

2022

10 Mar.

資料1
情報委員会（第24回）
令和4年3月10日

研究データと論文の公開に関する実態調査

池内 有為

学術調査官／文教大学

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客員研究官



『研究データと論文の公開に関する実態調査』と国際調査

	NISTEP			Springer Nature	State of Open Data
調査年	2016	2018	2020	2017	2021
対象	科学技術専門家ネットワーク (日本の産学官の研究者)			欧州・米国 豪・亜・アフリカ	192カ国
回答者数 (回答率)	1,406 (70.9%)	1,548 (70.5%)	1,349 (70.5%)	>7,700	>21,000
出典 (DOI)	10.15108/rm268	10.15108/rm289	10.15108/rm316	10.6084/m9.figshare.5975011.v1	10.6084/m9.figshare.17061347.v1

NISTEP(2020) 回答者の概要

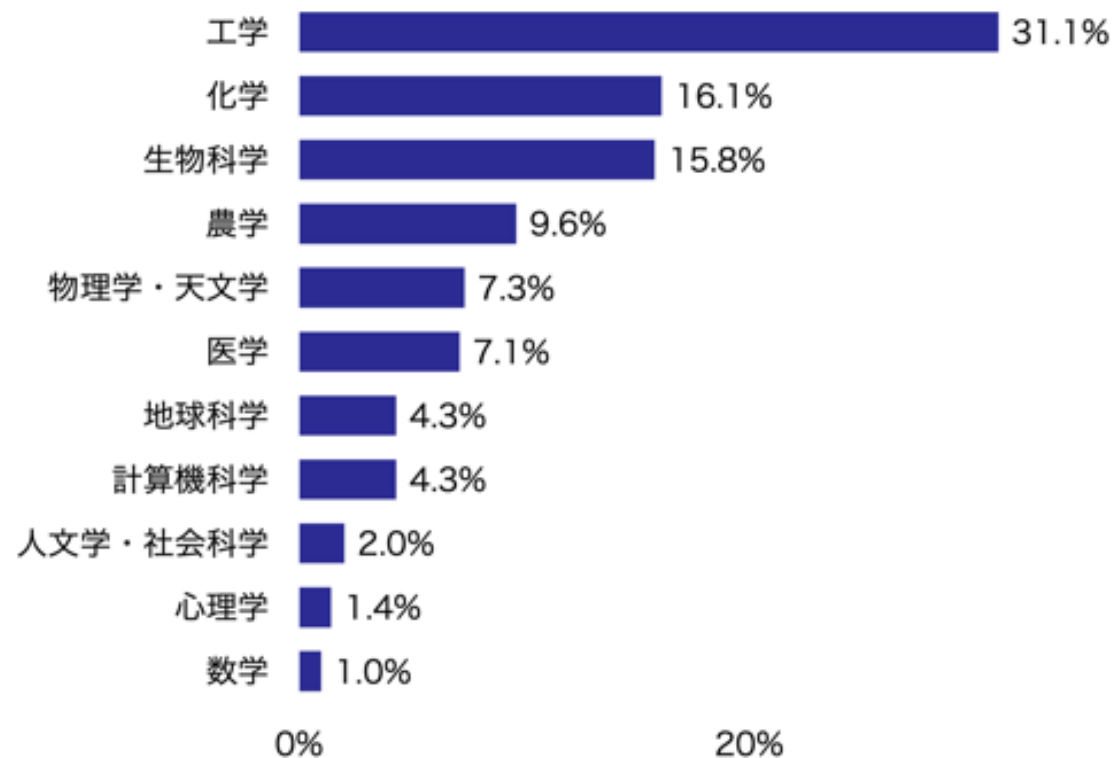
回答者の所属 (n=1,349)



回答者の年齢層

年代	回答者数	比率
30代以下	425	31.5%
40	659	48.9%
50	194	14.4%
60代以上	70	5.2%
不明	1	0.1%
合計	1,349	100.0%

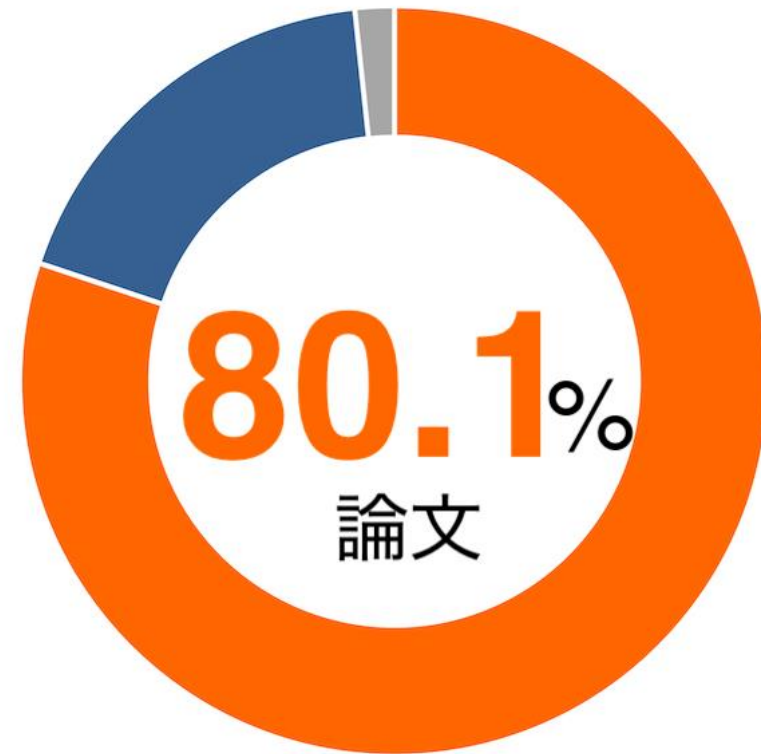
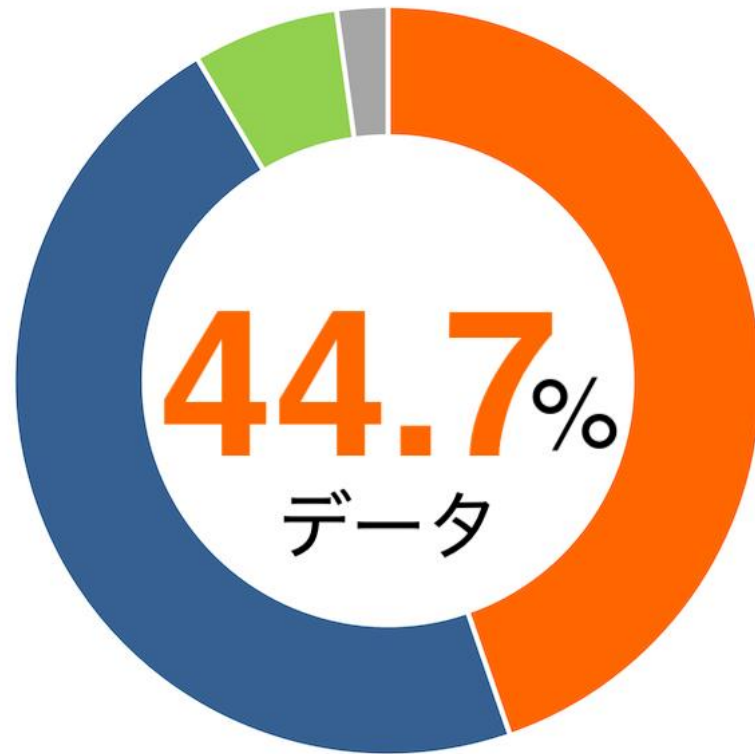
回答者の研究分野 (n=1,349)



調査項目

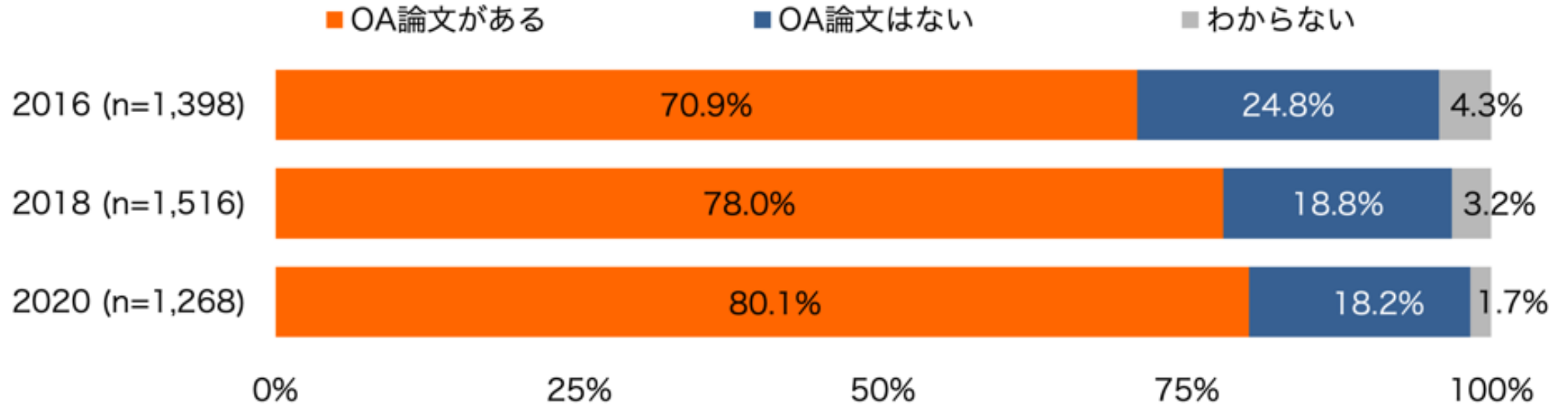
1. 論文と研究データの公開状況
2. 研究データの提供（共有）状況
3. 公開データの利用状況と課題
4. データマネジメントプラン（DMP）の作成状況
5. 研究データ公開の障壁
6. 研究データ公開のインセンティブ
7. 研究データの整備・公開・保存と支援の可能性
8. 研究成果公開の展望・問題点等（自由記述）

研究データと論文の公開経験 (2020, n=1,268)



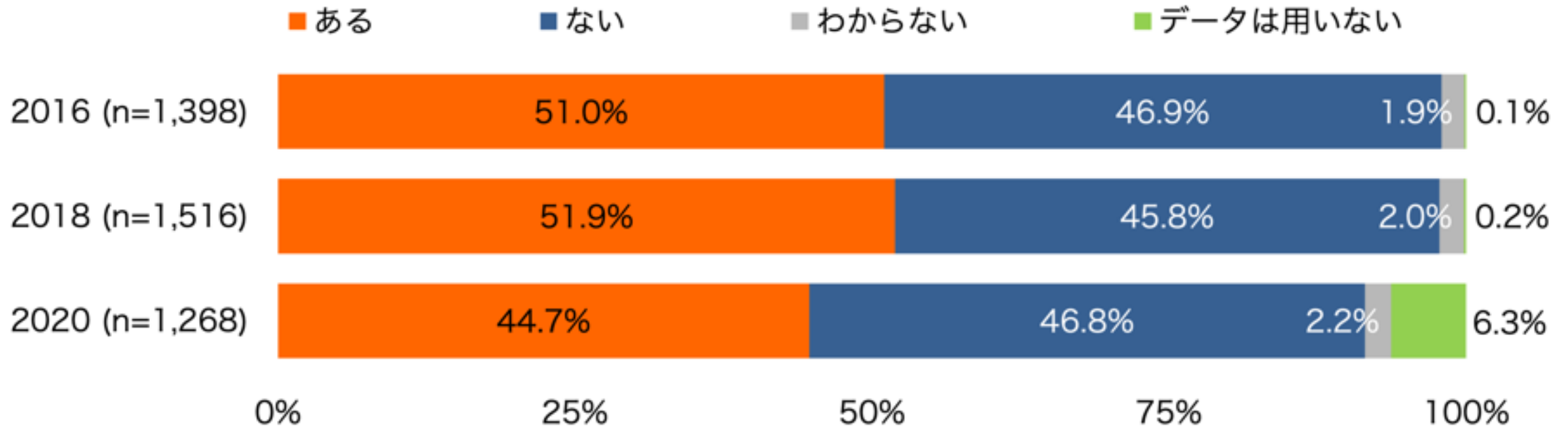
■ あり ■ なし ■ わからない ■ データは用いない

論文のオープンアクセス（OA）経験の経年変化



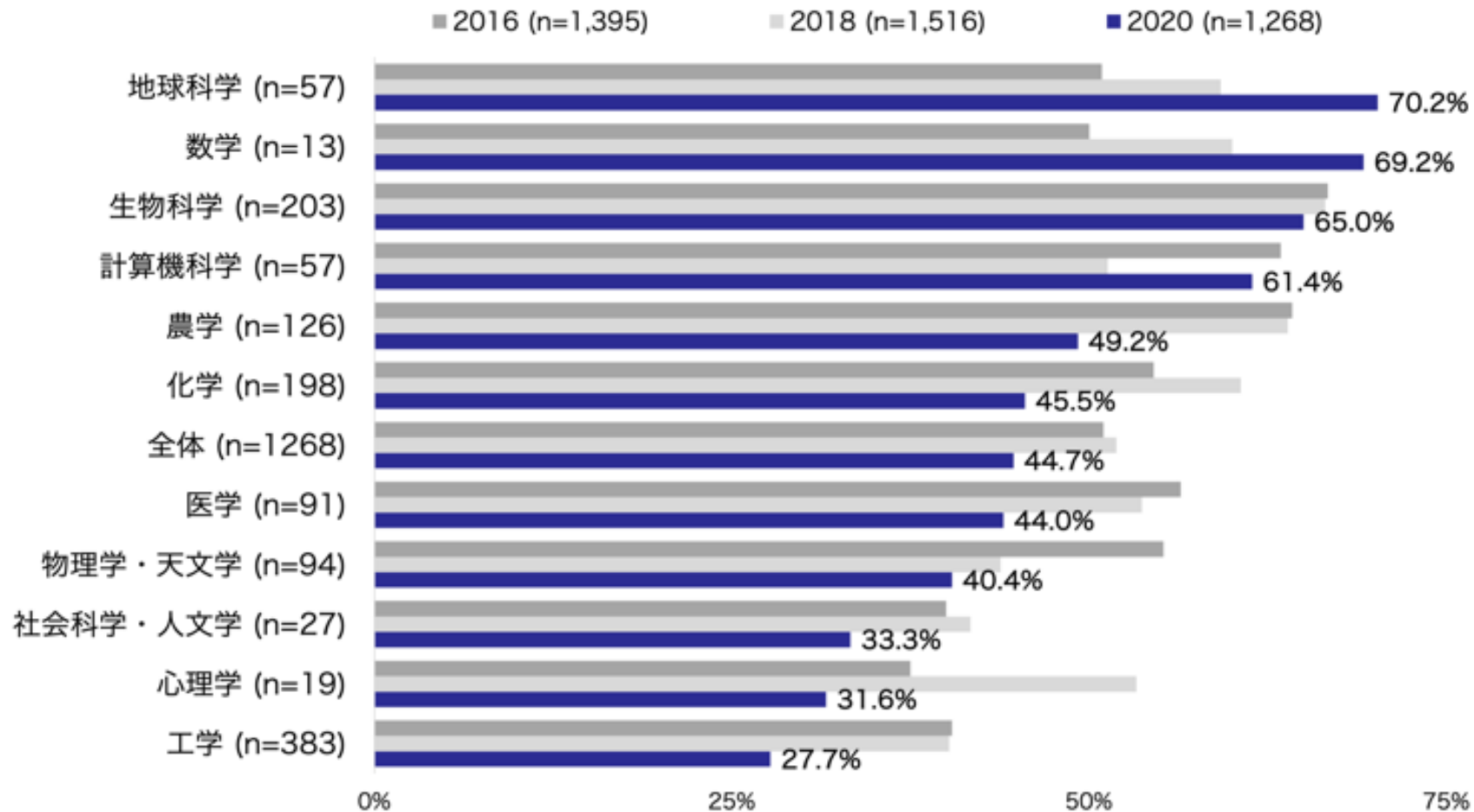
- オープンアクセス（OA）経験をもつ回答者は増加傾向
- 公開方法は「OA誌への投稿」（77.1%）、「雑誌が論文をOAにした」（32.6%）、「所属機関のリポジトリ」（30.7%）、「雑誌のOAオプション」（30.6%）など
- 公開理由は「投稿した雑誌がOA」（75.8%）、「研究成果を広く認知してもらいたい」（57.6%）に集中
- 未公開理由は「資金がないから」（57.6%）、「投稿したい雑誌がOAではない」（40.3%）に集中

データ公開経験の経年変化



- データ公開経験をもつ回答者の比率は、2016/2018年調査と比較してやや低下
- 調査対象の専門家ネットワークは毎年一部入れ替わりがあるためパネル調査ではない
- データ公開に対する理解が深まり、より正確に実態を把握できた可能性

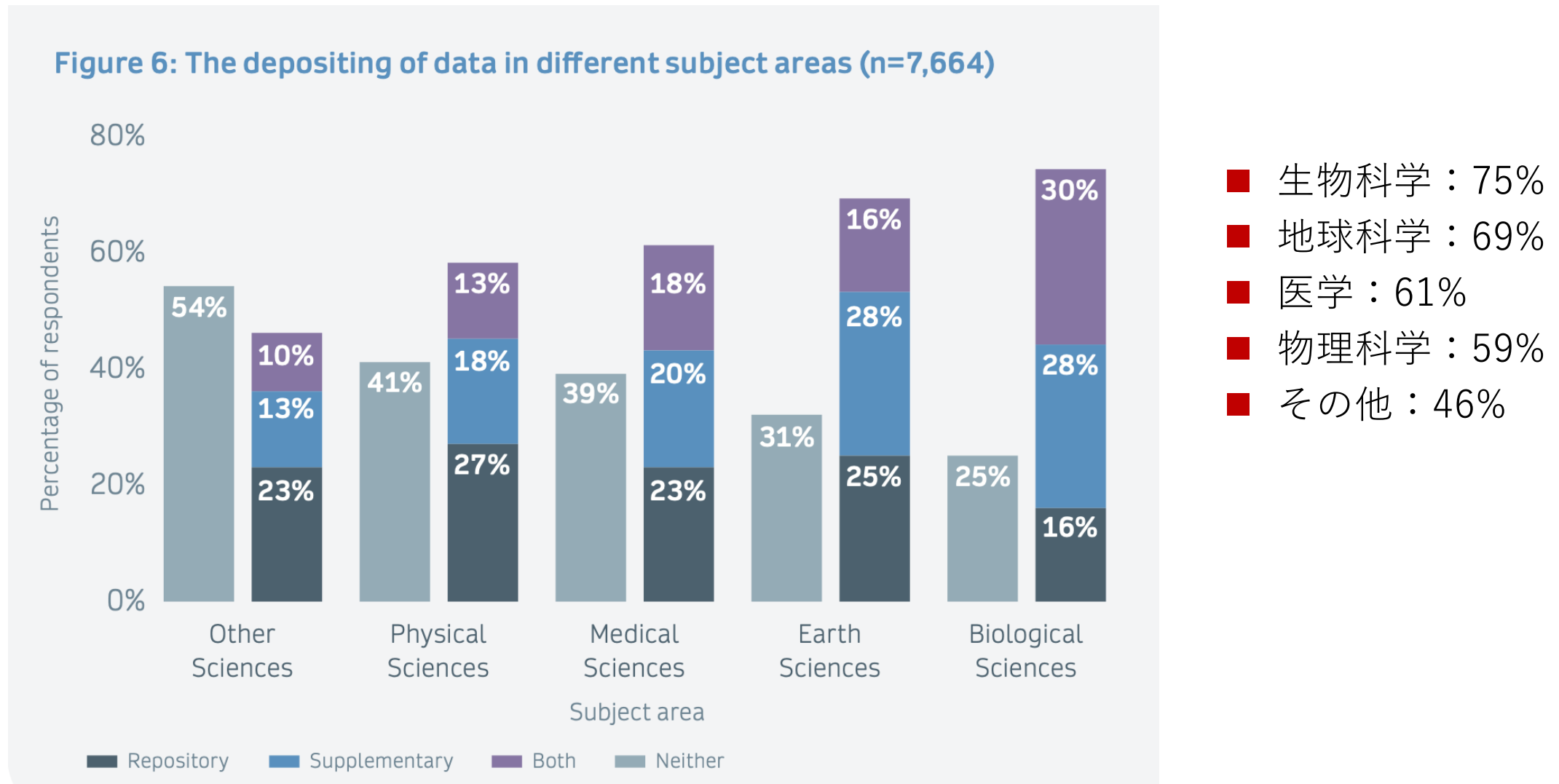
分野別データ公開経験



■ 分野による差が大きい

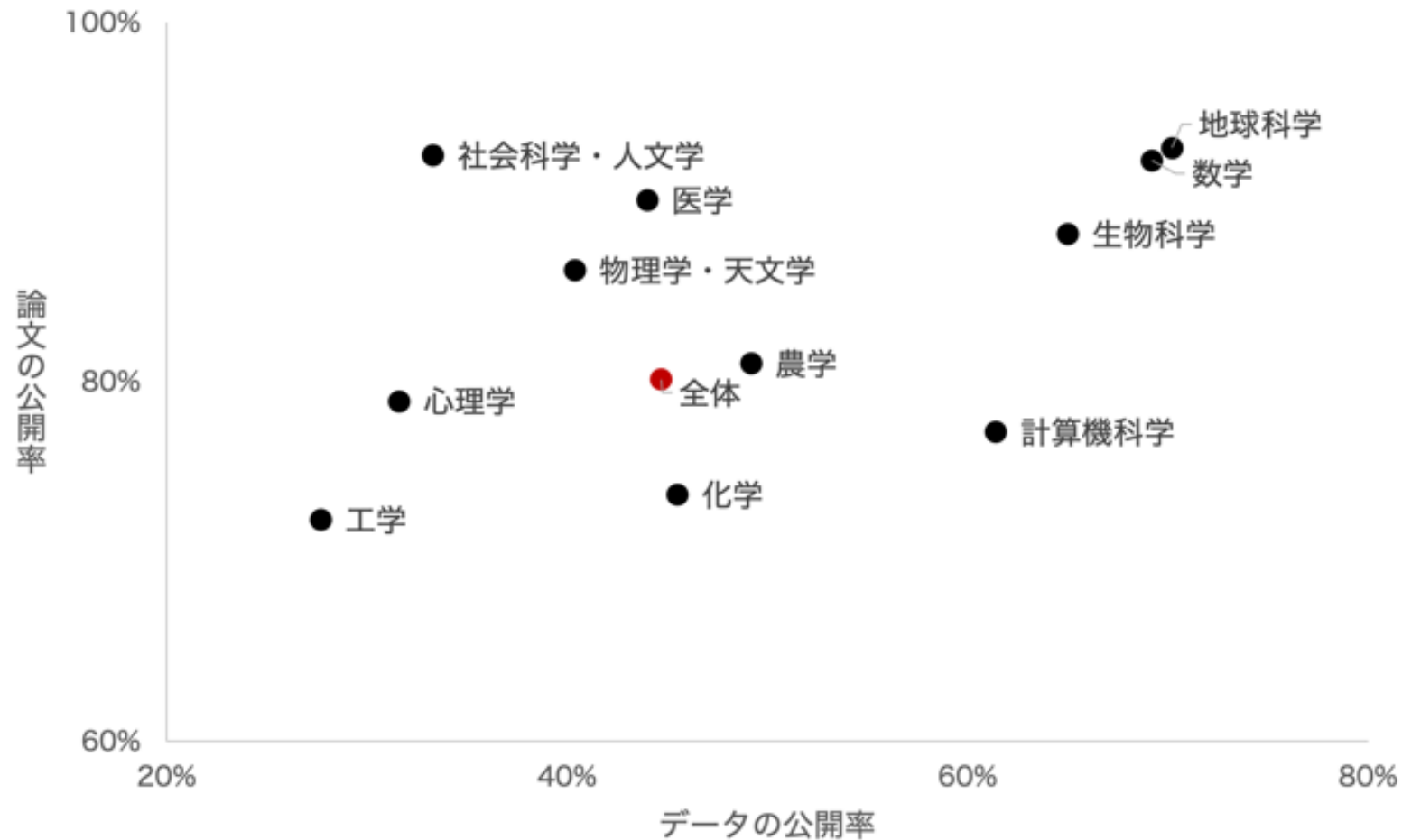
■ 地球科学（1位）と数学（2位）は前回調査から増加。ただし回答者数が比較的少ないため少数の回答が影響している可能性あり

参考：分野別データの寄託状況



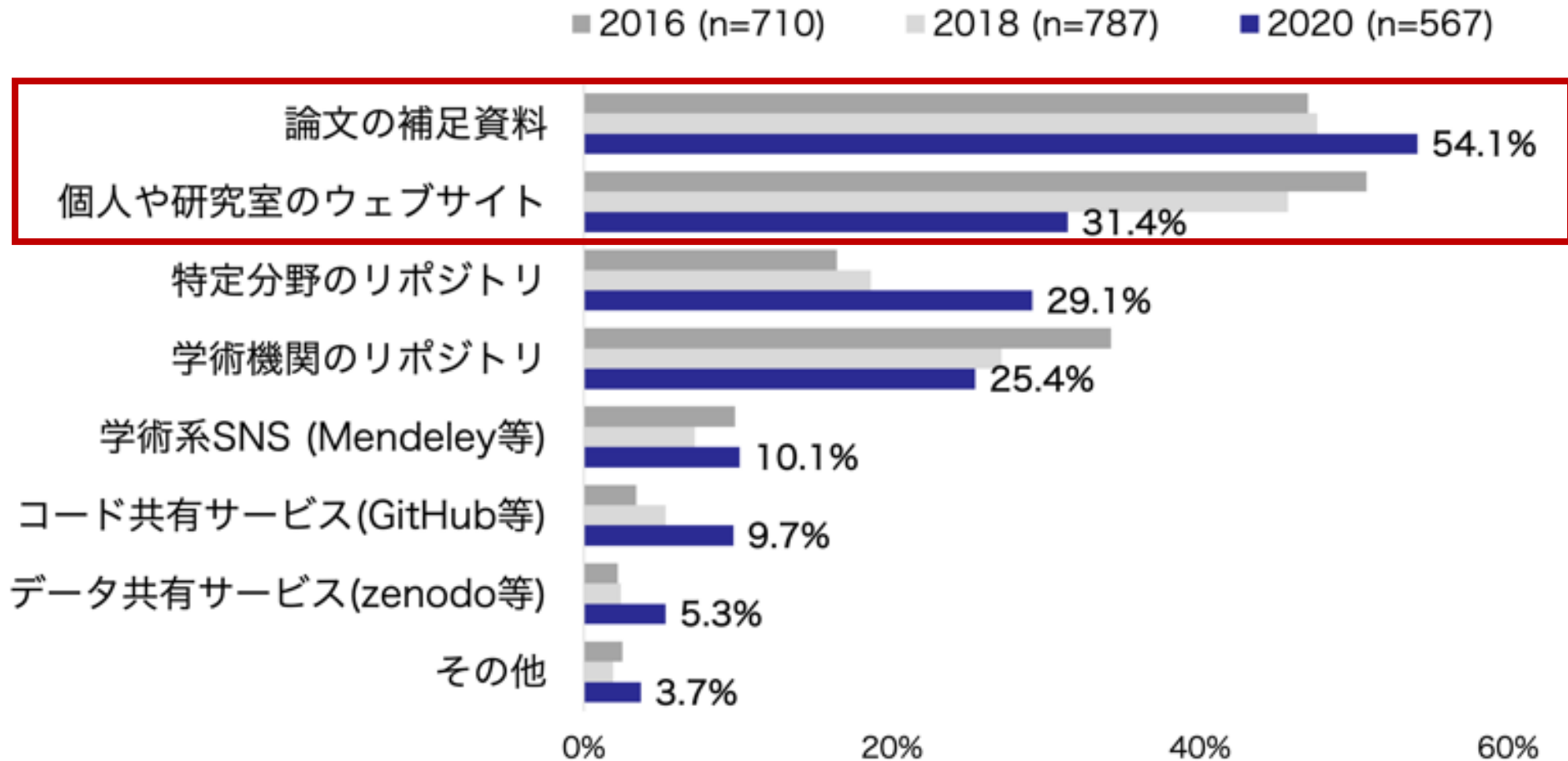
Whitepaper: Practical challenges for researchers in data sharing. SpringerNature.
<https://doi.org/10.6084/m9.figshare.5996786>

分野別データ公開経験と論文の公開経験



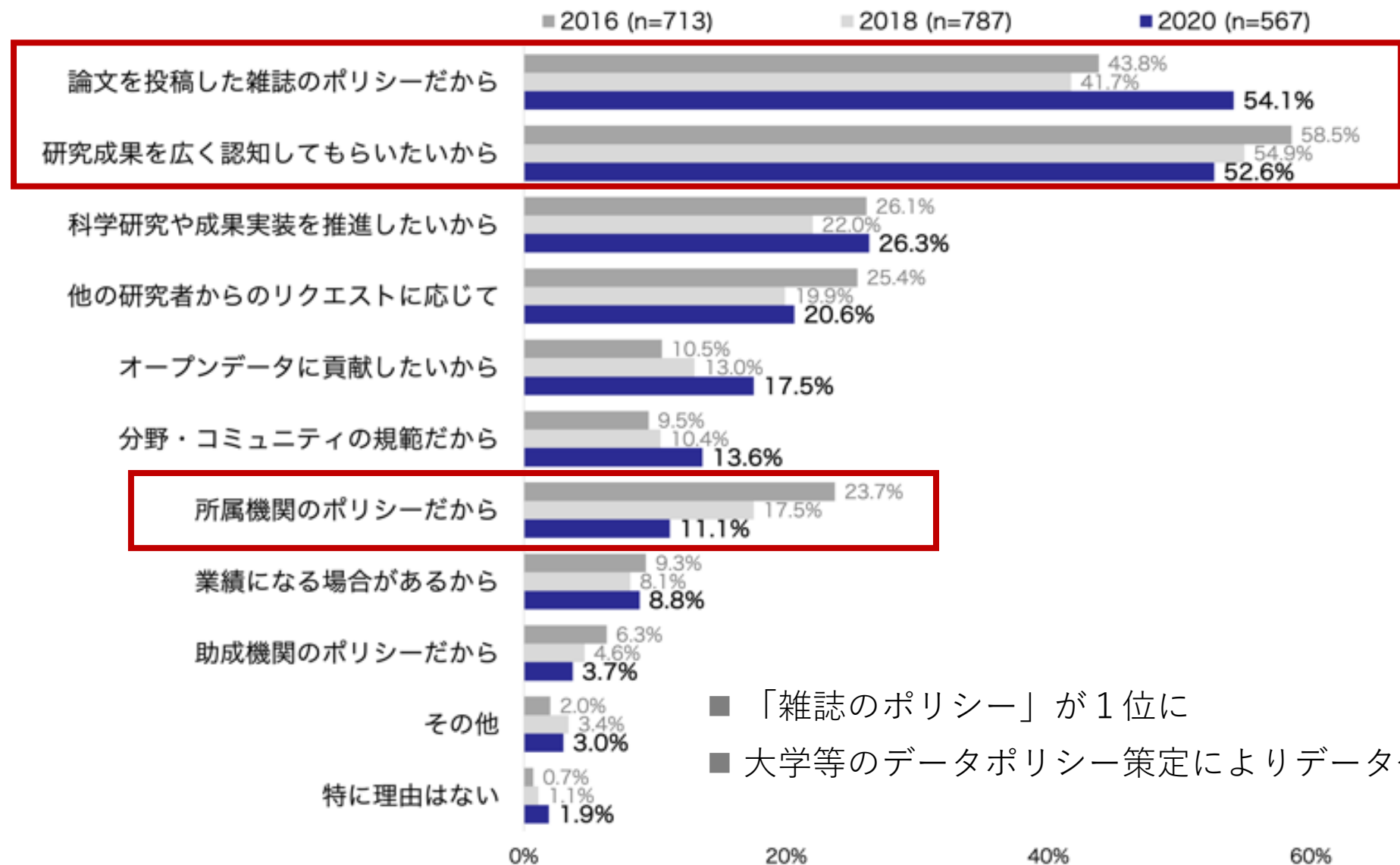
- データ公開経験とOA経験に有意な相関はなく、むしろ分野の特徴が出ている
- いずれも高いのは地球科学・数学、いずれも低いのは工学
- OA率が高いのは社会科学・人文学、データ公開率が高いのは計算機科学

データの公開方法



- 「論文の補足資料」が「個人や研究室のウェブサイト」を抜いて1位に→米国や欧州の調査でも同様の傾向
- 「特定分野のリポジトリ」が増加した一方で、「学術機関のリポジトリ」が減少

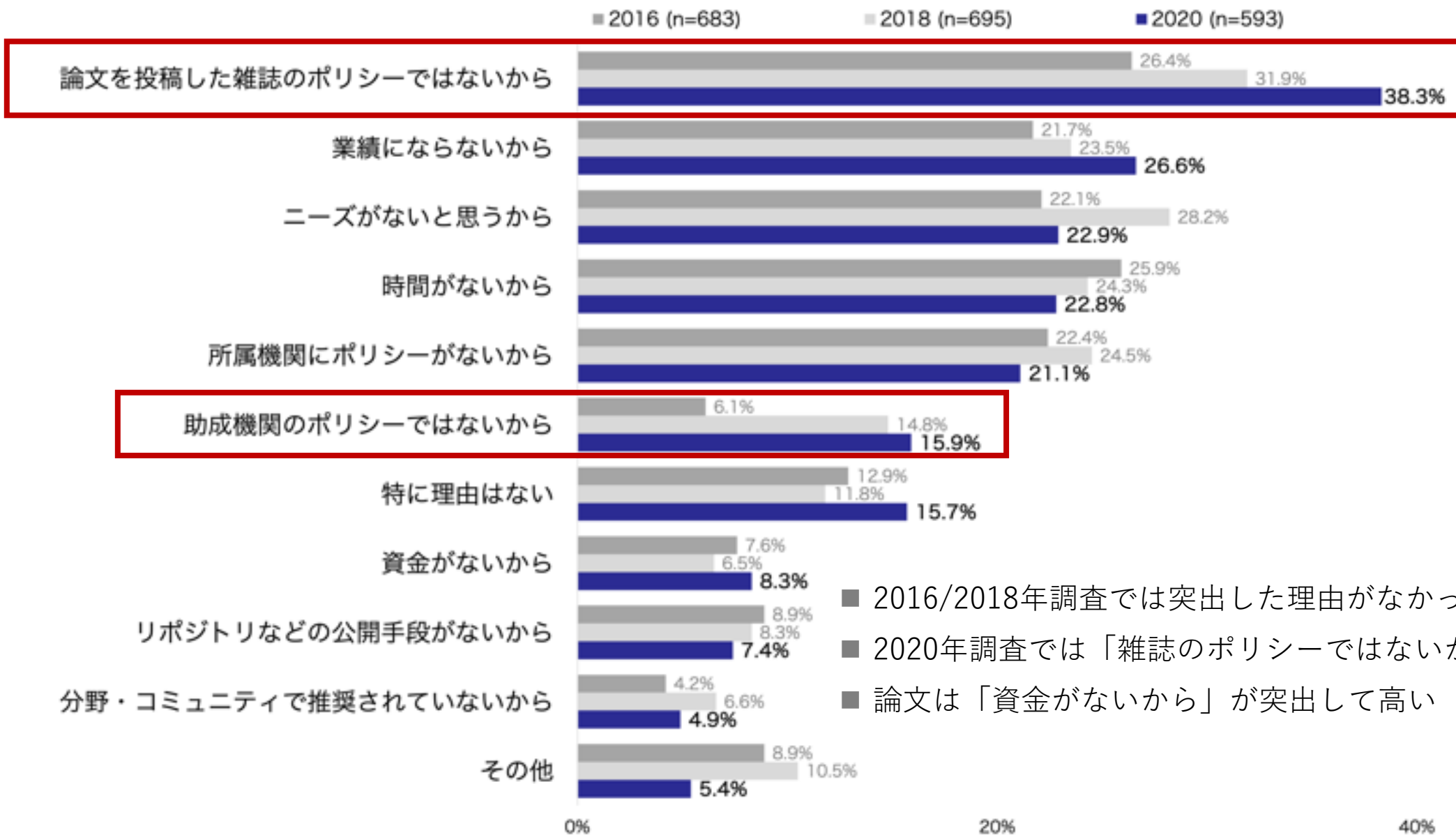
データの公開理由



■ 「雑誌のポリシー」が1位に

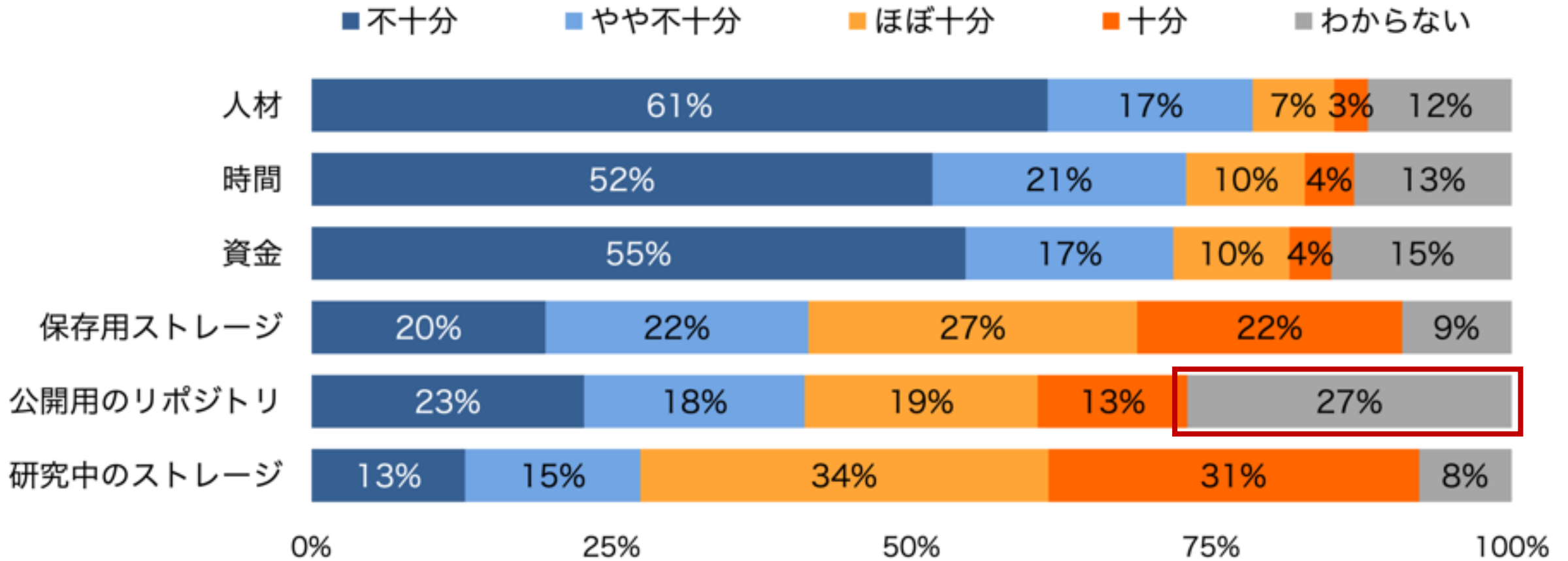
■ 大学等のデータポリシー策定によりデータ公開が進む可能性

データを公開していない理由



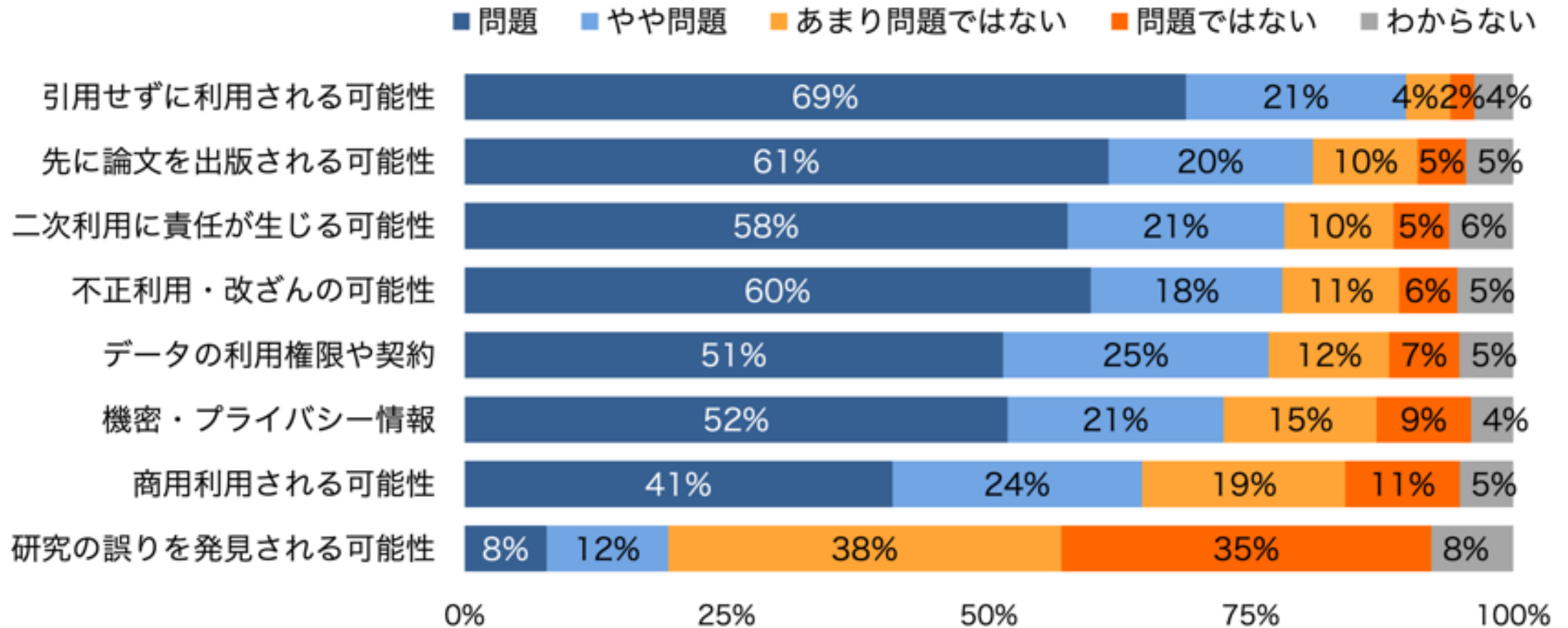
- 2016/2018年調査では突出した理由がなかった
- 2020年調査では「雑誌のポリシーではないから」が主な理由に
- 論文は「資金がないから」が突出して高い（57.6%）

データの整備・公開資源の充足度 (n=1,188)



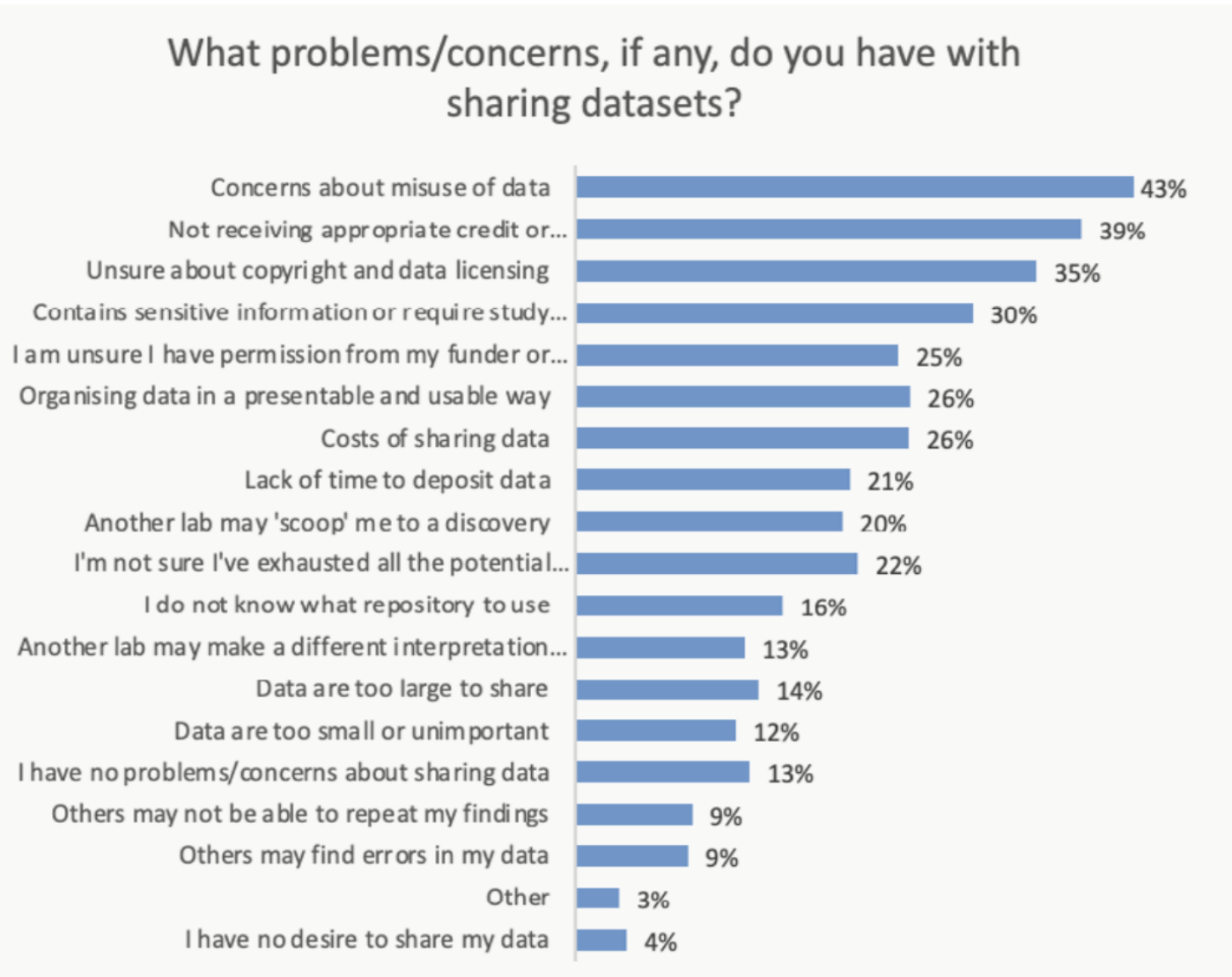
- 人材・時間・資金の不足感が強い
- ストレージやリポジトリは2016/2018年と比較して、やや改善しているものの十分とは言い難い
- 公開用のリポジトリは「わからない」の比率が高い

データ公開に対する懸念の強さ (n=1,188)



- 2016/2018年調査から継続して全体的に懸念が強い
- 「引用せずに利用される可能性」や「先に論文を出版される可能性」に対する懸念が強いのは国外の調査でも同様の傾向

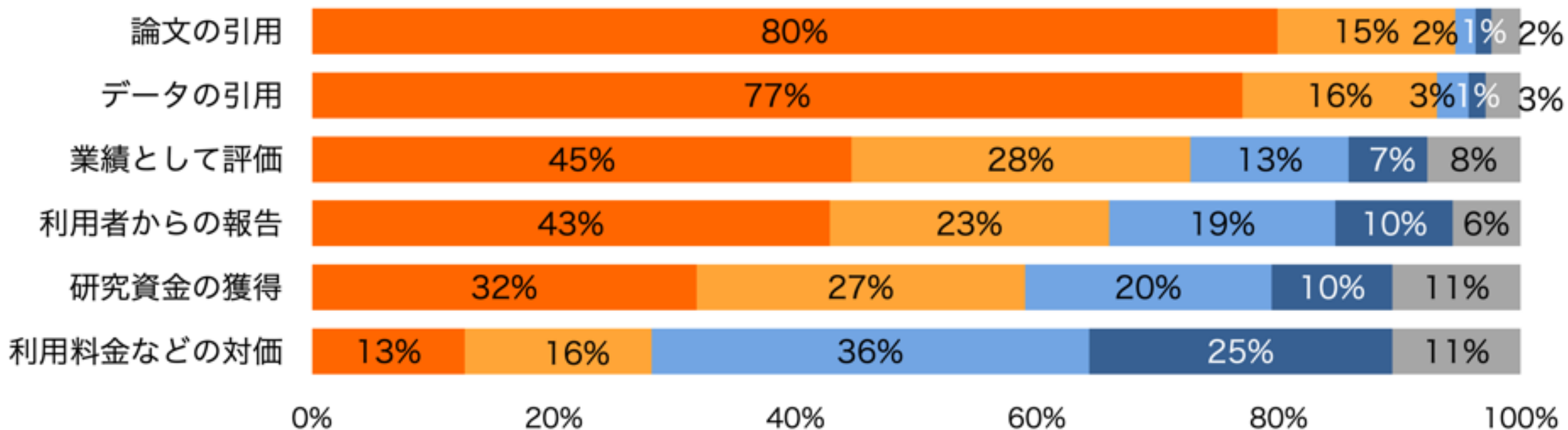
参考：データセットの共有における問題や懸念



- 誤用（43%）
- 適切なクレジットが得られない（39%）
- 著作権やライセンスがよくわからない（35%）
- センシティブ情報が含まれている、または同意を必要とする（30%）
- 資金提供者や研究機関の許可が得られているかどうか不安である（25%）
- データを見やすく、使いやすいように整理すること（26%）
- データ共有のコスト（26%）

データ公開によるインセンティブの重要性

■重要 ■やや重要 ■あまり重要ではない ■重要ではない ■わからない



- 「論文の引用」と「データの引用」が特に重視されている（2018年調査でも同様の傾向）
- Tenopir et al.(2020)による国際調査でも92.1%の回答者が「データを利用されたら引用されることが重要である」と回答 <https://doi.org/10.1371/journal.pone.0229003>
- 「利用料金などの対価」は相対的に重要性が低い

データ公開によって得られた良い結果（自由記述）

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%

	2018	2020
データ公開経験あり	787名	567名
記述あり	195名 (24.8%)	130名 (22.9%)

データ公開によって得られた良い結果<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。

35% of respondents
had been involved in a
**collaboration as a result of
data they had previously
shared**

The State of Open Data 2021
[10.6084/m9.figshare.17061347.v1](https://doi.org/10.6084/m9.figshare.17061347.v1)

■ 再利用・二次利用の比率は、2018年から10.3ポイント増加 (13.8%)

例：共同研究の契機

“**海外の研究者からの問い合わせ**があり、コラボレーションに繋がった。”

“タンパク質の機能に関する情報を論文投稿前に投稿規定になっていたのでデータベースに登録したら問い合わせがあり共同研究となった。予定していた雑誌に投稿するのをやめてお互いの実験結果を合わせて**よりインパクトのある研究として別の雑誌に投稿して採択**された。”

“**新たな共同研究の開拓**や、共同研究先の**学生の教育**などに役に立っていると考えている”

例：再利用・二次利用

“自身の研究データをもとにメタアナリシスなどによって新たな知見を掘り下げようとしてくれた方が現れ、**研究課題の掘り下げが進んだこと。**”

“実験結果を理論家が用いて論文を書くことにより、**分野の科学的理解が進む**”

“自分たちが苦労して作成したデータを使って、**見ず知らずの研究者が活用**してくれるのは、大変嬉しいことでした。”

データを公開したことによって起きた問題

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

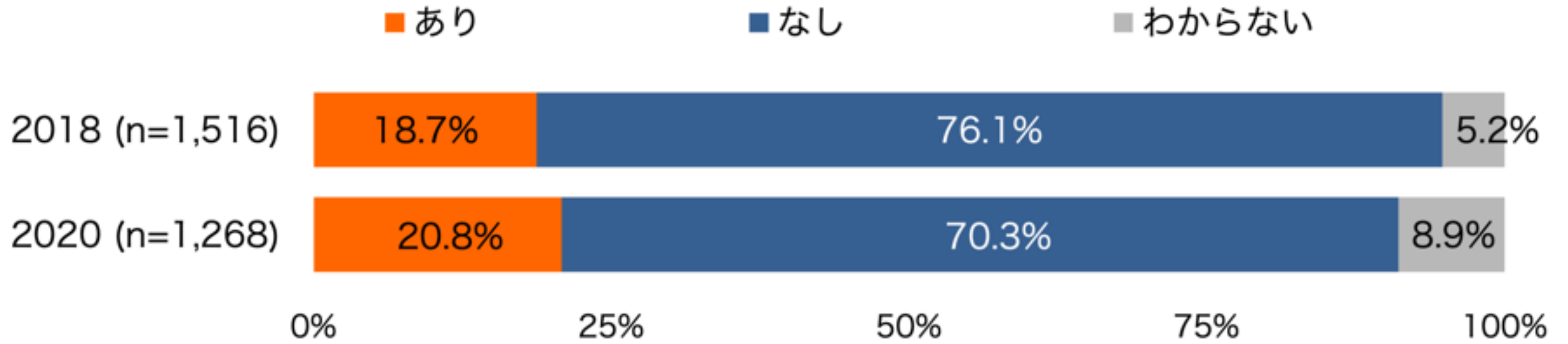
	2018	2020
データ公開経験あり	787名	567名
記述あり	54名 (6.9%)	33名 (5.8%)

例：引用せずに利用された

“論文に**引用も謝辞も書かれていないケースが多い**こと。公開に際しては、doi等の番号をつけて、データと言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。”

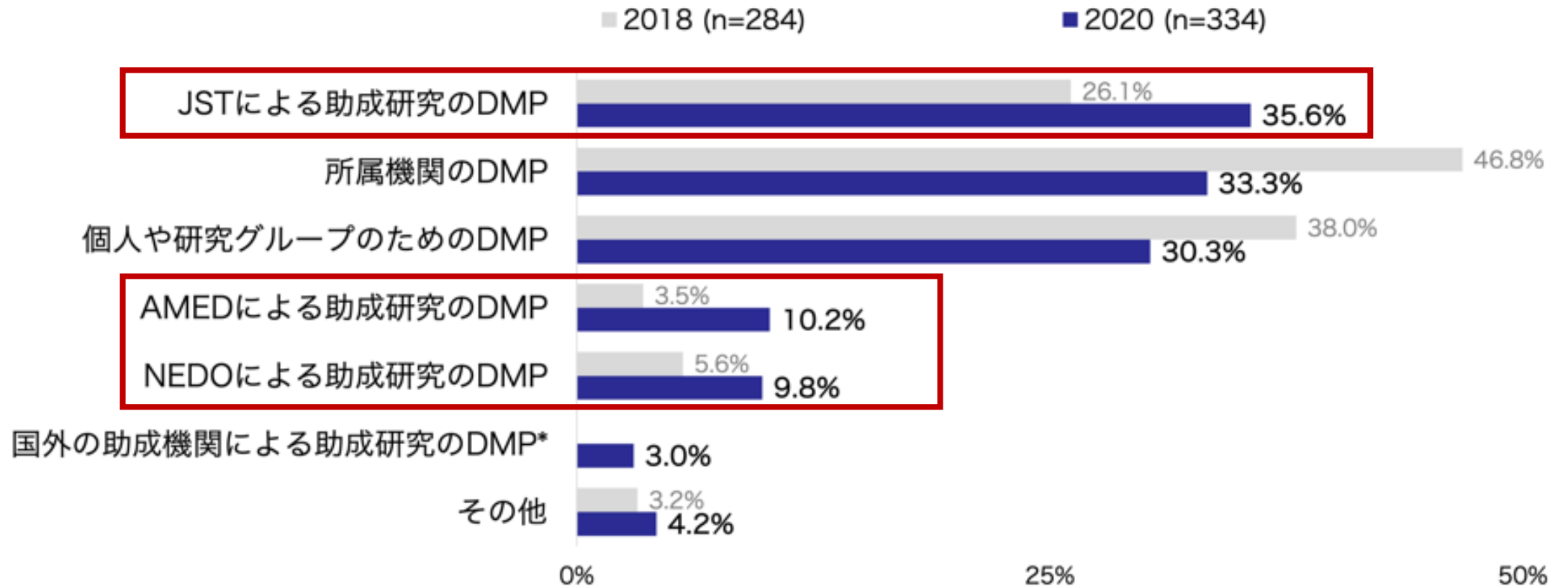
“自分が査読者となった論文原稿に、**ほぼそっくりそのまま内容が使われた**ことがあり、非常にびっくりしたことがある。”

データマネジメントプラン(DMP)の作成経験



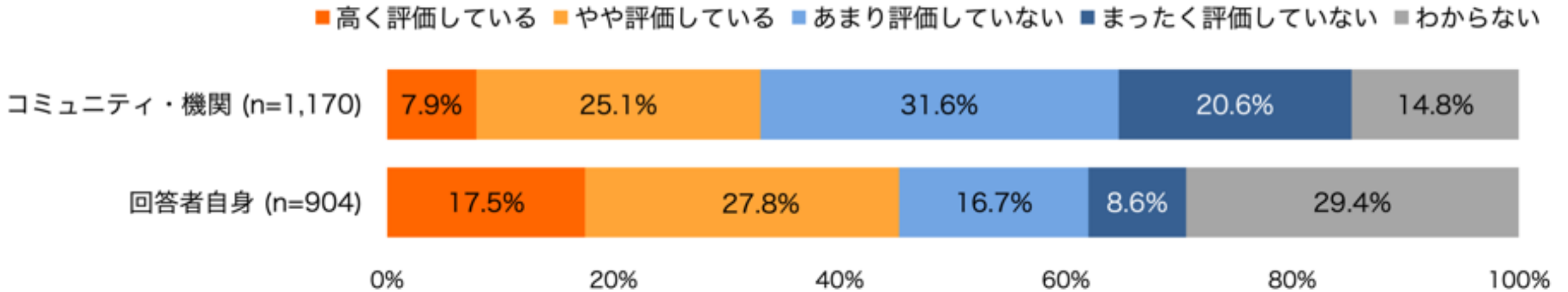
- 作成経験をもつ研究者の比率がわずかに増加
- 「わからない」の選択率が増加

作成経験があるDMP（複数回答）



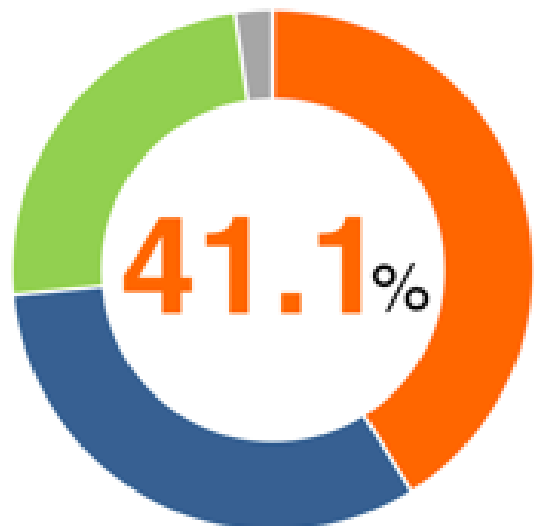
- 助成機関（JST、AMED、NEDO）の選択率は全て増加
- 作成理由の1位は「助成機関が要求しているから」（n=351, 51.5%）
→助成機関の要求がDMP作成を牽引。
- 未作成理由の1位は「DMPを知らなかったから」（n=891, 52.0%）

データ公開に対する評価

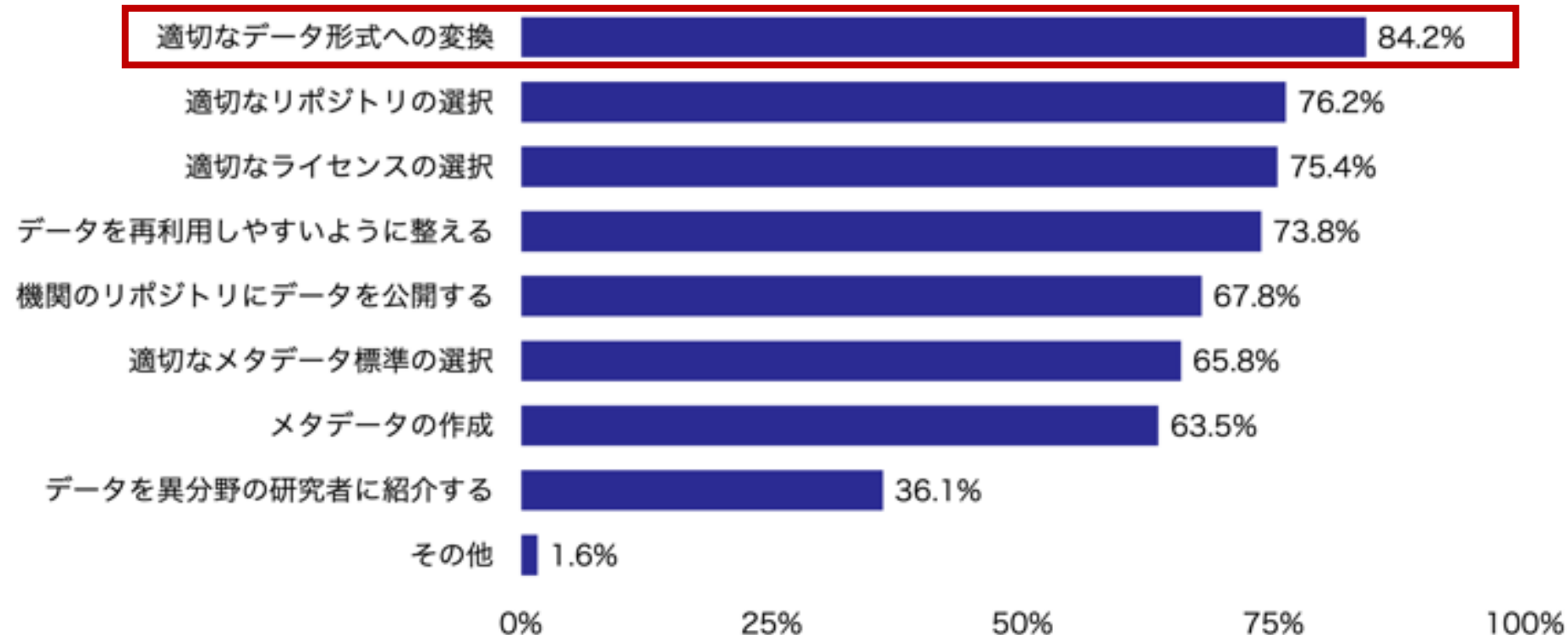


- コミュニティや機関よりも、回答者自身の方が評価していると認識している傾向
- 「わからない」の比率も高い
- データを公開していない理由の2位は「業績にならないから」(26.6%)

データの整備・公開・保存の依頼意思



■ はい ■ いいえ ■ わからない ■ 無回答



- データの整備・公開・保存プロセスを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は41.1% (n=1,188)
- 依頼したい項目 1位は「適切なデータ形式への変換」
- 2016/2018年調査では、第三者が支援する場合に専門性が必要であると考えられる項目の1位が「適切なデータ形式への変換」であった→専門性が高いことであっても依頼したいと考えている可能性

まとめ

- データの公開率は5割に満たない。分野による差がみられ、OAの公開率とは傾向が異なる。現時点では、雑誌の要求により論文の補足資料として公開する場合が最も多い。大学等のデータポリシー策定により、今後データ公開が進む可能性あり。
- データの整備・公開・保存のための資源は、特に人材・時間・資金に不足感が強い。公開用のリポジトリは認知度が低い。
- データ公開に対する懸念は依然として強く、特に引用されないことが問題視されているものの、実際にはそれほど問題が起きていない可能性あり。
- データ公開によるインセンティブとして論文やデータの引用が重視されているが、実際にはデータを公開しても引用に結びついていない可能性あり。
- DMPは、助成機関の要求により作成率が2割に達した。
- データ公開に対する評価は十分に行われておらず、「わからない」とする比率も高い。
- データの整備・公開・保存を第三者に依頼したいとする回答者は4割強。