

2023
22 Dec.


資料 2

科学技術・学術審議会
学術分科会
人文学・社会科学特別委員会
(第 20 回)
令和 5 年 12 月 22 日

人文学・社会科学分野におけるデータ管理と 利活用に資する専門人材の育成と課題

池内 有為

文教大学文学部 / 文部科学省科学技術・学術政策研究所客員研究官

 <https://orcid.org/0000-0002-5680-1881> <https://researchmap.jp/oui/>



C contents

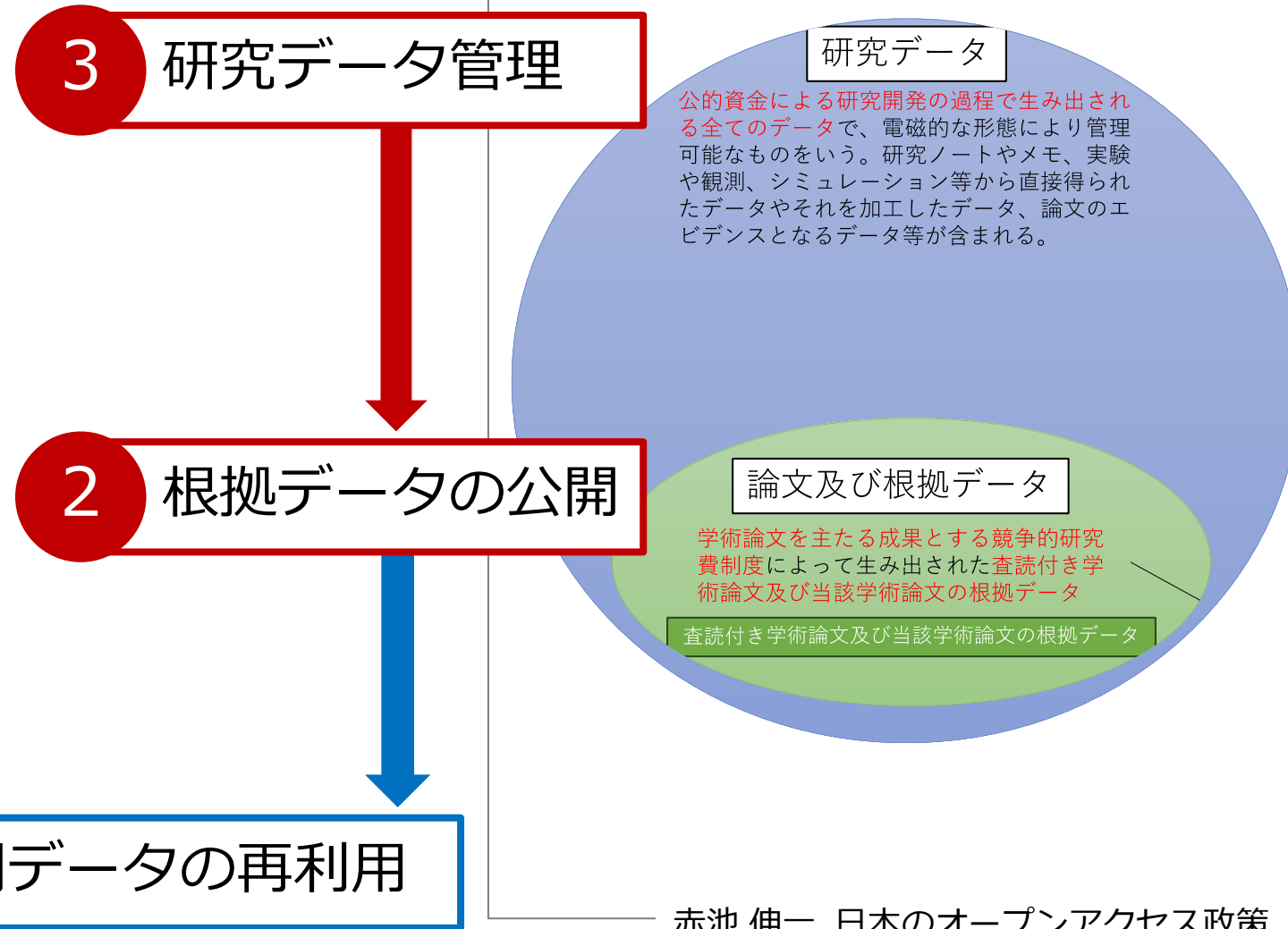
1. 研究データ公開と研究データ管理の要求
2. 日本の研究者による研究データの再利用・公開・管理状況
3. データ管理と利活用に資する支援人材の検討

1

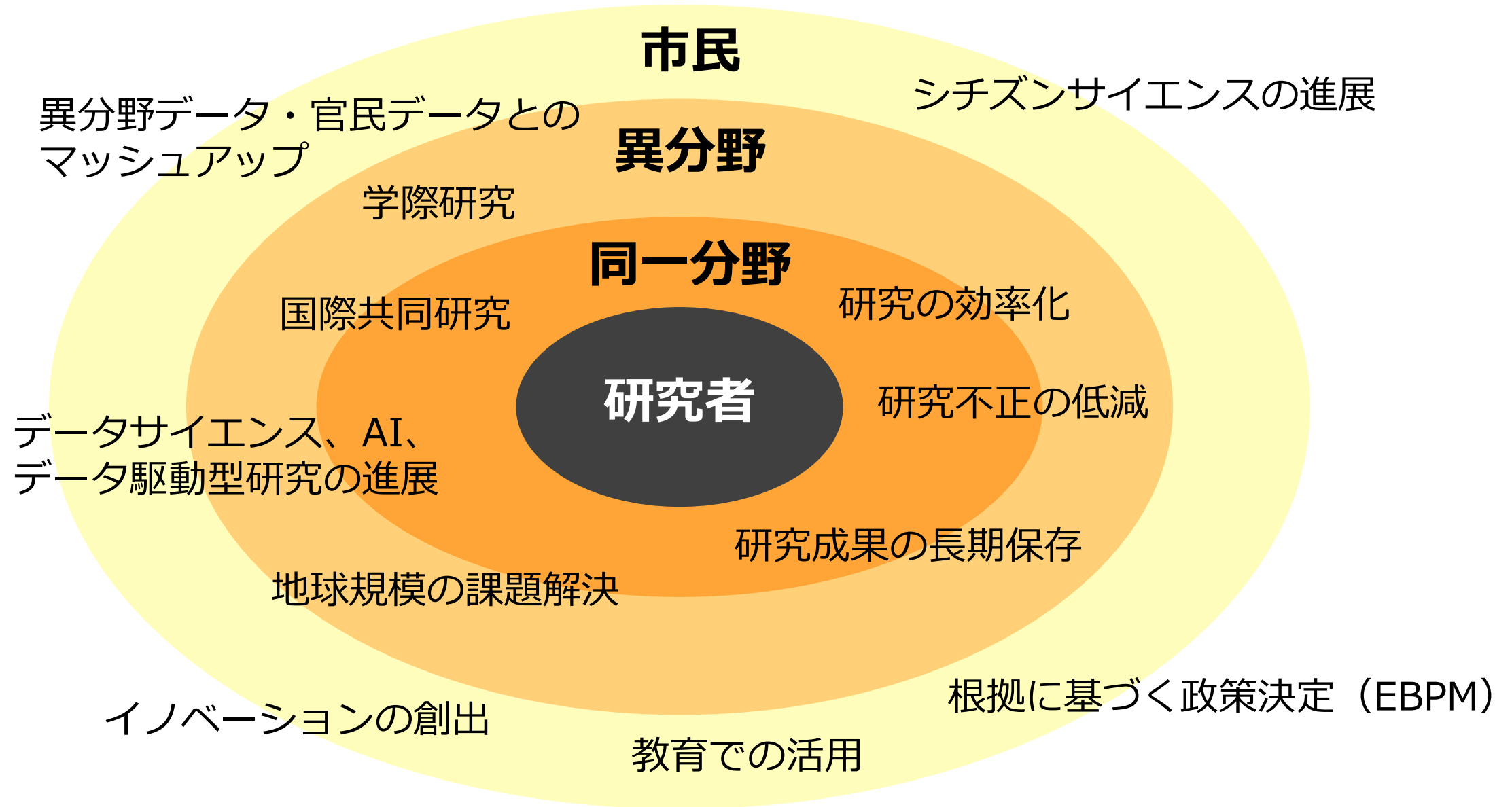
研究データ公開と研究データ 管理の要求

研究者・研究機関への要求とねらい

公的資金による学術論文等のオープンアクセスと研究データの管理・利活用の関係性



① 公開データの再利用と期待される効果



FAIR原則

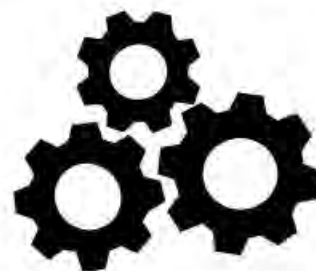
F
Findable



A
Accessible



I
Interoperable



R
Reusable



データ共有の対象は？（研究者・市民・非日本語話者・こども・AI…）

FORCE11: THE FAIR DATA PRINCIPLES (2016). <https://www.force11.org/group/fairgroup/fairprinciples>

NBDC研究チーム(訳), "FAIR原則（「THE FAIR DATA PRINCIPLES」和訳）" (2019). <https://doi.org/10.18908/a.2019112601>

人文学・社会科学データインフラストラクチャー構築推進事業（現在は「強化事業」）【研究者対象】

The Program for Constructing Data Infrastructure for the Humanities and Social Sciences (HSS)

aims to promote collaborate research

5つの拠点
データアーカイブ

culture among researchers and live institutions for HSS by funding and consultation.

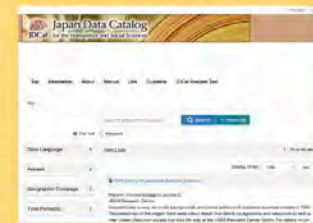


JDCat (Japan Data Catalog)



横断検索
JDCat

ガイダンス
動画



<https://jdcatalog.jsps.go.jp/>



<https://youtu.be/UVwqleOg03c>



"A Guide to Data Sharing in the Humanities and Social Sciences"



helps Japanese young researchers and graduate students to e... and... data sharing

データ共有の
手引き

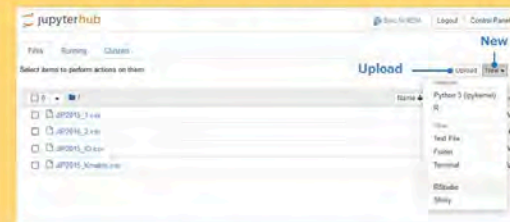
- Data format
- Privacy and anonymization
- Data copyrights, etc.

<https://www.jsps.go.jp/english/e-di/guide.html> *Available only in Japanese.



JDCat Analysis Tool

データ分析ツール



<https://jupyter.cs.rcos.nii.ac.jp/>

programs to analyze data... software or downloading... Jupyter Notebook... and shared.

- ✓ Collaborative research
- ✓ Education

JDCat and JDCat Analysis Tool were developed by JSPS and the National Institute of Informatics (NII).

*This tool is available only for the people who have account in a sort of Japanese research institutions.

<https://doi.org/10.17605/OSF.IO/GFR5H>

データ公開と研究評価

- i. **研究評価システムを変更し、多様な研究成果、活動、実践、およびキャリアパスを重視する評価を行う**ことで、オープンサイエンスを奨励すること
- ii. 変化を加速するために研究システム全体への関係者の関与を促すこと
- iii. 一貫性と調和を促進するために、グローバルな協力と相互学習を推進すること
- iv. オープンサイエンスの実践を促進するために意識を高め、リーダーシップを発揮すること
- v. 研究評価に使用するための科学のオープンかつ透明な指標を開発すること
- vi. オープンサイエンスの実践と研究評価のために必要なインフラと支援を提供すること
- vii. 研究者がオープンサイエンスを実践し、指標の開発と責任ある使用を支援するための教育と研修を提供すること。

国立大学図書館協会訳. G7 仙台科学技術大臣共同声明. 附属文書 1.

G7 オープンサイエンス作業部会 (OSWG). 研究評価とインセンティブに関するサブグループより.

https://www.janul.jp/sites/default/files/2023-06/g7annex1_jpn.pdf

②根拠データの公開

「公的資金による学術論文等のオープンアクセスの実現に向けた基本的な考え方」（2023年10月 CSTI有識者議員）→政府として決定予定

- 対象：学術論文を主たる成果とする競争的研究費制度によって生み出された**査読付き学術論文の根拠データ**
- 学術論文を主たる成果とする競争的研究費を受給する者（法人を含む。）に対し、**論文及び根拠データの学術雑誌への掲載後、即時に機関リポジトリ等の情報基盤への掲載を義務づける**。
- 研究成果を誰もが自由に利活用可能とするための発信手段として、研究DXプラットフォームを整備・充実させる。
- 大学等における機関リポジトリの価値向上、成果発信力の強化を行う。

③研究データ管理：DMP・メタデータ

「公的資金による研究データの管理・利活用に関する基本的な考え方」（2021年4月 統合イノベーション戦略推進会議決定）

- 対象：公的資金による研究開発の過程で生み出される**全てのデータ**
- 研究者は**管理対象データを特定、メタデータを付与**
 - ✓オープン・アンド・クローズ戦略に基づく公開・共有・非公開/非共有の設定
- NII RDCを中核的なプラットフォームとして位置づけ、メタデータを検索可能な体制を構築
- 大学等は機関におけるデータポリシーを策定

データ公開・管理と人材育成に関する論点

① 公開データの再利用

- ✓誰に、どのように再利用されることを目指すのか？
- ✓ターゲットによってデータ整備の粒度が異なる

② 根拠データの公開

- ✓データの公開状況と課題は何か？

③ 研究データ管理

- ✓研究データ管理の実践状況と課題は何か？



2

**日本の研究者による研究データの
再利用・公開・管理状況**

The State of Open Science in Japan



研究者



専門家ネットワーク



研究機関

2016

Open Data/RDM

OA

2018

Open Data/RDM

OA

2020

Open Data/RDM

OA

2022

Open Data/RDM

OA

Preprints

Preprints

第5期科学技術基本計画
第6期科学技術・イノベーション基本計画
統合イノベーション計画

RDM
Data Policy

JPCOAR/AXIES

RDM
Data Policy

JPCOAR/AXIES

The State of Open Science in Japan



| | Open Data/RDM + OA | | Preprints | OA + Preprints | Open Data RDM |
|-----|--------------------|------|-----------|----------------|---------------|
| 調査年 | 2016 | 2018 | 2020 | | 2022 |

対象 **科学技術・学術政策研究所 科学技術専門家ネットワーク**
産学官の研究者・技術者・マネージャ等

| | | | | | | |
|-----|------------------|------------------|------------------|------------------|------------------|------------------|
| 回答数 | 1,406 (70.9%) | 1,548 (70.5%) | 1,349 (70.5%) | 1,448 (75.7%) | 1,173 (70.2%) | 1,237 (73.9%) |
|-----|------------------|------------------|------------------|------------------|------------------|------------------|

報告書DOI [10.15108/rm268](https://doi.org/10.15108/rm268) [10.15108/rm289](https://doi.org/10.15108/rm289) [10.15108/rm316](https://doi.org/10.15108/rm316) [10.15108/rm301](https://doi.org/10.15108/rm301) [10.15108/rm327](https://doi.org/10.15108/rm327) [10.15108/rm335](https://doi.org/10.15108/rm335)

調査方法

対象 科学技術専門家ネットワーク（大学・公的機関・団体、企業）

方法 オンライン質問紙調査（Cuenote）

期間 2022年10月12日～2022年11月28日

有効回答 1,237（回答率 73.9%）



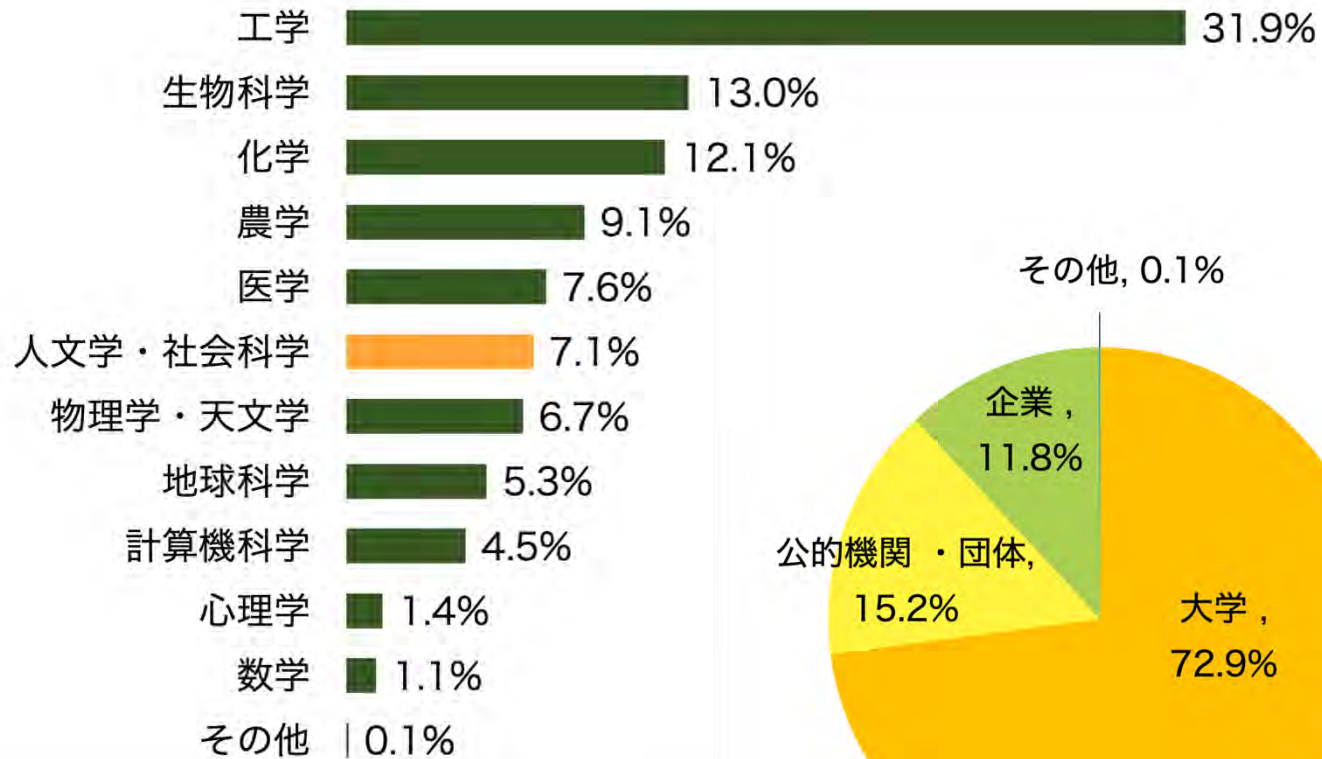
以下では、主に2022年調査の結果を示す。<https://doi.org/10.15108/rm335>

質問項目

| 章 | カテゴリ | 問題数 |
|---|---------------------|-----|
| 1 | 回答者に関する質問 | 2 |
| 2 | 研究データの提供（共有） | 2 |
| 3 | 公開データの入手 | 4 |
| 4 | データマネジメントプラン（DMP） | 4 |
| 5 | 研究データの公開 | 8 |
| 6 | 研究データの整備・公開・保存（RDM） | 7 |
| 7 | データ公開に対する評価 | 2 |
| 8 | 自由回答 | 1 |
| | 合計 | 30 |

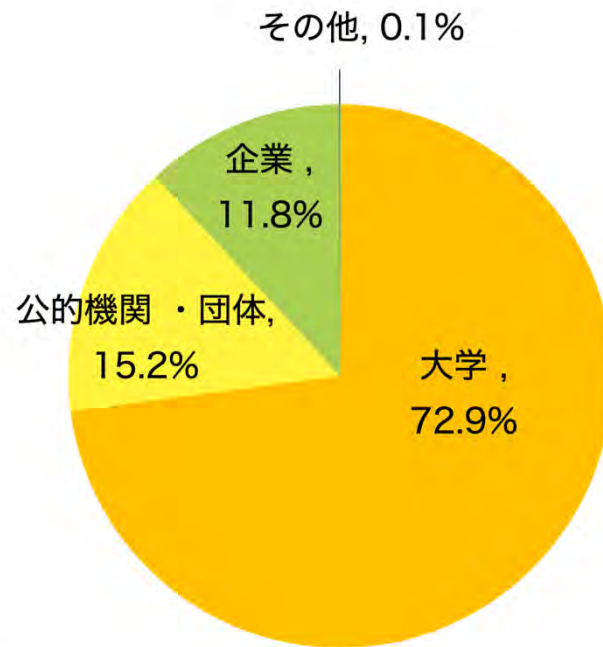
回答者の属性 (n=1,237)

研究分野



年齢層

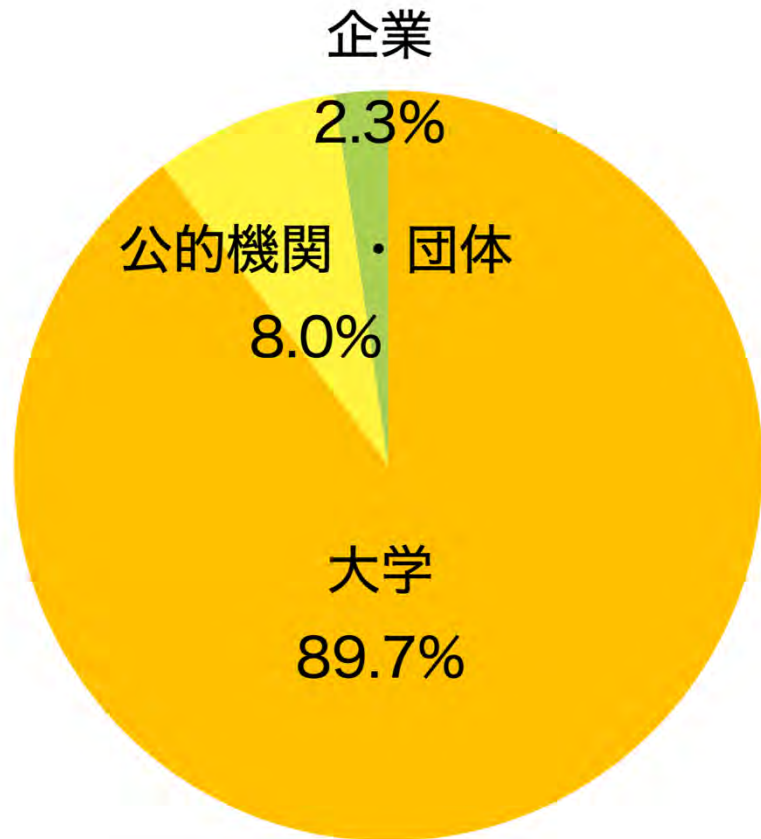
| 年代 | 回答者数 | 比率 |
|-------|-------|--------|
| 30代以下 | 263 | 21.3% |
| 40代 | 641 | 51.8% |
| 50代 | 255 | 20.6% |
| 60代以上 | 263 | 21.3% |
| 合計 | 1,237 | 100.0% |



所属機関

人文学・社会科学分野 (n=88)

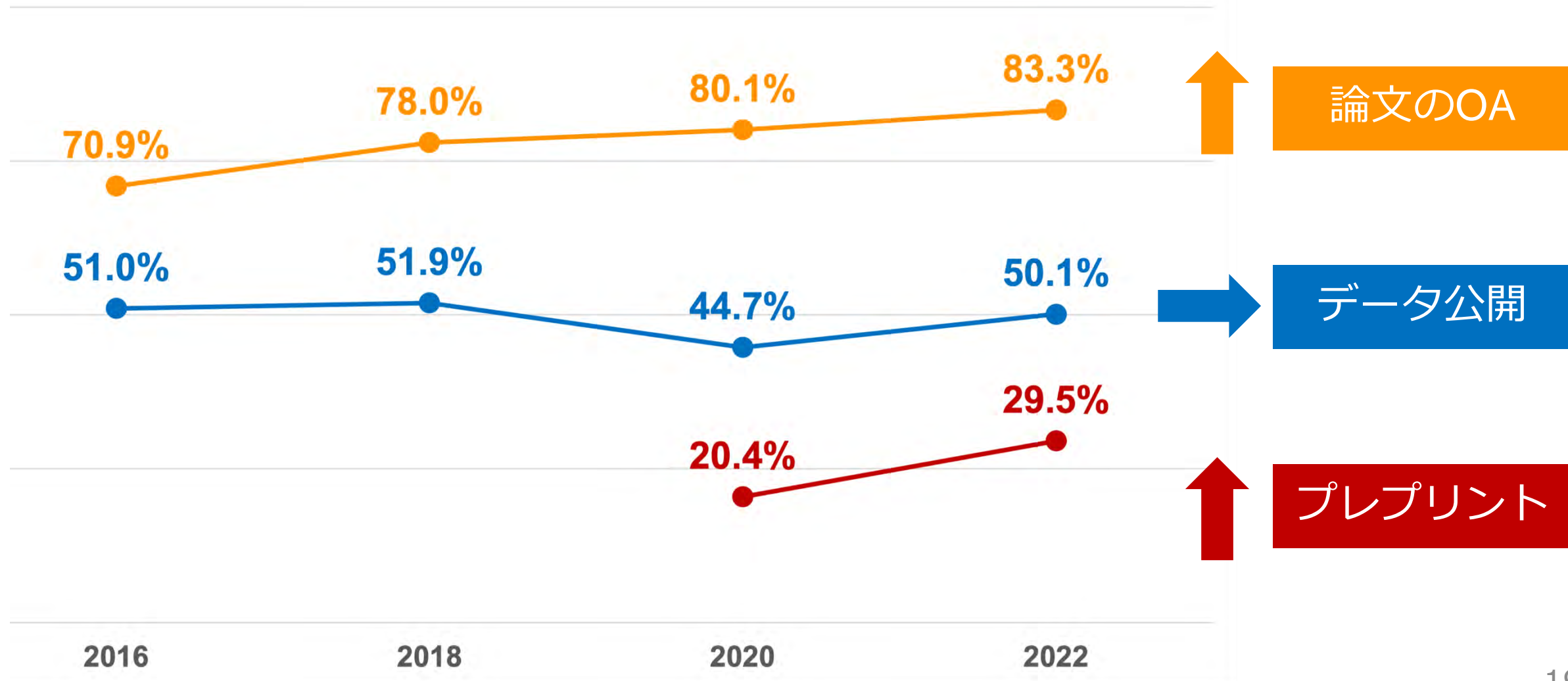
所属機関



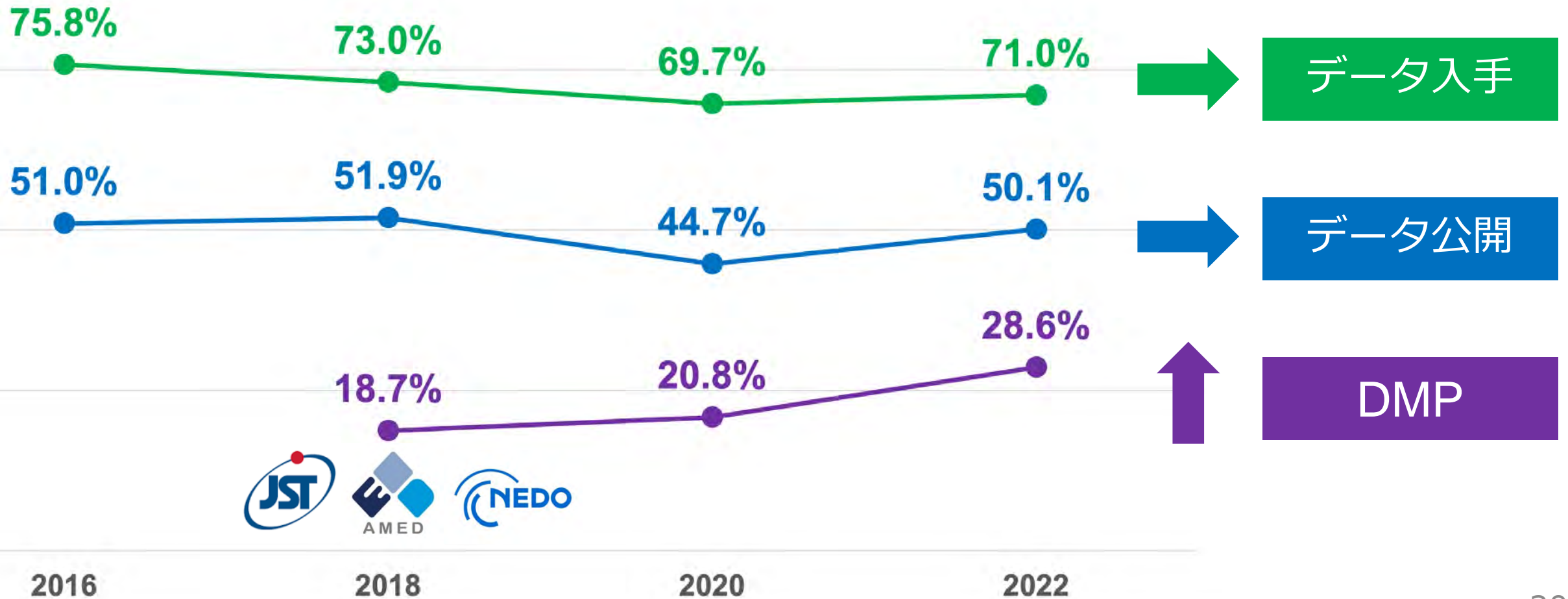
年齢層

| 年代 | 回答者数 | 比率 |
|-------|------|--------|
| 30代以下 | 14 | 15.9% |
| 40代 | 42 | 47.7% |
| 50代 | 22 | 25.0% |
| 60代以上 | 10 | 11.4% |
| 合計 | 88 | 100.0% |

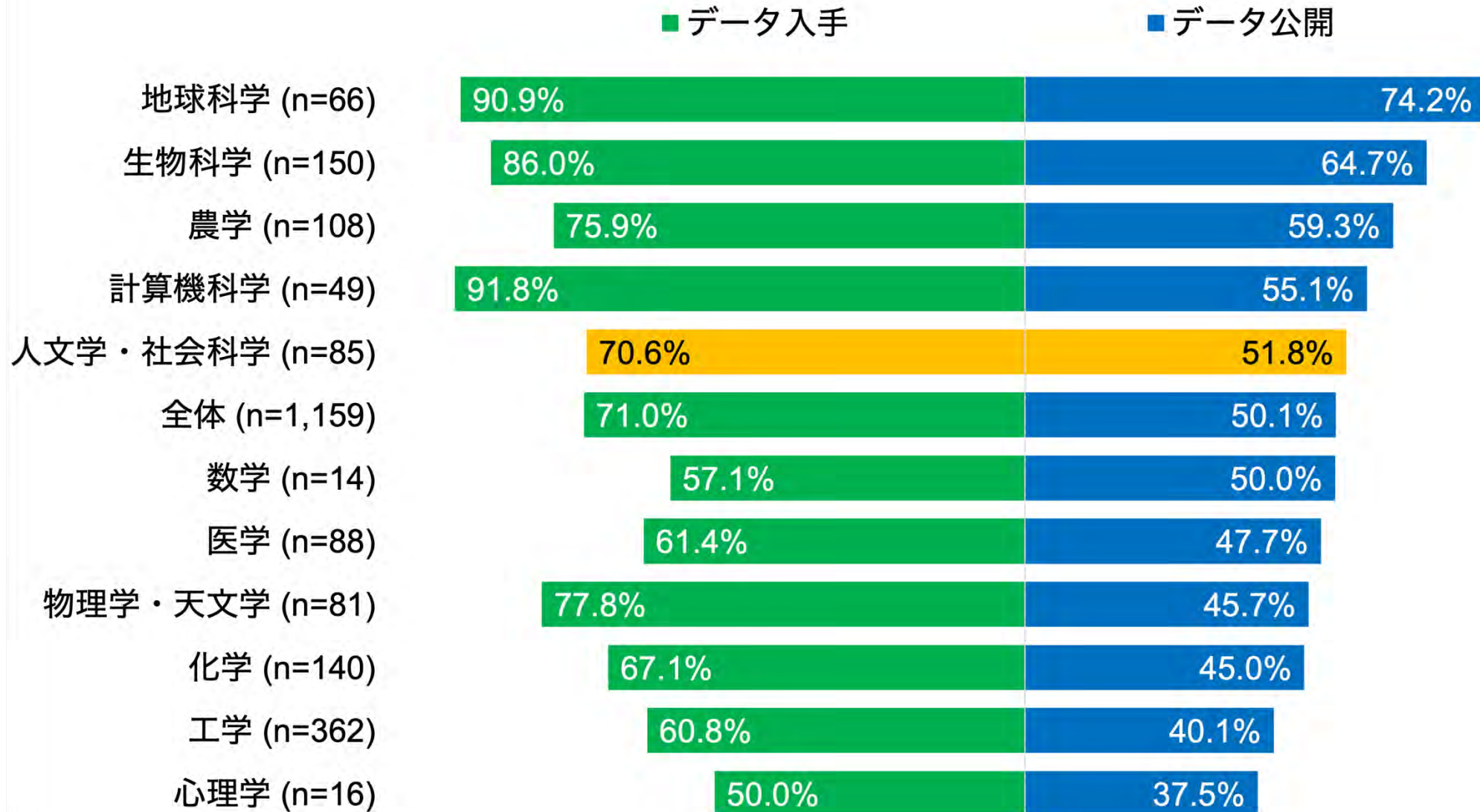
日本の研究者によるOSの実践状況 (2016-2022)



日本の研究者によるデータ共有関連の実践状況 (2016-2022)

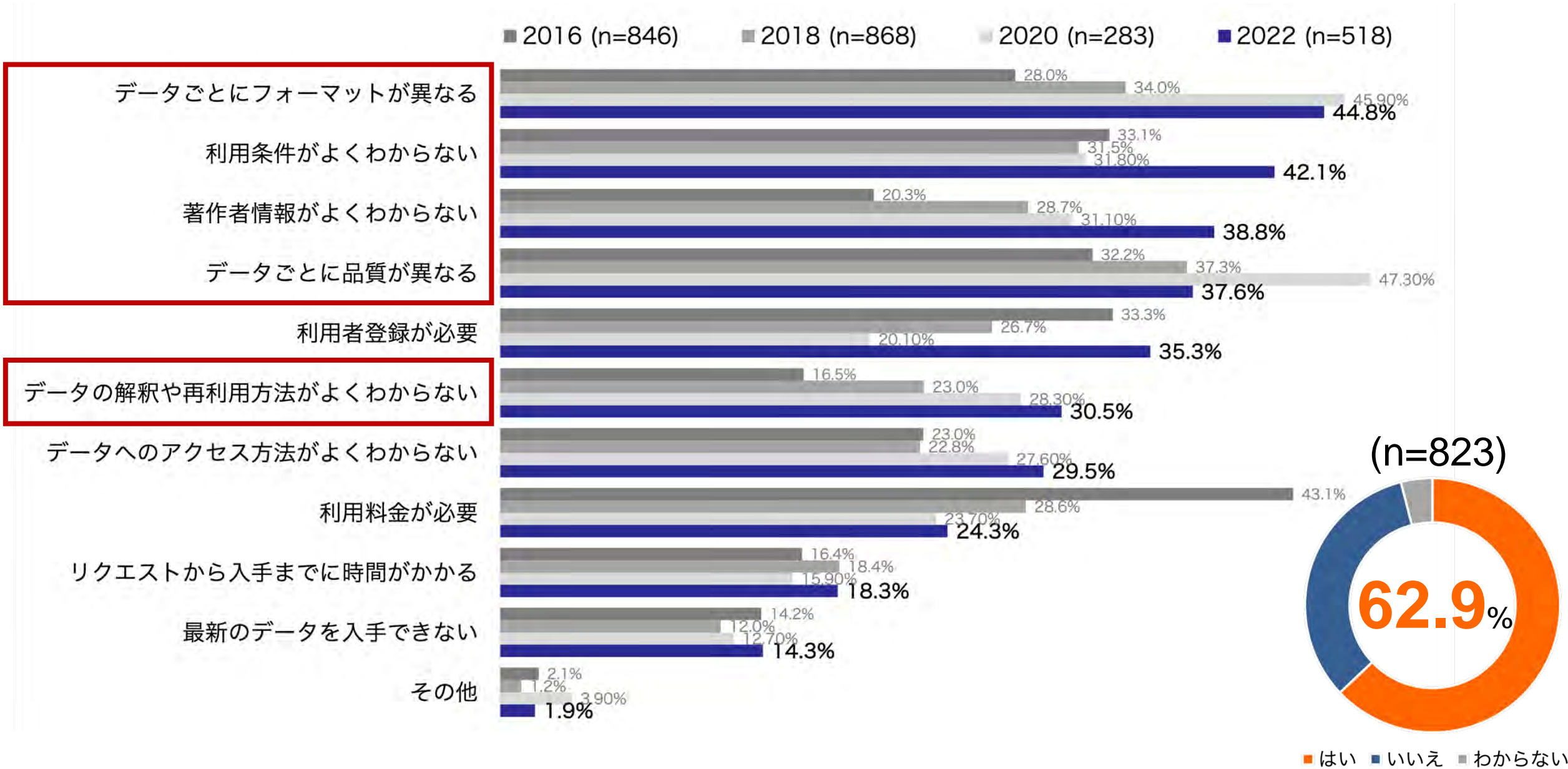


分野別：データ入手とデータ公開経験 (2022)



($r=0.797$, $p<0.01$)

データ入手における問題（複数選択）

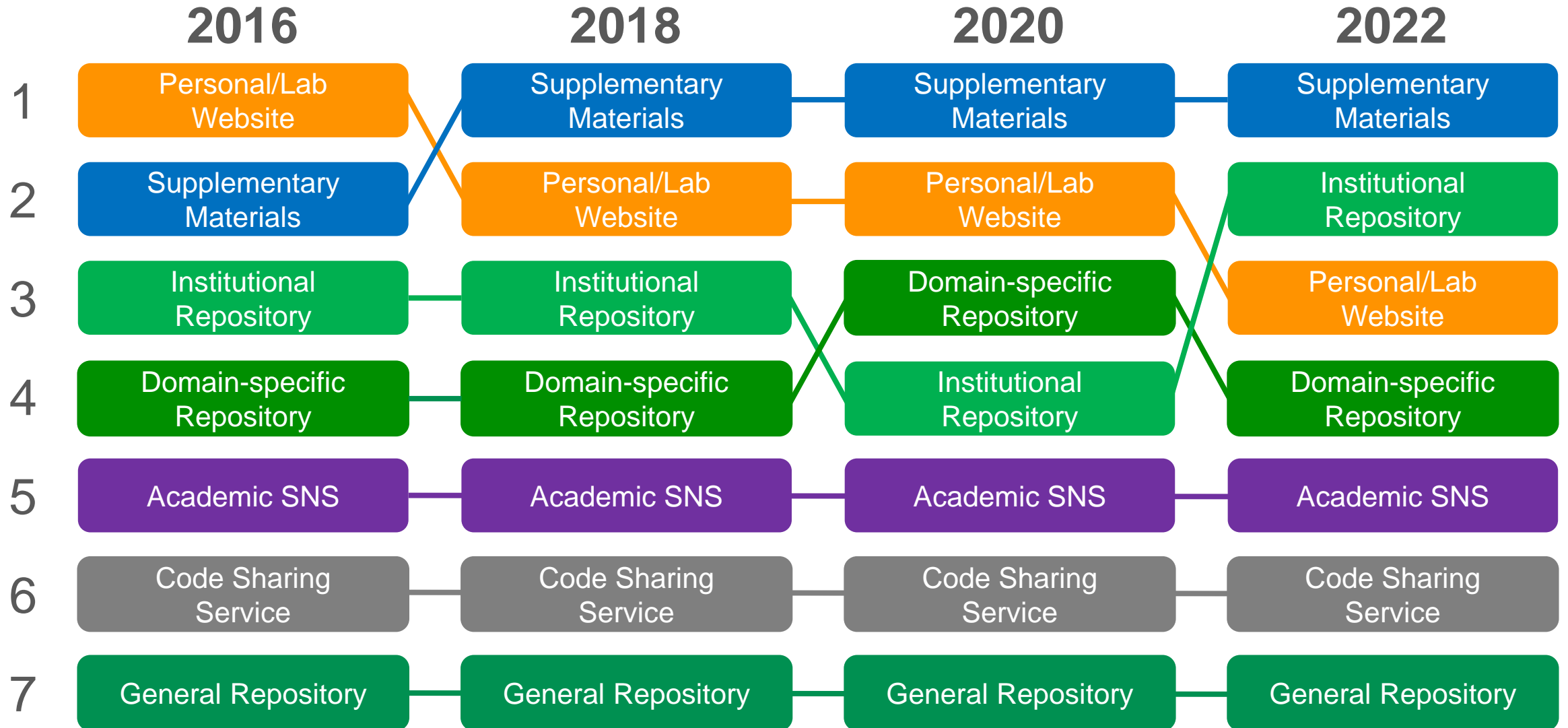


①公開データの再利用（現状まとめ）

- データの入手経験をもつ人文学・社会科学分野（HSS）の研究者は**70.6%**
- 公開データを入手する際に問題があると認識している研究者は**62.9%**
 - ✓データごとにフォーマットが異なる
 - ✓利用条件がよくわからない
 - ✓著作者情報がよくわからない
 - ✓データごとに品質が異なる
 - ✓データの解釈や再利用方法がよくわからない

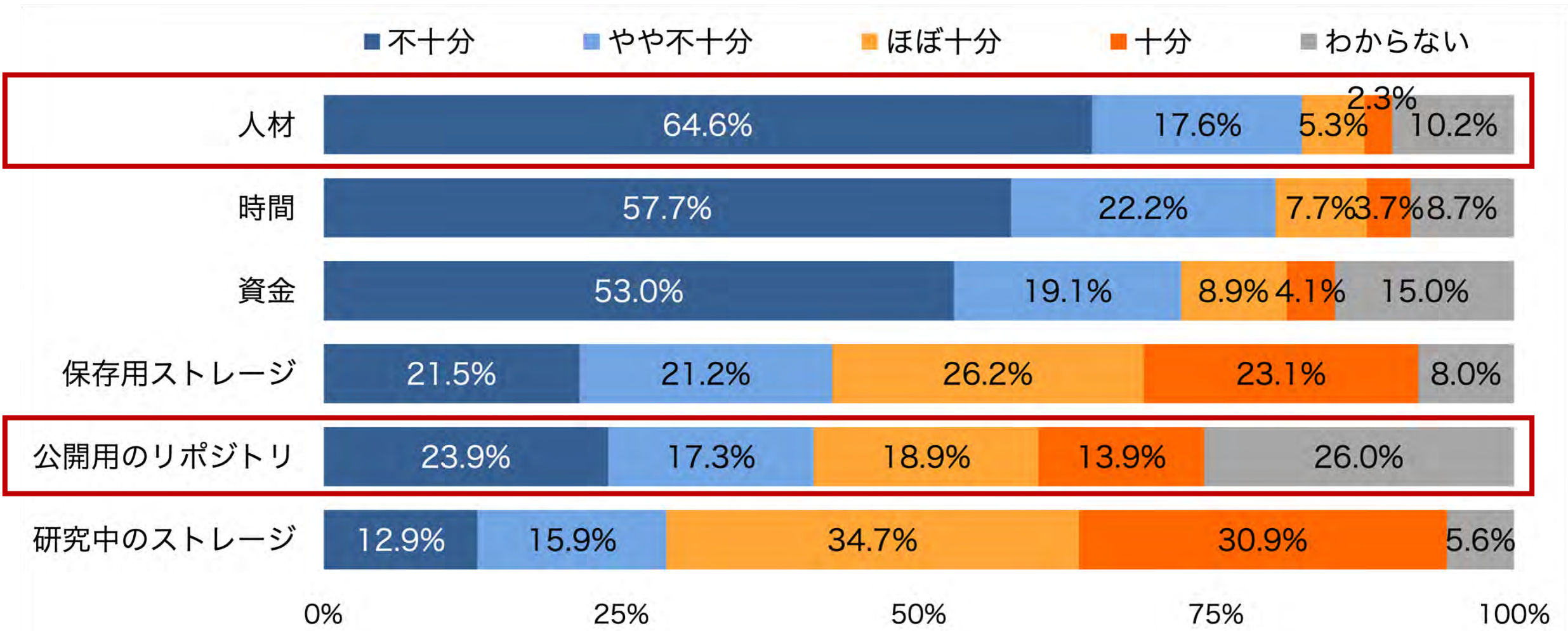
再利用を推進するためには適切なデータキュレーションが重要

データ公開の方法

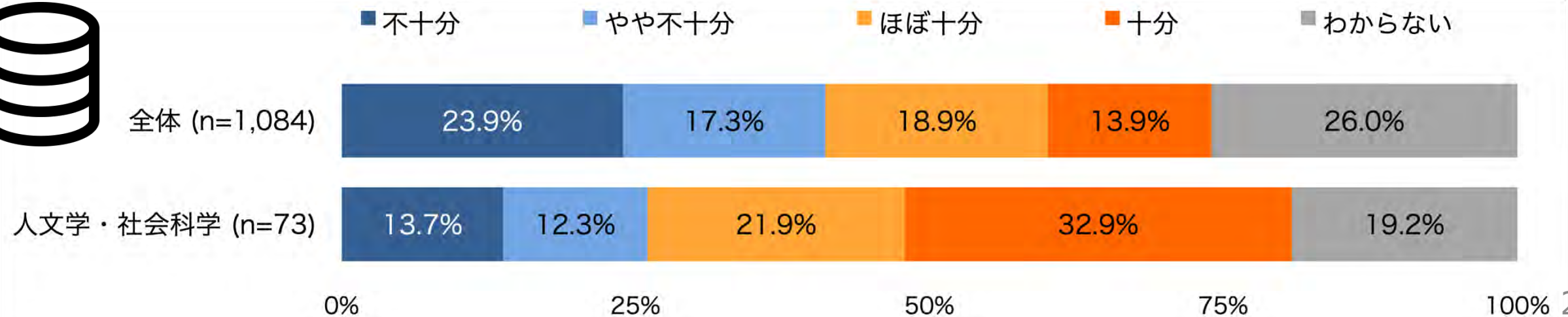
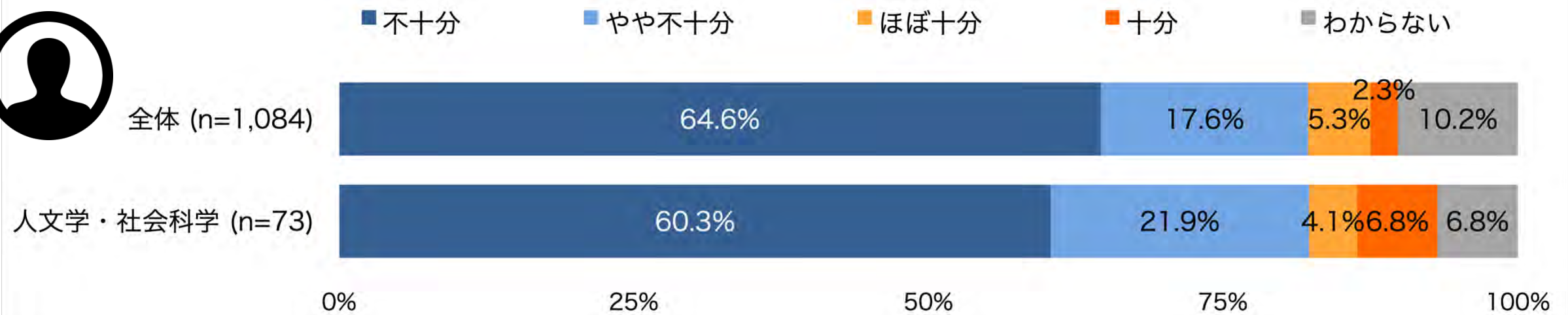
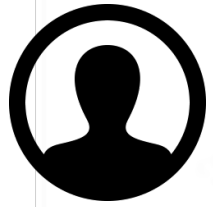


データ公開・整備に必要な資源の充足度

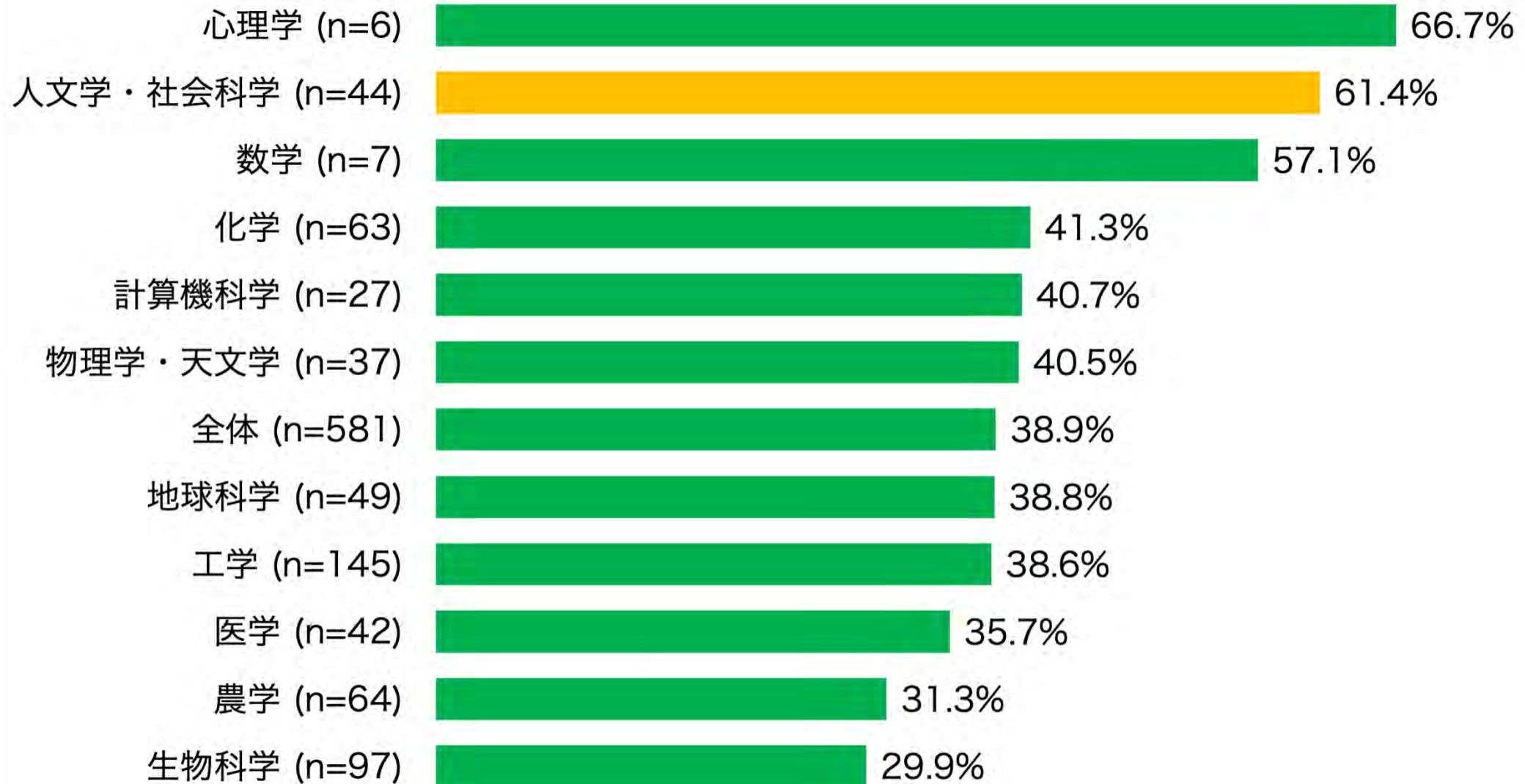
(n=1,084)



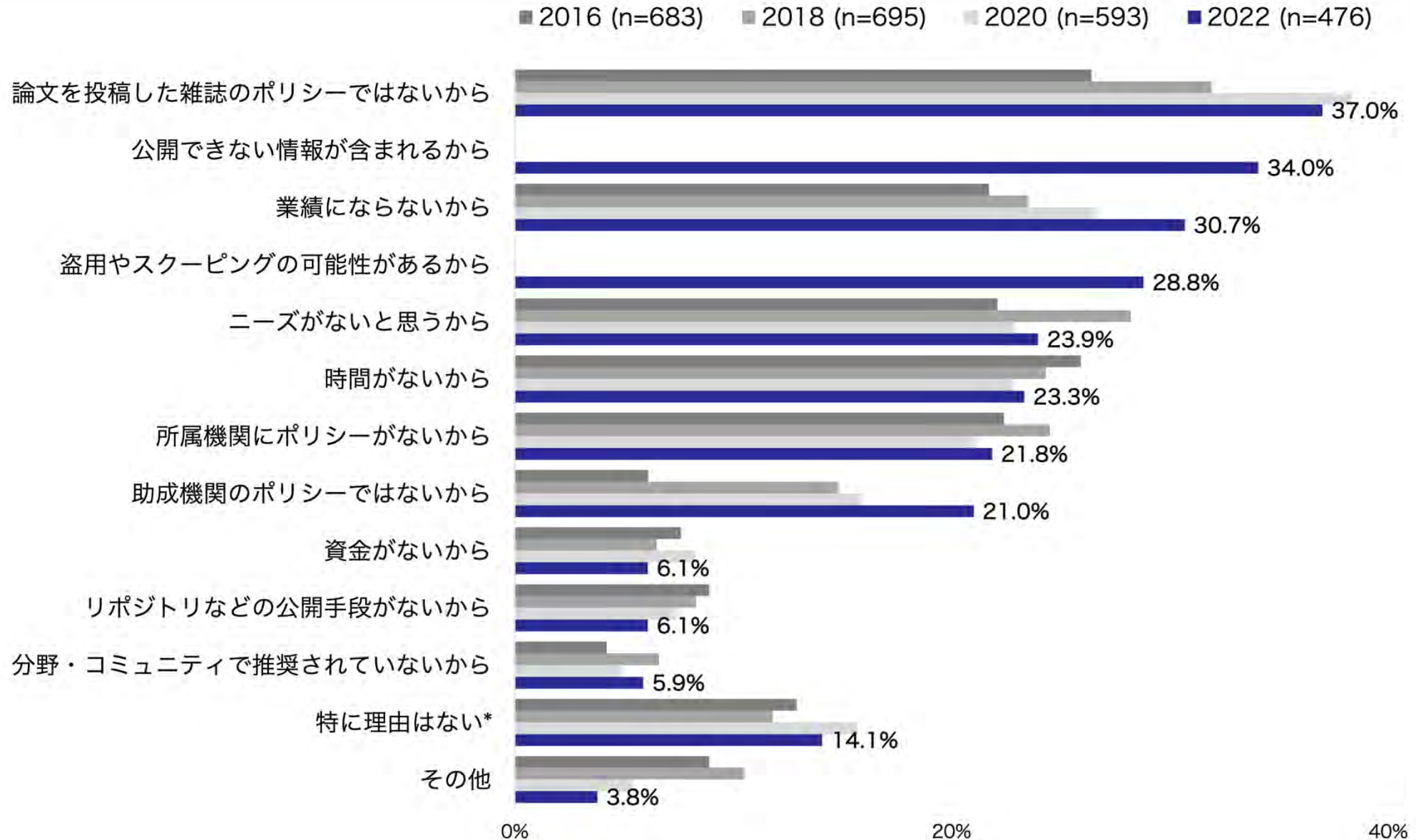
HSS: 人材と公開用リポジトリの充足度



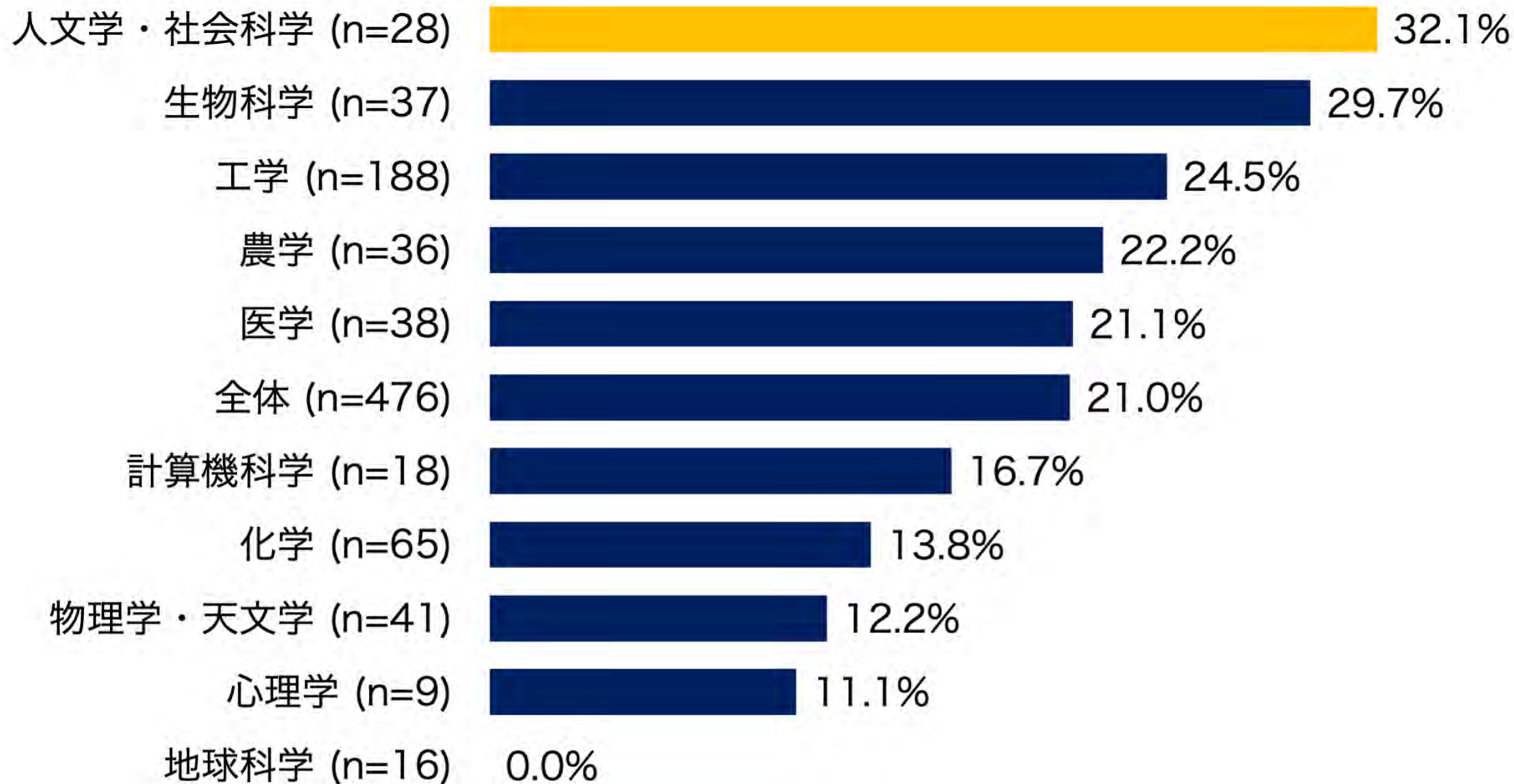
分野別：機関リポジトリでの公開 (2022)



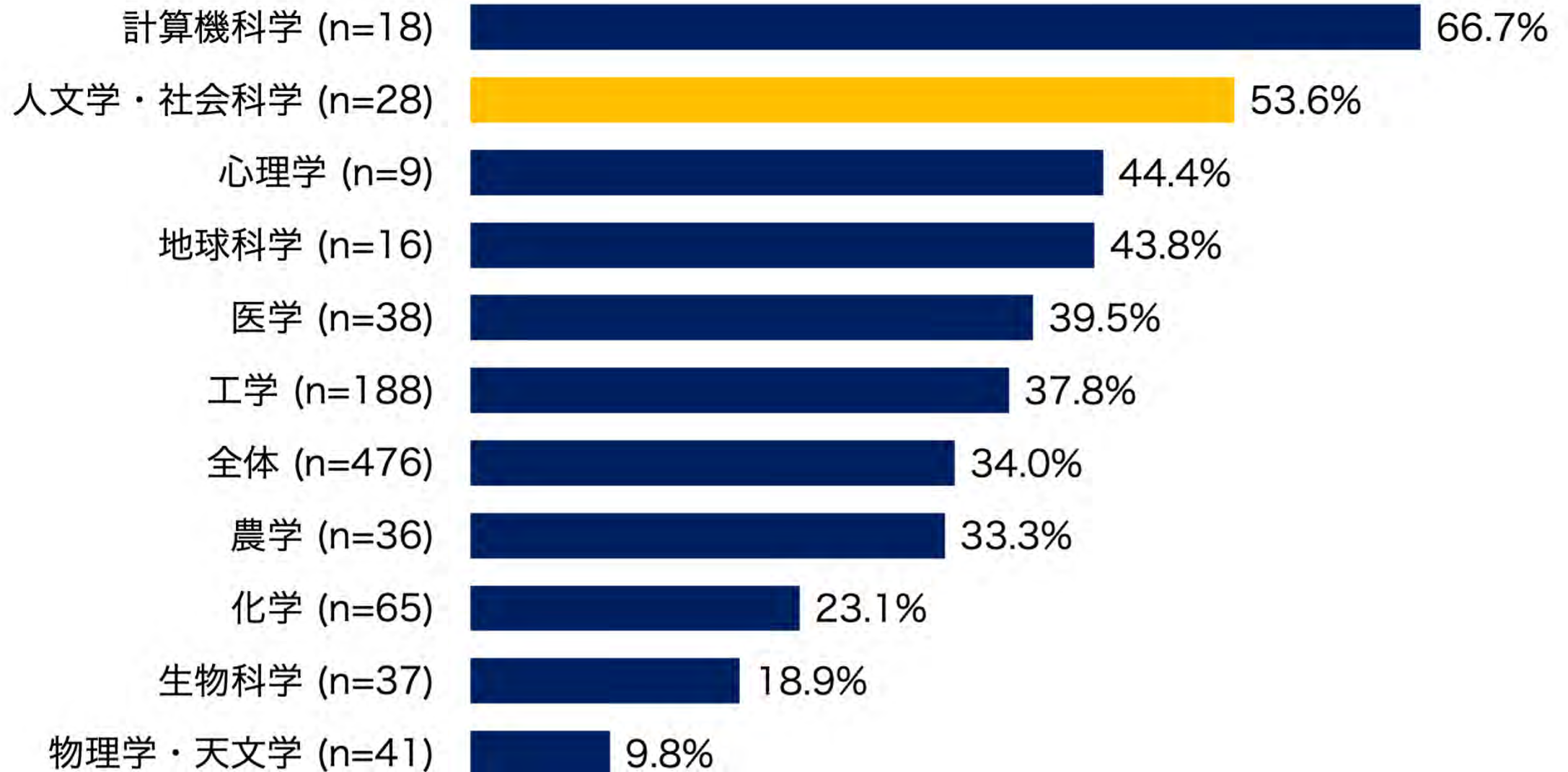
データの未公開理由（複数選択）



助成機関のポリシーではないから



公開できない情報が含まれているから

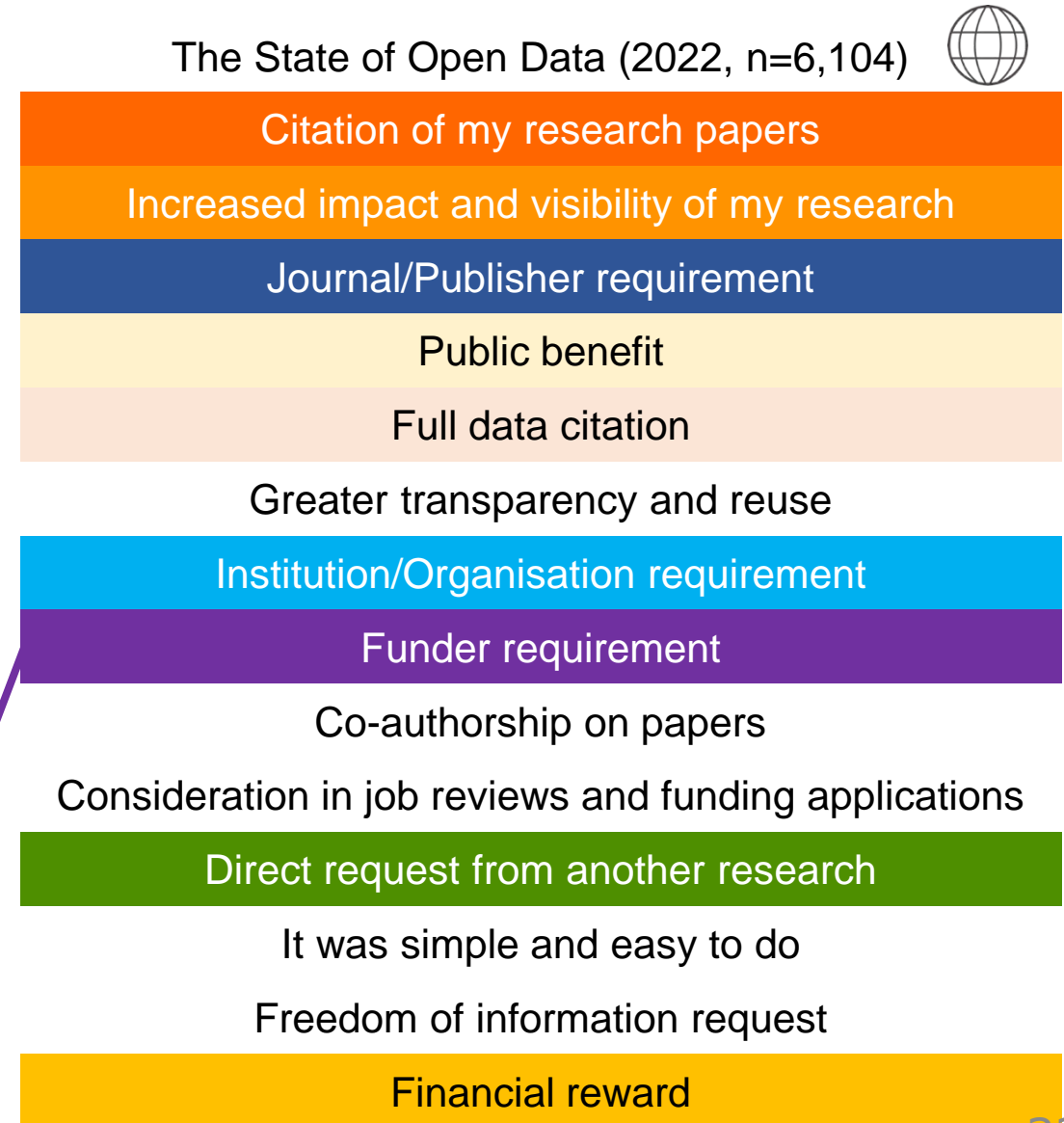


②根拠データの公開（現状まとめ）

- データ公開経験をもつHSSの研究者は**51.8%**（全体50.1%）
- 公開用リポジトリの充足感は全体と比較して高い
- 機関リポジトリによる公開経験をもつHSSの研究者は**61.4%**
- HSSの特徴的な未公開理由は“助成機関のポリシーではないから”、“公開できない情報が含まれるから”

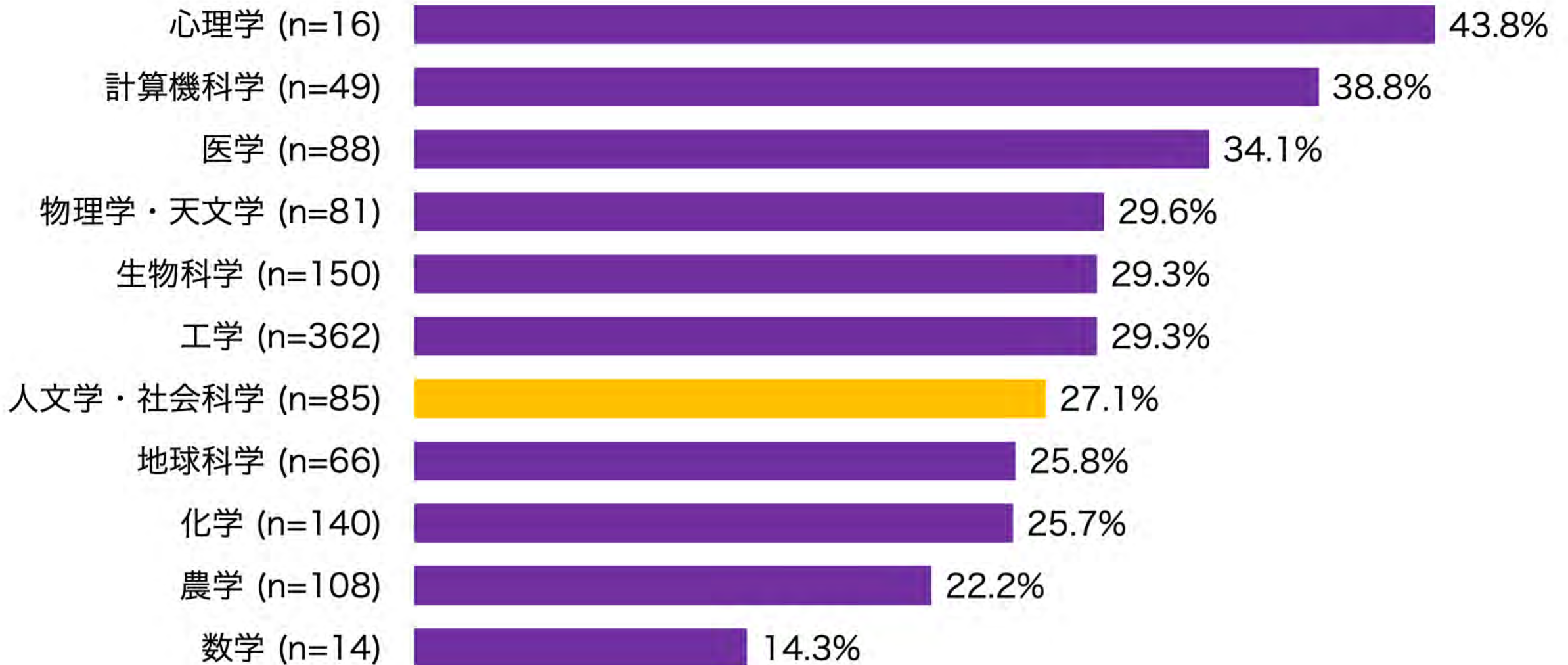
助成機関のDMP要求によりさらにデータ公開が進む可能性

データの公開理由 (国際調査との比較)

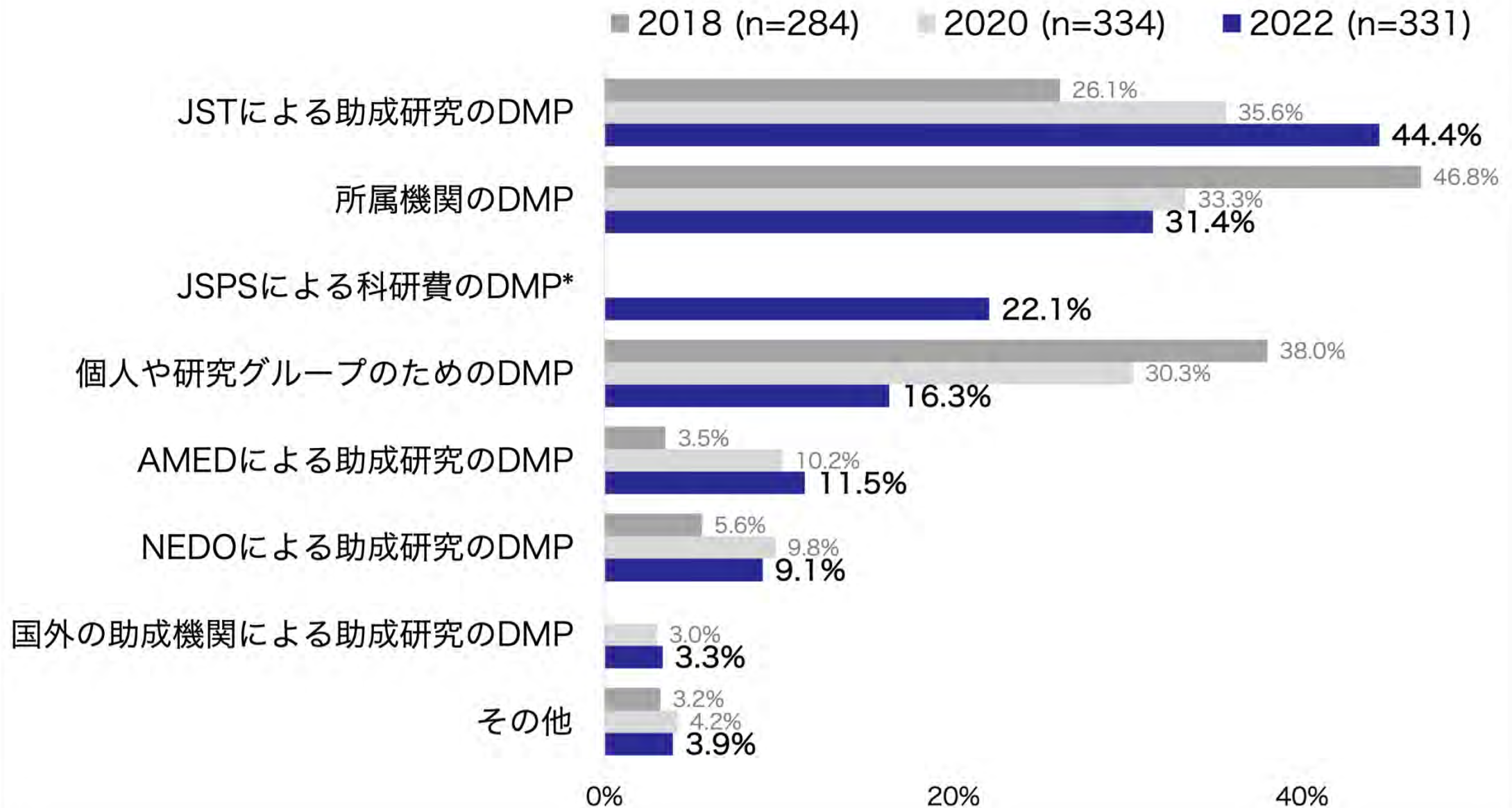


*15: Open data badge, 16: It was a field/industry expectation

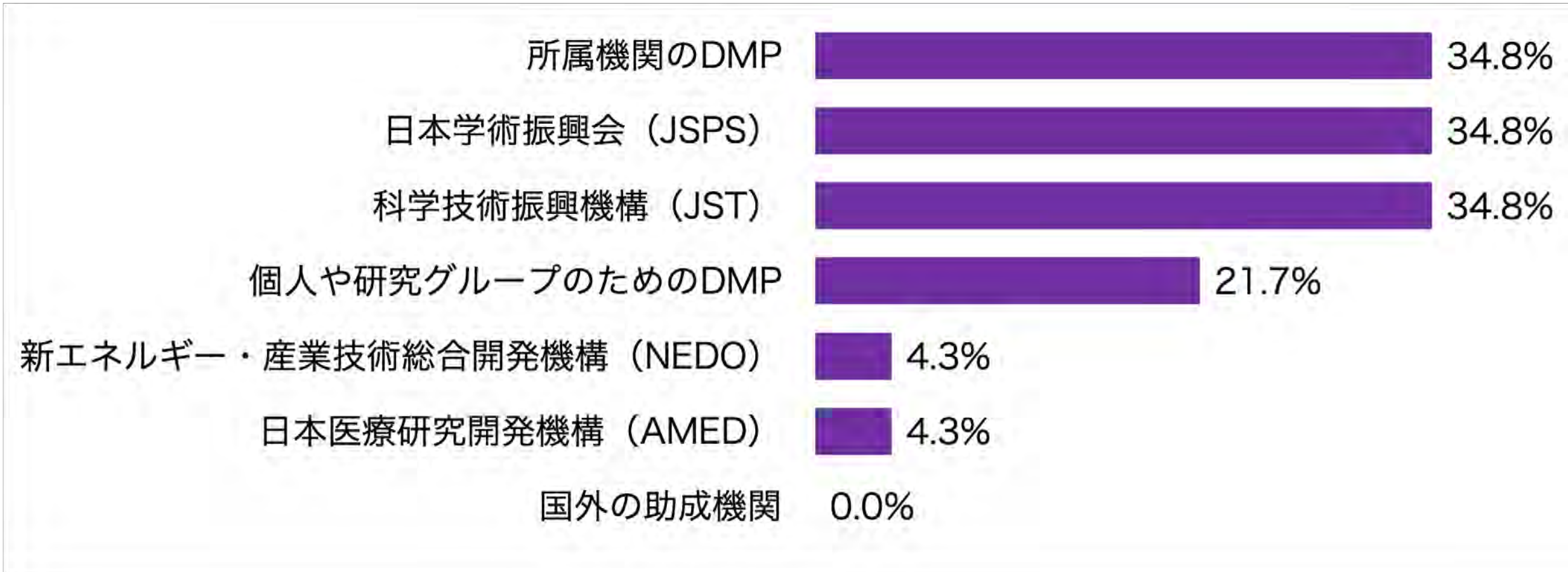
分野別：DMP作成経験



作成経験があるDMP



作成経験があるDMP : HSS (n=23)



“「DMP」は、研究経験や論文の掲載経験のないもの研究実態をまったく知らないで「研究不正対策」の視点だけから研究者を管理する側の「保身」に基づき生まれた発想と考える。研究開始から終了までの数年にわたって収集されたデータには、課題に直接関係しない資料や論文も数多く含まれ、膨大な数に上るが、ほとんどの（とくにアカデミアに属する）研究者は、このような膨大なデータを逐一整理・保存しているわけではなく、また、そのような行為にさく研究時間もないことから、**「DMP作成」は、研究そのものを阻害する**だけであり、必要悪と考える”

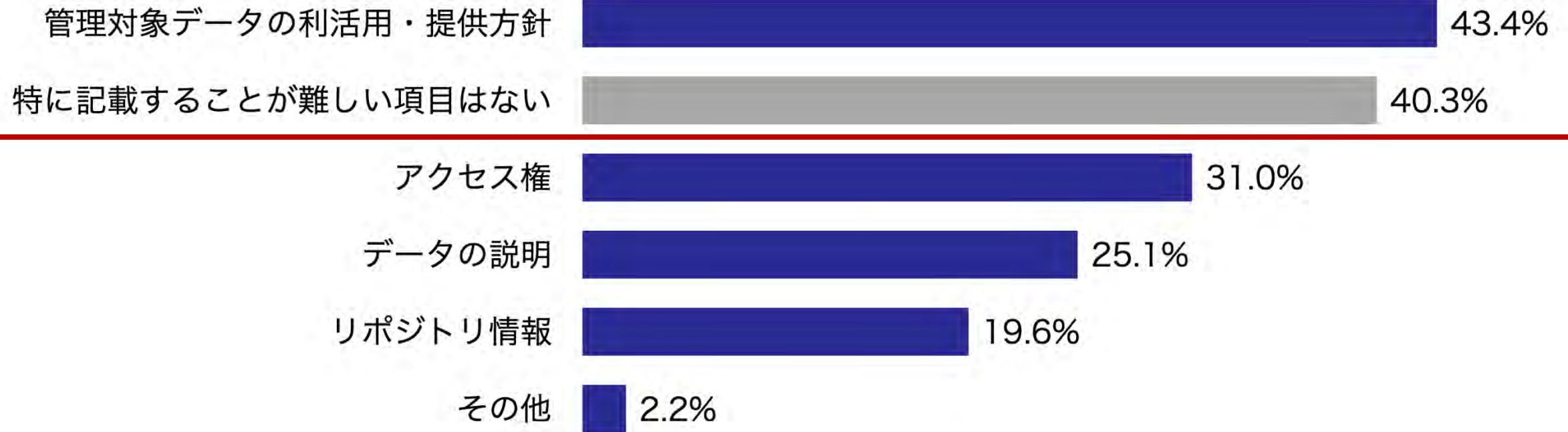
「メタデータの共通項目」 (内閣府)

- | | |
|----------------|-----------------------------------|
| (1) 資金配分機関情報 | (10) 概略データ量 |
| (2) e-Radの課題番号 | (11) 管理対象データの利活用・提供方針：アクセス権、公開予定日 |
| (3) プロジェクト名 | (12) リポジトリ情報 |
| (4) データNo. | (13) データ作成者のe-Rad研究者番号 |
| (5) データの名称 | (14) データ管理機関 |
| (6) 掲載日・掲載更新日 | (15) データ管理者のe-Rad研究者番号 |
| (7) データの説明 | (16) 連絡先 |
| (8) データの分野 | |
| (9) データ種別 | |

公的資金による研究データの管理・利活用に関する基本的な考え方
統合イノベーション戦略推進会議. 2021-04-27.

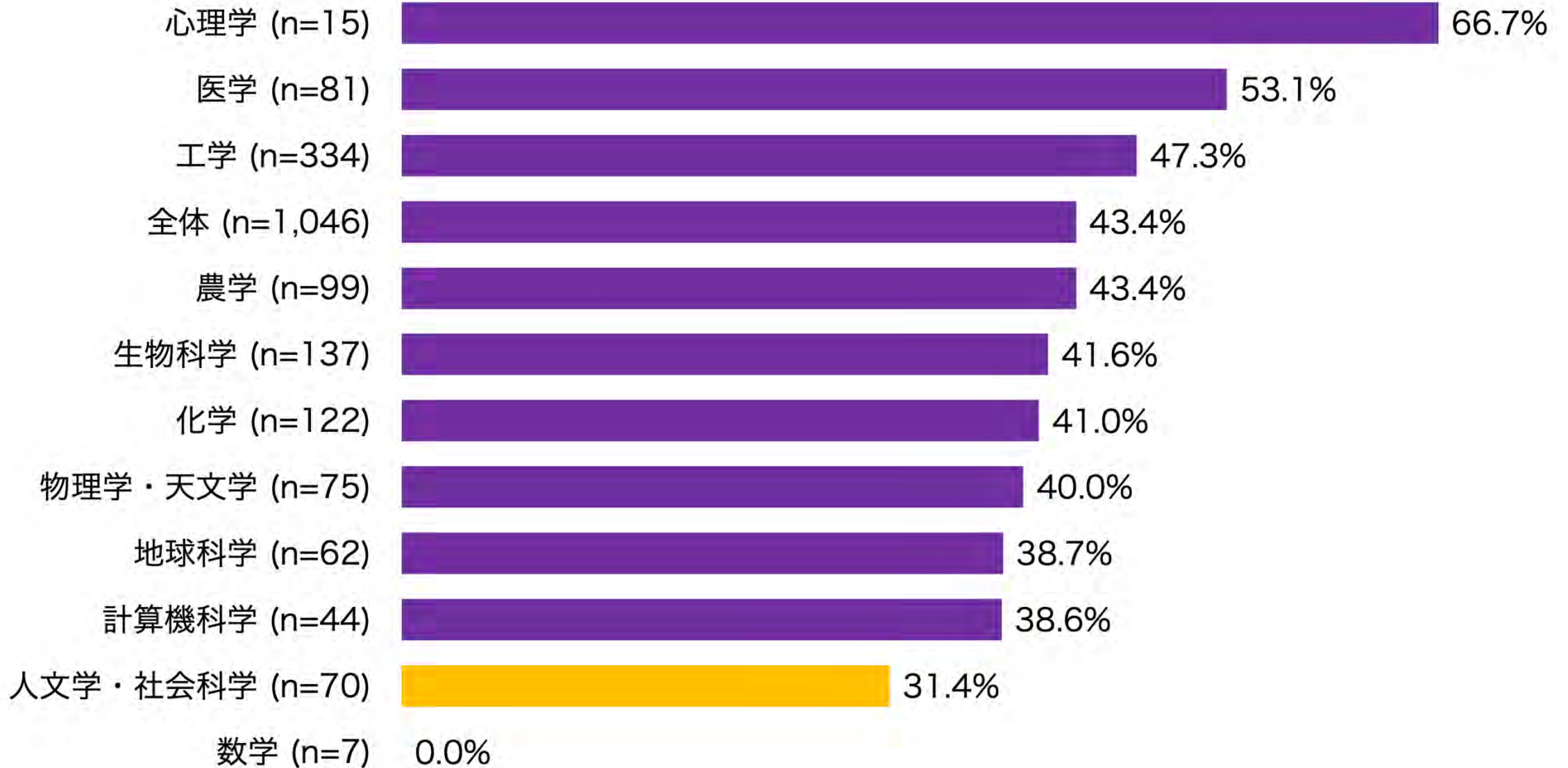
<https://www8.cao.go.jp/cstp/tyousakai/kokusaiopen/sanko1.pdf>

記述が困難な「メタデータ共通項目」*

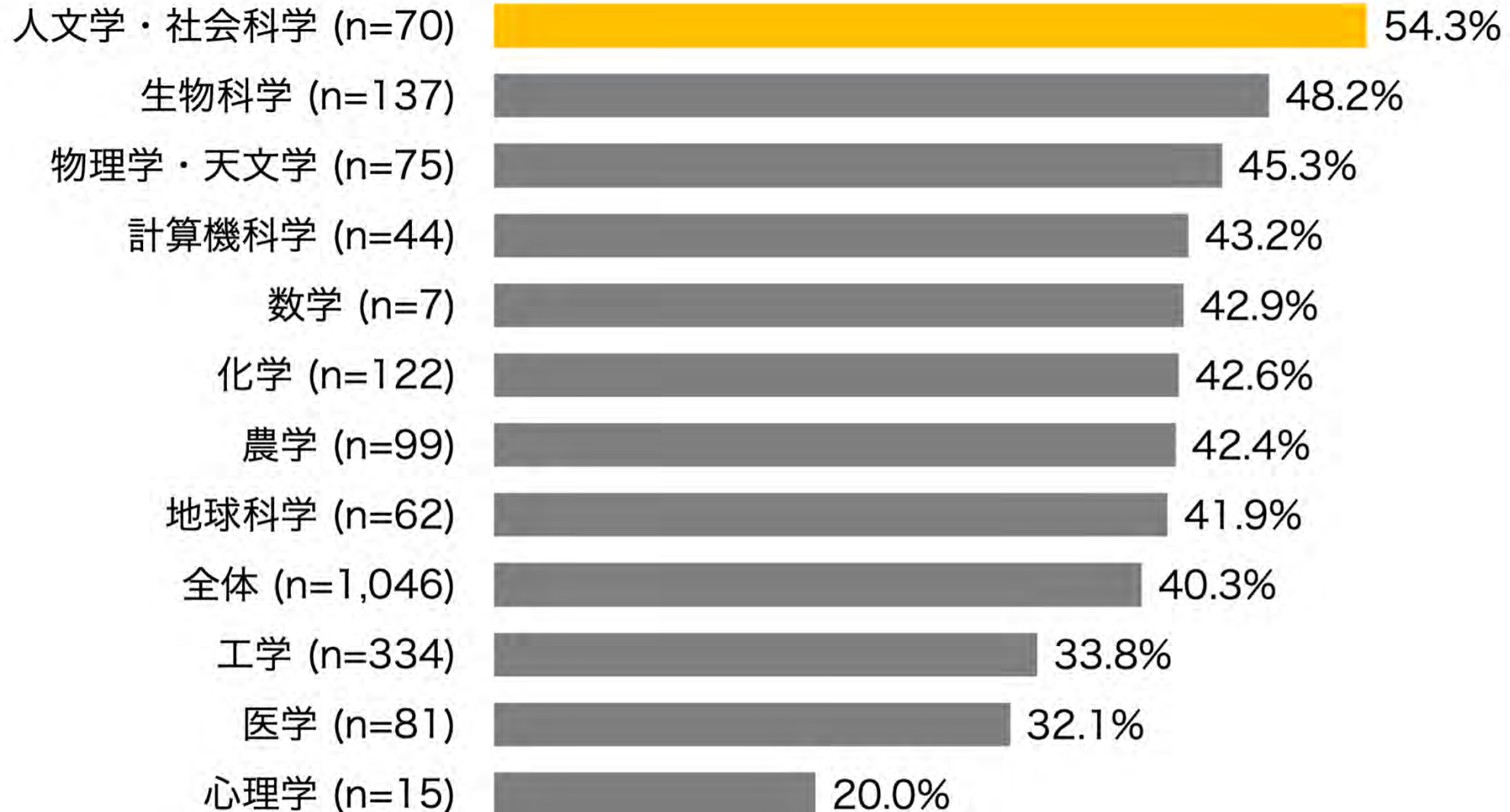


* 「公的資金による研究データの管理・利活用に関する基本的な考え方」で示された、DMPに掲載する「メタデータの共通項目」

分野別：利活用・提供方針



分野別：難しい項目はない



「作成は可能だが労力がかかる」

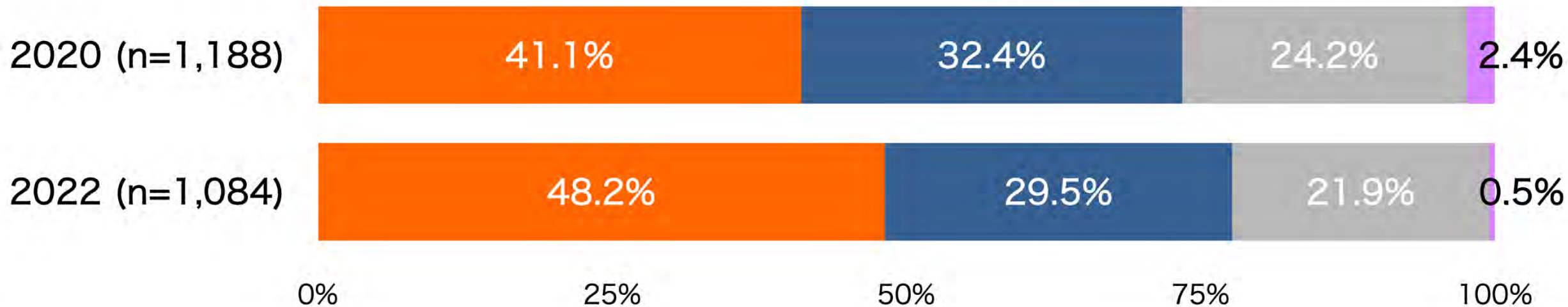
“**時間と精神的労力**さえかければ、上記に難しいことはありません。しかし、上記を作成・記載するのに非常に時間・精神的労力が必要で、その意味では全て「難しい」です”

“記載すること自体は可能と思われるが、**記載した経験がなく、ノウハウが無い**ため、いずれについても規準がわからない。そういう意味で、全て難しく感じる”

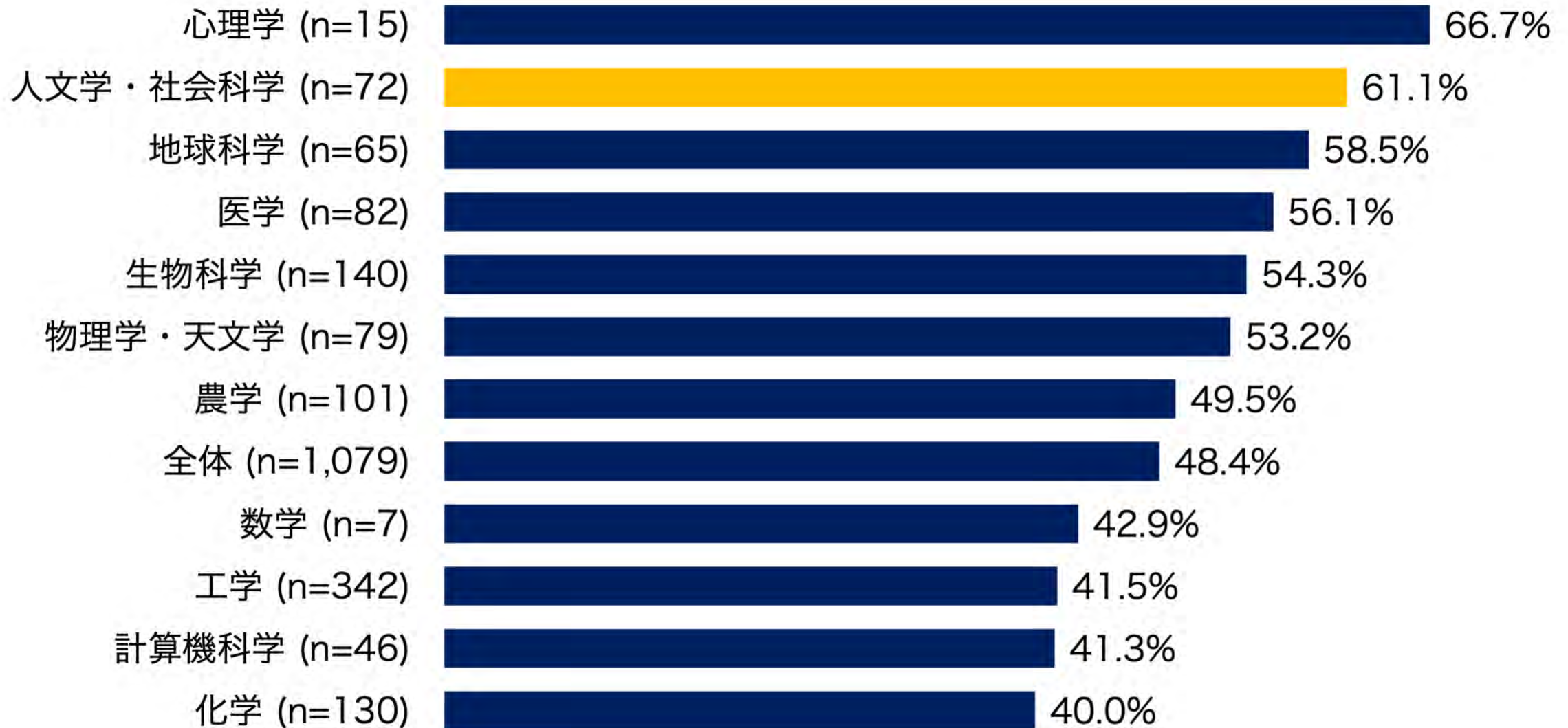
“項目が多すぎるので、本当に必要であれば、**自動生成**できる枠組みがほしい”

研究データ管理（RDM）の依頼意思

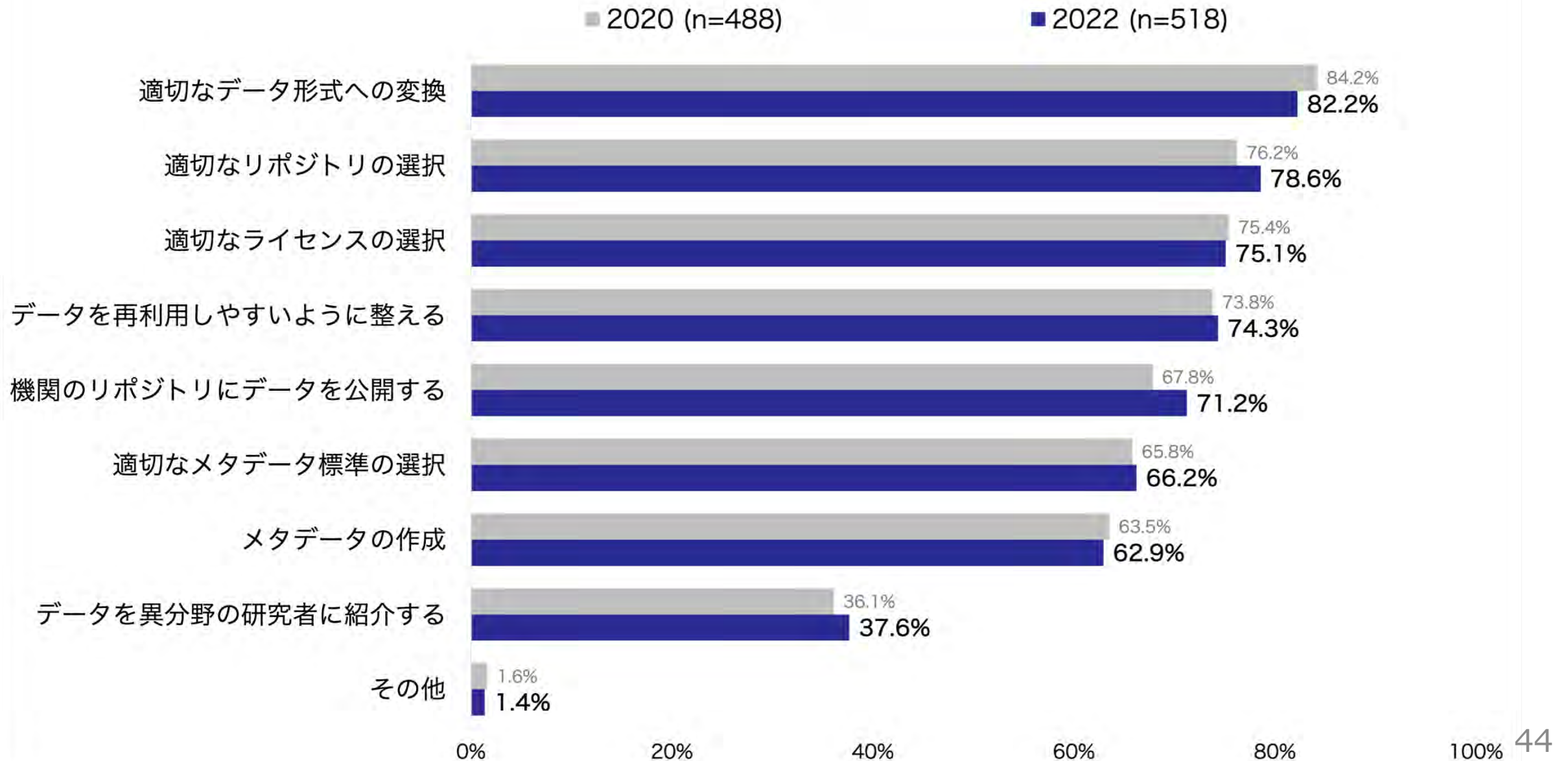
■ はい ■ いいえ ■ わからない ■ 無回答



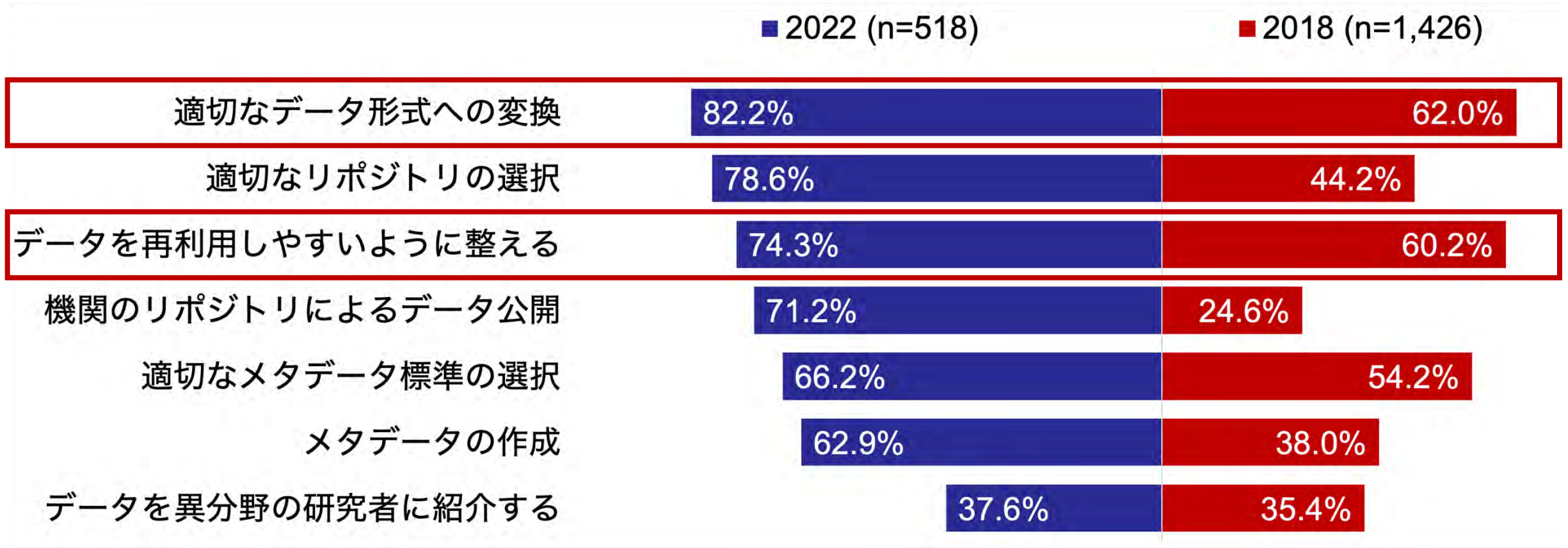
分野別：RDMの依頼意思



RDMに関して依頼したい項目（複数選択）

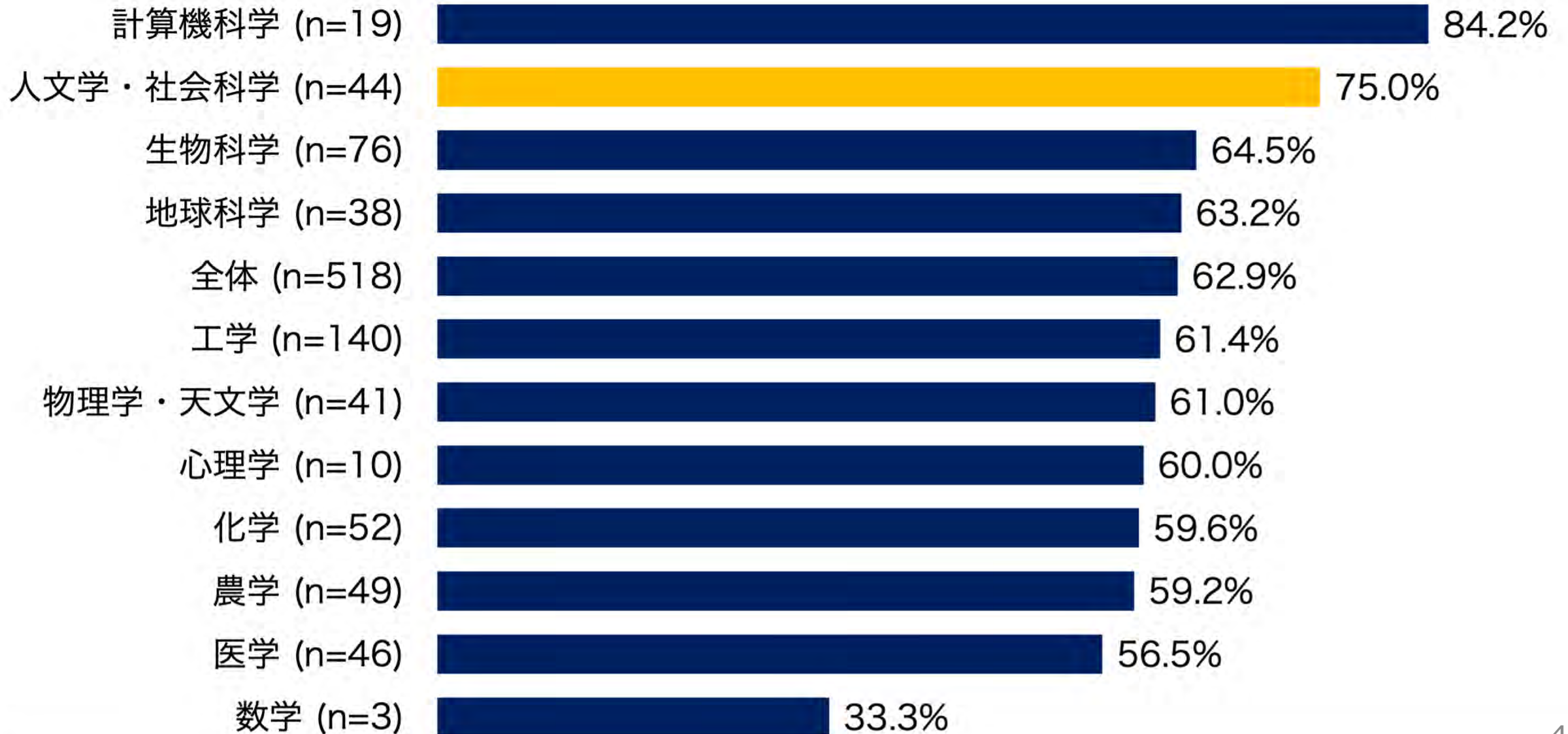


RDMを依頼したい項目と専門性の高さ

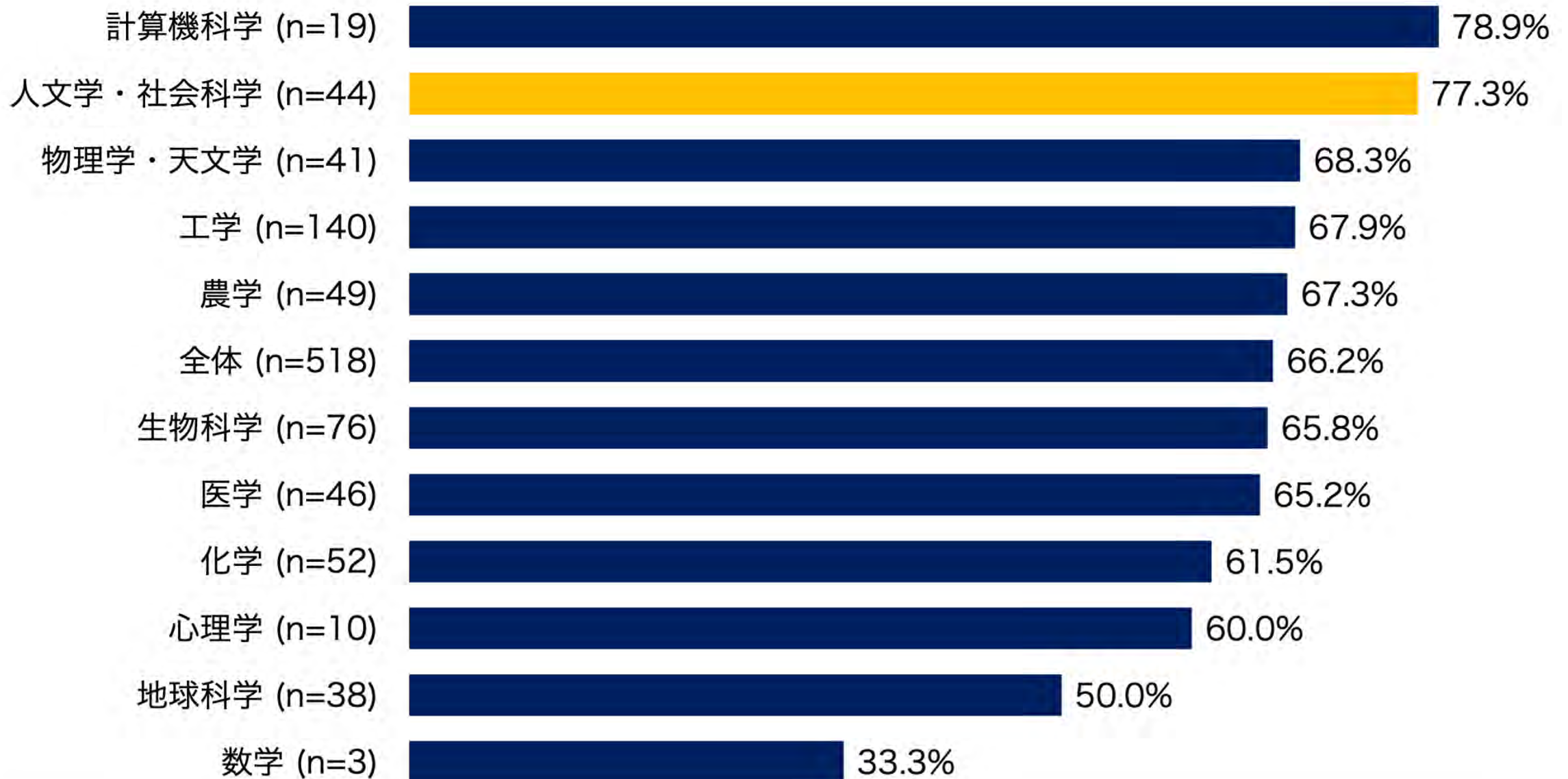


研究者が依頼を希望する項目は専門性が高い

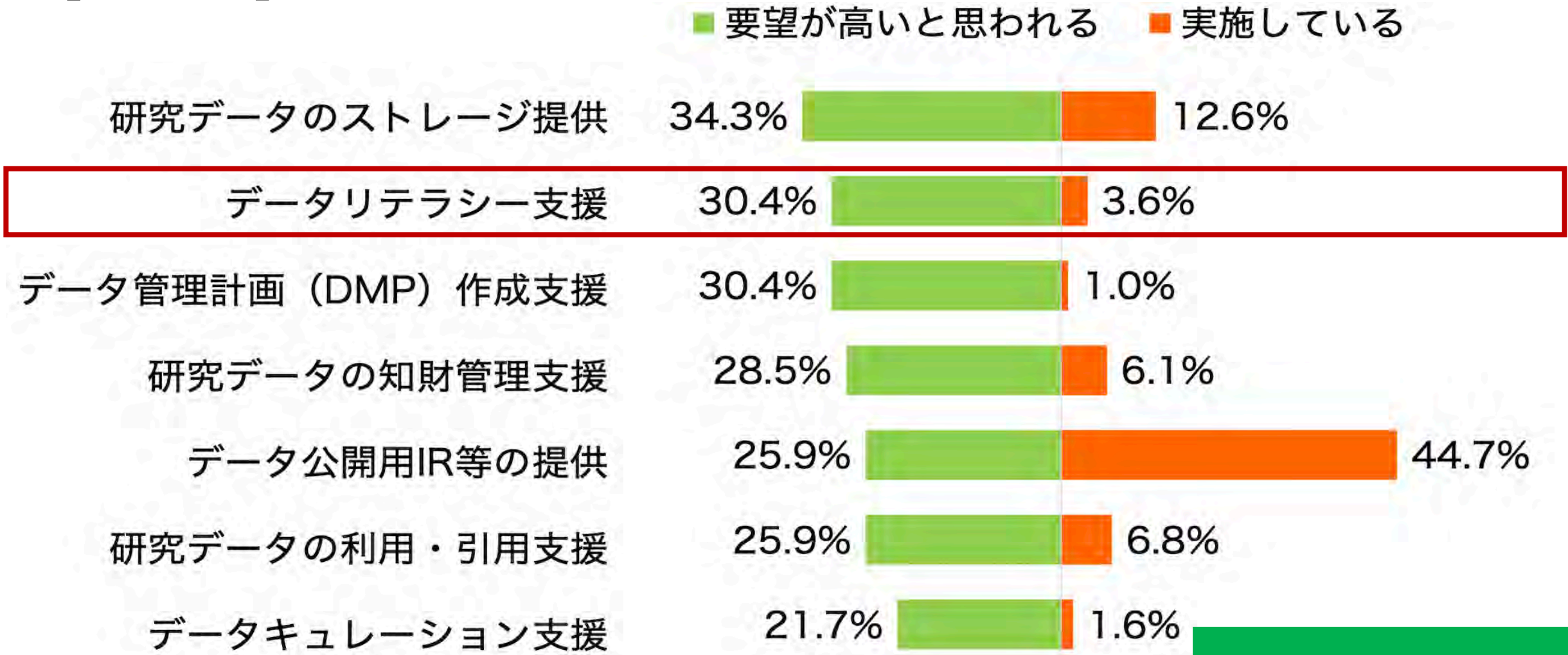
メタデータの作成



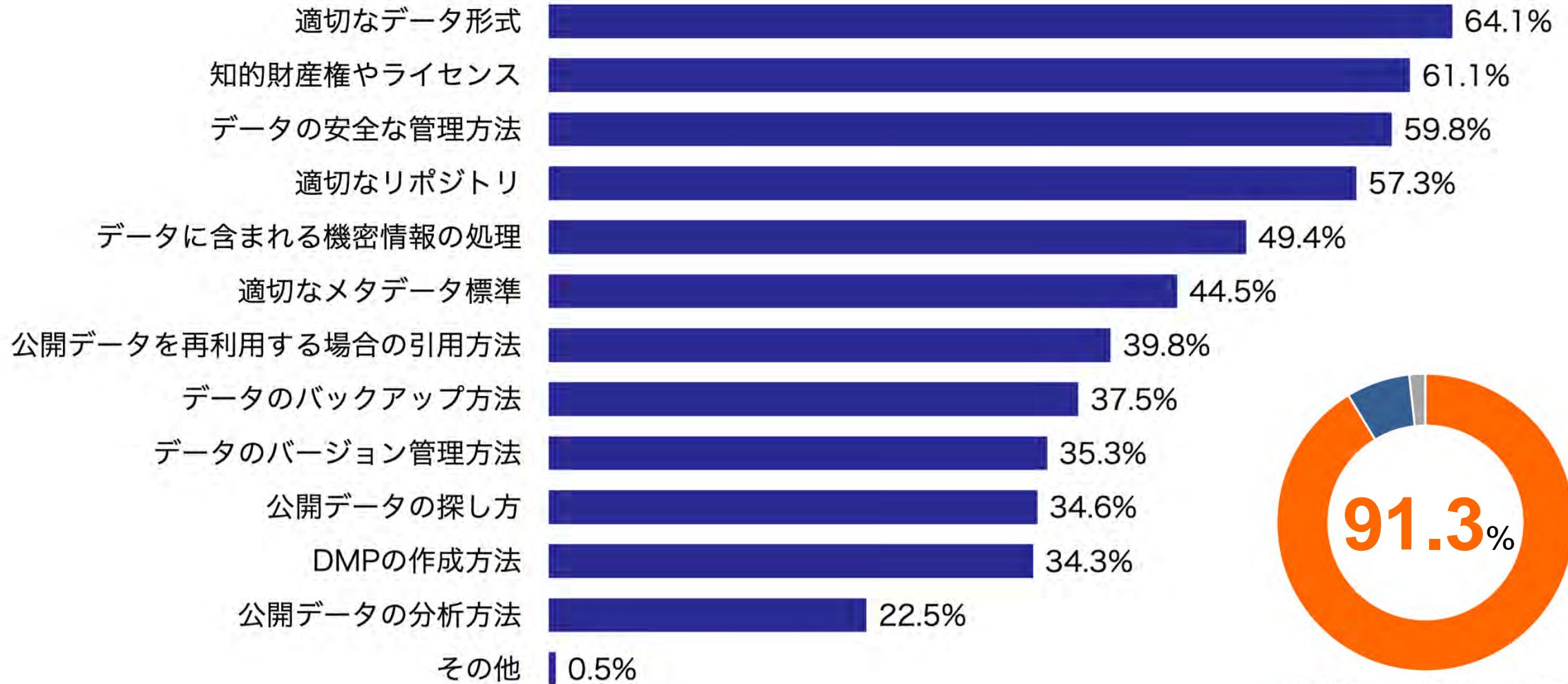
メタデータ標準の選択



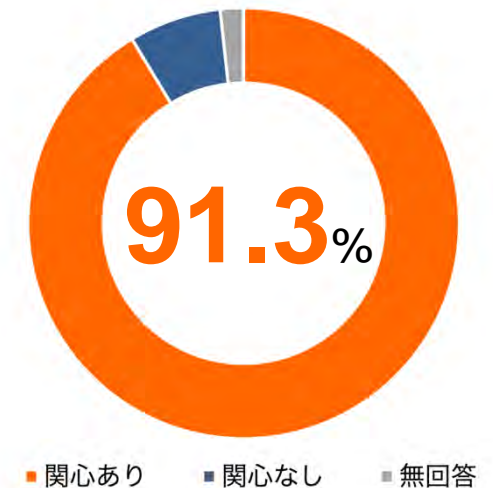
RDMサービスを展開するにあたり、自機関の研究者から要望が高いと思われるサービスと実施しているRDMサービス（複数選択可） 【n=309】



RDMに関して知りたい項目（複数選択）

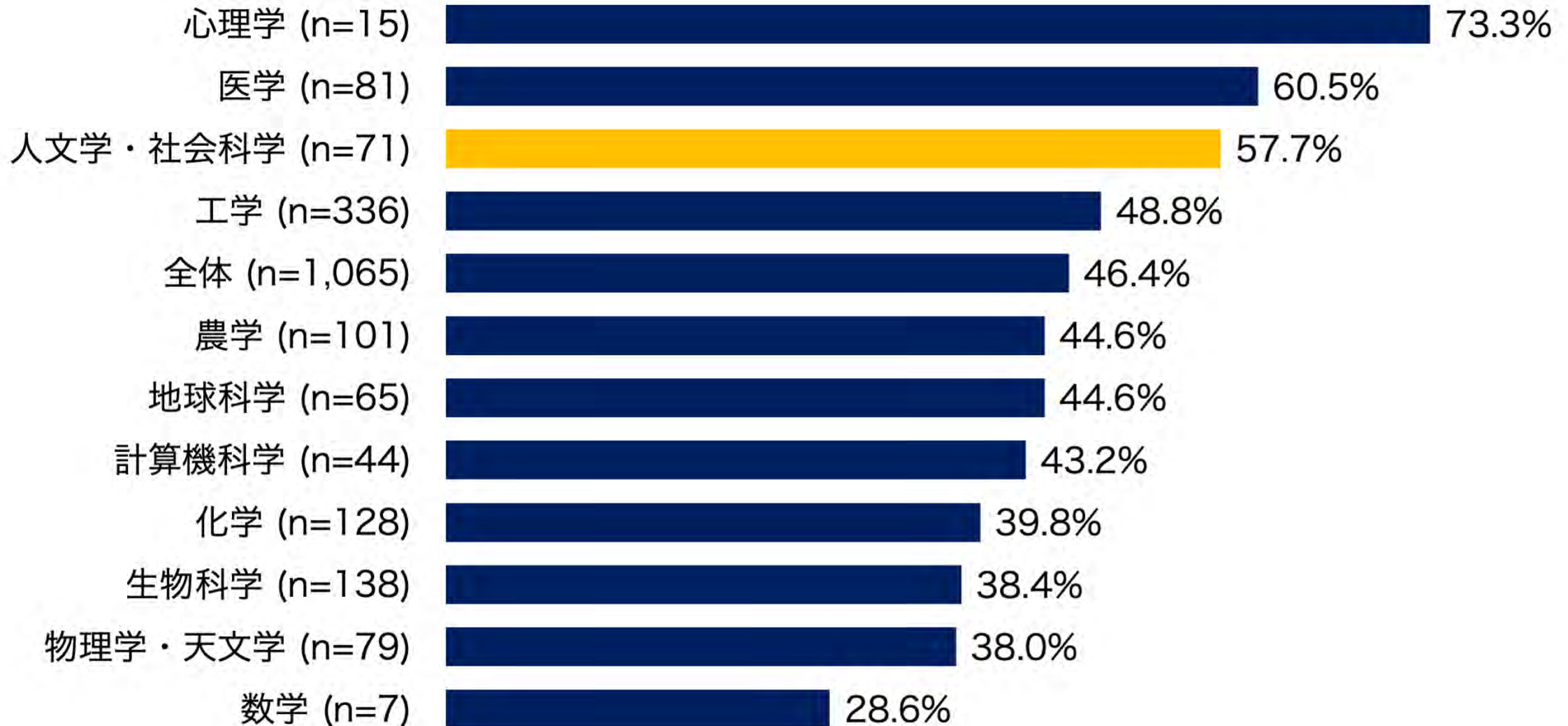


(n=990)



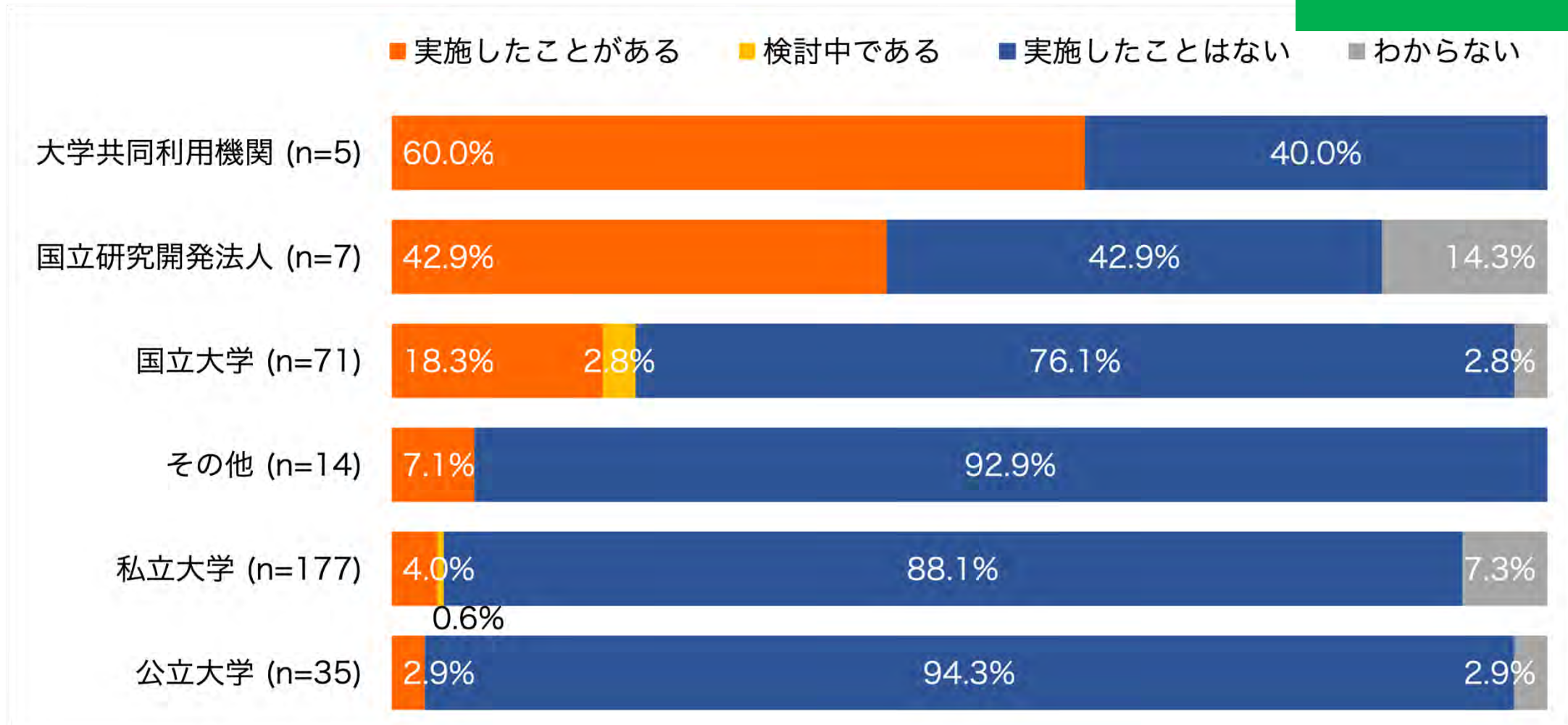
n=1,084

データに含まれる機密情報の処理



研究者向けの研究データ管理に関する研修会・ワークショップ・シンポジウム等のイベント実施状況 (2021~2022年) 【n=309】

JPCOAR/AXIES



教材「研究者のための研究データマネジメント」

オンライン教材



<https://jpcoar.repo.nii.ac.jp/records/294>

| 名前 / ファイル | ライセンス | アクション |
|--|--|---|
|  研究者のための研究データマネジメント_カテゴリー一覧 (14.1 kB) |  | ダウンロード Information |
|  研究者のための研究データマネジメント_研究前_外部資金の取得 (73.0 kB) |  | ダウンロード Information |

8120

 views

total ▾

[See details](#)

<https://jpcoar.repo.nii.ac.jp/records/294>

オンライン教材

Services ▾

Guides ▾

System ▾

About ▾

お知らせ

2023.12.07

研究データ管理用ストレージシステム"QRDM" サービス開始！

2023.11.30

「GakuNin RDMデータ活用セミナー：これからの研究データ管理を探る」の動画を公開

2023.10.20

GakuNin RDMデータ活用セミナー：これからの研究データ管理を探る(11/16)

2023.08.31

教材「はじめよう、研究データ管理」を公開しました

研究データ管理支援

DMPを作成したい



効率的なデータ管理方法を知りたい



ストレージを選択したい



『人文学・社会科学におけるデータ共有のための手引き』

オンライン教材



目次

1. はじめに
2. データを共有する意義
3. データ管理計画
4. メタデータ
5. データのフォーマット
6. データの保管
7. データ共有に関する倫理的側面
8. 個人情報と匿名化について
9. データに関する著作権
10. データアーカイブの役割

③研究データの管理（現状まとめ）

- DMPの作成経験をもつHSSの研究者は**27.1%**
- メタデータの共通項目に“記述が難しい項目はない”と考えるHSSの研究者は**54.3%**
 - ✓ただし、作成可能であっても労力はかかるため、支援が望ましい
- 研究データ管理の依頼意思をもつHSSの研究者は**61.1%**
 - ✓専門性・高：“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”
 - ✓専門性・低：“メタデータ作成”、“メタデータ標準の選択”
- データリテラシーに関心をもつ回答者は**91.3%**
- HSSの研究者は、“データに含まれる機密情報の処理”への関心が高い
 - ✓オンライン教材の活用等が有効か

RDM支援へのニーズが高く、比較的行いやすい

3

**データ管理と利活用に資する
支援人材の検討**

研究プロセスと研究データ管理サービス

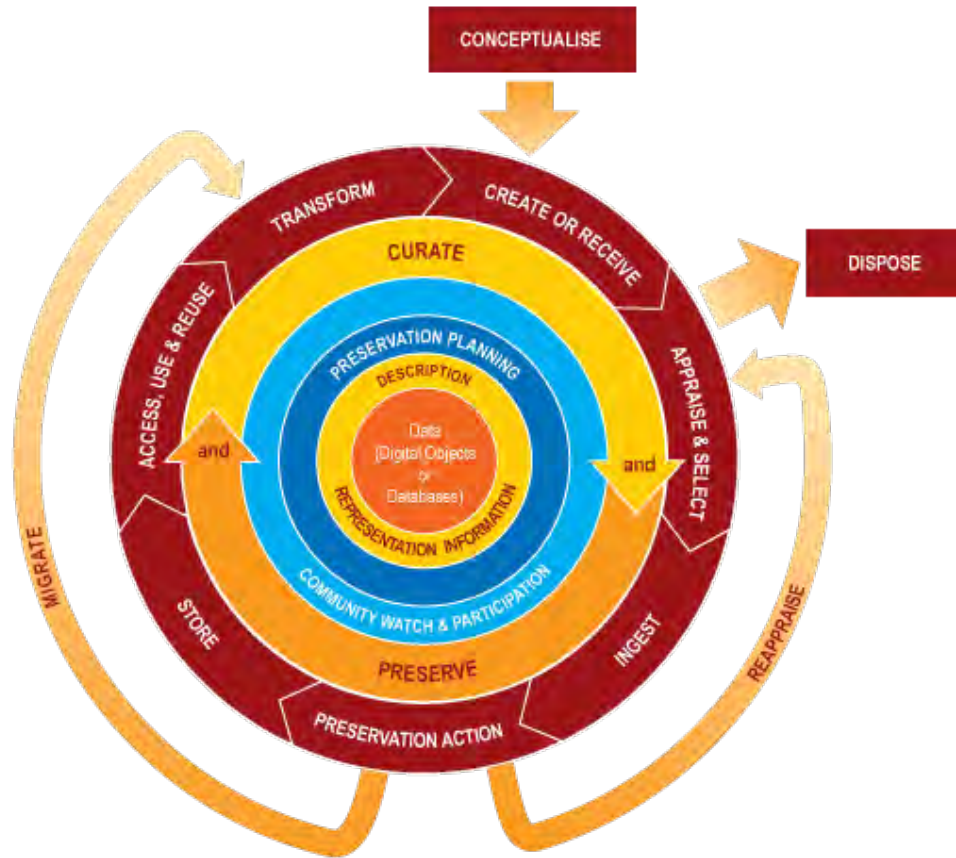


“研究支援”. 図書館情報学事典.
日本図書館情報学会編. 2023, p. 533より

NII RDC

要・支援人材

データキュレーションのライフサイクル



| DCC Curation Lifecycle Model | Steps |
|------------------------------|--------------|
| Conceptualize | 0. サービスの立ち上げ |
| Create or Receive | 1. データの受け取り |
| Appraise and Select | 2. 評価と選択 |
| Preservation Actions | 3. データの加工と処理 |
| Ingest Store | 4. 取り込みと格納 |
| n/a | 5. メタデータ記述 |
| Access, Use, Reuse | 6. アクセス |
| Transform | 7. 保存 |
| Reappraise and/or dispose | 8. 再利用 |

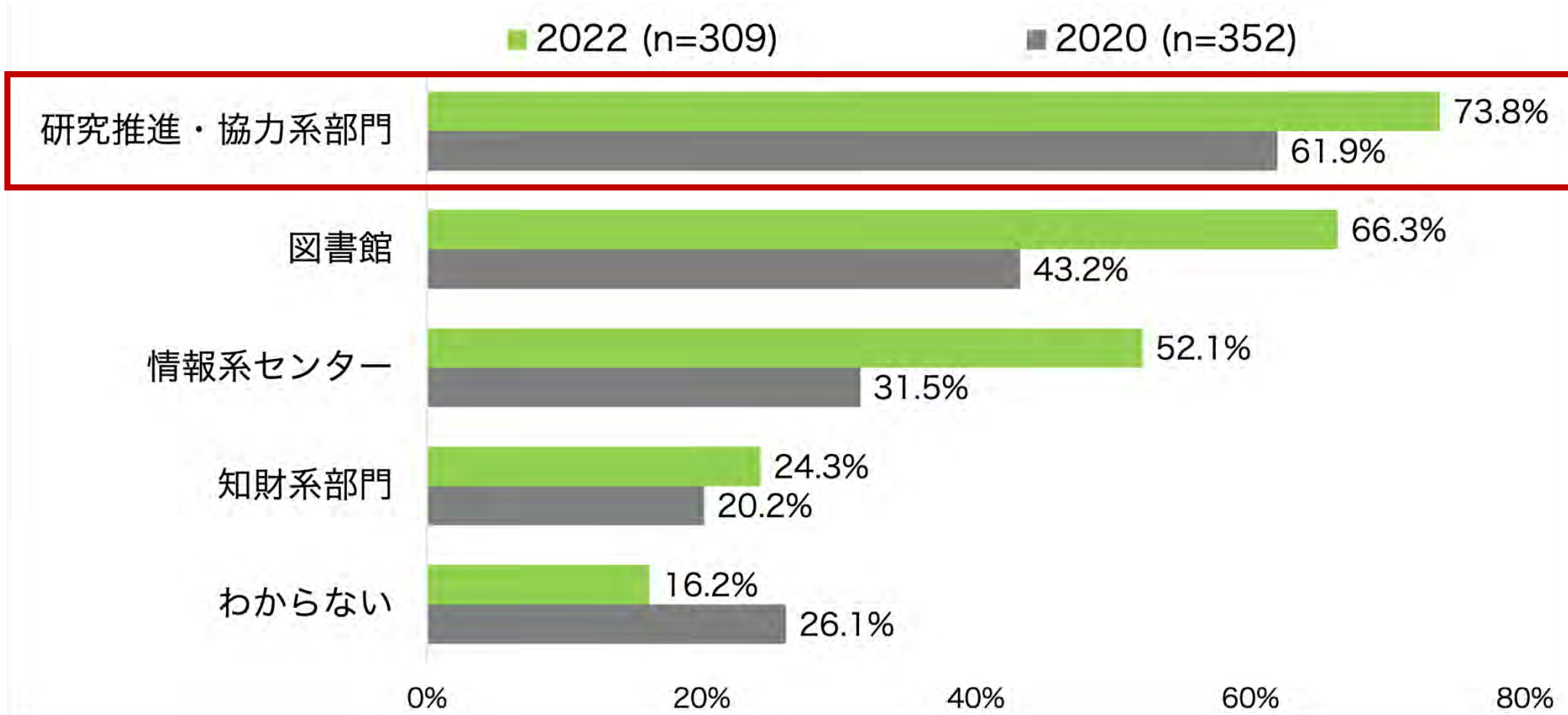
研究者とキュレーターの役割

| Steps | 各段階における役割と活動 |
|--------------|--|
| 0. サービスの立ち上げ | 適切な人員配置とビジネスモデルによって持続可能なデータキュレーションサービスを構築する。 |
| 1. データの受け取り | データの作成者がデータを機関または分野のデータリポジトリに寄託できるようにする。 |
| 2. 評価と選択 | データのレビューを行うことで、収集方針に合わないデータを選択・却下を可能にし、デジタルデータに固有の既知のリスクを軽減する。 |
| 3. データの加工と処理 | 確立された手順に従って、寄託された全てのデータを最適な方法で配置し、変換し、準備する。 |
| 4. 取り込みと格納 | 適切なリポジトリを用いて、データを安全な場所に格納する。 |
| 5. メタデータ記述 | データを容易に発見できるように、著者が登録したメタデータを拡張して適切なメタデータを記述する。 |
| 6. アクセス | 発見、発信、検索、ダウンロードの機能によって、アクセスを容易にする。 |
| 7. 保存 | データが有用である限り様々な形で保存する。データの保存期間については、ポリシーに基づいた決定を遵守する。 |
| 8. 再利用 | データの影響や価値を評価し、保存するか削除するかを決定する。 |

データキュレーションサービスと役割分担案（参考）

| キュレーションライフサイクル | 分野専門家 | 図書館 | 情報技術 専門家 | 研究者 |
|----------------|-------|-----|-------------|-----|
| 0. サービスの立ち上げ | ◎ | ◎ | ◎ | * |
| 1. データの受け取り | ◎ | | | * |
| 2. 評価と選択 | ◎ | | | |
| 3. データの加工と処理 | ◎ | | | |
| 4. 取り込みと格納 | | | ◎ | |
| 5. メタデータ記述 | | ◎ | | |
| 6. アクセス | ○ | ◎ | ○ | * |
| 7. 保存 | | ○ | ◎ | |
| 8. 再利用 | ○ | ○ | ◎ | * |

研究データ管理に関する議論でステークホルダーになり得る部署、または既に関与している部署（複数選択可）



その他：コンプライアンス室、大学事務部（教務、経理、総務）、委員会（教務委員会、研究関係の委員会）、検討中など。