方法を列挙した上で、排他的選択肢として「(9)公開したことはない」、「(10)わからない」、「(11)研究にデジタルデータは用いない」を示した。そして(1)から(8)を 1 つ以上選択した回答者は「公開経験あり」と判断した。理由は、先行研究やプレテストの結果から、データ公開の経験がない回答者や馴染みが薄い回答者が存在すると予想されたこと、データ公開の経験があっても回答者によって「データ公開」とみなす内容が異なると予想されたこと、データ公開の定義や具体例を示していても読み飛ばされる可能性があることから、公開経験の有無と方法を同一の質問で尋ねることによって、回答者の認識の違いによる齟齬を防ぎ、より正確にデータ公開経験を把握するためである。

## 5. 研究データの公開について

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

(再掲) 本調査における「データ」とは、研究のために収集・作成・観測したデジタルデータを指します。研究の成果である論文やスライドの根拠となるもので、テキスト、画像、音声、動画など、形式は限定しません。また、ゲノムデータ、地理情報、ソフトウェアコード、インタビューの録音と書き起こしなど、内容も限定しません。

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

\* (複数選択)

- (1) 個人や研究室のウェブサイト
- (2) 所属機関のリポジトリ・データアーカイブ
- (3) 特定分野のリポジトリ・データアーカイブ (DDBJやICPSRなど)
- (4) 無料のデータ共有サービス (figshare, zenodoなど)
- (5) コード共有サービス (GitHubなど)
- (6) 論文の補足資料 (supplementary materials)
- (7) 学術系SNS (Mendeley, ResearchGateなど)
- (8) その他 [具体的に]
- (9) 公開したことはない 【→Q20へ】
- (10) わからない 【→Q20へ】
- (11) 研究にデジタルデータは用いない 【→Q35へ】

図 15 2016年調査：アンケート入力画面

しかし、2020年調査では2016/2018年調査と同一の尋ね方ができなかったため、質問を2つにわけ、まずQ17でデータ公開経験の有無を尋ねた上で、Q18で公開方法を選択することとした(図16)。また、Q17の選択肢(2)と(3)は質問の意図が伝わるように2016/2018年調査よりも詳しく記述した。自由記述を伴う「その他」は最後にしか設定できなかったため、位置を入れ替えた。

なお、図16に示すように、アンケートの入力画面では研究データの公開経験を尋ねるQ17の直前に「研究データ」の定義を、直後にはデータ公開の方法を選択するQ18を表示して、同時に見られるようにした。

## 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データ、および研究成果そのものであるデータの両方を含みます。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、研究データを公開したご経験はありますか？ **必須**

- (1) はい
- (2) 研究にデジタルデータを用いるが、公開したことはない【→Q22. へ】
- (3) 研究にデジタルデータを用いないので、公開したことはない【→Q33. へ】
- (4) わからない【→Q24. へ】

Q18. 研究データを公開した際に利用したサーバやサービスなどを全てお選び下さい。

**【複数選択可】 必須**

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

図 16 2020 年調査：アンケート入力画面

このように排他的選択肢が設定できないために修正した質問は、表 5 の「変更等」の列に「修正」と記載した 12 項目であり、2016/2018 年調査の 6 項目を、それぞれ 2 問に分割した。具体的には、(1)研究分野と現在の研究状況、(2)論文の OA 経験の有無と OA にした方法、(3)公開データの入手経験の有無と入手先、(4)公開データの入手における問題の有無と具体的な問題点、(5)研究データの公開経験の有無と方法、(6)カレントデータの整備・公開の依頼意思の有無と項目である。なお、2016/2018 年調査では、(2)論文の OA 経験は有無のみを尋ねていたが、本調査では質問形式に慣れてもらうことを意図して「OA にした方法」も追加することとした。

質問方法の変更に伴う回答への影響を明らかにするために、2016/2018 年調査と同一の回答者群については、回答に矛盾がないかどうかを確認することとした。たとえば、データ公開経験を 2016/2018 年調査では「あり」としていた回答者が、本調査では「なし」や

「わからない」としていないかどうかを確認した。

### (3) アンケートシステムの変更に伴う修正：理由を尋ねる質問

2016/2018年調査では、論文や研究データを公開した、または未公開の理由を尋ねる質問で「あてはまるものはない」を排他的選択肢の1つとして挙げていた。2020年調査では、排他的選択肢の設定ができなかったため、他の選択肢と同時に選んでしまわないように「特に理由はない」という表現に修正した。DMPの作成または未作成の理由の選択肢についても同様に、「特に理由はない」という表現に改めた。

### (4) その他の修正

- ① 研究データを公開する際の問題について尋ねる項目で、2016/2018年調査では「データの所有権」という表現を用いていたが、経済産業省のガイドライン<sup>8</sup>を参考に「データの利用権限」という表現に改めた。
- ② 2018年調査の自由回答を参考に、Q5. 論文をオープンアクセスにした理由の選択肢に「雑誌や学会等がオープンアクセスにしたから」を追加した。
- ③ 2018年調査の自由回答を参考に、Q6. 論文をオープンアクセスにしていない理由の選択肢に「オープンアクセスではない発表方式の方が望ましいと思うから」を追加した。
- ④ プレテストにて、公開データの入手先や研究データの公開先の選択肢のうち、「機関リポジトリ」と「個人や研究室のウェブサイト」が紛らわしいという指摘を受けたため、「個人や研究室のウェブサイト」を選択肢の最後に移動した。
- ⑤ 論文のOAに関する質問や選択肢にプレプリントを含めていたが、プレプリントについては別途調査を実施した<sup>9</sup>ため削除した。
- ⑥ 研究データの整備や公開を図書館員やデータキュレーターに依頼できるかどうかを明らかにするために、2016/2018年調査では、各プロセスの専門性の高さを尋ねていたが、より直截的に依頼したいと思うかどうかを尋ねた。

### (5) 削除した項目

全体の質問数を抑制するために、2018年調査から8件の質問と1項目を削除した。以下のうち、a)～d)については経年変化がわずかであると推測した。e)は、それ以降の質問で「カレントデータ」について回答することを明確にするための導入として設定していたが、質問数を抑制するために削除した。f)は、カレントデータを公開しようとする場合の懸念の強さを9項目について尋ねていたが、項目数を抑制するために内容が近い「誤解や誤用の可能性」と「不正利用・改ざんの可能性」のうち、前者を削除した<sup>vii</sup>。

---

<sup>vii</sup> 2016年調査は国内外の先行研究を参照して「誤解や誤用の可能性」のみを挙げていた。2016年調査の自由回答では“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、2018年調査では「不正利用・改ざんの可能性」を追加した。その結果、「誤解や誤用の可能性」を「問題である」とした回答者は37%、「不正利

a) 論文の OA に関する質問（経年変化はわずかであると推測した）

- 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。
- 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。

b) 研究データを提供した方法（経年変化はわずかであると推測した）

- 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ください。

c) 公開データの利用について

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

d) カレントデータを公開した場合、他の研究者が理解できるかどうか

- 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せんか？
- 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せんか？

e) カレントデータの公開について

- 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？

f) カレントデータを公開しようとする場合の懸念

- 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？  
あてはまるものをそれぞれお選び下さい。  
「誤解や誤用の可能性」

---

用・改ざんの可能性は」40%であり、「不正利用・改ざんの可能性」の方がやや懸念が強かったことから、こちらを残すこととした<sup>3</sup>。

## 1.5 調査対象

調査対象は 2016/2018 年調査と同様に、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」<sup>10</sup>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、Q2 で「現在、論文出版や学会等での口頭発表を行っていますか」と尋ねて、行っていない回答者には自由回答のみ記入していただいた。

## 1.6 調査期間

プレテストは 2020 年 10 月 17 日から 10 月 23 日に実施して 9 名から回答を得た。回答に基づき、質問の順序やワーディング等の修正を行った (1.4 参照)。

本調査の期間は、2020 年 11 月 5 日から 11 月 15 日とした。調査への協力依頼は、11 月 5 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 11 月 12 日に送信した。

11 月 15 日以降も回答入力があったため、最終的に 11 月 17 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 28 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。本調査に用いたアンケートシステムでは排他的選択肢の設定ができなかったため、矛盾した回答もみられた。たとえば、Q22 では「特に理由はない」が別の選択肢とともに選ばれていた。こうした場合は、「特に理由はない」は選択していないものとして削除した。

## 1.7 回答率と回答者の属性

調査依頼の送付数は 1,914 名、最終回答数は 1,349 名 (回答率 70.5%) であった。このうち、1.5 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 1,268 名 (66.2%) 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。

### (1) 回答者の所属

回答者の所属は、大学が 935 名（全体の 69.3%）で最も多く、次いで公的機関・団体 251 名（18.5%）、企業 155 名（11.5%）、その他 10 名（0.7%）の順であった（図 17）。「その他」を選択した回答者は所属機関別の集計からは除き、検定も行わないこととした。



図 17 回答者の所属 (n=1,349)

### (2) 回答者の年齢層

回答者の年齢層は、40代が 659 名（全体の 48.9%）で最も多く、次いで 30代が 425 名（31.5%）、50代が 194 名（14.4%）の順であった（表 6）。非開示の回答者（1 名）は所属機関別の集計からは除き、検定も行わないこととした。

表 6 回答者の年齢層

年代	回答者数	比率
30代以下	425	31.5%
40代	659	48.9%
50代	194	14.4%
60代以上	70	5.2%
非開示	1	0.1%
合計	1,349	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 18 に示す。工学が 419 名（31.1%）で最も多く、次いで化学が 217 名（16.1%）、生物科学が 213 名（15.8%）であった。回答者数が 10 名以下であった天文学（9 名）は物理学と、人文学（3 名）は社会科学とあわせて「物理学・天文学」（合計 99 名、7.3%）、「人文学・社会科学」（合計 27 名、2.0%）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。

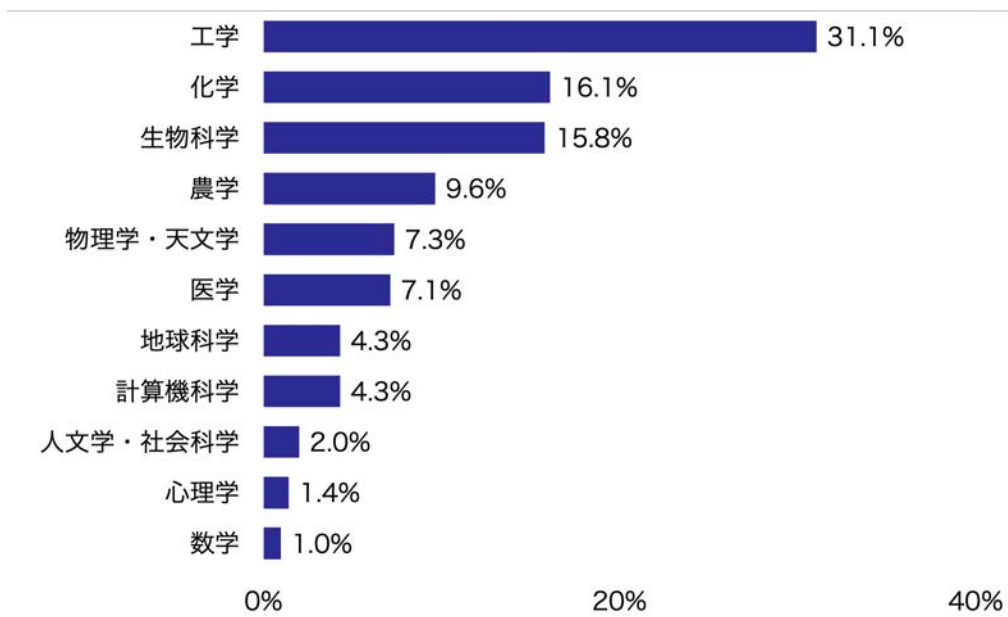


図 18 回答者の専門分野 (n=1,349)



## 2. 調査結果

調査結果を 2016/2018 年調査の結果と比較しながら、調査目的（表 4）のグループごとに示す。すなわち、「データと論文の公開状況」、「データの提供（共有）状況」、「公開データの利用状況と課題」、「データマネジメントプラン（DMP）の作成状況」、「データ公開の障壁」、「データ公開のインセンティブ」、「データの整備・公開に対する認識」の順に示す。

なお、Q1 から Q32 までの単純集計表は資料として報告書の末尾に掲載し、Q20、Q21、Q33 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を確認するために、「Q17. これまでに、研究データを公開したご経験はありますか？」と尋ねた（論文は Q3）。Q17 では選択肢として「はい」、「研究にデジタルデータは用いない」、「研究にデジタルデータを用いるが公開したことはない」、「わからない」を示した。また、続く Q18 には選択肢としてデータ公開の方法を列挙した（論文は Q4）。図 19 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。

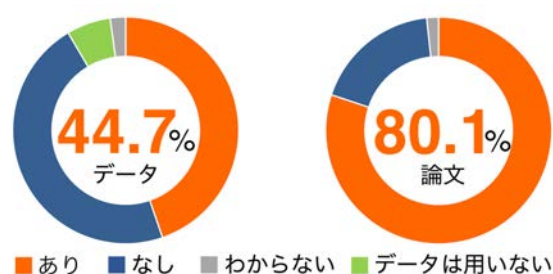


図 19 データと論文の公開経験（n=1,268）

データの公開経験を有する回答者（以下、「データ公開率」）は 44.7%、論文を OA にした経験を有する回答者（以下、「論文の OA 率」）は 80.1%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、2016/2018 年調査と本調査の結果を比較した。図 20 にそれぞれの結果を示す。

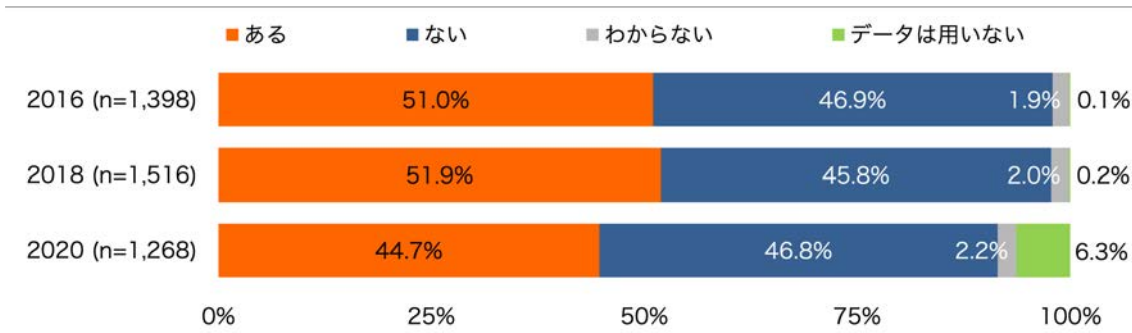


図 20 データ公開経験の経年変化（2016/18/20 年）

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 ポイント増加していたものの、本調査では 7.2 ポイント減少して 44.7%にとどまった。公開経験がない回答者は 2016 年調査が 46.9%、2018 年調査が 45.8%、2020 年調査が 46.8%であり、あまり変化はなかった。「わからない」を選択した回答者は 2.2%であり、2016 年調査の 1.9%、2018 年調査の 2.0%とほとんど変わらなかった。

一方、「研究にデジタルデータは用いない」とした回答者は 6.2%（80 名）であり、2016 年調査の 0.1%、2018 年調査の 0.2%と比較して 6 ポイントほど高かった。この 80 名を除いた 1,188 名のデータ公開率は 47.7%、公開経験がない回答者の比率は 49.9%、「わからない」が 2.4%となる。

2020 年調査でデータ公開率が減少した要因の一つとして、アンケートシステムの変更に伴い質問形式を変更したことの影響が考えられる（1.4(2)参照）。そこで 3 回の調査全てに共通する回答者 267 名の回答を調べて矛盾がないかどうかを確認した。その結果、72 名（27.0%）の回答に矛盾がみられた。具体的には、2016 年及び／または 2018 年調査ではデータ公開経験ありとしていた回答者のうち、「研究にデジタルデータは用いない」を 11 名が、「研究にデジタルデータを用いるが公開したことはない」を 61 名が、それぞれ選択していた。これらの回答をデータ公開経験ありとみなすと、回答者 1,268 名のデータ公開率は 50.4%、公開経験がない回答者は 42.0%、デジタルデータを用いない回答者は 5.4%、わからないは 2.2%となる。

同様に、以降の質問についても確認したところ、OA 経験は 24 名（8.9%）、データ提供経験は 34 名（12.6%）、データ入手経験は 72 名（26.8%）、DMP 作成経験は 39 名（14.5%）の回答に矛盾がみられた。そこで、アンケートシステムの変更に伴う質問形式の変更の影響を明らかにするために、2016 年調査と 2018 年調査の回答の矛盾を確認したところ、OA 経験は 12 名（4.5%）、データ提供経験は 44 名（16.5%）、データ入手経験は 48 名（18.0%）、データ公開経験は 38 名（14.2%）にみられた。つまり、同一のアンケートシステムで同一の質問を行ったとしても、データ公開や OA などに対する認識が変わったことによる回答の変更（例：自分では公開していると考えていたが、そうではなかった）や、入力ミスなどが一定数含まれると考えられる。特に前者については変更するべきではないと考えられることから、矛盾のある回答であっても、そのまま集計することとした。

論文の OA 率についても経年変化を確認するため、2016/2018 年調査の結果と比較した。

図 21 にそれぞれの結果を示す。

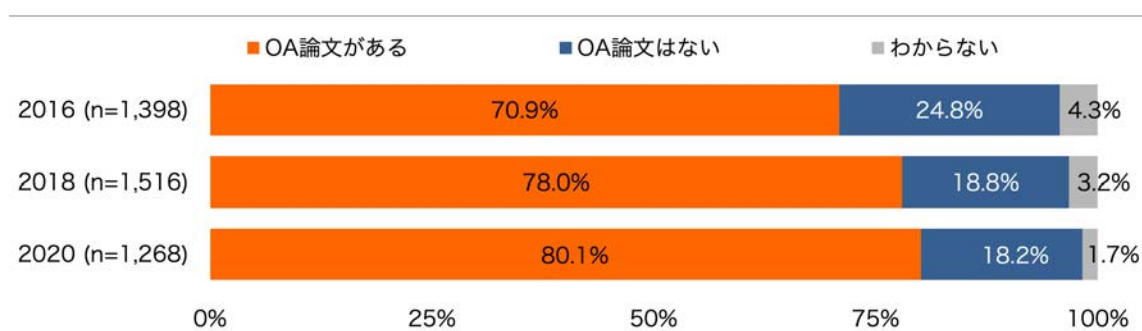


図 21 論文の公開経験の経年変化 (2016/18/20 年)

論文の OA 率は 70.9% (2016 年) から 78.0% (2018 年)、そして 80.1% (2020 年) まで 9.2 ポイント増加していた。OA 経験がない回答者は 24.8% (2016 年) から 18.2% (2020 年) まで 6.6 ポイント減少していた。「わからない」とした回答者も 4.3% (2016 年) から 1.7% (2020 年) まで 2.6 ポイント減少していた。

図 22 に、分野別のデータ公開率を 2016/2018 年調査とともに示す。配列は、2020 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 2020 年調査の値である。2016/2018 年調査の値については、2016 年の報告書 (RM268)<sup>2</sup>、及び 2018 年の報告書 (RM289)<sup>5</sup> 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、特に数学 (n=13) や心理学 (n=19) は少数であることにご留意いただきたい。

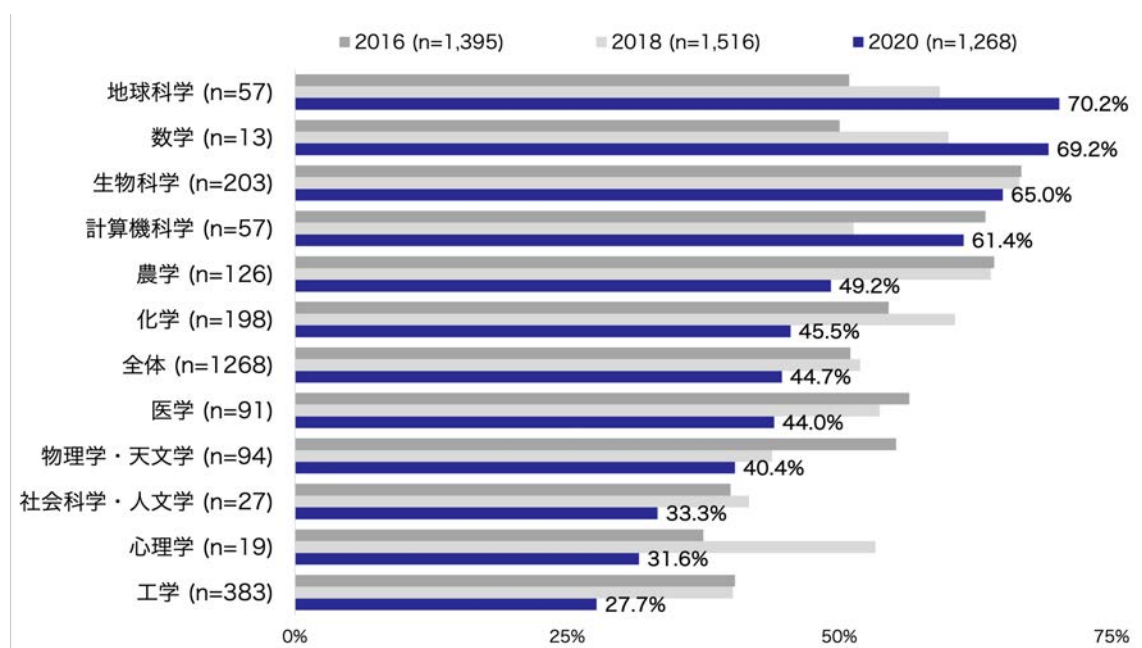


図 22 分野別データ公開率 (2016/18/20 年)

データの公開率が最も高かったのは地球科学 (70.2%)、次いで数学 (69.2%)、生物科学 (65.0%) の順であり、低かったのは工学 (27.7%)、心理学 (31.6%)、人文学・社会科学 (33.3%) であった。2018 年調査と 2020 年調査を比較して 10 ポイント以上増加していたのは、地球科学 (+11.0 ポイント) と計算機科学 (+10.1 ポイント) の 2 分野であり、減少していたのは心理学 (-21.7 ポイント)、化学 (-15.1 ポイント)、農学 (-14.7 ポイント)、工学 (-12.5 ポイント) の 4 分野であった。

図 23 に、分野別論文の公開率を 2016/2018 年調査の結果とともに示す。配列は、2020 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 2020 年調査の値である。

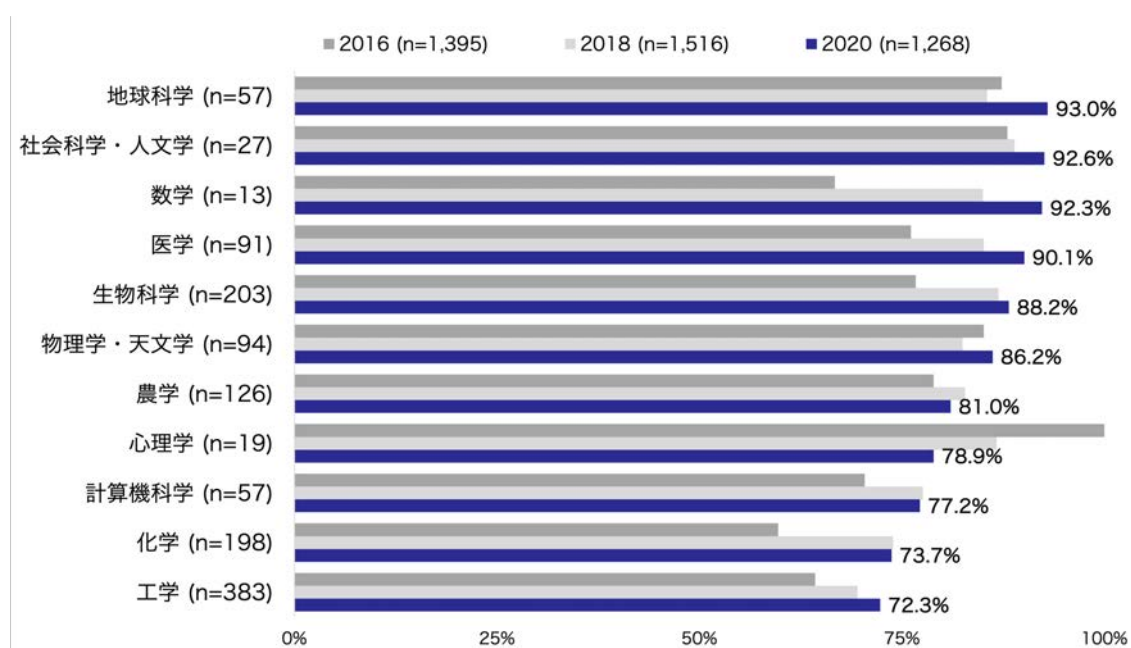


図 23 分野別論文の公開率 (2016/18/20 年)

論文の公開率が最も高かったのは地球科学 (93.0%)、次いで人文学・社会科学 (92.6%)、数学 (92.3%)、医学 (90.1%) の順であり、低かったのは工学 (72.3%)、化学 (73.7%)、計算機科学 (77.2%) であった。2018 年調査と 2020 年調査を比較して増加がみられたのは、地球科学 (+7.5 ポイント)、数学 (+7.3 ポイント) であり、減少していたのは心理学 (-7.8 ポイント) 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) 分野は論文の OA 率も高い (低い) とはいえなかった。

図 24 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。

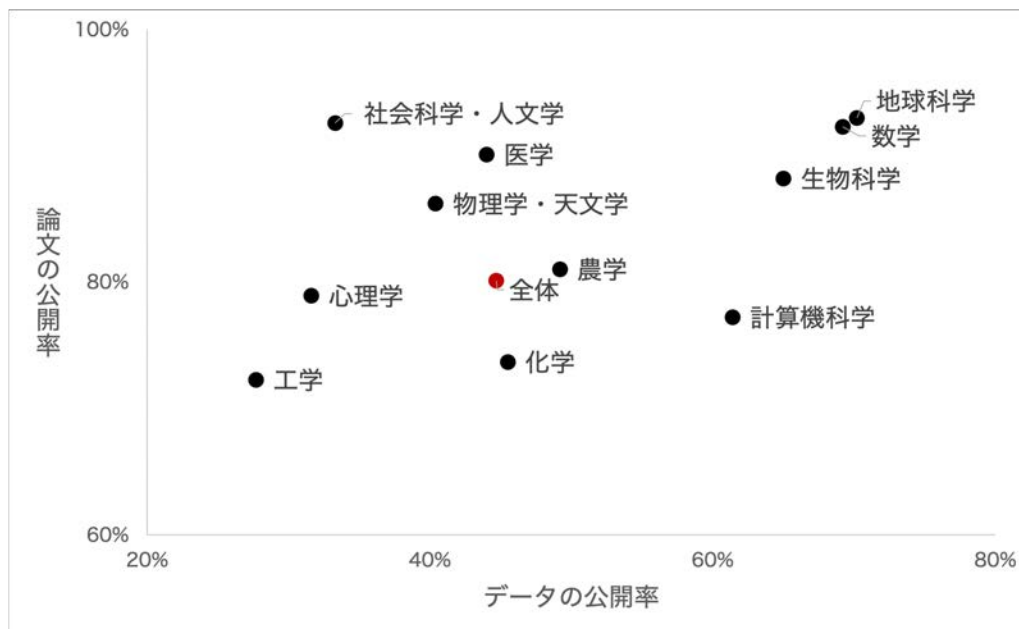


図 24 分野別データと論文の公開率 (n=1,268)

データと論文の公開率がいずれも高かったのは地球科学と数学、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計算機科学であり、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社会科学・人文学等であった。

## (2) データと論文の公開方法

データを公開した方法を確認するために、「Q18.研究データを公開した際に利用したサーバやサービスなどを全てお選び下さい」と複数選択方式で尋ねた。図 25 に、2016/2018 年調査及び本調査の結果を示す。配列は、2020 年調査において選択率が高かった公開方法の順とした。

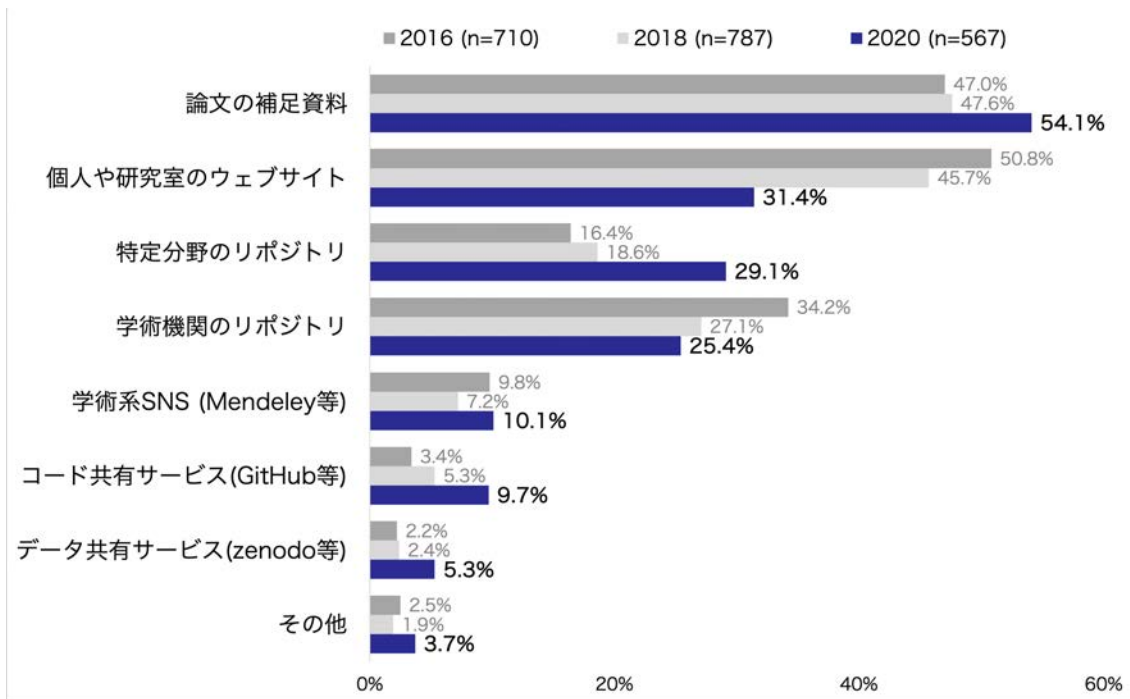


図 25 データの公開方法 (2016/18/20 年：複数回答)

データの公開方法として選択率が最も高かったのは、2018 年調査と同様に「論文の補足資料 (supplementary materials)」(54.1%)、次いで「個人や研究室のウェブサイトへの掲載」(31.4%)、「特定分野のリポジトリ・データアーカイブ (DDBJ<sup>viii</sup>や ICPSR<sup>ix</sup>など)」(29.1%)、「学術機関のリポジトリ・データアーカイブ(大学やNASA のリポジトリなど)」(25.4%)であった。「学術系 SNS (Mendeley<sup>x</sup>や ResearchGate<sup>xi</sup>など)」(10.1%)、「コード共有サービス (GitHub<sup>xii</sup>など)」(9.7%)、「データ共有サービス (figshare<sup>xiii</sup>や zenodo<sup>xiv</sup>など)」(5.3%)は、それほど選択率が高くなかった。

順位を確認すると、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

<sup>viii</sup> <https://www.ddbj.nig.ac.jp/index.html>

<sup>ix</sup> <https://www.icpsr.umich.edu/icpsrweb/>

<sup>x</sup> <https://www.mendeley.com/>

<sup>xi</sup> <https://www.researchgate.net/>

<sup>xii</sup> <https://github.co.jp/>

<sup>xiii</sup> <https://figshare.com/>

<sup>xiv</sup> <https://zenodo.org/>

文の補足資料」であったのが、2018年調査と2020年調査では逆転して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。また、「論文の補足資料」は3回の調査のうち、もっとも比率が高かった。2016/2018年調査では、3位が「学術機関のリポジトリ」、4位が「特定分野のリポジトリ」であったが、2020年調査では逆転して3位が「特定分野のリポジトリ」であり、2018年調査と比べて10.5ポイント増加していた。個人や研究室のウェブサイトが減少して、リポジトリやアーカイブの選択率がやや増えているという傾向がみられた。2020年に Hrynaszkiewicz らが北米と欧州の研究者を対象として実施した調査においても、もっともよく用いられているデータ共有の方法は論文の補足資料(67%)であり、次いで公開リポジトリ (publicly accessible repository) (59%) であった<sup>11</sup>。

「その他」には、プレプリントサーバー (5名)、企業のウェブサイト (3名)、YouTubeなどの動画共有サイト (2名)、データペーパー<sup>xv</sup> <sup>12</sup> (2名) などの記述があった。

論文を公開した方法を確認するために、「Q4.論文をオープンアクセスにした際に利用したサーバやサービスなどを全てお選び下さい」と複数選択方式で尋ねた。図25に、集計結果を示す。配列は、選択率が高かった公開方法の順とした。

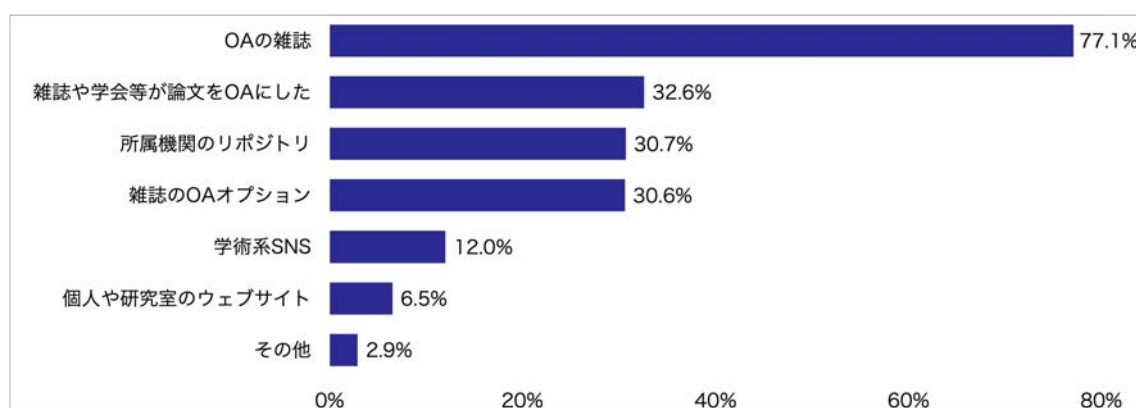


図 26 論文の公開方法 (n=1,016: 複数回答)

論文の公開方法として選択率が最も高かったのは、「オープンアクセスの雑誌 (Scientific Reports や PLOS など)」(77.1%)、次いで「雑誌や学会等が論文をオープンアクセスにした」(32.6%)、「所属機関のリポジトリ」(30.7%)、「雑誌のオープンアクセスオプション (Springer Open Choice など)」(30.6%)であった。「学術系 SNS (Mendeley や ResearchGate など)」(12.0%)、「個人や研究室のウェブサイト」(6.5%) は、それほど選択率が高くなかった。

<sup>xv</sup> 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 567 名を対象として、公開理由を複数選択方式で尋ねた。図 27 に 2016/2018 年調査とあわせて集計結果を示す。配列は、2020 年調査において選択率が高かった公開方法の順とした。

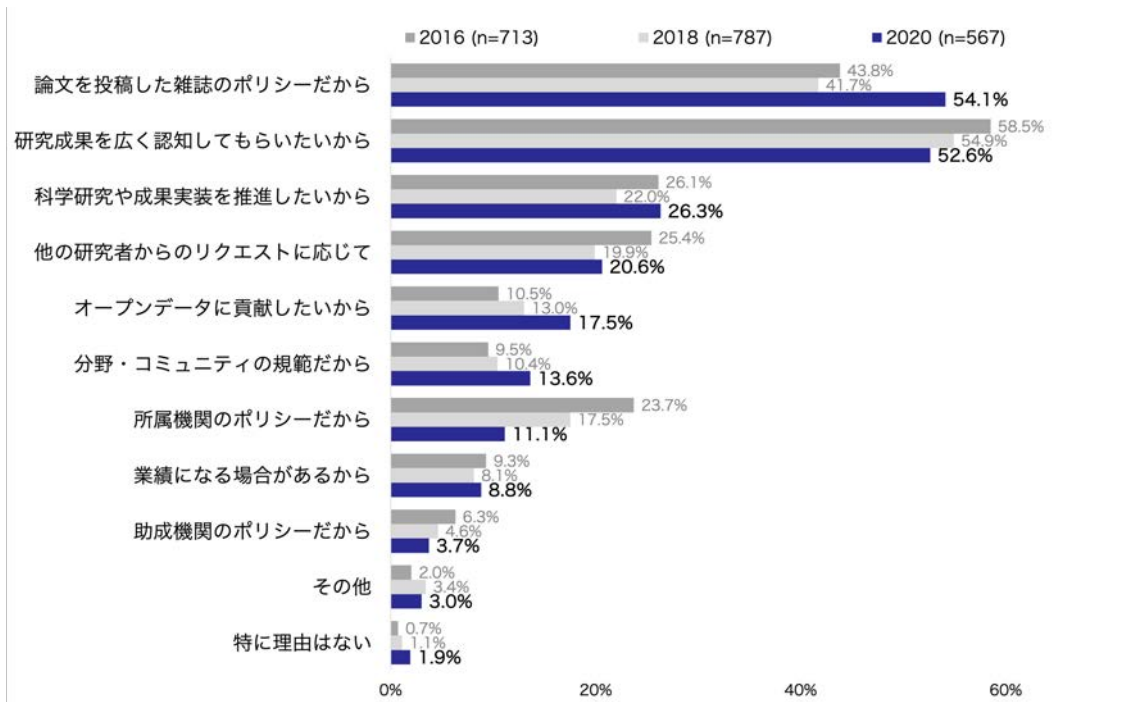


図 27 データの公開理由 (2016/18/20 年：複数回答)

データの公開理由として選択率が最も高かったのは、「論文を投稿した雑誌のポリシー（投稿規定）だから」(54.1%)、次いで「研究成果を広く認知してもらいたいから」(52.6%)であり、この 2 項目はいずれも過半数の回答者が選択していた。続く「科学研究や成果実装を推進したいから」(26.3%)や「他の研究者からのリクエストに応じて」(20.6%)は 2016/2018 年調査とほぼ同様であり、「オープンデータに貢献したいから」(17.5%)と「分野・コミュニティの規範だから」(13.6%)はわずかながら経年的に増加していた。「所属機関のポリシーだから」(11.1%)、及び「助成機関のポリシー（助成条件）だから」(3.7%)は、2016/2018 年調査と比較すると減少傾向にあった。「業績になる場合があるから」(8.8%)は、ほとんど変化がみられなかった。

「その他」には、17 名 (3.0%) による記述があった。論文の信頼性や再現性に関連した記述として、論文のエビデンスを示すため (6 名)、査読者からのリクエストや勧めに応じて (4 名) といった理由がみられた。また、データ公開を要求された例として、“企業の方針に沿って”、“業務上必要だから”、“データを公開することが条件である研究だったから”などの記述がみられた。自発的な理由として、“研究成果の社会還元のため”、“引用数が増えるから”といった記述もみられた。



#### (4) 論文の公開理由

論文の公開経験がある回答者 1,016 名を対象として、公開理由を複数選択方式で尋ねた。図 28 に 2016/2018 年調査とあわせて集計結果を示す。配列は、2020 年調査において選択率が高かった順とした。なお、「雑誌や学会等がオープンアクセスにしたから」(図 28, \*印) は 2020 年調査で新規に追加した選択肢である。

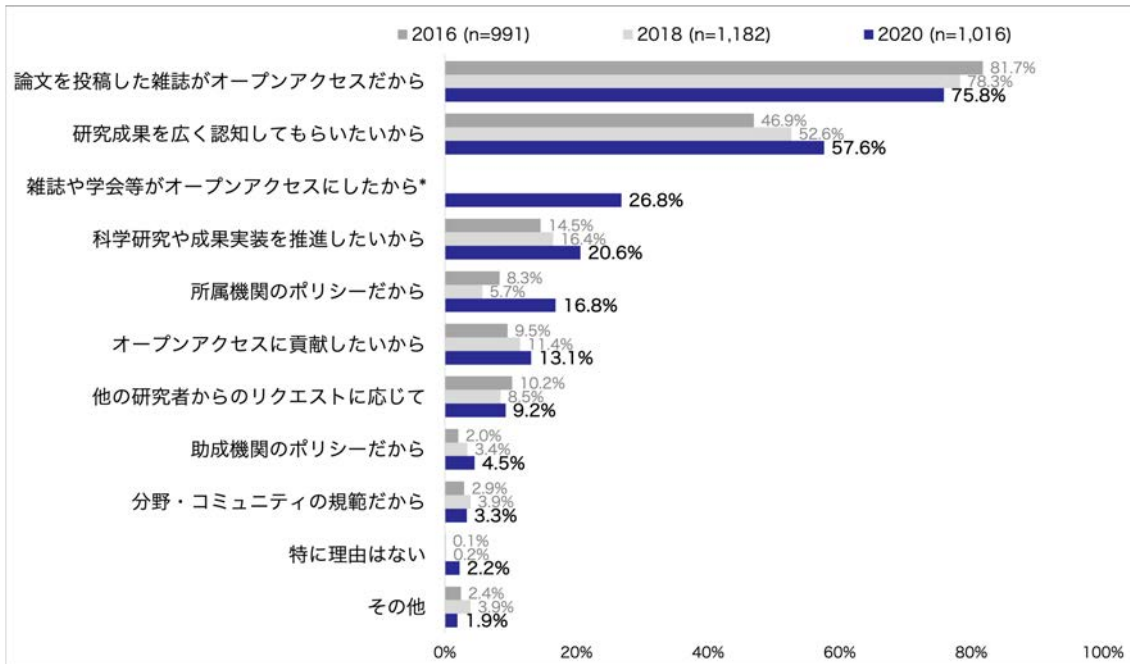


図 28 論文の公開理由 (2016/18/20 年：複数回答)

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」(75.8%)であり、次いで「研究成果を広く認知してもらいたいから」(57.6%)であった。この 2 項目が突出している状況は、2016/2018 年調査と同様であるといえる。新規に追加した「雑誌や学会等がオープンアクセスにしたから」は 26.8%であった。「科学研究や成果実装を推進したいから」(20.6%)、「オープンアクセスに貢献したいから」(13.1%)は、わずかながら経年的に増加していた。「所属機関のポリシーだから」(16.8%)は 2016 年 (8.3%)、2018 年 (5.7%) と比較して増加していた。「助成機関のポリシーだから」(4.5%)についても、わずかながら増加傾向がみられた。「他の研究者からのリクエストに応じて」(9.2%)や「分野・コミュニティの規範だから」(3.3%)は、それほど変化がみられなかった。

所属別に「所属機関のポリシーだから」の選択率を集計した結果、大学は 18.9%、企業は 12.8%、公的機関・団体は 8.8%であり差がみられた。大学による OA ポリシーの策定が影響していると考えられる。

## (5) データの未公開理由

Q.17で「研究にデジタルデータを用いるが、公開したことはない」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 29 に、2016/2018 年調査とあわせて集計結果を示す。配列は、2020 年調査において選択率が高かった公開方法の順とした。

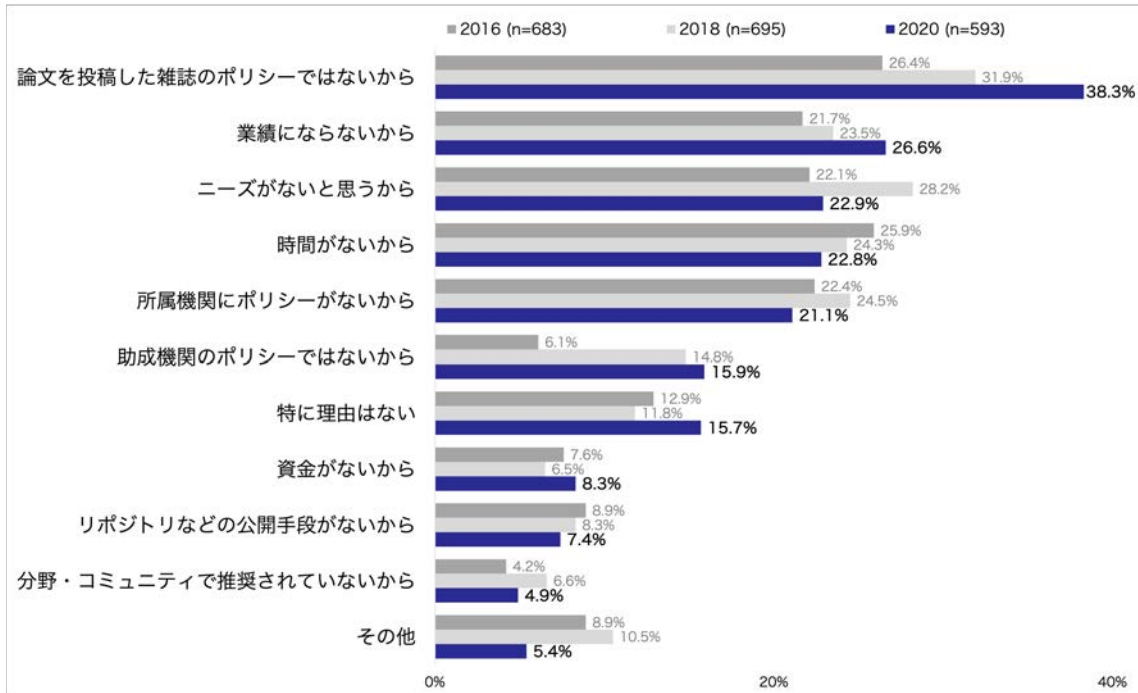


図 29 データの未公開理由（2016/18/20 年：複数回答）

データの未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のポリシーではないから」（38.3%）、次いで「業績にならないから」（26.6%）、「ニーズがないと思うから」（22.9%）、「時間がないから」（22.8%）、「所属機関にポリシーがないから」（21.1%）の順であった。2016/2018 年調査では、これらの 5 項目が約 20%～30%を占めており、突出した理由がみられなかったが、2020 年調査では「雑誌のポリシーではないから」が増加していた。「助成機関のポリシー（助成条件）ではないから」（15.9%）、「資金がないから」（8.3%）、「リポジトリなどの公開手段がないから」（7.4%）、「分野・コミュニティで推奨されていないから」（4.9%）は大きな変化はなかった。

所属別に「所属機関にオープンアクセス方針がないから」の選択率を算出したところ、大学は 3.5%、企業は 31.9%、公的機関・団体は 7.3%であり、差がみられた。特に企業は OA 方針を策定していないと推測される。

本研究の質問とはやや異なるが、Tenopir らが 2017 年から 2018 年にかけて、米国地球物理学連合 (AGU) の研究者を対象として実施した調査によれば、「データの全てまたは一部を共有していない理由」の 1 位は「まず論文を出版する必要がある」（35.1%）、「十分な時間がない」（26.4%）、「データを公開する権利を持っていない」（22.6%）であった<sup>13</sup>。

「その他」には 32 名 (5.4%) による記述があった。表 7 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 33 件である。

表 7 データの未公開理由：「その他」の回答 (n=32)

内容と分類
1. データ公開によるリスクがある (8) 再利用率に対する懸念 (6)、アイデアを盗用される可能性 (2)
2. データ公開のための環境が不十分 (7) 公開のためのコストがかかる・面倒 (3)、対応が面倒 (2)、データのサイズが大きい (1)、倫理的配慮の基準がはっきりしない (1)
2. データ公開の慣習がない (7) 求められれば公開できるようにしているがリクエストがない (3)、公開によるメリットがない (2)、必要ならば直接連絡が来る (1)、倫理審査の際にデータ公開について記載していない (1)
4. 自分で利用するため (6) 今後、自分でデータを利用する可能性がある (4)、データペーパーを書いてから公開する (1)、他の研究者に先んじられたくない (1)
5. データ公開は必要ない (3) 論文で十分 (1)、必要ない (1)、公開しても取り扱える人は極めて限定的 (1)
6. データの性質上公開できない (1) 知的財産権 (1)

()内の数字は人数を示す。1 人で複数の理由を記述している場合もある。

未公開理由のうち最も多かったのは、「データ公開によるリスクがある」(合計 8 名) という意見であった。“不特定多数の人間に公開した場合、どのような用途に使われるか分からず不安なため”など、再利用に対する懸念 (6 名) やアイデアを盗用される可能性がある (2 名) といった意見がみられた。

次いで、「データ公開のための環境が不十分である」(合計 7 名) という意見が多く、公開のためのコストがかかる・面倒である (3 名)、公開後の対応が面倒である (2 名)、データのサイズが大きい (1 名)、倫理的配慮の基準がはっきりしない (1 名) といった回答がみられた。同じく 2 番目に多かったのは、「分野等にデータ公開の慣習がない」(合計 7 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 (2 名)、必要ならば直接連絡が来る (1 名)、倫理審査の際にデータ公開について記載していない (1 名) といった記述がみられた。

4 番目は、「自分で利用するため」(合計 6 名) であった。「その他」の記述は 2018 年調

査の分類に準じてまとめたが、こうした回答は 2018 年調査ではみられなかったため、新規に見出しを作成した。具体的には、今後、自分でデータを利用する可能性がある（4 名）、データペーパーを書いてから公開する（1 名）、他の研究者に先んじられたくない（1 名）といった回答がみられた。

5 番目は、「データ公開は必要ない」（合計 3 名）という意見であった。論文への記載で十分である（1 名）、データ公開は必要ない（1 名）、“取得直後のデータ（生データ）を取り扱える人は極めて限られているから（解析し、グラフ化しないと価値は極めて低いから）”（1 名）といった記述がみられた。

2018 年調査では、「データの性質上公開できない」という意見が最も多かったが、本調査で言及していたのは 1 名のみであった。具体的には、“知財を守るため”であった。

## (6) 論文の未公開理由

Q.3 で論文をオープンアクセスにした経験がないと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 30 に、2016/2018 年調査とあわせて集計結果を示す。配列は、2020 年調査において選択率が高かった順とした。なお、「オープンアクセスではない発表方式の方が望ましいと思うから」（図 30, \*印）は 2020 年調査で新規に追加した選択肢である。

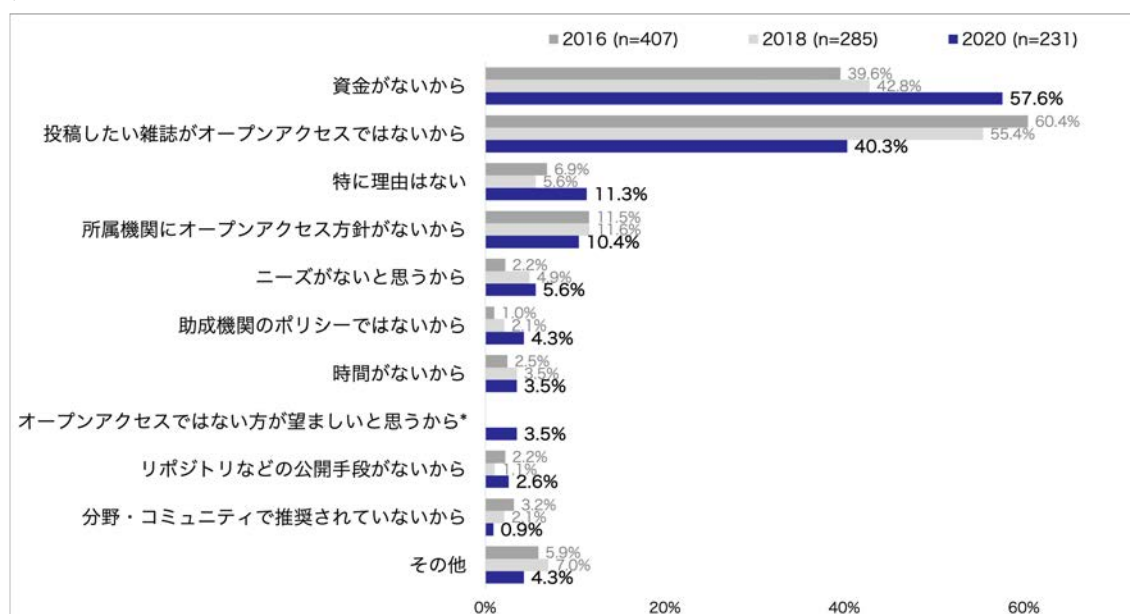


図 30 論文の未公開理由（2016/18/20 年：複数回答）

2018 年調査において選択率が最も高かったのは「資金がないから」（57.6%）、2 位は「投稿したい雑誌がオープンアクセスではないから」（40.3%）であった。2016/2018 年調査は、いずれも「投稿したい雑誌がオープンアクセスではないから」が 1 位であったが、本調査で初めて「資金がないから」が 1 位となった。これら 2 項目の選択率が高いという傾向は変わらず、3 位の「特に理由はない」（11.3%）と 4 位の「所属機関にオープンアクセス方針がないから」（10.4%）は選択率がやや高いものの、「ニーズがないと思うから」（5.6%）

など 6 項目の選択率は 6%未満であった。新規項目の「オープンアクセスではない発表方式の方が望ましいと思うから」は 3.5%に留まった。

「その他」には 10 名 (4.3%) による記述があった。プレプリントを公開しているから (2 名)、企業に関連する研究だから (2 名)、“費用削減のため”、“費用に見合うとは思わない”、“共著者の了解を取るのに時間がかかる”、“著作権の扱いが不明であるため”、“得られた成果の大きさによる”といった記述がみられた。

なお、2018 年調査で最も多かったのは、OA ジャーナルの信頼性が低いとする記述 (7 名) であり、うち 3 名はハゲタカジャーナル (predatory journal) <sup>xvi</sup> 14) について言及していた。一方、2020 年調査では、“そもそもオープンアクセスの論文誌には価値がないから”という記述のみであった。

## (7) データの公開意思

データの未公開理由について、「特に理由はない」以外を選択した回答者 500 名に、「Q23. Q22 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？」と尋ねた。図 31 に、2016/2018 年調査とあわせて集計結果を示す。

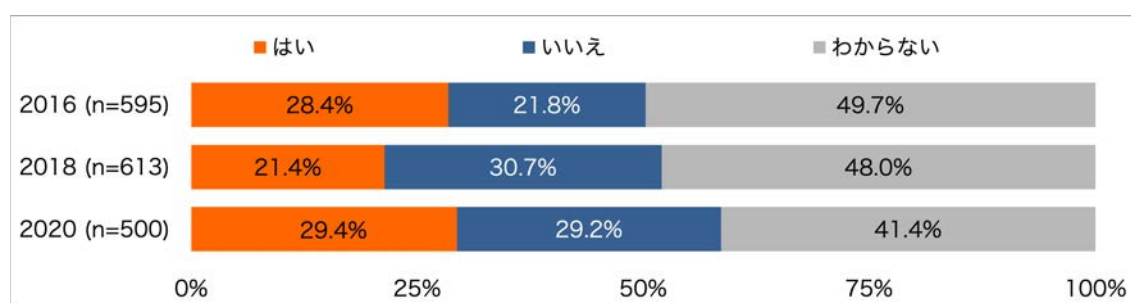


図 31 データの公開意思 (2016/18/20 年)

2016/2018 年調査と 2020 年調査の結果を比較すると、「わからない」が 2016 年 (49.7%)、2018 年 (48.0%)、2020 年 (41.4%) と、経年的に減少していた。「はい」を選択した回答者は 29.4%であり、2018 年 (21.4%) からは 8 ポイント増加しているものの、3 割未満に留まっている。「いいえ」を選択した回答者は 29.2%であり、同じく 3 割程度であった。

## (8) 論文の公開意思

同じく論文の未公開理由について、「特に理由はない」以外を選択した回答者 205 名に、「Q7. Q6 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いますか？」と尋ねた。図 32 に、2016/2018 年調査とあわせて集計結果を示す。

<sup>xvi</sup> 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。

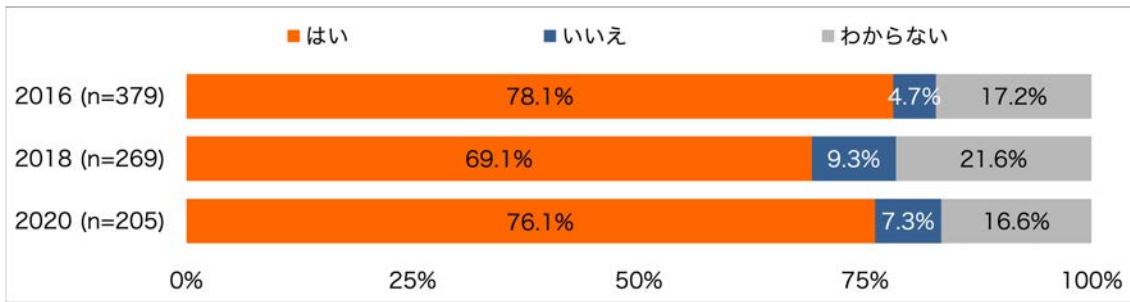


図 32 論文の公開意思 (2016/18/20 年)

2016/2018 年調査と 2020 年調査の結果を経年的に比較すると、「わからない」の選択率は 2016 年が 17.2%、2018 年が 21.6%、2020 年が 16.6%であり、もっとも選択率が低かった。「はい」を選択した回答者は 76.1%であり、2018 年 (69.1%) からは 7 ポイント増加していた。「いいえ」を選択した回答者は 7.3%であり、2018 年 (9.3%) よりは 2 ポイント減少していたものの、2016 年 (4.7%) よりは 2.6 ポイント増加していた。

データと論文の公開意思を比較すると、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は 2016 年調査から 2018 年調査にかけて減少していたが、2020 年調査では増加したという傾向はデータと論文に共通していた。

## 2.2 データの共有 (提供) 状況

データの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 33 に、2016/2018 年調査とあわせて集計結果を示す。

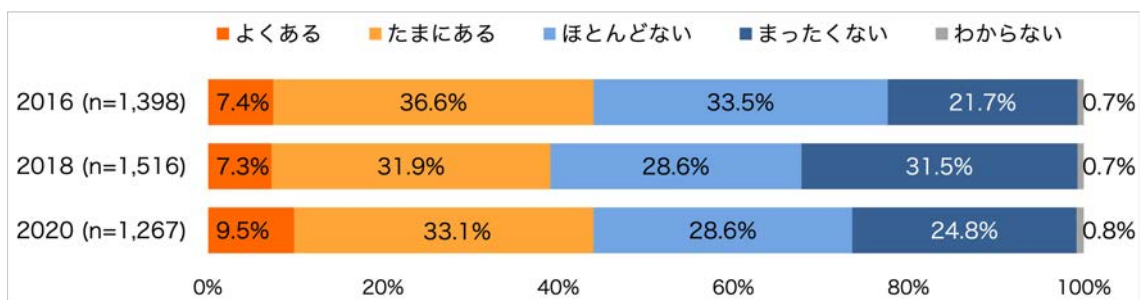


図 33 データの共有経験 (2016/18/20 年)

全体的な傾向として、2016 年調査は「よくある」と「たまにある」の合計が 44.0%、2018 年調査では 39.1%に 3.9 ポイント減少し、2020 年調査では 42.6%に 3.5 ポイント増加した。また、2016 年調査は「まったくない」と「ほとんどない」の合計が 55.2%、2018 年

調査では 60.2%に 5.0 ポイント増加し、2020 年調査では 53.4%に 6.8 ポイント減少した。2020 年調査は「よくある」の選択率が 9.5%であり、2016/2018 年調査と比較してやや高かったものの、データの共有経験はそれほど変化がなかったといえる。

図 34 に、分野別データ共有（提供）経験を有する回答者の比率、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 2016/2018 年調査の結果<sup>xvii</sup>とともに示す。分野に付した「n」、及びグラフの右に付した比率（%）は 2020 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%、2020 年調査は 71.2%であった。

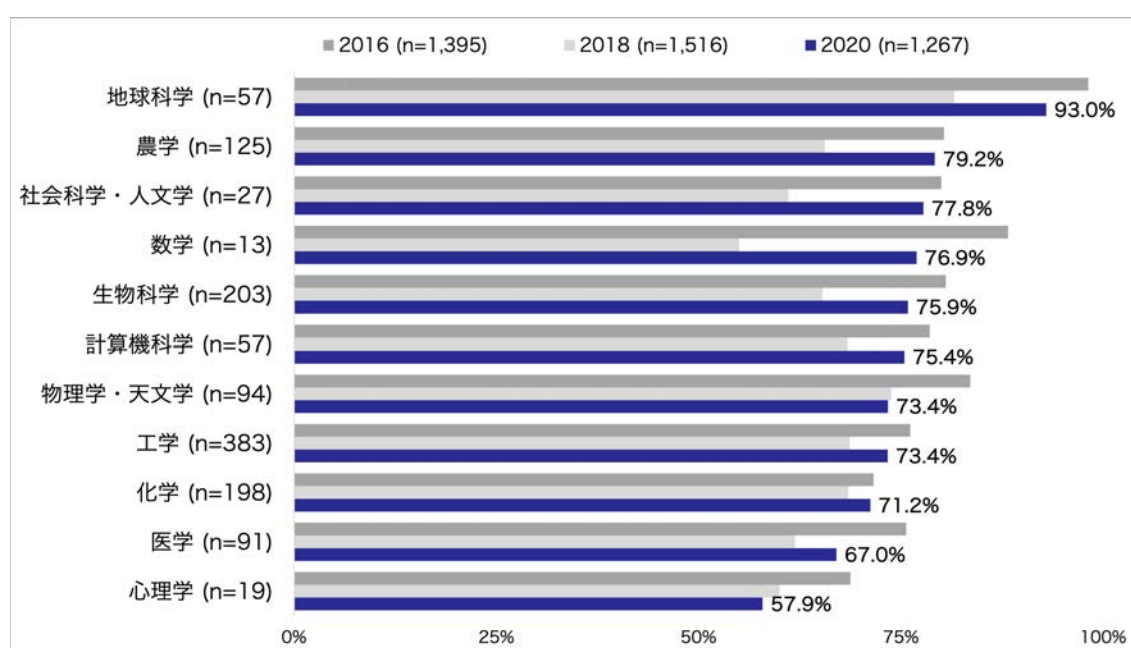


図 34 分野別データ共有経験（2016/18/20 年）

データ提供経験を有する回答者の比率が最も高かったのは地球科学（93.0%）、次いで農学（79.2%）、社会科学・人文学（77.8%）の順であり、低かったのは心理学（57.9%）、医学（67.0%）、化学（71.2%）であった。2018 年調査と 2020 年調査を比較して、10 ポイント以上増加がみられたのは、数学（+21.9 ポイント）、社会科学・人文学（+16.7 ポイント）、農学（+13.6 ポイント）、地球科学（+11.4 ポイント）、生物科学（+10.6 ポイント）であった。大幅に減少した分野はみられなかった。

分野別のデータ提供経験は、データ公開経験と正の相関がみられた（Pearson の相関係数  $r=0.604, p<0.05$ ）。つまり、データをよく共有している分野はデータを公開しているという傾向がみられた。

<sup>xvii</sup> 2016 年調査では、データの提供経験と被提供経験を尋ねた。ここでは提供経験の結果を示す。なお、提供経験の方が被提供経験よりも頻度が高かった。

## 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、「Q9. これまでに、公開データを入手したご経験はありますか?」と尋ねた。選択肢は「はい」、「いいえ」、「わからない」として、続く Q10 には選択肢としてデータの入手方法の例を示した。図 35 に、2016/2018 年調査とあわせて集計結果を示す。

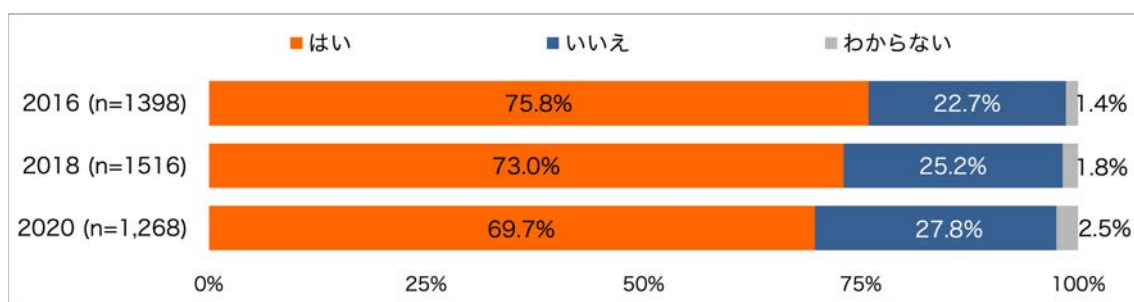


図 35 公開データの入手経験の有無 (2016/18/20 年)

2016/2018 年調査と 2020 年調査の結果を経年的に比較すると、「はい」の選択率は 2016 年が 75.8%、2018 年が 73.0%、2020 年が 69.7%であり、減少傾向にあった。逆に「いいえ」は 2016 年が 22.7%、2018 年が 25.2%、2020 年が 27.8%であり、増加傾向にあった。

図 36 に、分野別データ入手経験を 2016/2018 年調査の結果とあわせて示す。分野に付した「n」、及びグラフの右に付した比率 (%) は 2020 年調査の値である。



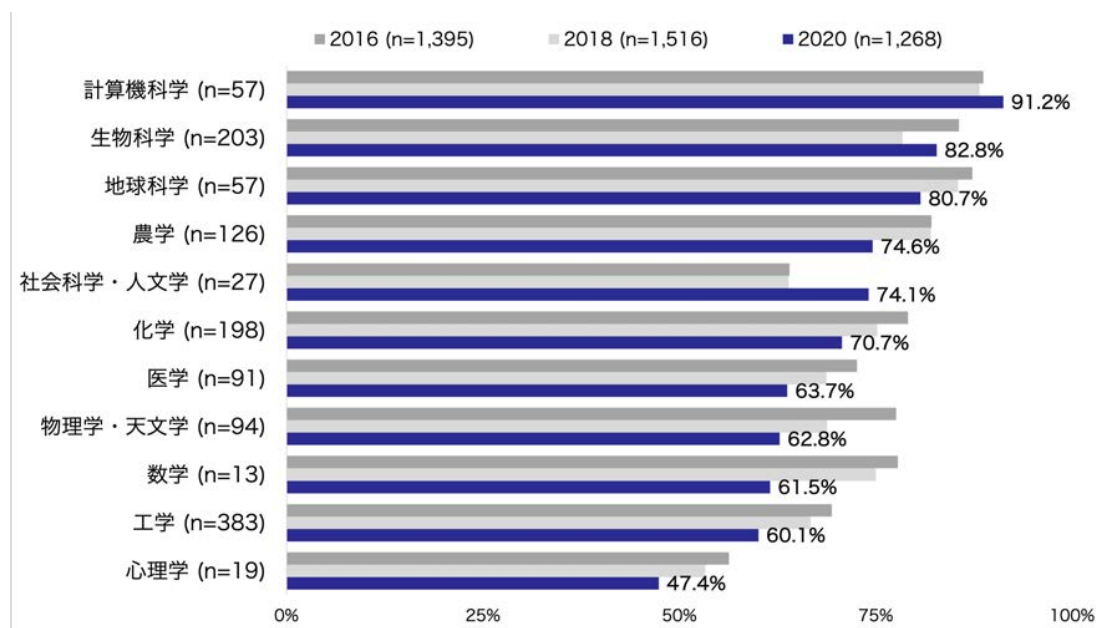


図 36 分野別公開データの入手経験 (2016/18/20 年)

公開データの入手経験を有する回答者の比率が最も高かったのは計算機科学 (91.2%)、次いで生物科学 (82.8%)、地球科学 (80.7%) の順であり、低かったのは心理学 (47.4%)、工学 (60.1%)、数学 (61.5%) であった。2018 年調査と 2020 年調査を比較すると、社会科学・人文学は 10.2 ポイント増加しており、数学は 13.5 ポイント減少していた。ただし、両分野とも回答者数が少ないことがポイントの増減に強く影響していると考えられる。

分野別のデータ入手経験は、データ公開経験と正の相関がみられた (Pearson の相関係数  $r=0.604$ ,  $p<0.05$ )。同様に、データ提供経験とも正の相関がみられた (Pearson の相関係数  $r=0.651$ ,  $p<0.05$ )。つまり、データをよく入手している分野は、データ公開や共有を行っているという傾向がみられた。

## (2) 公開データの入手方法

データ入手経験がある回答者 884 名を対象として、公開理由を複数選択方式で尋ねた。図 37 に 2016/2018 年調査とあわせて集計結果を示す。配列は、2020 年調査において選択率が高かった公開方法の順とした。

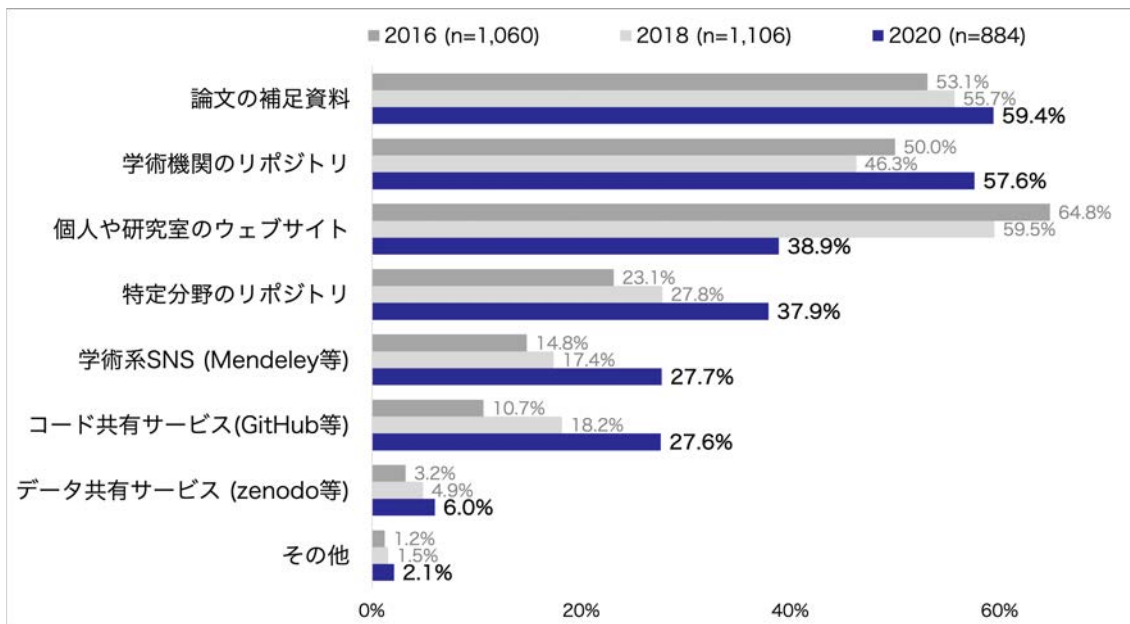


図 37 公開データの入手方法（2016/18/20 年：複数回答）

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「論文の補足資料」（59.4%）、次いで「学術機関のリポジトリ（大学や NASA のリポジトリなど）」（57.6%）、「個人や研究室のウェブサイト」（38.9%）、「特定分野のリポジトリ・データアーカイブ（DDBJ や ICPSR など）」（37.9%）の順であった。「学術系 SNS（Mendeley や ResearchGate など）」（27.7%）や「コード共有サービス（GitHub など）」（27.6%）は、それぞれ約 10 ポイント増加していた。「データ共有サービス（figshare や zenodo など）」（6.0%）はやや増加しているものの、選択率は低かった。

表 8 に、データの公開方法と入手方法の順位を示す。2016/2018 年調査では「個人や研究室のウェブサイト」が 1 位であったが、2020 年調査ではデータの公開先と同様に、「論文の補足資料」が 1 位であった。

表 8 データの公開方法と入手方法

順位	データの公開方法	データの入手方法
1 位	論文の補足資料	論文の補足資料
2 位	個人や研究室のリポジトリ	学術機関のリポジトリ
3 位	特定分野のリポジトリ	個人や研究室のリポジトリ
4 位	学術機関のリポジトリ	特定分野のリポジトリ
5 位	学術系 SNS	学術系 SNS
6 位	コード共有サービス	コード共有サービス
7 位	データ共有サービス	データ共有サービス

「その他」には、19 名（2.1%）による記述があった。官公庁や自治体が公開しているオ

ーブンデータ(12名)、“国立国会図書館デジタルコレクション”、“NGOのウェブサイト”、“Kaggle<sup>xviii</sup>”などが挙げられていた。

### (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、データの入手経験がある回答者(884名)を対象として、「Q11. 公開データを入手した際に、問題があると感じたご経験はありますか?」と尋ねた。図 38 に、2016/2018年調査とあわせて集計結果を示す。

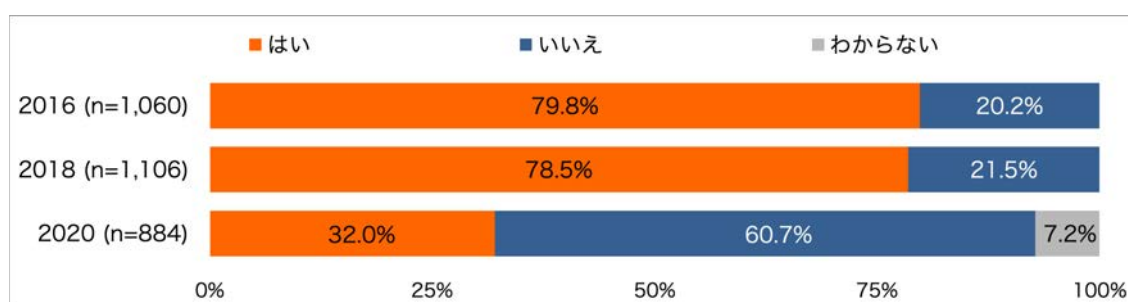


図 38 データ入手の障壁の有無 (2016/18/20年)

2016/2018年調査では、8割近くの回答者がデータを入手する際に問題があるとしていたが<sup>xix</sup>、2020年調査では32.0%にとどまった。「いいえ」は60.7%、「わからない」は7.2%であった。

続く Q12 では、公開データを入手する際に問題があったとした回答者 816 名を対象として、データ入手の際の障壁を複数選択方式で尋ねた。図 39 に 2016/2018年調査とあわせて集計結果を示す。配列は、2020年調査において選択率が高かった公開方法の順とした。

<sup>xviii</sup> <https://www.kaggle.com/>

<sup>xix</sup> 1.4(2)に示したように、2016/2018年調査では、次項に挙げる選択肢を1つ以上選択した場合は「はい」とみなした。「わからない」という選択肢は設けなかった。

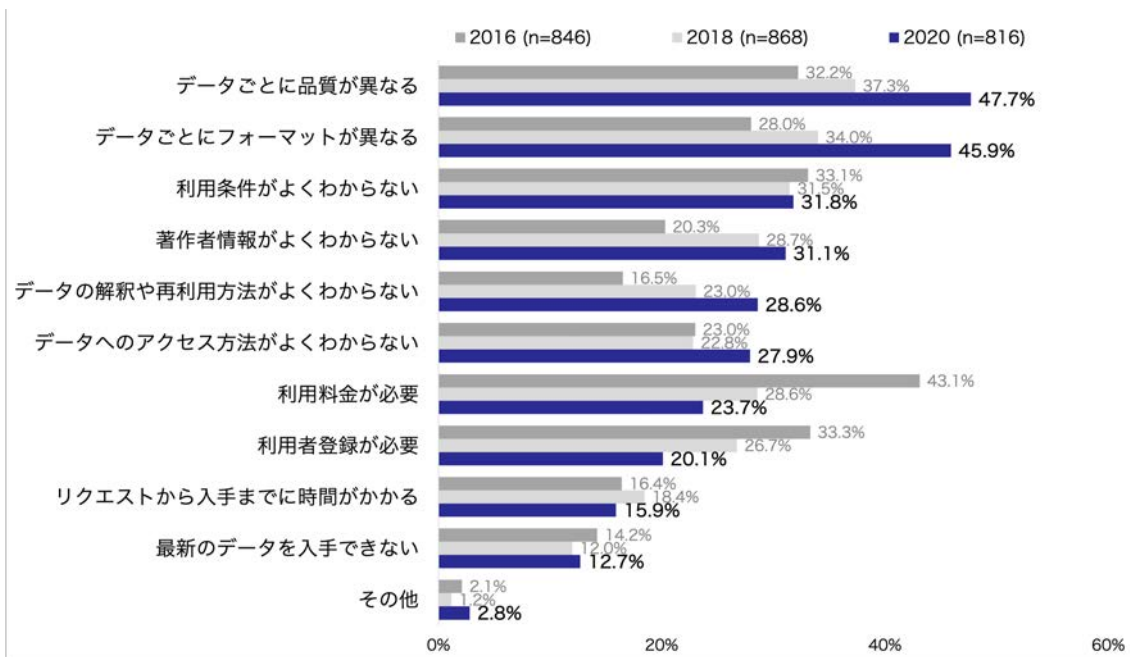


図 39 データ入手の障壁 (2016/18/20 年：複数回答)

データ入手の際の問題のうち、最も選択率が高かったのは「データごとに品質が異なる」(47.7%)、2位は「データごとにフォーマットが異なる」(45.9%)、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」(31.8%)、4位は「著作者情報がよくわからない」(31.1%)、5位は「データの解釈や利用方法がよくわからない」(28.6%)、6位は「利用したいデータへのアクセス方法がよくわからない」(27.6%)であった。以上6項目のうち「利用条件がよくわからない」を除いた5項目は、全て選択率が増加していた。

7位の「利用料金が必要」(23.1%)、8位の「利用者登録が必要」(20.1%)はいずれも選択率が減少していた。9位の「リクエストから入手までに時間がかかる」(15.9%)と10位の「最新のデータを入手できない」(12.7%)は選択率が低いものの、3回の調査で常に一定の割合で選択されていた。

表 9 に、3回の調査におけるデータ入手の障壁の順位を示す。2016年の時点では、データ入手の時点での障壁、すなわち利用料金がかかることや利用者登録が必要であることを問題であるとした回答者が多かった(表の網掛け部分)。しかし、これらは2018年、2020年と順位を落とし、データの利活用における障壁の順位が高くなったといえる。

表 9 データ入手の障壁の経年変化 (2016/18/20 年)

順位	2016 年	2018 年	2020 年
1 位	利用料金	データの品質が異なる	データの品質が異なる
2 位	利用者登録	データのフォーマット	データのフォーマット
3 位	利用条件がわからない	利用条件がわからない	利用条件がわからない
4 位	データの品質が異なる	著作者情報がわからない	著作者情報がわからない
5 位	データのフォーマット	利用料金	データの解釈・再利用
6 位	アクセス方法	利用者登録	アクセス方法
7 位	著作者情報がわからない	データの解釈・再利用	利用料金
8 位	データの解釈・再利用	アクセス方法	利用者登録
9 位	入手までの時間	入手までの時間	入手までの時間
10 位	最新データの入手	最新データの入手	最新データの入手

「その他」には、8 名 (2.8%) による記述があった。データそのものに関する問題として、“信頼性が第三者によって確認されていない”、“急に公開が停止されたり、プライバシーを侵害していたなどで利用することが問題となることがある”、“所属組織の方針（フリープログラムの利用など）に沿うものか判断できない”、“機微情報として取り扱うべきかどうか迷う場合がある”といった記述がみられた。また、データ入手や管理における問題として、“自動ダウンロード用の API が未整備の場合、入手に手間が掛かる”、“入手後の扱いが煩雑”という記述もみられた。また、再利用した際の扱いについて、“先行研究や類似研究として、自身の研究の参考にでき有用であるが、論文の引用には使いづらい”という指摘がみられた。

## 2.4 データマネジメントプラン（DMP）の作成状況

### (1) DMP の作成経験

データマネジメントプラン（DMP）の作成経験を確認するために、「Q13. これまでに、DMP を作成したご経験はありますか?」と尋ねた。選択肢は「はい」、「いいえ」、「わからない」として、続く Q14 には選択肢として DMP の例を示した。図 40 に集計結果を示す。

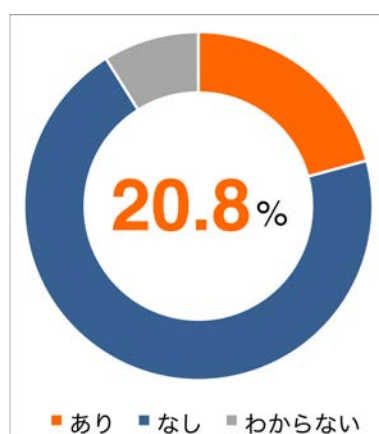


図 40 DMP の作成経験 (n=1,268)

DMP の作成経験を有する回答者は 20.8% (264 名)、「なし」は 70.3% (891 名)、「わからない」は 8.9% (113 名) であった。

DMP 作成経験の経年変化を確認するために、2018 年調査と本調査を比較した。図 41 にそれぞれの結果を示す。

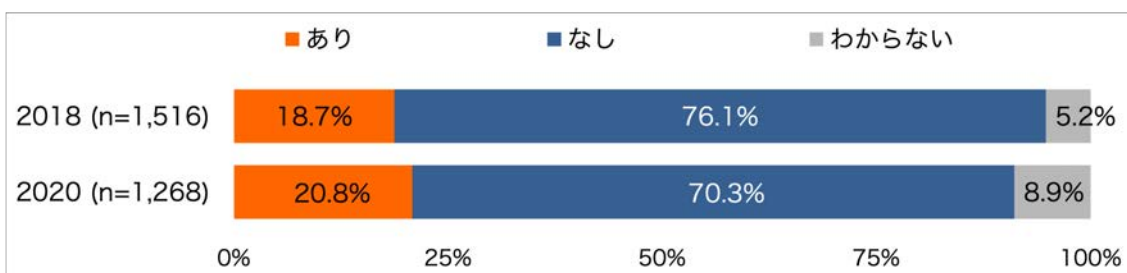


図 41 DMP 作成経験の経年変化 (2018/20 年)

DMP 作成経験を有する回答者は、18.7% (2018 年調査) から 20.8% (2020 年調査) まで 2.1 ポイント増加していた。作成経験をもたない回答者は 76.1% から 70.3% まで 5.8 ポイント減少していた。一方、「わからない」とした回答者は 5.2% から 8.9% まで 3.7 ポイント増加していた。

図 42 に分野別の DMP の作成経験を有する回答者の比率を示す。

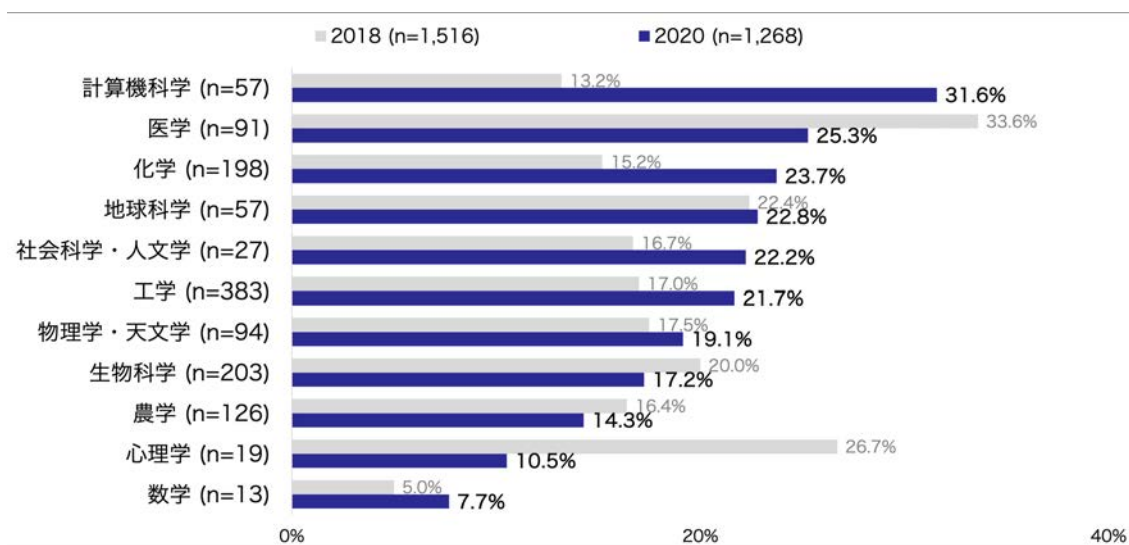


図 42 分野別 DMP の作成経験 (2018/20 年)

DMP の作成経験を有する回答者の比率が最も高かったのは計算機科学 (31.6%)、次いで医学 (25.3%)、化学 (23.7%)、地球科学 (22.8%) の順であり、低かったのは数学 (7.7%)、心理学 (10.5%)、農学 (14.3%) であった。2018 年調査と比較して増加していたのは計算機科学 (+18.4 ポイント) や化学 (+8.5 ポイント) であり、減少していたのは心理学 (-16.2 ポイント) や医学 (-8.3 ポイント) であった。

## (2) 作成経験がある DMP の種類

DMP 作成経験がある回答者 334 名を対象として、作成経験がある DMP の種類を複数選択方式で尋ねた。図 43 に、2018 年調査とあわせて集計結果を示す。配列は、2020 年調査において選択率が高かった DMP の種類順とした。なお、「国外の助成機関による助成研究の DMP」(図 43, \*印) は 2020 年調査で新規に追加した選択肢である。

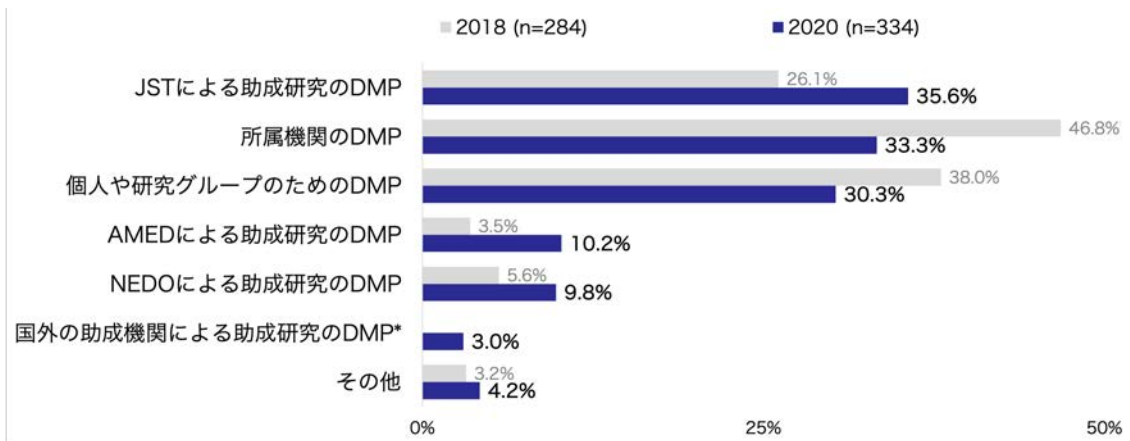


図 43 作成経験がある DMP (2018/20 年：複数回答)

例示した DMP のうち選択率が最も高かったのは「科学技術振興機構 (JST)」(35.6%)、2 位は「所属機関の DMP」(33.3%)、3 位は「個人や研究グループのための DMP」(30.3%)、4 位は「日本医療研究開発機構 (AMED)」(10.2%)、5 位は「新エネルギー・産業技術総合開発機構 (NEDO)」(9.8%)、6 位は「国外の助成機関による助成研究の DMP」(3.0%) であった。国内の助成機関の DMP の作成経験を有する回答者の比率が増えており、JST は 9.5 ポイント、AMED は 6.7 ポイント、NEDO は 4.2 ポイント増加していた。

「その他」には、11 名 (4.2%) による記述があった。具体的には、高エネルギー加速器研究機構 (KEK) の共同利用実験に関する DMP (3 名)、受託研究のための DMP (2 名) や聞き取り調査先への説明のための DMP (1 名)、ジャーナルの DMP (1 名)、臨床研究計画書 (1 名) などであった。

### (3) DMP の作成理由

DMP の作成経験がある回答者 351 名を対象として、作成理由を複数選択方式で尋ねた。図 44 に、2018 年調査とあわせて集計結果を示す。



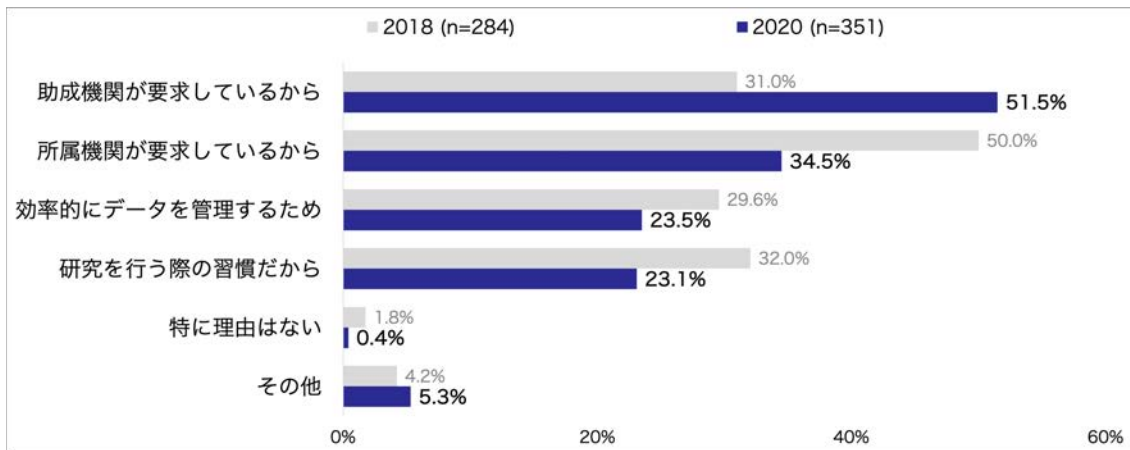


図 44 DMP の作成理由 (2018/20 年：複数回答)

DMP の作成理由のうち、選択率が最も高かったのは、「助成機関が要求しているから」(51.5%)、2 位は「所属機関が要求しているから」(34.5%)、3 位は「効率的にデータを管理するため」(23.5%)、4 位は「研究を行う際の習慣だから」(23.1%) であった。2018 年調査では所属機関の要求が 1 位であったが、2020 年調査では助成機関の要求が 17.0 ポイント上回った。

「その他」には、14 名 (5.3%) による記述があった。第三者からの要求として、研究機関の要求<sup>xx</sup> (3 名) やジャーナルの要求 (2 名) があった。第三者とのデータ共有に関連して、“企業のデータを扱う必要が出てきたから”、“共同研究者とデータの取り扱いについて事前に決めておき将来のトラブルを未然に防ぐため”、“無用のトラブルを避けるため”といった記述がみられた。また、大学生とのデータ共有に関連して、“卒業していく学生のデータを引き継ぐため”、“学生にデータを共有する際のルール徹底のため”という記述がみられた。データの内容に関連して、知的財産権保護のため (2 名)、“個人情報を取り扱うことがあるから”という回答がみられた。

<sup>xx</sup> 所属機関以外の研究機関であると判断した。

#### (4) DMP の未作成理由

DMP の作成経験がない回答者 891 名を対象として、未作成理由を複数選択方式で尋ねた。図 45 に 2018 年調査とあわせて集計結果を示す。配列は、2020 年調査において選択率が高かった未作成理由の順とした。

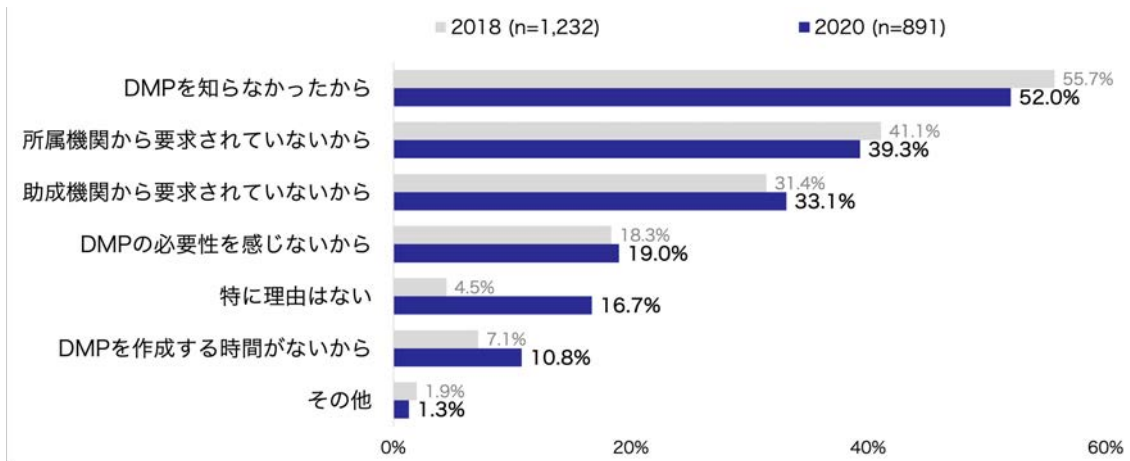


図 45 DMP の未作成理由 (2018/20 年：複数回答)

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」(52.0%)、2 位は「所属機関から要求されていないから」(39.3%)、3 位は「助成機関から要求されていないから」(33.1%)、4 位は「DMP の必要性を感じないから」(19.0%)、5 位は「特に理由はない」(16.7%)、6 位は「DMP を作成する時間がないから」(10.8%) であった。2018 年調査とほぼ同じ順位であったが、2020 年調査では「特に理由はない」の選択率が高く、12.2 ポイント増加していた。

「その他」には、12 名 (1.3%) による記述があった。DMP の作成に近いものとして、所属機関の規則に従っているから (3 名)、研究代表者や専門の DMP 担当者が作成しているから (2 名)、“天文学分野では確立した公開フォーマットがある程度のレベルに達しており、国内で事務手続きを増やされるのは迷惑”といった記述がみられた。また、そもそもの事情や認識として、“該当するようなデータを使用しない”、“DMP を作成しなくとも、必要なデータ、試料は全て保存しているから”、“民間研究者であり助成を受けていない。そもそも民間においては研究成果をすべて公開することが前提とされていない”、“所属機関に保存するサーバーが存在しないから”という記述もみられた。DMP に対する印象として、“管理が大変そうなので”、“DMP が、自己目的化しそうだから”という記述がみられた。

## 2.5 データの整備や公開の障壁

### (1) カレントデータの量

回答者が研究に用いているカレント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「Q27. 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 46 に 2016/2018 年調査の結果とともに示す。

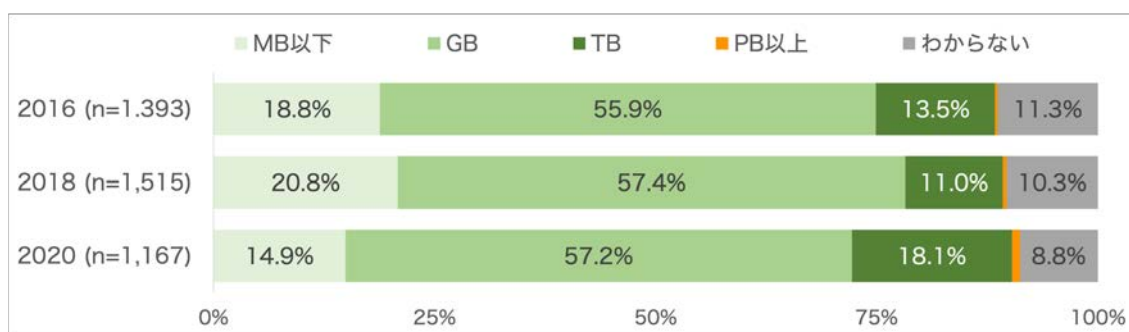


図 46 カレントデータの量 (2016/18/20 年)

最も選択率が高かったのは、「GB (ギガバイト)」(57.2%)、次いで「TB (テラバイト)」(18.1%)、「MG (メガバイト) 以下」(14.9%)、「PB (ペタバイト) 以上」(0.9%) の順であった。データ量を「わからない」とする回答者は 8.8%であった。2016/2018 年調査と比較すると、ややデータの量が増えている傾向がみられた。

### (2) カレントデータの整備・公開に必要な資源の状況

データの整備や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「Q28. カレントデータを整備・公開しようとする場合、次の資源は十分に整っていますか (いましましたか) ? 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 47)。

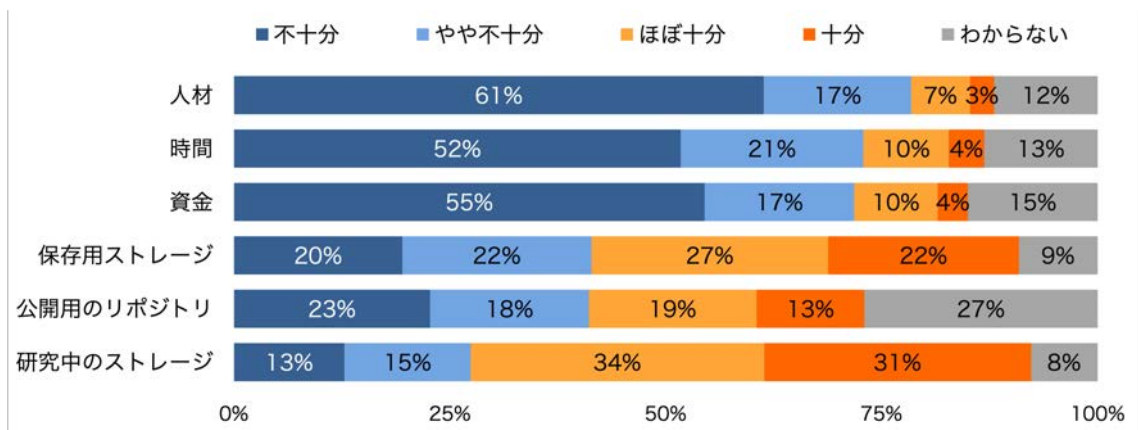


図 47 データの整備や公開に関する資源の充足度 (n=1,188)

人材、時間、資金については「不十分」とする回答者が過半数であった。特に人材は78.5%が、時間は72.8%が、資金は71.8%が、それぞれ「不足」または「やや不足」していると認識していた。「わからない」とする回答者の比率も高く、人材は12.0%、時間は13.1%、資金は15.0%であった。

保存用ストレージ、公開用のリポジトリ、研究中のストレージについては、「不十分」と「やや不十分」の合計選択率が5割を下回っていた。研究中のストレージは比較的充足度が高く、「十分」が30.9%、「ほぼ十分」が34.0%であった。公開用のリポジトリについては「わからない」とした回答者が27.0%であり、認知度が低いことがうかがえた。

資源の状況について経年変化を確認するために、2016/2018年調査と比較した。図48にそれぞれの結果から、「不十分」と「やや不十分」の合計選択率を示す。

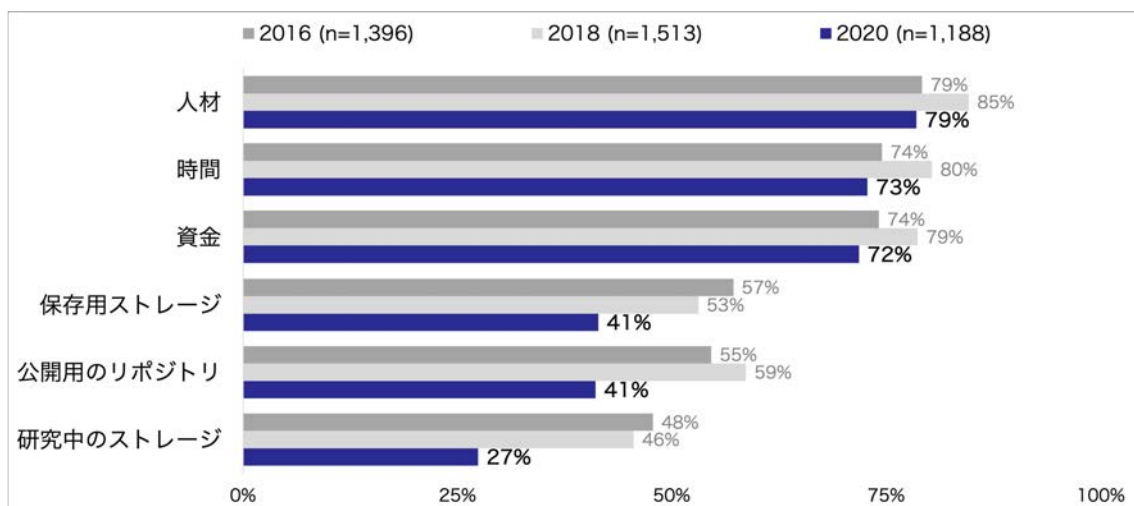


図 48 データ公開資源の不足度 (2016/18/20年)

人材、時間、資金については、2016年、2018年、2020年調査において、ほとんど変化がみられなかった。データ公開を実践している回答者についても、資源が不足している状

況がうかがえる。一方、保存用ストレージ、公開用のリポジトリ、研究中のリポジトリについては、やや不足感が低減している傾向がみられた。

### (3) カレントデータの公開に対する懸念

データ公開に対する懸念の強さを確認するために、8項目を挙げて「Q29. 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 54）。

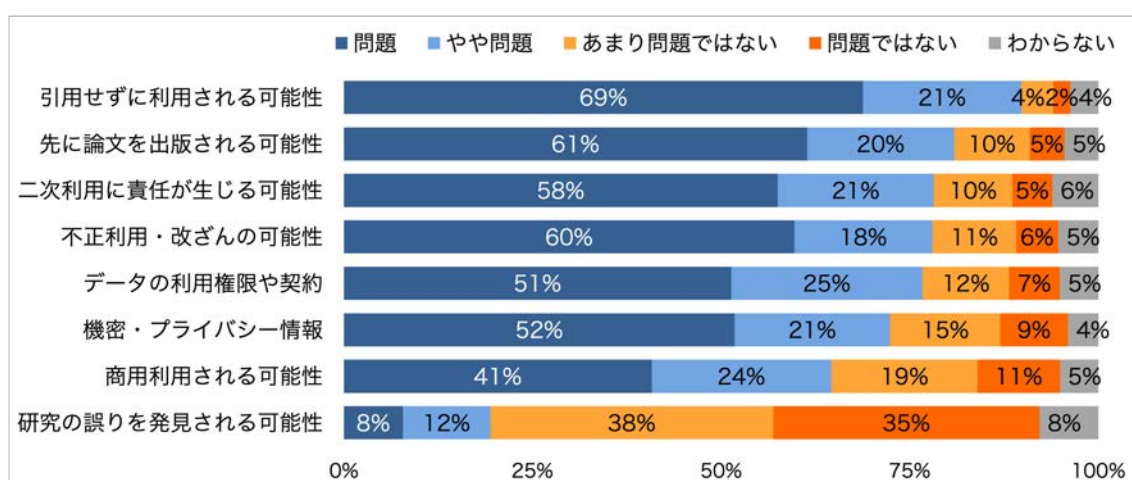


図 49 カレントデータの公開に関する懸念 (n=1,188)

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 89.8%）であった。次いで「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 80.9%）、「二次利用に関して責任が生じる可能性」（同 78.2%）、「不正利用・改ざんの可能性」（同 78.0%）、「データの利用権限や契約」（同 76.6%）の順に懸念が強かった。「研究の誤りを発見する可能性」（同 19.4%）は懸念をもつ回答者が比較的少なく、「問題ではない」（35.3%）と「あまり問題ではない」（37.5%）を選択する回答者の比率が高かった。

資源の状況について経年変化を確認するために、2016/2018 年調査と比較した。図 50 にそれぞれの結果から、「問題である」と「やや問題である」の合計選択率を示す。

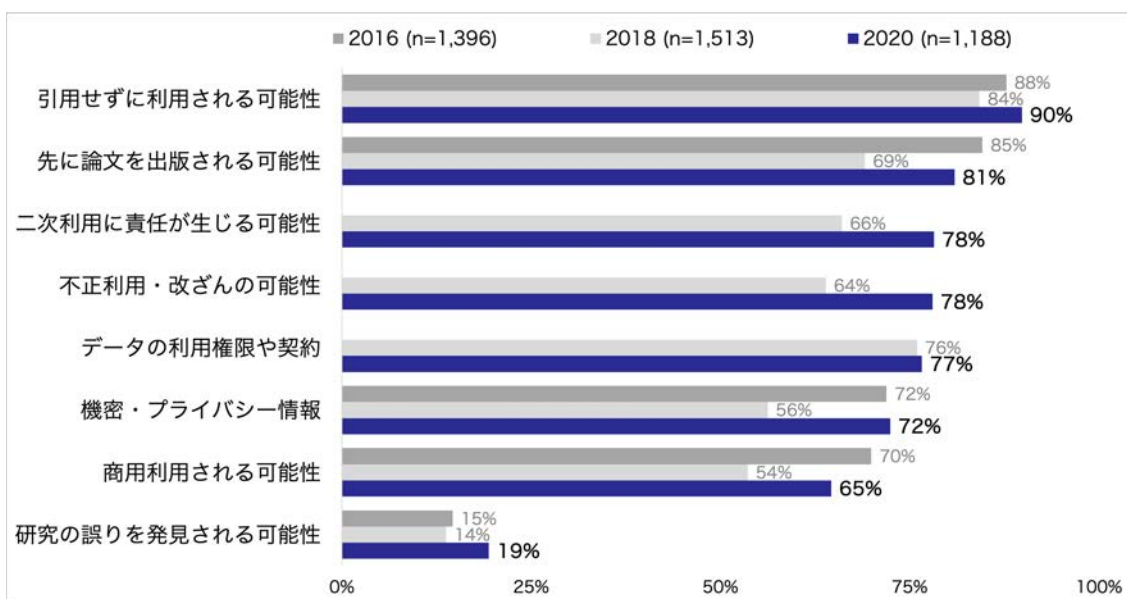


図 50 データ公開に対する懸念（2016/18/20 年）

データ公開に対する懸念は、2016 年調査から 2018 年調査にかけてはやや低減していたものの、2020 年調査では再び強くなっており、全体としてあまり変化が見られなかった。懸念が強い項目の順位はほとんど変化がなく、「データの利用権限や契約」のみ、やや順位が下がっていた（2018 年調査は 9 項目中 2 位、2020 年調査は 8 項目中 5 位）。「研究の誤りを発見される可能性」は 3 調査とも強い懸念を抱く回答者の比率が低かったが、2020 年調査ではやや高まっていた。

#### (4) データ公開のネガティブな影響

2016/2018 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、「Q21. 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、データ公開経験を有する回答者 567 名のうち 89 名から回答が得られた。うち、4 名は質問の趣旨とは異なる回答であったため除外して、85 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 33 名であり、52 名は特に問題が起きていないという旨を記述していた。つまり、データ公開経験を有する回答者（567 名）のうち、何らかの問題が起きたと記述していたのは 5.8%（33 名）であった（図 51）。なお、記述がなかった 478 名の中には、問題が起きたことがあるものの記述しなかった回答者が含まれている可能性もある。



図 51 データ公開によって問題が起きた経験の有無 (n=567)

表 11 に 33 名の回答の内容をタグ付けして集計し、11 項目に分類した結果を示す。

表 10 データ公開によるデメリット

内容	人数	比率
問い合わせ等への対応	6	18.2%
誤用された	5	15.2%
引用せずに利用された	4	12.1%
データの権利に関する問題	4	12.1%
懸念が生じた	3	9.1%
ミスを発見した	2	6.1%
公開に手間がかかる	2	6.1%
先取権の喪失	2	6.1%
長期公開・保存のコストがかかる	2	6.1%
不正アクセス	2	6.1%
情報流出	1	3.0%
合計	33	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」(18.2%)であり、研究者、企業、一般の方から問い合わせがあり、労力を割かれる、迷惑メールが届く、不適切なコメントが送られてくるといった記述がみられた。2位は「誤用された」(15.2%)であり、“データ作成者の意図しない使い方をした例があった”、“マスコミからの問い合わせで、先方の都合よい解釈をされることがある”といった記述がみられた。3位は「引用せずに利用された」と「データの権利に関する問題」(12.1%)であった。引用せずに利用されたことについては、謝辞への記載もなかった(2名)という記述や、“公開に際しては、doi等の番号をつけて、データといえどもきちんと引用して

欲しい。そのデータを取得するための労力をリスペクトする仕組みは必要”という指摘がみられた。データの権利に関する問題として、特許の侵害、所属機関を移動した後の著作権の扱い、ジャーナルのコード公開ポリシーの変更により論文を取り下げた、“助成金との関係、特に商業利用やこれを起点とした起業などの制約について”といった記述がみられた。

5位は「懸念が生じた」(9.1%)であり、“論文投稿時に二重投稿とみなされる恐れ”や、情報の流出により競争相手を利することになるのではないかとといった記述がみられた。懸念は回答者の内面の問題であり、実際に起きたことではないが、研究や公開に影響を与えているため回答に含めた。

6位は「ミスを発見した」、「公開に手間がかかる」、「先取権の喪失」、「長期公開・保存のコストがかかる」、「不正アクセス」であり、それぞれ2名(6.1%)が記述していた。ミスの発見について、1名は修正しているものの、もう1名は修正できずにそのまま掲載され続けているとしていた。公開の手間について、1名は手順が煩雑、もう1名はリポジトリの使い勝手が良くないと指摘していた。長期公開・保存については、いずれもデータ作成者の退職による問題を指摘していた。

11位は「情報流出」であり、1名(3.0%)が記述していた。これは実際に起きた事実として、5位の「懸念が生じた」とは区別した。

本調査の結果を2018年調査と比較するために、表11に各年の結果を回答者が多い順に示した。

表 11 データ公開によるデメリット (2018/20年)

2018年調査	人数	2020年調査	人数
問い合わせ等への対応	17	問い合わせ等への対応	6
引用せずに利用された	14	誤用された	5
先取権の喪失	9	引用せずに利用された	4
誤用された	7	データの権利に関する問題	4
更新のコストがかかる	2	懸念が生じた	3
徒労感	2	ミスを発見した	2
不正アクセス	1	公開に手間がかかる	2
商用利用された	1	先取権の喪失	2
研究者以外に利用された	1	長期公開・保存のコストがかかる	2
		不正アクセス	2
		情報流出	1
合計	54	合計	33

2018年調査、本調査ともに「問い合わせ等への対応」が1位であった。以下、「引用された」や「誤用された」が上位であるのも共通していた。2018年調査で指摘されていたが、本調査では指摘されていなかった項目として、「更新のコストがかかる」と「徒労感」がある。徒労感とは、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であった。



## 2.6 データ公開のインセンティブ

### (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、「Q20. 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、データ公開経験を有する回答者（567名）のうち168名から回答が得られた。うち、8名は質問の趣旨とは異なる回答であったため除外して、160名の回答を分類して集計した。

その結果、何らかの良い結果が得られたと記述していたのは22.9%（130名）であった（図52）。なお、記述がなかった399名の中には、何らかの良い結果が得られたものの記述しなかった回答者が含まれている可能性もある。

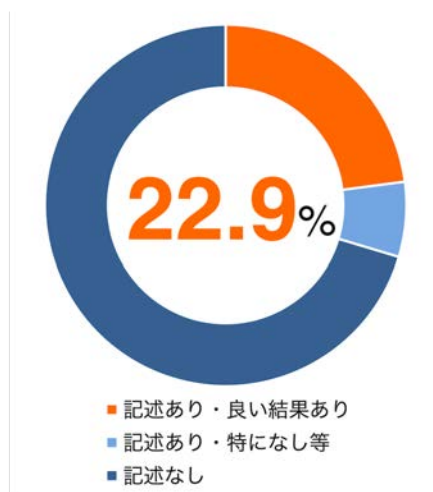


図 52 データ公開によって良い結果が得られた経験の有無（n=567）

表 12 に 7 項目に分類した集計結果を 2018 年調査の結果とともに示す。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 152 件となっている。

表 12 データ公開によって得られた良い結果 (2018/20 年)

内容	2018 年		2020 年	
	件数	比率	件数	比率
研究上の利点	104	40.6%	54	35.5%
研究・データ・研究者のビジビリティ向上	66	25.8%	41	27.0%
科学・分野の進展	27	10.5%	27	17.8%
人とのつながり	26	10.2%	14	9.2%
評価	11	4.3%	7	4.6%
個人的な利点	9	3.5%	6	3.9%
その他	13	5.1%	3	2.0%
合計	256	100.0%	152	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」(35.5%)、次いで「研究・データ・研究者のビジビリティ向上」(27.0%)、「科学・分野の進展」(17.8%)、「人とのつながり」(9.2%)、「評価」(4.6%)、「個人的な利点」(3.9%)の順であった。2018 年調査と順位に変化はなかったものの、「科学・分野の進展」は 7.3 ポイント増加していた。一方、「研究上の利点」は 5.1 ポイント減少していた。具体的な内容を表 13 に示す。

表 13 データ公開によって得られた良い結果の詳細 (n=152)

内容と詳細
1. 研究上の利点 共同研究の契機 (40)、研究の進展 (6)、研究の信頼性の向上 (5)、補足資料(2)、先取権を証明できた (1)
2. 研究・データ・研究者のビジビリティ向上 引用が増加した (20)、認知度向上 (19)、参照・引用された (2)
3. 科学・分野の進展 再利用・二次利用 (21)、社会貢献 (5)、オープンサイエンスへの貢献 (1)
4. 人とのつながり 交流の契機 (14)
5. 評価 研究・研究者・組織の評価(5)、業績 (2)
6. 個人的な利点 成果共有の手間の軽減 (6)
7. その他 教育 (3)

( )内の数字は人数を示す。

2018年調査と比較すると、再利用・二次利用されたことを挙げる回答が9名（256件の回答比で3.5%）から21名（152件の回答比で13.8%）まで10.3ポイント増加していた。新たな回答としては、“論文の補足資料として、より分かりやすく説明できたと思われる”など、補足資料の公開についての言及が2件みられた。

## (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「Q24. 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54に結果を示す。

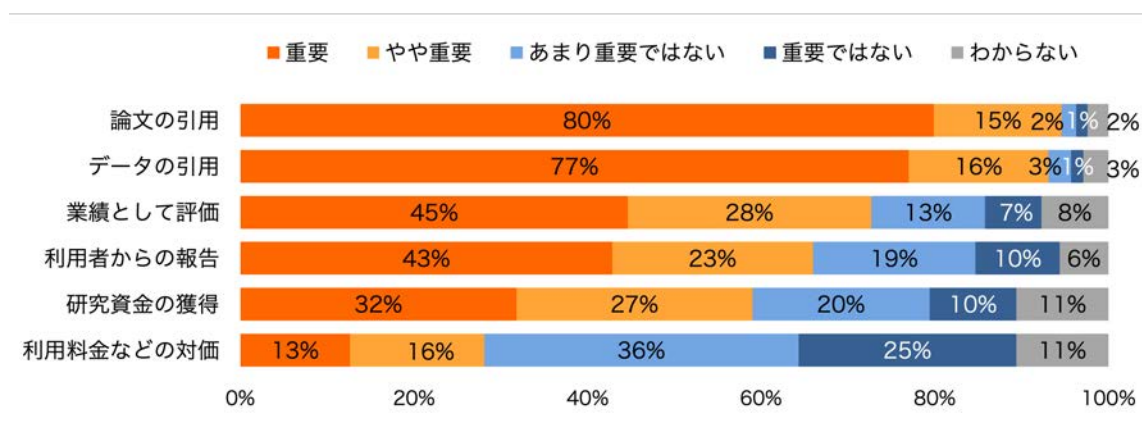


図 53 データ公開によるインセンティブの重要度 (n=1,180)

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94.5%）、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93.0%）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72.6%）、「利用者からの報告」（同65.9%）、「研究資金の獲得」（同59.2%）であった。一方、「利用料金などの対価」（同28.1%）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。

データ公開によるインセンティブの重要度の順位は、2018年調査と同様であった（図54）。回答者の多くが重要であると考えている上位5件については、「重要である」と「やや重要である」の合計が増加しており、「利用料金などの対価」についてはポイントが減少していた。また、「わからない」という回答が増えている項目もみられ、「利用料金などの対価」（10.6%）は5.0ポイント、「研究資金の獲得」（10.6%）は4.4ポイント増加していた。

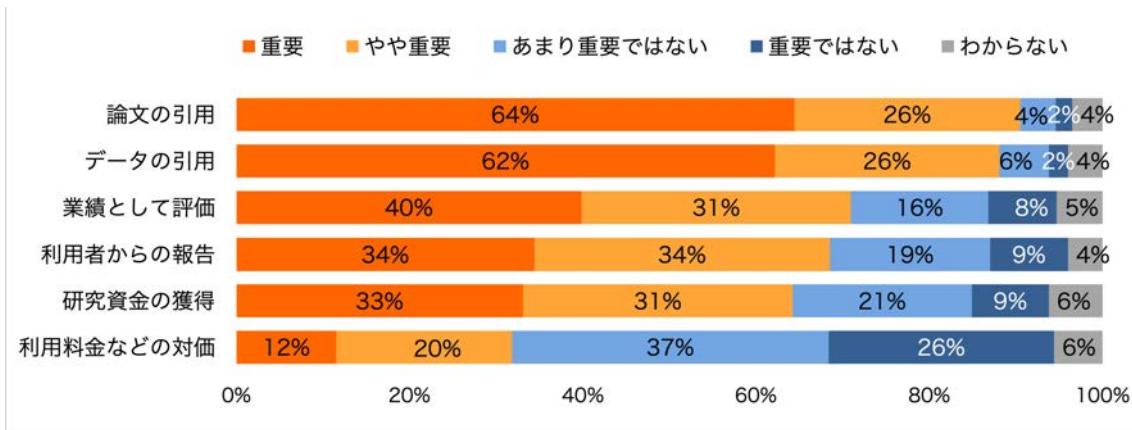


図 54 データ公開によるインセンティブの重要度（2018年：n=1,513）

2018年調査、及び2020年調査において、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最も重要視されていた。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」である点も共通していた（図49）。研究者は、引き続き引用を重要視していると推察される。データを公開している論文は引用が増加することを示す複数の研究がある<sup>15</sup><sup>16</sup>ことから、データ公開の推進においては、引用の慣習化や徹底が重要であると考えられる。

### (3) コミュニティや研究機関によるデータ公開に対する評価【新規】

データ公開に対する周囲の評価の状況を明らかにするために、「Q25.研究データを公開することは、ご自身が所属するコミュニティや機関で評価されていると思えますか？」と尋ねた。選択肢は、「評価されていると思う」、「ある程度評価されていると思う」、「ほとんど評価されていないと思う」、「まったく評価されていないと思う」の4件法として、「わからない」と「その他」も設けた。図55に結果を示す。

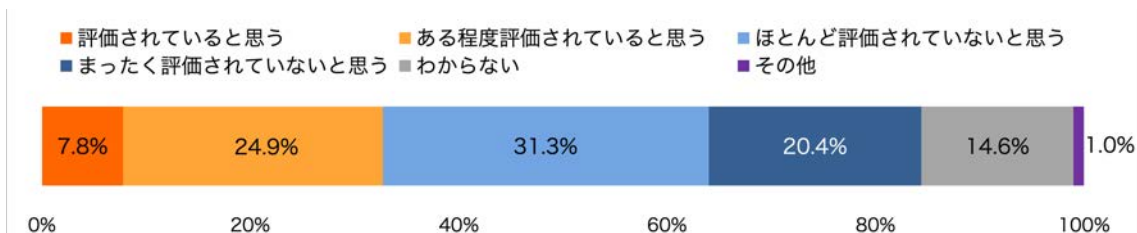


図 55 コミュニティや機関によるデータ公開に対する評価（n=1,182）

「評価されていると思う」は7.8%、「ある程度評価されていると思う」は24.9%であり、約1/3の回答者（32.7%）が評価されていると考えていることがわかった。一方、「ほとんど評価されていないと思う」は31.3%、「まったく評価されていないと思う」は20.4%であり、過半数の回答者（51.7%）は評価されていないと考えていることが明らかになった。「わからない」を選択した回答者は14.6%であった。

「その他」は12名が選択していた。これらのうち3名は“規範であり、評価云々の問題ではない”など、データ公開は当然のことであると指摘していた。一方、“オープンデータに非常に積極的に重要視している研究者とオープンデータの動向や重要性に気づいていない研究者の2極化”がみられるという指摘もあった。

また、公開と評価を切り離すという認識もみられ、“必要に応じて要求があればデータを要求者に公開するというのが一般的であり、公開そのものに評価という概念はない”、“評価されるから公開するのだとは思わない。必要・需要があればその都度公開すればよい”といった記述がみられた。そもそも、公開そのものが禁止されているといった記述もみられた(3名)。

評価そのものについて、“データの精度を測る指標がないため、すでに公開されているオープンデータ等と比較すると評価されにくい”という指摘や、“論文として公表したデータと査読を受けていないデータが同じ評価を受けるとは限りません”など、データにより評価が異なるという指摘もみられた。

#### (4) 回答者自身によるデータ公開に対する評価【新規】

回答者自身がデータ公開を評価しているかどうかを明らかにするために、「Q26.ご自身が指導する研究者(大学院生やポスドクなど)が研究データを公開していることを、評価していらっしゃいますか?」と尋ねた。選択肢は、「高く評価している」、「やや評価している」、「あまり評価していない」、「まったく評価していない」の4件法として、「わからない」、「指導する研究者がいない」、「その他」も設けた。図56に結果を示す。

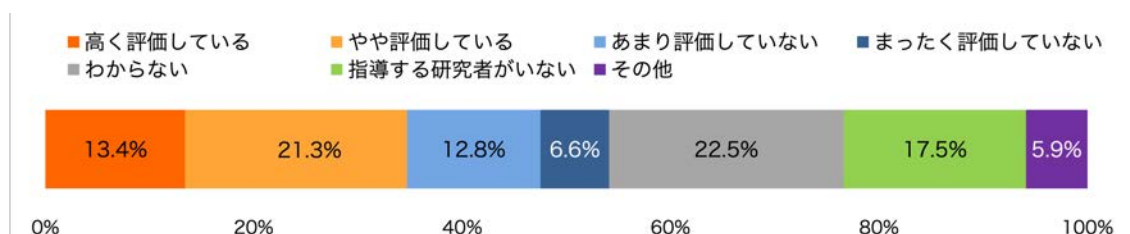


図 56 回答者自身によるデータ公開に対する評価 (n=1,180)

「高く評価している」は13.4%、「やや評価している」は21.3%であり、合計34.7%の回答者が評価していることが明らかになった。「あまり評価していない」は12.8%、「まったく評価していない」は6.6%であり、合計19.4%の回答者は評価していないことがわかった。「わからない」を選択した回答者は22.5%、そもそも「指導する研究者がいない」回答者は17.5%であった。「指導する研究者がいない」と「その他」を除く比率については後述する。

「その他」は70名が選択しており、うち「データ公開経験がない」という回答者が53名であった。データを公開することは当然であり、特に評価はしないという意見が3名の回答にみられた。たとえば、“取得した遺伝子データは公開することが大前提なので、公開設定をしていない場合に注意することはありますが、公開を評価することはありません”と

述べられていた。

また、公開すること自体ではなく、その内容によって評価するという旨の意見もみられた。たとえば、“公開している研究者はいないが、公開するという場合、公開そのものを評価するというよりは、公開できるほど価値がある、という考え方に評価する可能性はある”、“既に出版された論文等のデータの補足となるものの公開については評価する”といった記述がみられた。なお、データ公開自体が必要ないという意見もみられた。

データ公開に対する評価について、コミュニティや機関による評価と回答者自身による評価を比較するために、「その他」と「指導する研究者がない」を除いて再集計した。結果を図 57 に示す。

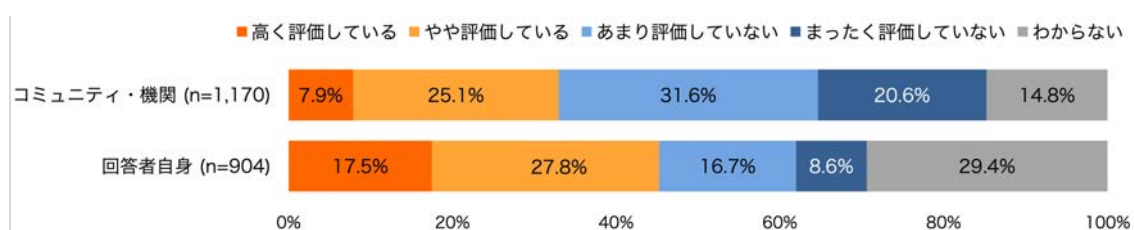


図 57 データ公開に対するコミュニティ・機関と回答者自身の評価

コミュニティや機関よりも、回答者自身の方が「評価している」という回答者の比率が高かった。「高く評価している」と「やや評価している」の合計は、コミュニティや機関が 33.0%であったのに対して、回答者自身は 45.2%であった。「評価していない」という回答者の比率はコミュニティや機関の方が回答者自身よりも高く、「あまり評価していない」と「まったく評価していない」の合計は前者が 52.2%、後者が 25.3%であった。「わからない」とする回答者は、コミュニティや機関が 14.8%、回答者自身が 29.4%といずれも高かった。

## 2.7 研究データ管理（RDM）に対する認識

### (1) RDM に関するリテラシーへの関心

研究データ管理（Research Data Management, RDM）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、「特に知りたい項目はない」を示した。

2016/2018 年調査では、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」を示し、それ以外の項目を 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。しかし、2020 年調査に用いたアンケートシステムでは排他的選択肢の設定ができなかったため、「特に知りたい項目はない」と別の選択肢を同時に選んでいる回答者もみられた。そこで、「特に知りたい項目はない」を単独で選んだ回答者のみを

「関心なし」とみなし、その他の選択肢を同時に選んだ回答者は「関心あり」とみなして、関心の有無を集計した。図 58 に集計結果を示す。

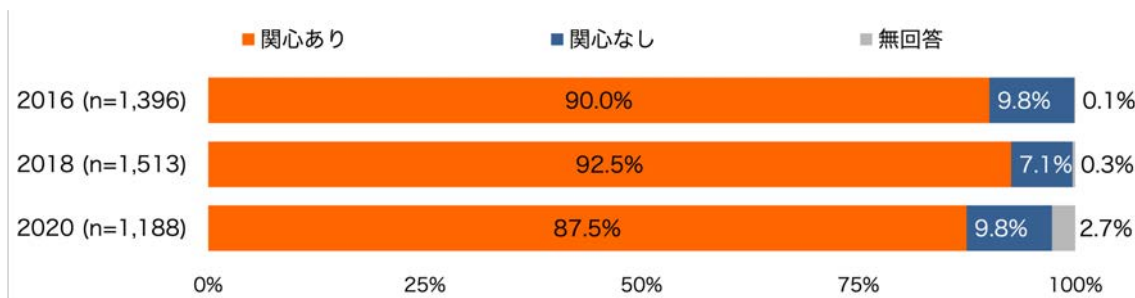


図 58 研究データ管理（RDM）に関するリテラシーへの関心の有無

RDM に関する項目に関心をもつ回答者の比率は 87.5%、「なし」は 9.8%、無回答は 2.7%であった。2018 年調査よりも関心をもつ回答者の比率が 5 ポイント減少しているものの、2016/2018 年調査と比較してそれほど変化はみられなかった。

## (2) RDM に関してより詳しく知りたい項目

図 59 に、RDM に関してより詳しく知りたい項目の集計結果を 2016/2018 年調査の結果とともに示す。配列は、2020 年調査において選択率が高かった公開方法の順とした。

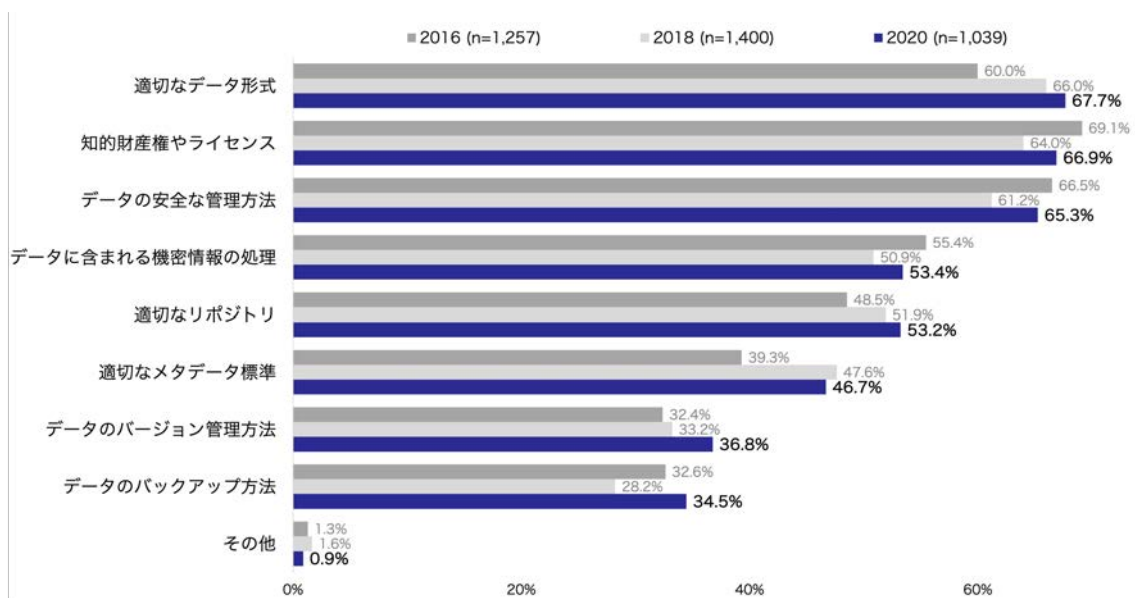


図 59 研究データ管理（RDM）に関して知りたい項目（2016/18/20 年：複数回答）

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」（67.7%）、次いで「知的財産権やライセンス」（66.9%）、「データの安全管理方法」（65.3%）、「データに含まれる機密情報の処理」（53.4%）の順であった。

2016/2018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様であった。選択率で比較すると、「適切なデータ形式」、「適切なりポジトリ」、「データのバージョン管理方法」は3回連続でポイントが増加していた。

「その他」には9件の回答があり、“非ホワイト国への技術提供（指導）にあたるかどうか”、“参加者から取得したデータを公開する際の倫理的配慮事項（インフォームドコンセント）”といった記述がみられた。また、データを整備・公開しても評価されないためメリットがない（2名）、データを公開することで研究の妨げになるのではないかと（2名）といったコメントもみられた。

### (3) RDM プロセスの依頼

RDMのプロセスを研究支援者に依頼したいかどうかを明らかにするために、「Q31. カレントデータの整備・公開・保存プロセスを、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に依頼したいと思われますか?」と尋ねた。図 60 に集計結果を示す。

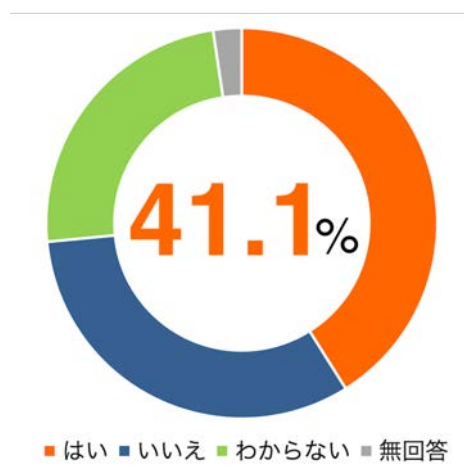


図 60 研究データ管理（RDM）プロセスの依頼意思（n=1,188）

RDM プロセスを研究支援者に依頼したいと考えている回答者は 41.1%、考えていない回答者は 32.4%、わからないを選択した回答者は 24.2%、無回答は 2.4%であった。

続いて、具体的に依頼したいと考える RDM プロセスを明らかにするために、Q31 で「はい」を選択した回答者を対象として、「Q32. 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に依頼したいと思われる項目を全てお選び下さい」と複数選択方式で尋ねた。図 61 に集計結果を示す。



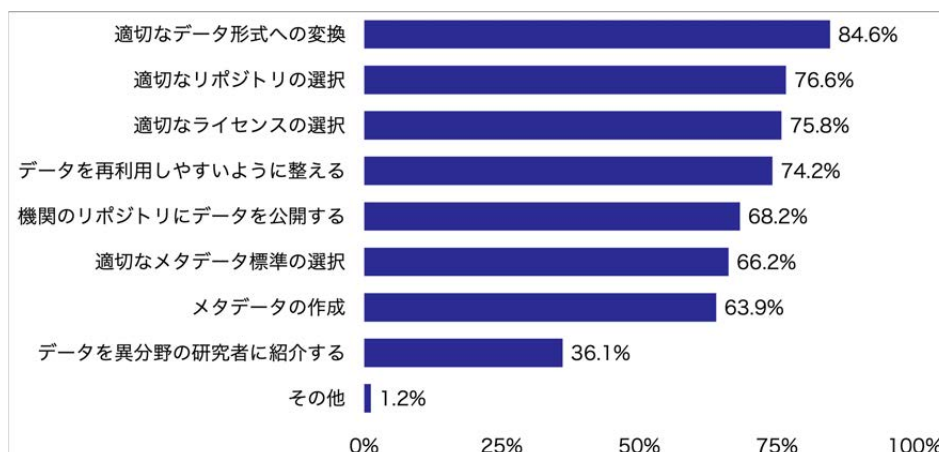


図 61 研究データ管理（RDM）に関して依頼したい項目（n=488：複数回答）

RDM プロセスを依頼したい項目のうち、選択率が最も高かったのは、「適切なデータ形式への変換」（84.6%）、次いで「適切なりポジトリの選択」（76.6%）、「適切なライセンスの選択」（75.8%）、「データを再利用しやすいように整える」（74.2%）の順であった。「データを異分野の研究者に紹介する」（36.1%）以外の項目はすべて 6 割以上の回答者が選択していた。

「その他」には 8 名による記述があり、具体的に“研究プロジェクトの成果に関する動画編集”、“二次利用された際の倫理的配慮事項のクリア”が挙げられていた。また、“研究以外の作業に時間を取られたくない”（3 名）、“公開に関する全ての作業”といった記述がみられた。

2016/2018 年調査では、これらのプロセスのうち専門性が必要であると考えられる項目を選択していただいた。そして選択率が低い項目、すなわち多くの回答者が「専門性が必要であるとは考えていない項目」を明らかにして、第三者による支援の可能性について検討した。結果を図 62 に示す。

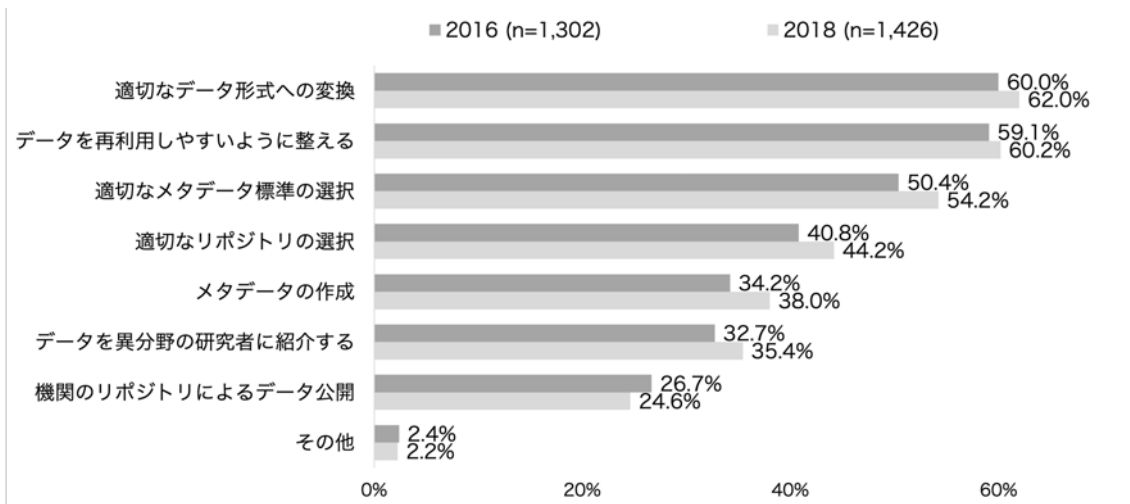


図 62 データ整備・公開に専門性を必要とする項目（2016/18年：複数回答）

2016/2018年調査では、過半数の回答者が「適切なデータ形式への変換」、「データを再利用しやすいように整える」、「適切なメタデータ標準の選択」には専門性が必要であると考えていることがわかった。一方、「機関リポジトリによるデータ公開」、「データを異分野の研究者に紹介する」、「メタデータの作成」は選択率が低かったため、第三者による支援に適しているのではないかと考察した。

しかし本調査によれば、専門性が高いと考えられていたプロセスは、図書館員やデータキュレーターに依頼したいと考える回答者の比率も比較的高かった。以上の結果から、第三者によるRDMに関する支援について検討する際には、専門性の高さや研究者のニーズの両方を勘案する必要があることが示唆された。

### 3. まとめ

---

本調査は、日本の研究者によるデータと論文の公開状況や認識を明らかにするために、2016年、2018年に科学技術専門家ネットワークを対象として実施したウェブ質問紙調査の後続調査である。3回目となる本調査では、大学、企業、公的機関・団体に所属する研究者1,349名（回答率70.5%）による回答を分析した結果を2016/2018年調査や先行研究の結果と比較した。

データの公開率は、2016年調査が51.0%、2018年調査が51.9%であったが、2020年調査では前回から7.2ポイント減少して44.7%であった。論文のOA率は、2016年調査が70.9%、2018年調査が78.0%、2020年が80.1%であり、2.1ポイント増加していた。

分野別のデータ公開率を確認すると、地球科学（70.2%）から工学（27.7%）まで42.5ポイントの差がみられた。論文のOA率も地球科学（93.0%）から工学（72.3%）まで20.7ポイントの差がみられたものの、データ公開ほど大きな差ではなかった。データ公開率とOA率には有意な相関がみられなかった。つまり、データ公開率が高い（低い）分野は論文の公開率も高い（低い）とはいえなかった。

データの公開理由は1位が「論文を投稿した雑誌がOA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、2016/2018年調査とは1位と2位の順位が逆転していた。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、2016/2018年調査と同様であった。データ、論文のいずれも雑誌のポリシーが主たる理由であった。

データの未公開理由は1位が「論文を投稿したい雑誌がOAではないから」、2位が「業績にならないから」であった。論文の未公開理由は1位が「資金がないから」、2位が「論文を投稿した雑誌のポリシーではないから」であり、2016/2018年調査とは1位と2位の順位が逆転していたものの、この2つの理由が突出しているという点は同様であった。未公開理由が解決された場合の公開意思について、データは「はい」が約3割、「いいえ」が約4割、「わからない」が約4割であり、「わからない」の選択率が減少していた。論文の公開意思は「はい」が76.1%、「いいえ」が7.3%、「わからない」が16.6%であり、データと比較して「はい」の選択率が高いものの、経年的な変化はそれほどみられなかった。

データの提供（共有）経験は71.2%が有していた。分野別にみると、地球科学（93.0%）から心理学（57.9%）まで35.1ポイントの差がみられた。なお、分野別のデータ提供経験はデータ公開経験と正の相関がみられ、よくデータを提供している分野はよくデータを公開しているという傾向がみられた。

公開データの入手経験は69.7%が有していた。分野別にみると、計算機科学（91.2%）から心理学（47.4%）まで43.8ポイントの差がみられた。分野別のデータ入手経験はデータ公開経験及びデータ提供経験と正の相関がみられ、よくデータを入手している分野は、よくデータを公開している可能性や、よくデータを提供している傾向がみられた。

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 32.0%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、2018 年調査、本調査ではこれらの順位が徐々に低下しており、データの再利用における問題の順位が徐々に高まっていた。

データマネジメントプラン（DMP）の作成経験を有する回答者は 20.8%であり、2018 年調査と比較して 2.1 ポイント増加していた。分野別にみると、計算機科学（31.6%）から数学（7.7%）まで 23.9 ポイントの差がみられた。作成経験がある DMP の 1 位は JST の DMP、2 位は所属機関の DMP であった。また、DMP の作成理由の 1 位は「助成機関が要求しているから」、2 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、2018 年調査と同様であった。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ものの、前回調査と比べると 3.7 ポイント減少していた。

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認識されていた。研究データ管理（RDM）のプロセスを図書館員やデータキュレータに依頼したいと思うかどうかを尋ねたところ、「はい」と回答したのは 41.1%であった。依頼したい項目の 1 位は「適切なデータ形式への変換」、2 位は「適切なりポジトリの選択」、3 位は「適切なライセンスの選択」であった。2016/2018 年調査では、同じ項目について専門性が必要であるかどうかを尋ねた結果、1 位は「適切なデータ形式への変換」であった。つまり、「適切なデータ形式への変換」には専門性が高いと考えられており、かつ、第三者に依頼したいとも考えられていた。

データ公開に対する懸念も依然として強く、1 位は「引用せずに利用される可能性」、2 位は「先に論文を出版される可能性」、3 位は「二次利用に責任が生じる可能性」であった。一方で、実際にデータを公開している回答者に問題が起きたかどうかを自由記述方式で尋ねた結果、具体的に記述していた回答者は 5.8%であり、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（ただし、問題が起きたことがあるが記述しなかった回答者が存在する可能性はある）。実際に、引用せずに利用された経験を有する回答者は 4 名であった。データ公開のポジティブな影響についても自由記述方式で尋ねた結果、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挙げられていた。データ公開によって起きた事柄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への理解が深まると考えられる。

新規の調査項目として、データ公開を回答者が所属しているコミュニティや機関、及び本人が評価しているかどうかを尋ねた。データ公開を「高く評価している」、「やや評価している」の合計は、コミュニティや機関が 33.0%、回答者自身が 45.2%であった。「あまり

評価していない」、「まったく評価していない」の合計は、コミュニティや機関が 52.2%、回答者自身が 25.3%であった。「わからない」の選択率も高く、コミュニティや機関が 14.8%、回答者自身が 29.4%であった。データを公開することが評価につながるようになるのかどうか、今後も継続的に調査していきたい。

最後に本調査の課題を述べる。調査対象である科学技術専門家ネットワークは、毎年入れ替えがあるため、同一回答者によるパネル調査ではない。また、一部の分野については、回答者数が十分に確保できなかった。2016/2018年調査と本調査ではウェブアンケートシステムが変更されたため、質問の方法を変更せざるを得なかった。このことが回答に影響を及ぼした可能性がある。今後は、日本の研究者の実態をより正確に捉えるために、さらに広範な調査を実施する方法を模索したい。そしてオープンサイエンスの進展状況と課題、及び支援の可能性を検討するための基礎資料を提供することによって、証拠に基づく政策立案（EBPM）に貢献したい。

## 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。

2021年4月から、大森悠生氏（筑波大学大学院 人間総合科学学術院 人間総合科学研究群 情報学学位プログラム）に RA（リサーチアシスタント）として調査の集計や分析等にご尽力いただきました。厚くお礼申し上げます。大森氏は、2016/2018年調査と本調査の結果を二次分析して、オープンアクセスに対する研究者の認識の経年変化を明らかにされました。本調査から新たな知見を導き出して下さったことに対して、重ねて感謝申し上げます。

## 参考文献

---

- <sup>1</sup> 内閣府. 第6期科学技術・イノベーション基本計画. 2021.  
<https://www8.cao.go.jp/cstp/kihonkeikaku/index6.html>, (accessed 2021-10-18).
- <sup>2</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>, (accessed 2021-10-18).
- <sup>3</sup> 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>, (accessed 2021-10-18).
- <sup>4</sup> 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](http://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00003152-00000079-0021), (accessed 2021-10-18).
- <sup>5</sup> 池内有為, 林和弘. 研究データ公開と論文のオープンアクセスに関する実態調査 2018. 文部科学省科学技術・学術政策研究所, 2020, NISTEP RESEARCH MATERIAL No.289, 96p. <https://doi.org/10.15108/rm289>, (accessed 2021-10-18).
- <sup>6</sup> 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—オープンサイエンスの進展状況と課題—. STI Horizon. 2020, vol. 6, no. 1, p. 37-42.  
<https://doi.org/10.15108/stih.00207>, (accessed 2021-10-18).
- <sup>7</sup> Open Science Definition. FOSTER. <https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>, (accessed 2021-10-18).
- <sup>8</sup> 経済産業省. AI・データの利用に関する契約ガイドライン 1.1 版. 2019, 28p.  
<https://www.meti.go.jp/press/2019/12/20191209001/20191209001-1.pdf>, (accessed 2021-10-18).
- <sup>9</sup> 池内有為, 林和弘. プレプリントの利活用と認識に関する調査. 文部科学省科学技術・学術政策研究所, 2021, NISTEP RESEARCH MATERIAL No. 301, 94p.  
<https://doi.org/10.15108/rm301>, (accessed 2021-10-18).
- <sup>10</sup> “科学技術専門家ネットワーク”, 科学技術・学術政策研究所.

<http://www.nistep.go.jp/activities/st-experts-network>, (accessed 2021-10-18).

- <sup>11</sup> Hrynaszkiewicz, I.; Harney, J; Cadwallader, L. A survey of researchers' needs and priorities for data sharing. *Data Science Journal*, 2021, vol. 20, no. 1, p.31.  
<https://doi.org/10.5334/dsj-2021-031>, (accessed 2021-10-18).
- <sup>12</sup> 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>, (accessed 2021-10-18).
- <sup>13</sup> Tenopir, C.; Rice, N. M. et al. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, 2020, vol. 15, no. 3, e0229003.  
<https://doi.org/10.1371/journal.pone.0229003>, (accessed 2021-10-18).
- <sup>14</sup> 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*. 2019, no. 341, p. 12-14. <https://doi.org/10.11501/11359093>, (accessed 2021-10-18).
- <sup>15</sup> Colavizza, G.; Hrynaszkiewicz, I.; Staden, I.; Whitaker, K.; McGillivray, B. The citation advantage of linking publications to research data. *PLOS ONE*, 2020, vol. 15, no. 4, e0230416. <https://doi.org/10.1371/journal.pone.0230416>, (accessed 2021-10-18).
- <sup>16</sup> Piwowar, H. A.; Vision, Todd. J. Data reuse and the open data citation advantage. *PeerJ*. 2013, vol. 1, e175. <https://doi.org/10.7717/peerj.175>, (accessed 2021-10-18).



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、学術論文やその根拠となるデータを公開する、あるいは、研究データそのものの利活用を促進して新たな知識やイノベーションを創出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。日本においても第5期科学技術基本計画（2016～2020年度）においてオープンサイエンスの推進が明示された後、2018-2020年に毎年閣議決定された統合イノベーション戦略においては、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・公開等を促進する施策が盛り込まれております。さらに2020年に入って、COVID-19など世界的な緊急課題に対応するべく、プレプリントや研究データの迅速な公開と共有が浸透しつつあり、データ駆動型科学の発展や科学研究のデジタル転換そのものが加速しております。これらの状況を勘案しながら、現在、第6期科学技術・イノベーション計画の策定が進んでおり、研究データの利活用は引き続き重要なトピックとして取り扱われる方向で検討が進んでおります。

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の提出や公開を要求する学術雑誌も増加しています。一方、研究データには分野を中心とする様々な特性があり、また、公開するためには人的・物的資源が必要であるため、その特性を踏まえた支援体制の構築は喫緊の課題であり、併せて、公開しようとする場合の障壁を明らかにして除去することも必要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。文部科学省科学技術・学術政策研究所（NISTEP）では2016年、2018年にも同様の調査を行い、その結果は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」というご回答も大変参考になりますので、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。

調査分析結果は、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。

2020年11月5日

文部科学省  
科学技術・学術政策研究所  
科学技術予測センター  
上席研究官 林 和弘  
客員研究官 池内有為

調査期間：2020年11月5日～11月15日

質問数：全33問 [(1)研究分野について、(2)論文の公開について、(3)研究データの提供について、

(4)公開データの入手について、(5)データマネジメントプラン (DMP) について、  
(6)研究データの公開について、(7)研究データの整備や公開について、(8)自由回答]

必須項目：23問 (回答により1～15問)

所要時間：5～20分程度

<アンケートサイトに関する問い合わせ>

科学技術予測センター 田中

stfc@nistep.go.jp

**【留意事項】**

- ・ アンケートは画面下の「次へ」をクリックして進めてください。
- ・ 前に戻る場合は、Webブラウザの「戻る」（「←」ボタン等）で戻ってください。
- ・ 回答を一時保存する機能はありません。回答を途中で中止した場合、回答者様のブラウザの仕様によっては回答状況が維持されず、再度最初の質問から回答いただく必要があります。
- ・ 回答中にページを再読み込み（リロード）した場合、リロードをされたページの回答は全て初期化されます。回答中のリロードはお控えください。
- ・ 回答データは全ての質問に回答した後、「登録」を押さない限り送信されないため、回答終了後忘れずに「登録」ボタンを押してください。なお、「登録」後の修正はできません。

本調査はオープンサイエンス（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。 **必須**

- (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]

Q2.現在、論文出版や学会等での口頭発表について行っていますか。 **必須**

- 行っている
- 行っていない【→Q33. へ】

次 へ

## 2. 論文の公開について

ご自身の論文のオープンアクセス（公開）状況についておうかがいします。  
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。

Q3. これまでに、論文をオープンアクセスにしたご経験はありますか？（Q4もご参照下さい） **必須**

- (1) はい
- (2) いいえ【→Q6. へ】
- (3) わからない【→Q8. へ】

Q4. 論文をオープンアクセスにした際に利用したサーバやサービスなどを全てお選び下さい。【複数選択可】 **必須**

- (1) オープンアクセスの雑誌（Scientific ReportsやPLOSなど）
- (2) 雑誌のオープンアクセスオプション（Springer Open Choiceなど）
- (3) 所属機関のリポジトリ
- (4) 学術系SNS（MendeleyやResearchGateなど）
- (5) 雑誌や学会等が論文をオープンアクセスにした
- (6) 個人や研究室のウェブサイト
- (7) その他【具体的に】

Q5. 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】【→回答後Q8. へ】 **必須**

- (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) 雑誌や学会等が論文をオープンアクセスにしたから
- (10) 特に理由はない
- (11) その他【具体的に】

Q6.論文をオープンアクセスに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】 **必須**

- (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) オープンアクセスではない発表方式の方が望ましいと思うから
- (10) 特に理由はない【→Q8. へ】
- (11) その他【具体的に】

Q7.Q6.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？

**必須**

- (1) はい
- (2) いいえ
- (3) わからない

次 へ

### 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データ、および研究成果そのものであるデータの両方を含みます。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。特定の人以上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
- (5) わからない

### 4. 公開データの入手について

これまでに、公開データをインターネットで入手したご経験についておうかがいします。「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。

Q9.これまでに、公開データを入手したご経験はありますか？ **必須**

- (1) はい
- (2) いいえ 【→Q13. へ】
- (3) わからない 【→Q13. へ】

Q10.公開データを入手した際に利用したサーバやサービスなどを全てお選び下さい。

**【複数選択可】 必須**

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



Q11.公開データを入手した際に、問題があると感じたご経験はありますか？ **必須**

- (1) はい
- (2) いいえ【→Q13.へ】
- (3) わからない【→Q13.へ】

Q12.公開データを入手した際に、問題だと感じたことがある項目を全てお選び下さい。【複数選択可】 **必須**

- (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】

次へ

## 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)

Q13. これまでに、DMPを作成したご経験はありますか？ **必須**

- (1) はい
- (2) いいえ 【→Q16. へ】
- (3) わからない 【→Q17. へ】

Q14. 作成したことがあるDMPの種類を全てお選び下さい。【複数選択可】 **必須**

- (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) 国外の助成機関による助成研究のDMP
- (5) 所属機関のDMP
- (6) 個人や研究グループのためのDMP
- (7) その他 [具体的に]

Q15. DMPを作成した理由として、あてはまるものを全てお選び下さい。【複数選択可】 【→回答後Q17. へ】 **必須**

- (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) 特に理由はない
- (6) その他 [具体的に]



Q16.DMPを作成していない理由として、あてはまるものを全てお選び下さい。【複数  
選択可】 **必須**

- (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) 特に理由はない
- (7) その他 [具体的に]

次へ

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データ、および研究成果そのものであるデータの両方を含みます。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、研究データを公開したご経験はありますか? **必須**

- (1) はい
- (2) 研究にデジタルデータを用いるが、公開したことはない【→Q22. へ】
- (3) 研究にデジタルデータを用いないので、公開したことはない【→Q33. へ】
- (4) わからない【→Q24. へ】

Q18. 研究データを公開した際に利用したサーバやサービスなどを全てお選び下さい。

**【複数選択可】 必須**

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

Q19. 研究データを公開した理由として、あてはまるものを全てお選び下さい。【複数

**選択可】 必須**

- (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) 特に理由はない
- (11) その他【具体的に】

Q20. 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。

Q21. 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【→回答後Q24. へ】

Q22.研究データを公開していない理由として、あてはまるものを全てお選び下さい。

【複数選択可】 **必須**

- (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) 公開できない情報が含まれるから（機密やプライバシー情報等）
- (11) 特に理由はない【→Q24. へ】
- (12) その他 [具体的に]

Q23.Q22.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？ **必須**

- (1) はい
- (2) いいえ
- (3) わからない

次 へ

Q24.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。

「データの引用」とは、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ことを指します。 **必須**

	重要ではない	あまり重要ではない	やや重要	重要	わからない
利用者からの報告	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
利用料金などの対価	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
データに紐づいた論文の引用	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
データの引用	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
業績として論文と同様に評価されること	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
研究資金獲得の選考要因に加味されること	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q25.研究データを公開することは、ご自身が所属するコミュニティや機関で評価されていると思われますか？

- (1) 評価されていると思う
- (2) ある程度評価されていると思う
- (3) あまり評価されていないと思う
- (4) まったく評価されていないと思う
- (5) わからない
- (6) その他 [具体的に]

Q26.ご自身が指導する研究者（大学院生やポスドクなど）が研究データを公開していることを、評価していらっしゃいますか？

- (1) 高く評価している
- (2) やや評価している
- (3) あまり評価していない
- (4) まったく評価していない
- (5) わからない
- (6) 指導する研究者がいない
- (7) その他 [具体的に]

### 7. 研究データの整備・公開・保存について

ここからは、研究データの整備や公開、保存に必要な支援について検討するための質問です。研究によって扱うデータの種類や量は異なる場合が多いと考えられるため、「論文などで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」についてお答えをお願いします（以下、「カレントデータ」と記します）。データ公開のご経験がない場合は、カレントデータのうち、論文等の根拠となる部分を公開すると想定してお答え下さい。

【例：研究1件のために質問紙調査とインタビュー調査を実施した場合】

<カレントデータ>

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

<データ公開>

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

Q27.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

Q28.カレントデータを整備・公開・保存しようとする場合、次の資源は十分に整っていますか？ あてはまるものをそれぞれお選び下さい。 **必須**

	十分	ほぼ十分	やや不十分	不十分	わからない
研究中のデータ用ストレージ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
データ公開用のリポジトリ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
研究終了後のデータ保存用ストレージ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
データの整備・公開のための時間	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
データの整備・公開のための人材	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
データの整備・公開のための資金	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(つづく)

Q29.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？ あてはまるものをそれぞれお選び下さい。【必須】

	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	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
機密・プライバシー情報	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
データの利用権限や契約	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
不正利用・改ざんの可能性	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
二次利用に関して責任が生じる可能性	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
引用せずに利用される可能性	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
研究の誤りを発見される可能性	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
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	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q30.カレントデータを整備・公開・保存する上で、より詳しく知りたいと思われる項目をお選び下さい。【複数選択可】

「メタデータ」とは、データの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のことを指します。また、メタデータにはさまざまなフォーマットや標準があります。（参考：Research Data Allianceによる分野別メタデータ標準の一覧 <https://rd-alliance.github.io/metadata-directory/standards/>）

- (1) 適切なデータ形式
- (2) 適切なリポジトリ
- (3) 適切なメタデータ標準
- (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) 特に知りたい項目はない
- (10) その他【具体的に】

Q31.カレントデータの整備・公開・保存プロセスを、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に依頼したいと思われますか？

- (1) はい
- (2) いいえ【→Q33. へ】
- (3) わからない【→Q33. へ】

Q32.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に依頼したいと思われる項目を全てお選び下さい。【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) 適切なメタデータ標準の選択
- (4) 適切なライセンスの選択
- (5) メタデータの作成
- (6) データを再利用しやすいように整える
- (7) 機関のリポジトリにデータを公開する
- (8) データを異分野の研究者に紹介する
- (9) その他【具体的に】

次 へ



8. 自由回答

Q33.論文や研究データの公開に関するご自身の研究分野の展望、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。

---

次 ^

## (2) 単純集計表

### 回答者の属性

表 14 回答者の所属機関

機関	人数	比率
大学	935	69.3%
公的機関・団体	249	18.5%
企業	155	11.5%
その他	10	0.7%
合計	1,349	100.0%

表 15 回答者の年齢層

年代	人数	比率
30 代以下	425	31.5%
40 代	659	48.9%
50 代	194	14.4%
60 代以上	70	5.2%
非開示	1	0.1%
合計	1,349	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 16 回答者の研究分野

研究分野	人数	比率
工学	419	31.1%
化学	217	16.1%
生物科学	213	15.8%
農学	130	9.6%
医学	96	7.1%
物理学	90	6.7%
コンピュータサイエンス	58	4.3%
地球科学	58	4.3%
社会科学	24	1.8%
心理学	19	1.4%
数学	13	1.0%
天文学	9	0.7%
人文学	3	0.2%
合計	1,349	100.0%

Q2. 現在、論文出版や学会等での口頭発表を行っていますか。\*

表 17 研究活動の状況

研究活動	人数	比率
行なっている	1,268	94.0%
行っていない	81	6.0%
合計	1,349	100.0%

## 2. 学術論文について

Q3. 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 18 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	1,016	80.1%
OA の論文はない	231	18.2%
わからない	21	1.7%
合計	1,268	100.0%

Q4. 論文をオープンアクセスにした際に利用したサーバやサービスなどを全てお選び下さい。[複数選択可] \*

表 19 論文の公開方法 (n=1,016 : 複数回答)

公開方法	人数	比率
オープンアクセスの雑誌 (Scientific Reports や PLOS など)	783	77.1%
雑誌や学会等が論文をオープンアクセスにした	331	32.6%
所属機関のリポジトリ	312	30.7%
雑誌のオープンアクセスオプション (Springer Open Choice など)	311	30.6%
学術系 SNS (Mendeley や ResearchGate など)	122	12.0%
個人や研究室のウェブサイト	66	6.5%
その他	29	2.9%

Q5. 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 20 論文を OA にした理由 (n=1,016 : 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	770	75.8%
研究成果を広く認知してもらいたいから	585	57.6%
雑誌や学会等が論文をオープンアクセスにしたから	272	26.8%
科学研究や成果実装を推進したいから	209	20.6%
所属機関のポリシーだから	171	16.8%
オープンアクセスに貢献したいから	133	13.1%
他の研究者からのリクエストに応じて	93	9.2%
助成機関のポリシー (助成条件) だから	46	4.5%
分野・コミュニティの規範だから	34	3.3%
特に理由はない	22	2.2%

Q6. 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 21 論文が OA ではない理由 (n=231 : 複数回答)

理由	人数	比率
資金がないから	133	57.6%
投稿したい雑誌がオープンアクセスではないから	93	40.3%
オープンアクセスではない発表方式の方が望ましいと思うから	26	11.3%
所属機関にオープンアクセス方針がないから	24	10.4%
ニーズがないと思うから	13	5.6%
助成機関のポリシー (助成条件) ではないから	10	4.3%
時間がないから	8	3.5%
リポジトリなどの公開手段がないから	6	2.6%
分野・コミュニティで推奨されていないから	2	0.9%
その他	8	3.5%
特に理由はない	10	4.3%

Q7. Q6 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 22 オープンアクセスの意思

OA 意思	人数	比率
はい	156	76.1%
いいえ	15	7.3%
わからない	34	16.6%
合計	205	100.0%

### 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 23 データ提供経験

頻度	人数	比率
よくある	120	9.5%
たまにある	420	33.1%
ほとんどない	403	31.8%
まったくない	314	24.8%
わからない	10	0.8%
無回答	1	0.1%
合計	1,268	100.0%

#### 4. 公開データの利用について

Q9. これまでに、公開データを入手したご経験はありますか？ \*

表 24 公開データの入手経験

入手経験	人数	比率
あり	884	69.7%
なし	352	27.8%
わからない	32	2.5%
合計	1,268	100.0%

Q10. 公開データを入手した際に利用したサーバやサービスなどを全てお選び下さい。  
[複数選択可] \*

表 25 公開データの入手方法 (n=884：複数回答)

入手方法	人数	比率
論文の補足資料	525	59.4%
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)	509	57.6%
個人や研究室のウェブサイト	344	38.9%
特定分野のリポジトリ・データアーカイブ (DDBJ や ICPSR など)	335	37.9%
学術系 SNS (Mendeley や ResearchGate など)	245	27.7%
コード共有サービス (GitHub など)	244	27.6%
データ共有サービス (figshare や zenodo など)	53	6.0%
その他	19	2.1%

Q11. 公開データを入手した際に、問題があると感じたご経験はありますか？ [複数選択可] \*

表 26 公開データ入手における問題の有無

問題	人数	比率
はい	283	32.0%
いいえ	537	60.7%
わからない	64	7.2%
合計	884	100.0%

Q12. 公開データを入手した際に、問題だと感じたことがある項目を全てお選び下さい。 [複数選択可] \*

表 27 公開データを入手する際の問題 (n=816：複数回答)

問題	人数	比率
データごとに品質が異なる	135	47.7%
データごとにフォーマットが異なる	130	45.9%
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	90	31.8%
著作者情報がよくわからない	88	31.1%
データの解釈や利用方法がよくわからない	81	28.6%
利用したいデータへのアクセス方法がよくわからない	79	27.9%
利用料金が必要	67	23.7%
利用者登録が必要	57	20.1%
リクエストから入手までに時間がかかる	45	15.9%
最新のデータを入手できない	36	12.7%
その他	8	2.8%



## 5. データマネジメントプラン（DMP）について

Q13. 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 28 DMP 作成経験の有無

DMP 作成経験	人数	比率
はい	20.8%	18.7%
いいえ	70.3%	76.1%
わからない	8.9%	5.2%
合計	1,268	100.0%

Q14. 作成したことがある DMP の種類を全てお選び下さい。 [複数選択可] \*

表 29 作成経験がある DMP (n=264 : 複数回答)

作成した DMP	人数	比率
科学技術振興機構（JST）による助成研究の DMP	94	35.6%
所属機関の DMP	88	33.3%
個人や研究グループのための DMP	80	30.3%
日本医療研究開発機構（AMED）による助成研究の DMP	27	10.2%
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	26	9.8%
国外の助成機関による助成研究の DMP	8	3.0%

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

表 30 DMP を作成した理由 (n=264 : 複数回答)

理由	人数	比率
助成機関が要求しているから	136	51.5%
所属機関が要求しているから	91	34.5%
効率的にデータを管理するため	62	23.5%
研究を行う際の習慣だから	61	23.1%
その他	14	5.3%
特に理由はない	1	0.4%

Q16. DMP を作成していない理由として、あてはまるものを全てお選び下さい。[複数選択可] \*

表 31 DMP を作成していない理由 (n=891：複数回答)

理由	人数	比率
DMP を知らなかったから	463	52.0%
所属機関から要求されていないから	350	39.3%
助成機関から要求されていないから	295	33.1%
DMP の必要性を感じないから	169	19.0%
DMP を作成する時間がないから	96	10.8%
その他	12	1.3%
特に理由はない	149	16.7%

## 6. 研究データの公開について

Q17. これまでに、研究データを公開したご経験はありますか？ \*

表 32 データ公開経験の有無

データ公開経験	人数	比率
はい	567	44.7%
研究にデジタルデータは用いない	80	6.3%
研究にデジタルデータは用いるが公開したことはない	593	46.8%
わからない	28	2.2%
合計	1,268	100.0%

Q18. 研究データを公開した際に利用したサーバやサービスなどを全てお選び下さい。  
[複数選択可] \*

表 33 データを公開した方法 (n=567:複数回答)

データ公開方法	人数	比率
論文の補足資料	307	54.1%
個人や研究室のウェブサイト	178	31.4%
特定分野のリポジトリ・データアーカイブ (DDBJ や ICPSR など)	165	29.1%
学術機関のリポジトリ・データアーカイブ (大学や NASA のリポジトリなど)	144	25.4%
学術系 SNS (Mendeley や ResearchGate など)	57	10.1%
コード共有サービス (GitHub など)	55	9.7%
データ共有サービス (figshare や zenodo など)	30	5.3%
その他	21	3.7%

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

表 34 データを公開した理由 (n=567:複数回答)

理由	人数	比率
論文を投稿した雑誌のポリシー (投稿規定) だから	307	54.1%
研究成果を広く認知してもらいたいから	298	52.6%
科学研究や成果実装を推進したいから	149	26.3%
他の研究者からのリクエストに応じて	117	20.6%
オープンデータに貢献したいから	99	17.5%
分野・コミュニティの規範だから	77	13.6%
所属機関のポリシーだから	63	11.1%
業績になる場合があるから	50	8.8%
助成機関のポリシー (助成条件) だから	21	3.7%
その他	17	3.0%
特に理由はない	11	1.9%

Q20. 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。

Q21. 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。

※Q20、21 の自由回答は電子付録として掲載する。

Q22. 研究データを公開していない理由として、あてはまるものを全てお選び下さい。  
 [複数選択可] \*

表 35 データを公開していない理由 (n=593：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	227	38.3%
公開できない情報が含まれるから（機密やプライバシー情報等）	187	31.5%
業績にならないから	158	26.6%
ニーズがないと思うから	136	22.9%
時間がないから	135	22.8%
所属機関にポリシーがないから	125	21.1%
助成機関のポリシー（助成条件）ではないから	94	15.9%
資金がないから	49	8.3%
リポトリなどの公開手段がないから	44	7.4%
分野・コミュニティで推奨されていないから	29	4.9%
その他	32	5.4%
特に理由はない	93	15.7%

Q23. Q22 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 36 データ公開意思

データ公開意思	人数	比率
はい	147	29.4%
いいえ	146	29.2%
わからない	207	41.4%
合計	500	100.0%

※データを公開していない 593 名から、Q22 で「特に理由はない」を選択した回答者以外に尋ねた。

Q24. 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。\*

表 37 データ公開のインセンティブの重要度 (n=1,180)

	0	1	2	3	4
データに紐づいた論文の引用	2.4%	1.3%	1.7%	14.7%	79.8%
データの引用*	2.9%	1.4%	2.6%	16.1%	76.9%
業績として論文と同様に評価されること	7.7%	6.5%	13.1%	28.0%	44.6%
利用者からの報告	5.6%	9.7%	18.7%	23.1%	42.8%
研究資金獲得の選考要因に加味されること	10.6%	10.0%	20.4%	27.2%	31.8%
利用料金などの対価	10.6%	25.1%	36.2%	15.5%	12.6%

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

\*「データの引用」とは、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ことを指します。

Q25. 研究データを公開することは、ご自身が所属するコミュニティや機関で評価されていると思われますか？

表 38 コミュニティや機関によるデータ公開の評価

コミュニティや機関によるデータ公開の評価	人数	比率
評価されていると思う	92	7.8%
ある程度評価されていると思う	294	24.9%
ほとんど評価されていないと思う	370	31.3%
まったく評価されていないと思う	241	20.4%
わからない	173	14.6%
その他	12	1.0%
合計	1,180	100.0%

Q26. ご自身が指導する研究者（大学院生やポスドクなど）が研究データを公開していることを、評価していらっしゃいますか？

表 39 回答者自身によるデータ公開の評価

回答者自身によるデータ公開の評価	人数	比率
高く評価している	158	13.4%
やや評価している	251	21.2%
あまり評価していない	151	12.8%
まったく評価していない	78	6.6%
わからない	266	22.5%
指導する研究者がない	206	17.4%
その他	70	5.9%
無回答	2	0.2%
合計	1,182	100.0%

## 7. 最近の研究のためのデータについて

Q27. 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 40 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	174	14.9%
GB（ギガバイト）	668	57.2%
TB（テラバイト）	211	18.1%
PB（ペタバイト）以上	11	0.9%
わからない	103	8.8%
合計	1,167	100.0%

Q28. カレントデータを整備・公開・保存しようとする場合、次の資源は十分に整っていますか？ あてはまるものをそれぞれお選び下さい。\*

表 41 カレントデータの管理資源の充実度 (n=1,188)

	0	1	2	3	4
データの整備・公開のための人材	12.0%	61.4%	17.1%	6.8%	2.8%
データの整備・公開のための資金	15.0%	54.5%	17.3%	9.7%	3.5%
データの整備・公開のための時間	13.1%	51.7%	21.1%	9.9%	4.1%
データ公開用のリポジトリ	27.0%	22.7%	18.4%	19.4%	12.5%
研究終了後のデータ保存用ストレージ	9.1%	19.5%	21.9%	27.4%	22.1%
研究中のデータ用ストレージ	7.7%	12.8%	14.6%	34.0%	30.9%

0=わからない、1=不十分、2=やや不十分、3=ほぼ十分、4=十分

Q29. 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？ あてはまるものをそれぞれお選び下さい。\*

表 42 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	3.7%	68.8%	21.0%	4.2%	2.3%
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%	61.4%	19.5%	10.0%	4.6%
不正利用・改ざんの可能性	5.3%	59.7%	18.3%	11.1%	5.6%
二次利用に関して責任が生じる可能性	6.1%	57.5%	20.7%	10.4%	5.3%
機密・プライバシー情報	4.0%	51.8%	20.6%	14.6%	9.0%
データの利用権限や契約	5.1%	51.3%	25.3%	11.5%	6.7%
商用利用される可能性	5.1%	40.8%	23.8%	19.4%	10.9%
研究の誤りを発見される可能性	7.8%	7.8%	11.6%	37.5%	35.3%

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

Q30. カレントデータを整備・公開・保存する上で、より詳しく知りたいと思われる項目をお選び下さい。[複数選択可]

表 43 研究データ管理 (RDM) に関して知りたい項目 (n=1,039 : 複数回答)

項目	人数	比率
適切なデータ形式	703	60.8%
知的財産権やライセンス	695	60.1%
データの安全な管理方法 (セキュリティ)	678	58.7%
データに含まれる機密情報の処理	555	48.0%
適切なリポジトリ	553	47.8%
適切なメタデータ標準	485	42.0%
データのバージョン管理方法	382	33.0%
データのバックアップ方法	358	31.0%
その他	9	0.8%
特に知りたい項目はない	117	10.1%

\* 「メタデータ」とは、データの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のことを指します。また、メタデータにはさまざまなフォーマットや標準があります。(参考: Research Data Alliance による分野別メタデータ標準の一覧 <https://rd-alliance.github.io/metadata-directory/standards/>)

Q31. カレントデータの整備・公開・保存プロセスを、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に依頼したいと思われませんか？

表 44 研究データ管理 (RDM) プロセスの依頼意思

依頼意思	人数	比率
はい	488	41.1%
いいえ	385	32.4%
わからない	287	24.2%
無回答	28	2.4%
合計	1,188	100.0%



Q32. 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に依頼したいと思われる項目を全てお選び下さい。[複数選択可]

表 45 研究データ管理（RDM）に関して依頼したい項目（n=488：複数回答）

項目	人数	比率
適切なデータ形式への変換	413	84.6%
適切なリポジトリの選択	374	76.6%
適切なライセンスの選択	370	75.8%
データを再利用しやすいように整える	362	74.2%
機関のリポジトリにデータを公開する	333	68.2%
適切なメタデータ標準の選択	323	66.2%
メタデータの作成	312	63.9%
データを異分野の研究者に紹介する	176	36.1%

\*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

Q33. 論文や研究データの公開に関するご自身の研究分野の展望、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。

※Q33 の自由回答は電子付録として掲載する。なお、調査に対する謝辞や「特にない」といった回答は削除した。

### (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
地球科学	57	4.5%	70.2%	93.0%	80.7%	93.0%	22.8%
数学	13	1.0%	69.2%	92.3%	61.5%	76.9%	7.7%
生物科学	203	16.0%	65.0%	88.2%	82.8%	75.9%	17.2%
計算機科学	57	4.5%	61.4%	77.2%	91.2%	75.4%	31.6%
農学	126	9.9%	49.2%	81.0%	74.6%	79.2%	14.3%
化学	198	15.6%	45.5%	73.7%	70.7%	71.2%	23.7%
医学	91	7.2%	44.0%	90.1%	63.7%	67.0%	25.3%
物理学・天文学	94	7.4%	40.4%	86.2%	62.8%	73.4%	19.1%
社会科学・人文学	27	2.1%	33.3%	92.6%	74.1%	77.8%	22.2%
心理学	19	1.5%	31.6%	78.9%	47.4%	57.9%	10.5%
工学	383	30.2%	27.7%	72.3%	60.1%	73.4%	21.7%
全体	1,268	100.0%	44.7%	80.1%	69.7%	74.4%	20.8%

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

「入手」のみ n=1,267

調査資料-316

研究データ公開と論文のオープンアクセスに関する実態調査 2020

2021年11月

文部科学省 科学技術・学術政策研究所  
データ解析政策研究室  
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 2020

November 2021

IKEUCHI, Ui and HAYASHI, Kazuhiro  
Research Unit for Data Application  
National Institute of Science and Technology Policy (NISTEP)  
Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan



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