

京都大学での研究データ公開に関する アンケートと個別調査

家森俊彦^{1,2}・青木学窓^{1,2}・梶田将司^{1,2}・
元木環^{1,2}・川口朋子^{1,3}・天野絵里子^{1,4}

¹ 京都大学アカデミックデータ・イノベーションユニット

² 京都大学学術情報メディアセンター

³ 京都大学文書館

⁴ 京都大学学術研究支援室

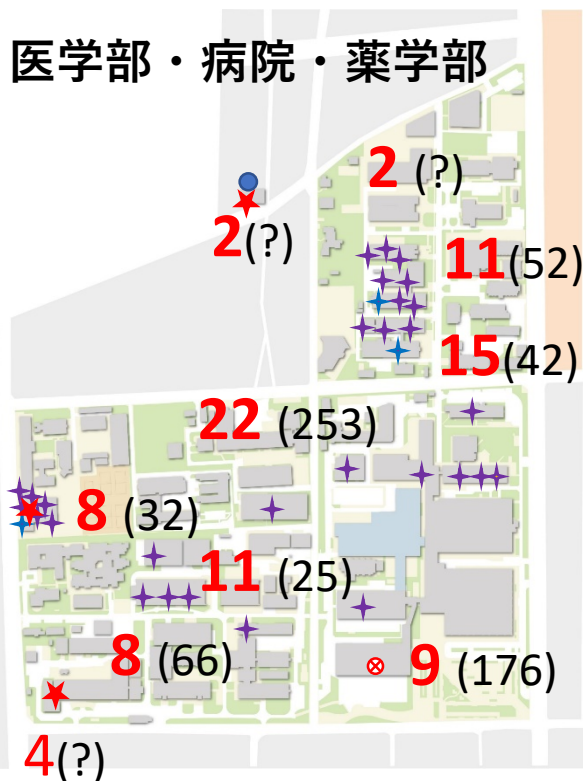
第1回 2018.12 – 2019.01 どこに、誰が、どのようなデータセットを保有しているのか
(有効回答: **244**)

第2回 2019.11 – 2019.12 RDMおよびOpen dataに関する意識調査
(有効回答: **329**)

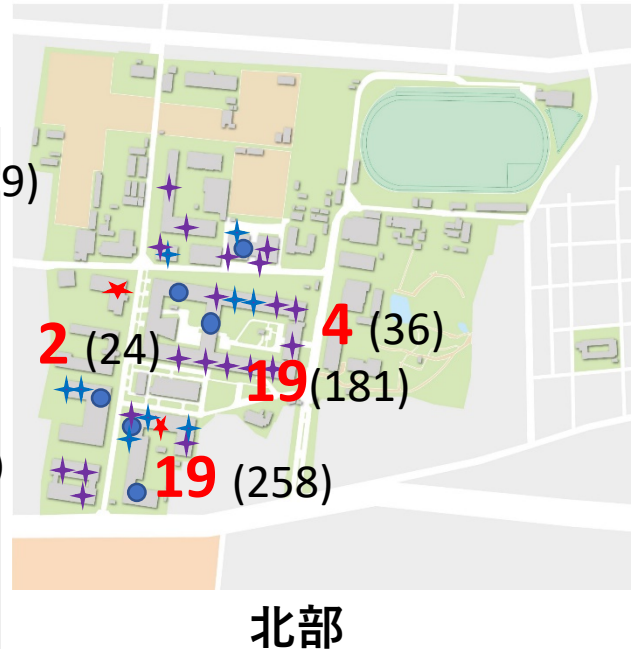
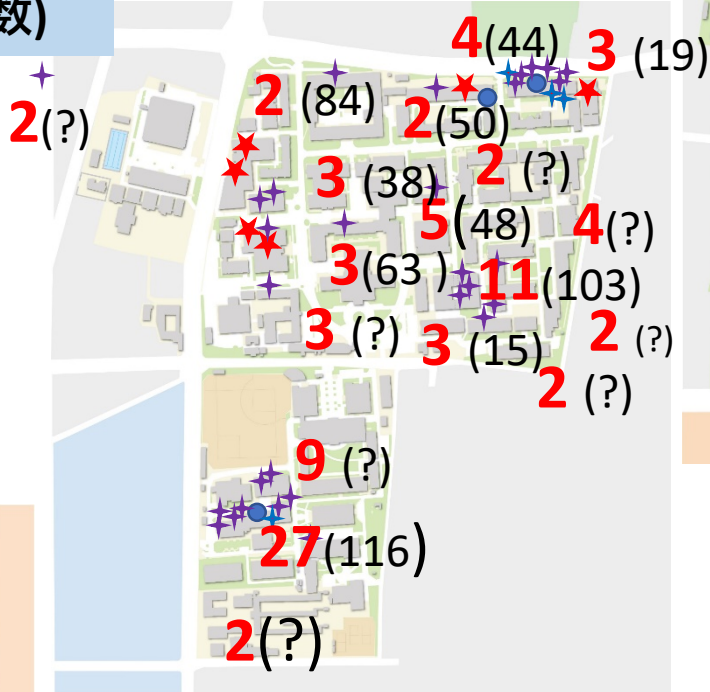
1回目回答者メールアドレス(**239件**)と2回目(**75件**)に**共通アドレス無し!!**

➔ 2回の回答者集合はかなり独立?

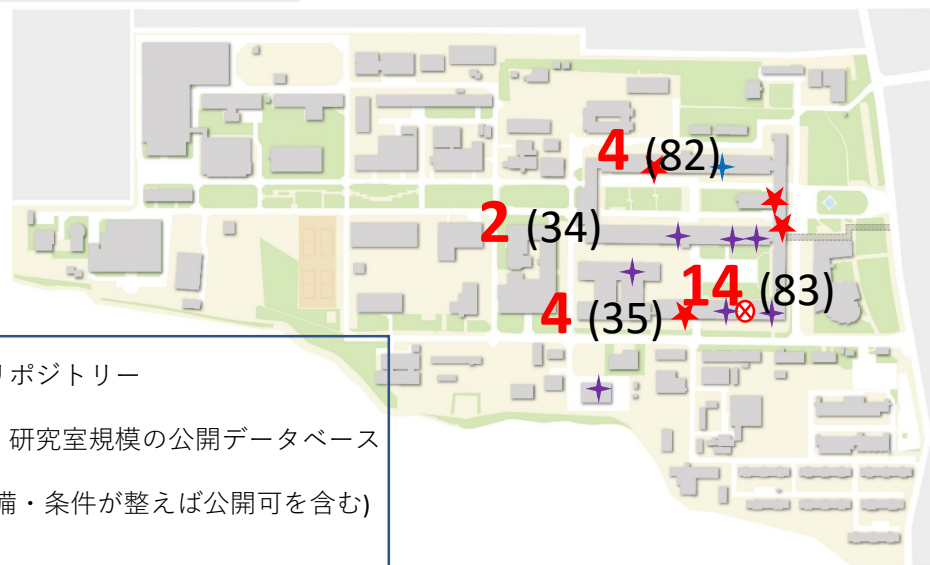
アンケート(2019) 回答数 (部局別 括弧内は教員数)



吉田キャンパス 本部・西部・吉田南



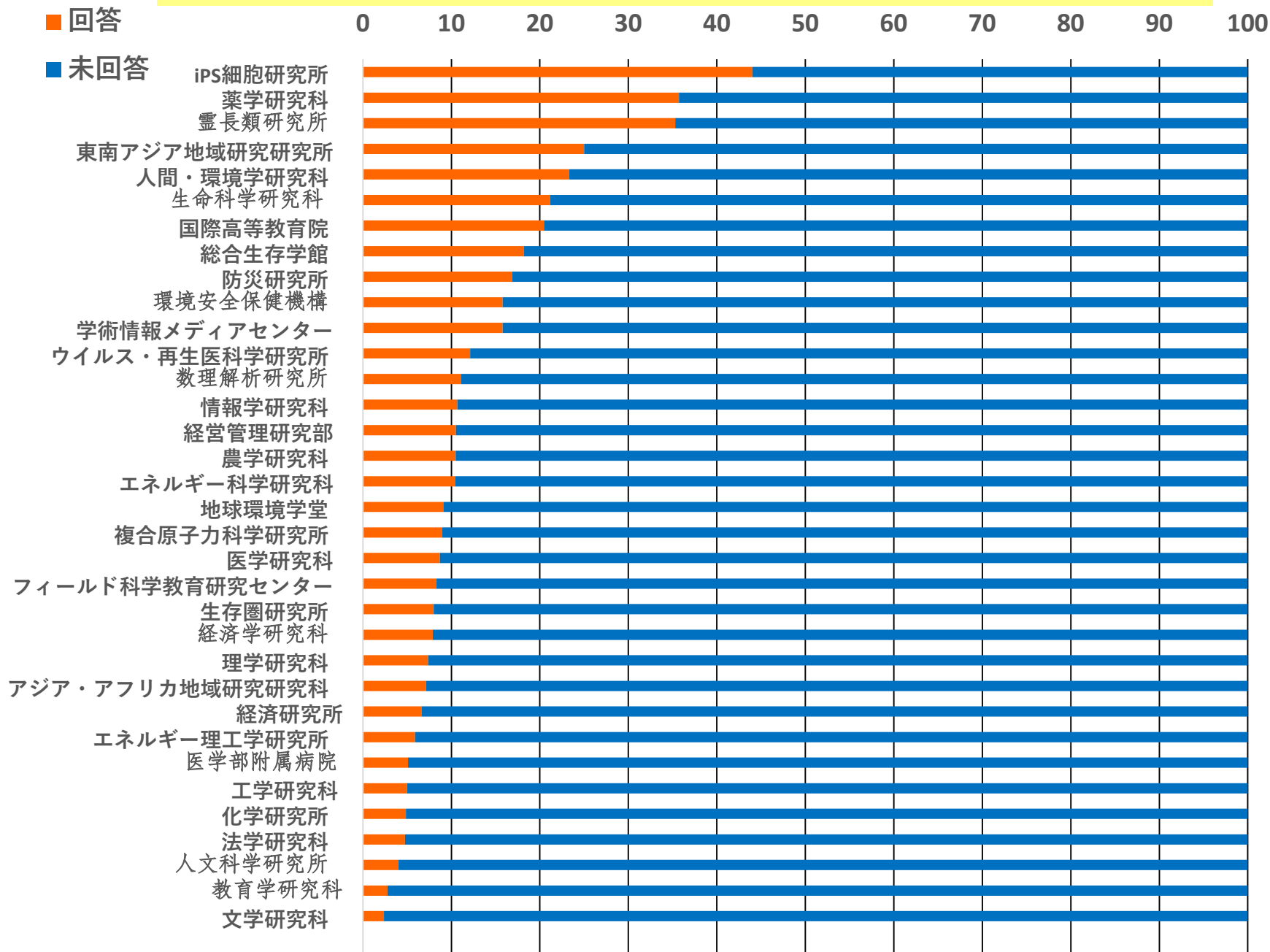
宇治キャンパス



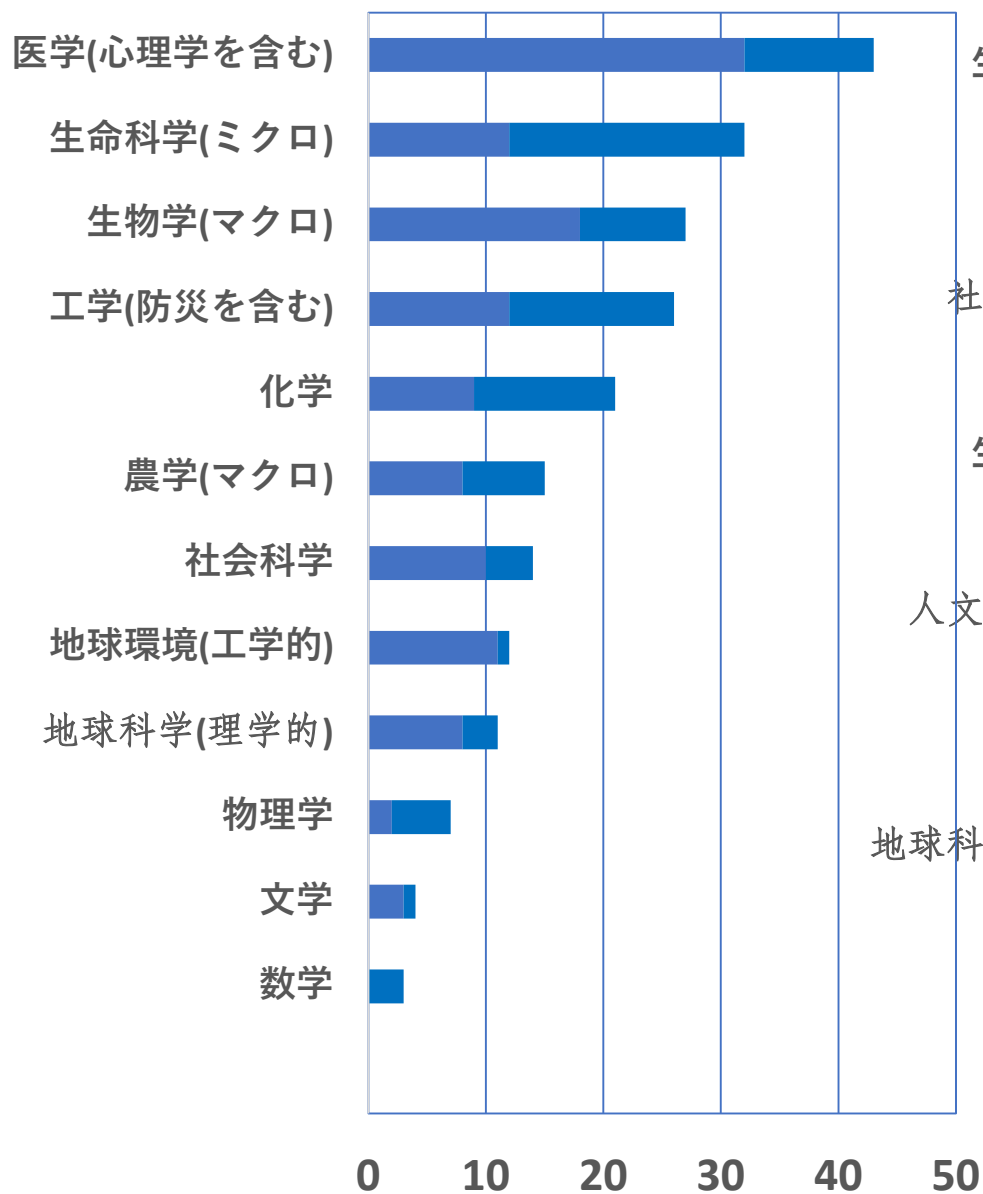
回答者の所属を示す。
研究者の居る建物は
正確ではないので注
意。

- ★ 組織・部門の公開リポジトリ
- 特定プロジェクト・研究室規模の公開データベース
- ✦ 非公開データ (準備・条件が整えば公開可を含む)
- ⊗ 公開不可データ

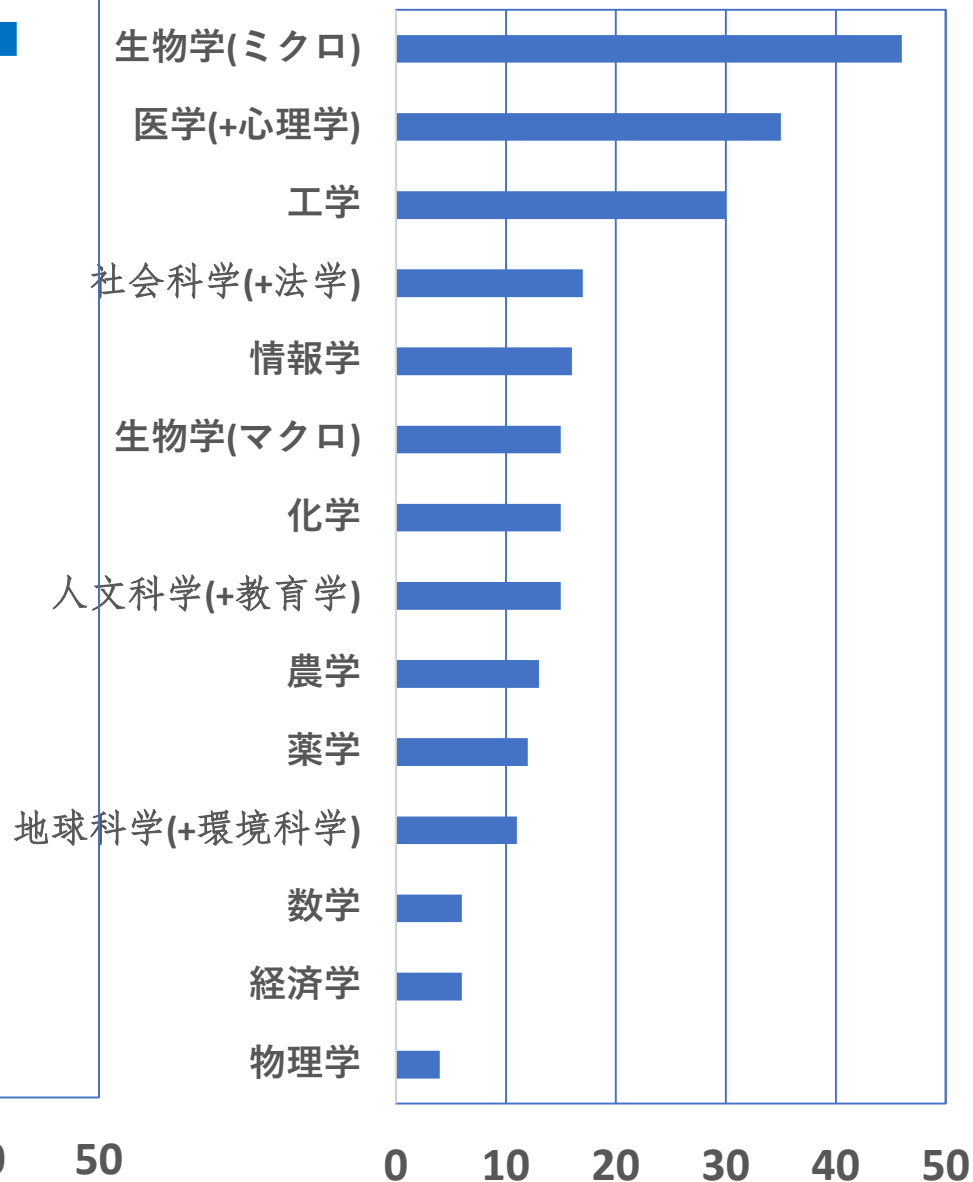
アンケート回答者・未回答者の部局別割合 (第2回調査)



分野別回答数 **1**回目調査



分野別回答数 **2**回目調査



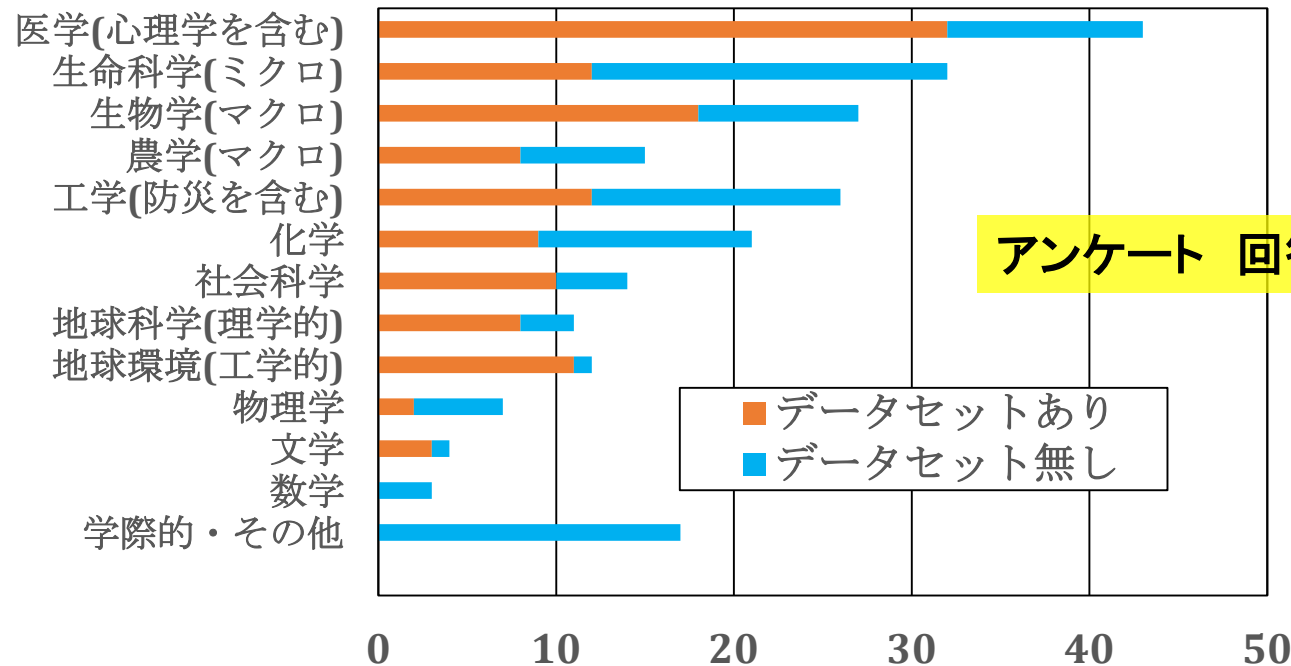
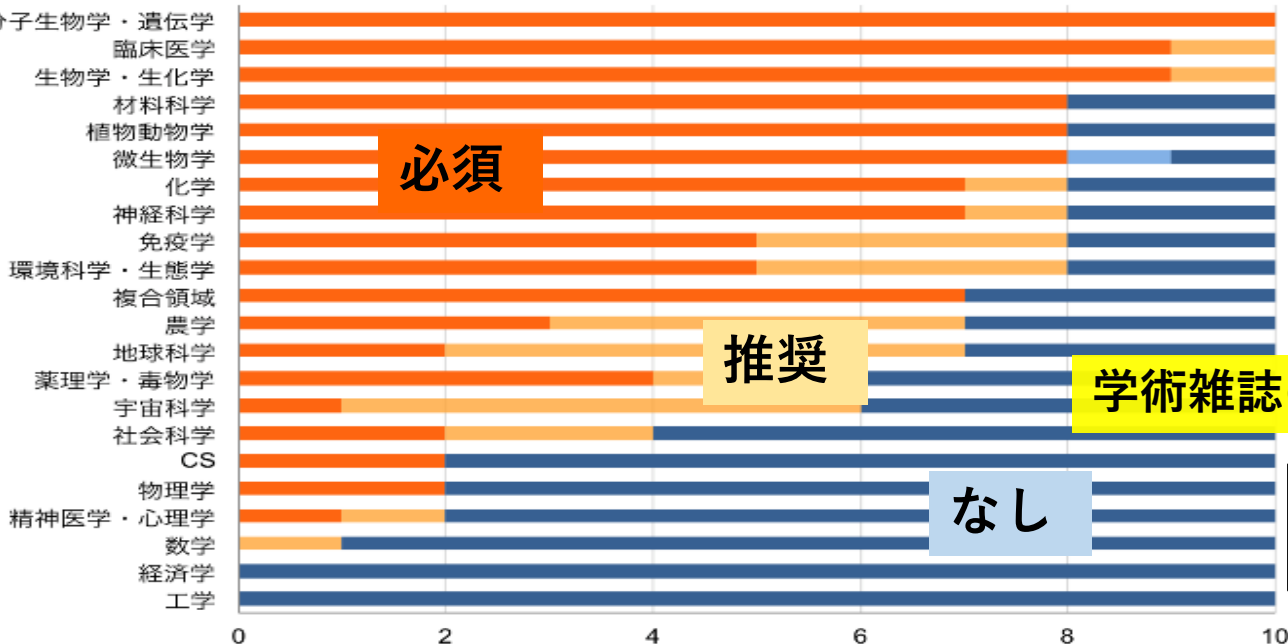
■ 必須 ■ 推奨 ■ 受諾 ■ なし

2014年

学術雑誌のデータ公開ポリシー

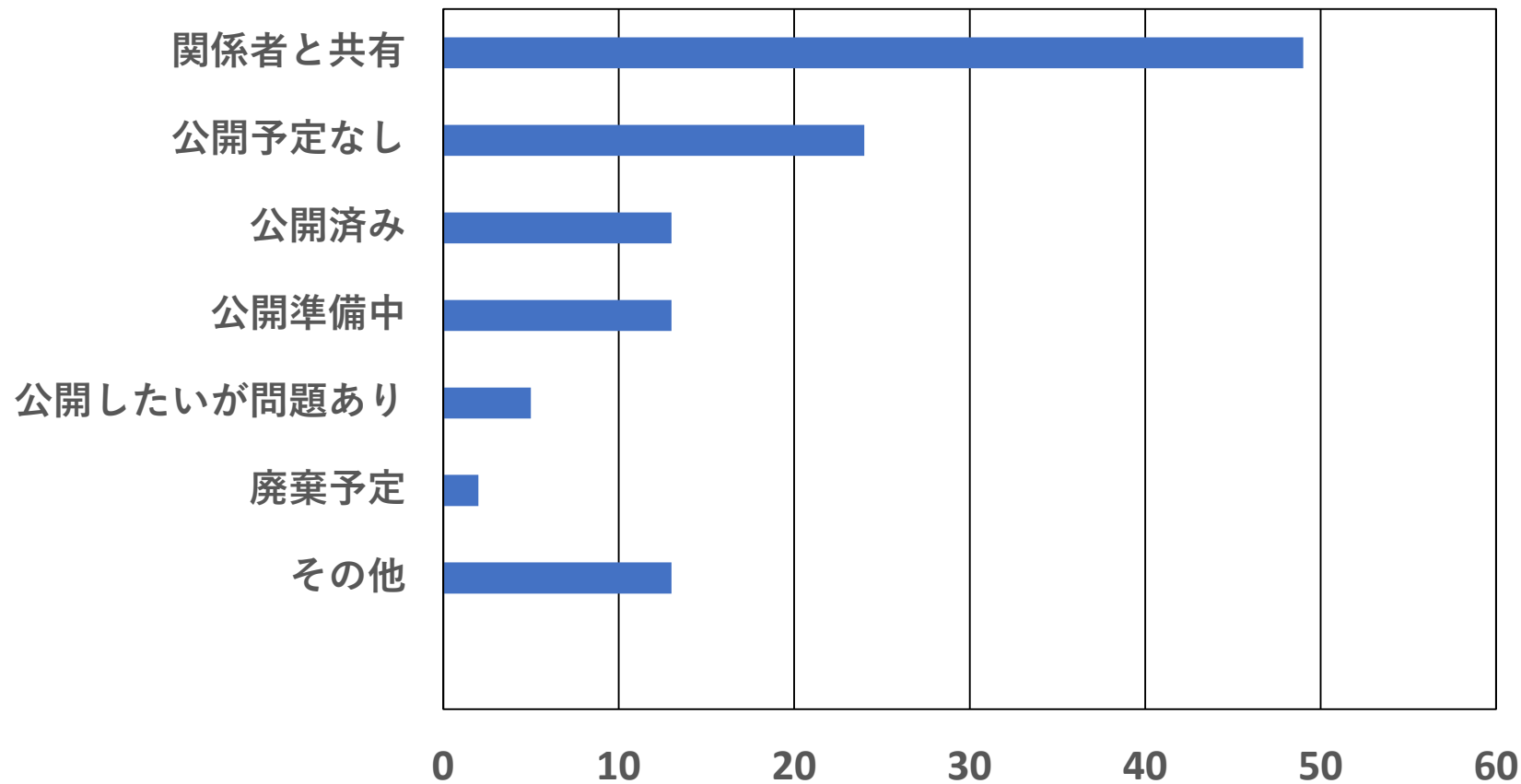
池内有為「学術雑誌のデータ公開ポリシー：経年変化とデータ引用ポリシーの状況」

<http://id.nii.ac.jp/1351/00007454/>



アンケート 回答数と回答者の研究分野

データの公開/非公開 (1回目調査)



どのようなシステム上でデータセットを公開していますか

(回答数)

- 28 学外組織により構築されたデータ公開システム等にデータを登録している。
- 11 サーバー機器を購入し、独自に構築したデータ公開システムにデータを登録している。
- 8 学内組織により構築されたデータ公開システム等にデータを登録している
- 4 学内・学外クラウド基盤と契約し、独自に構築したデータ公開システムにデータを登録している。
- 1 研究室でどうなっているのかよく知らないです。

計 52 (有効回答の16%)

学外組織のシステム: 計19件の内訳

- **JPOSTDB** (Japan ProteOme STandard Repository/Database) (実験の生データや解析結果)
- **NCBI** (National Center for Biotechnology Information) 3件
- **DDBJ** (DNA Database of Japan) 4件
- **Protein Data Bank** 2件
- **OSF** (心理実験の要約データ) 2件
- **PIG Dataset** (ピアノ運指のデータ)
- **???** (コンピュータプログラム)
- **JAMSTEC** (アンサンブル大気再解析データ)
- **???** (放射線測定データ)
- **GITHUB** (ソースコードおよびテスト用データ) 2件
- **arXiv** (プレプリント)

「公開が困難」な研究データに関する質問

(質問) あなたが (できれば公開したいが) 「公開が困難」な研究データをお持ちの場合、

公開できない理由はどのようなものですか(複数選択可)

回答者数：114名 (全回答者数329名)

選択数	理由
60	<u>時間がない</u>
59	<u>データ公開の動機が不十分である</u>
44	<u>データ提供者との間でデータ公開に関する事前合意が得られていない</u>
42	<u>公開したデータへの問い合わせ対応、フォローアップ、継続性に不安がある</u>
28	<u>データ公開に必要な技術的手続きがわからない</u>
25	<u>公開可能・不可能なデータを選別できない</u>
22	<u>適切なデータ公開基盤が見つからない</u>
21	<u>ライセンス適用等、公開したデータ利用に制限をかける方法がわからない</u>
11	<u>データの質保証に不安がある</u>
10	その他
6	共同研究実施者等と連絡がつかず、データ公開への合意形成ができない

RDMS+教育で対応可?

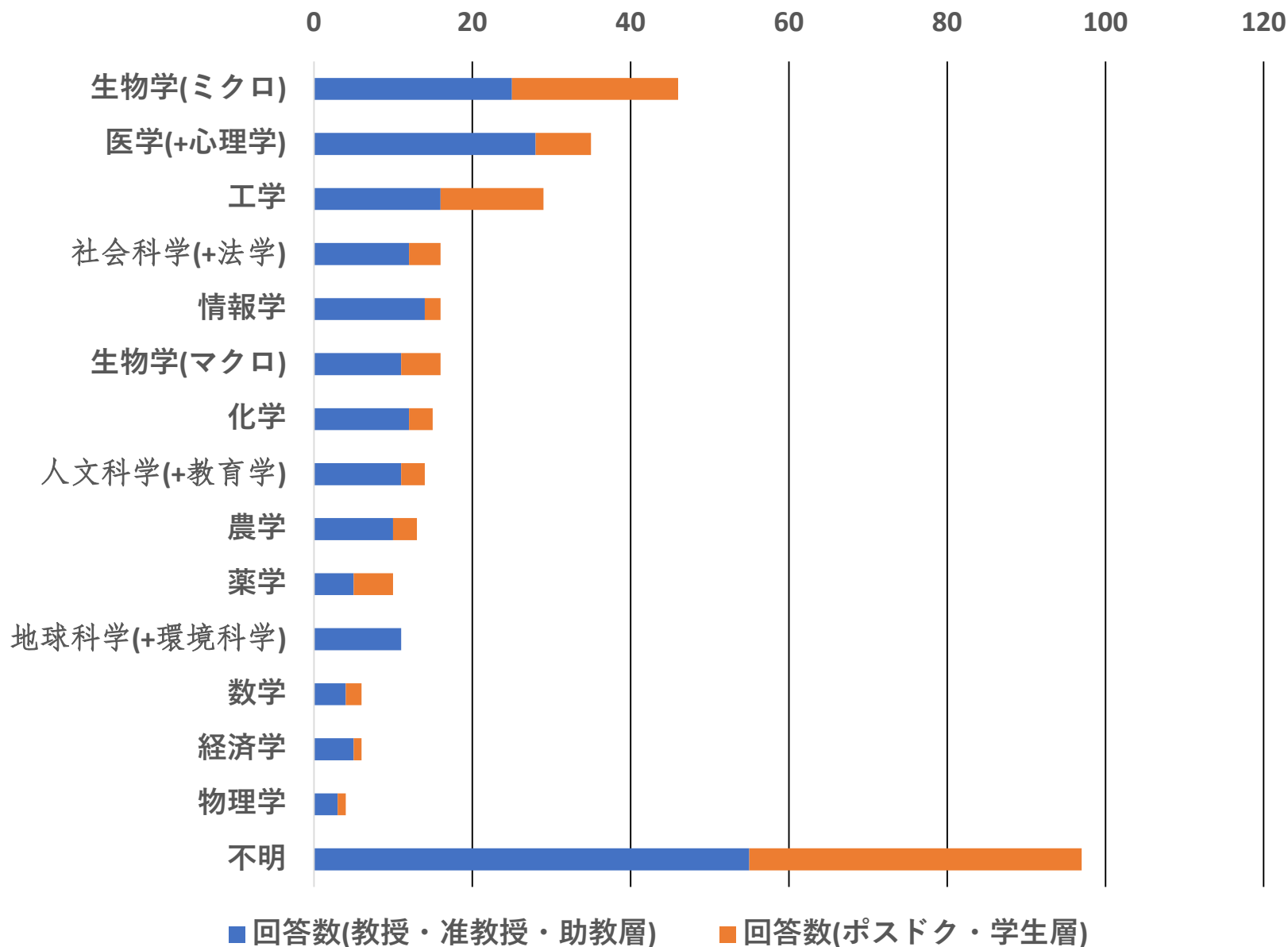
「公開できない」研究データに関する質問

あなたが「(積極的に)公開できない」研究データをお持ちの場合、以下の質問にお答えください

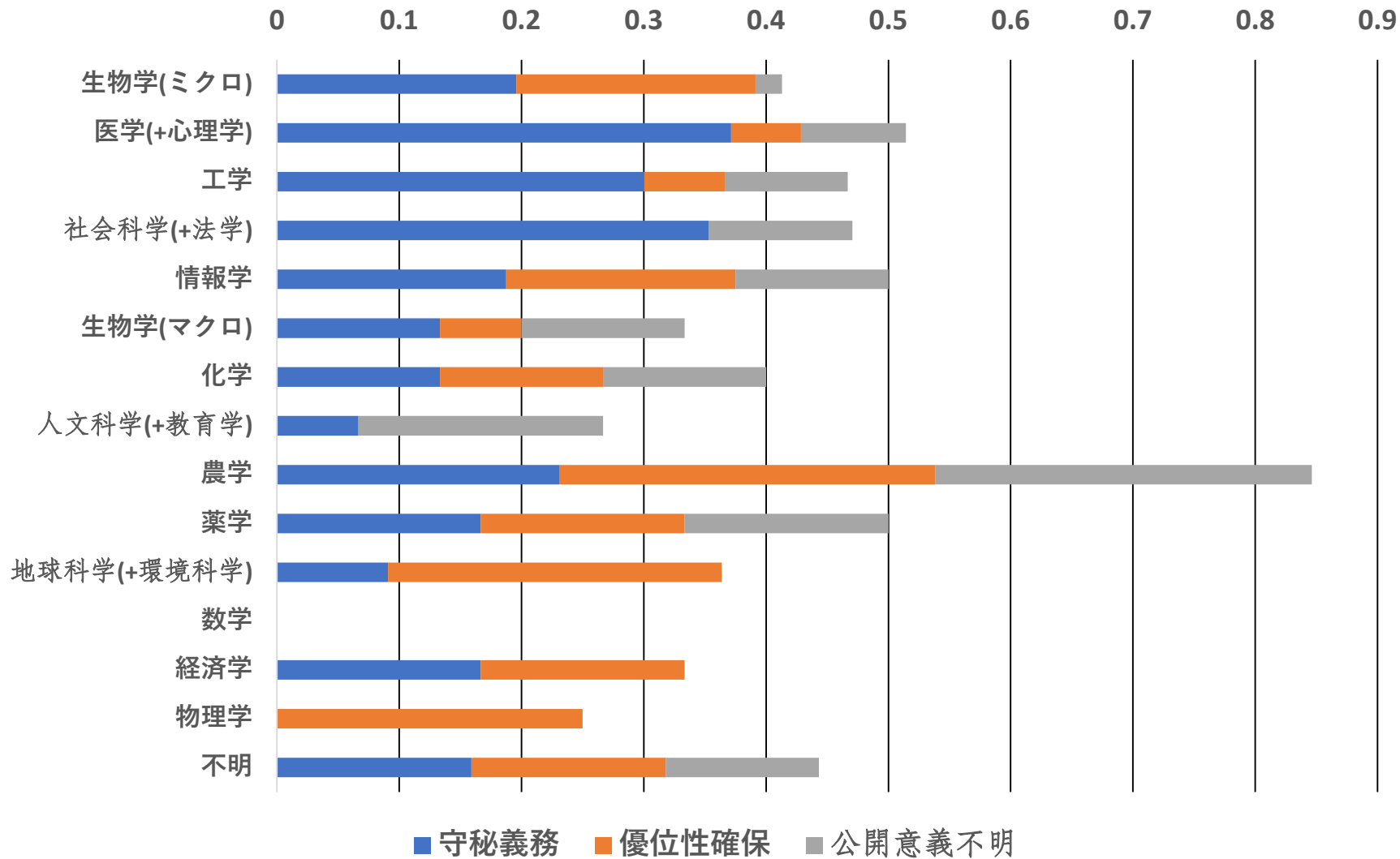
20。データセットを公開しない理由について、当てはまるものをすべてお答えください。
当てはまるものをすべて選択してください。

- ・ データの取得・生成・利用に際し、守秘義務が伴うため
- ・ 他の研究者に対する、研究上の優位性を維持するため
- ・ データを公開することに意義が見いだせない
- ・ その他:

非公開の理由(層別・各分野別回答数) (第2回調査)



非公開の理由 (各分野アンケート回答数に占める割合)



研究データマネージメント環境に関する希望調査

現在保有する、あるいはこれから収集・生成するデータセットに関し、大学が何らかのサポートをする場合、どのようなサポートを希望しますか？

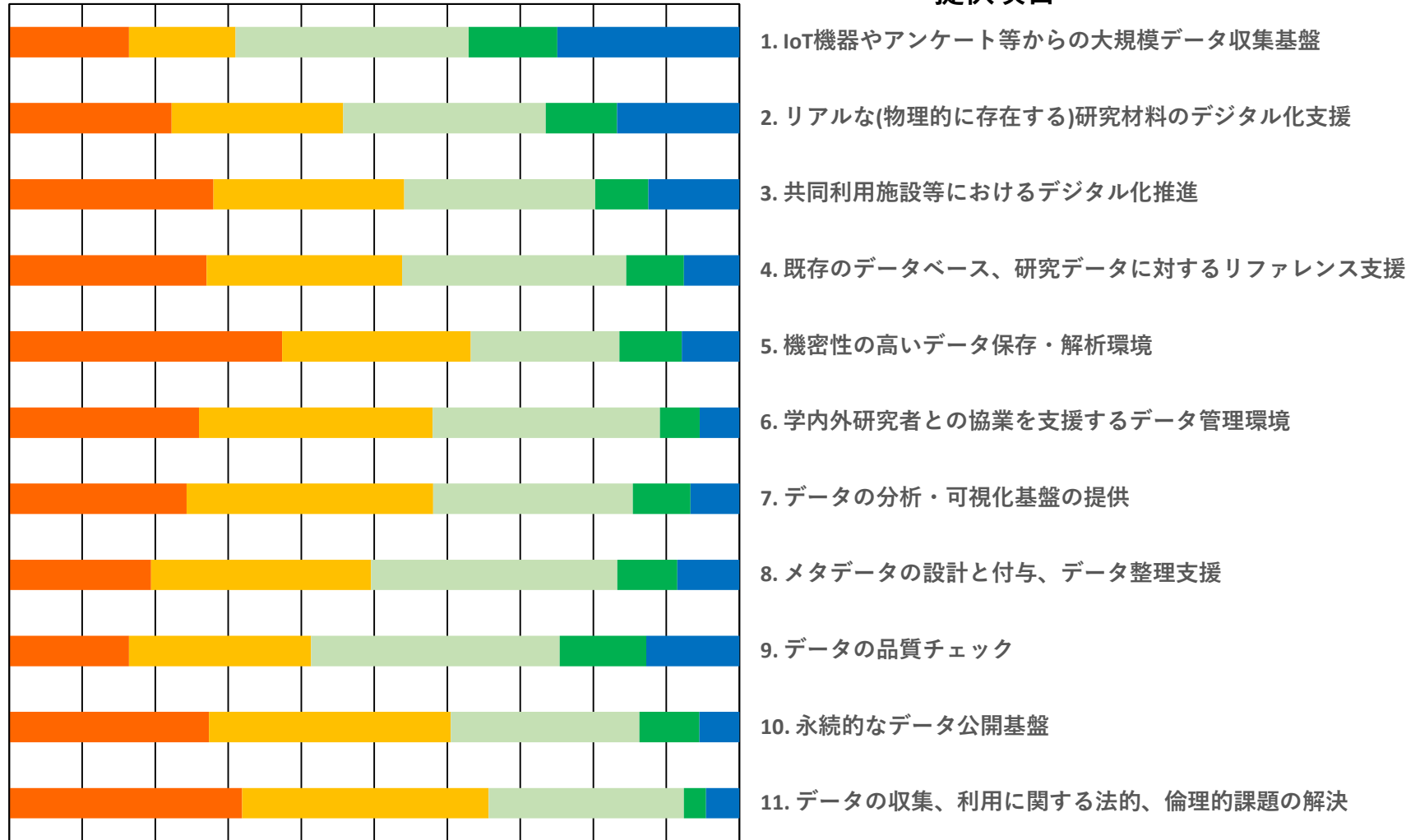
1 つだけマークしてください。(あまり必要でない 1,2,3,4,5 大いに必要)

1. IoT機器やアンケート等からの大規模データ収集基盤 *
2. リアルな(物理的に存在する)研究材料のデジタル化支援 *
3. 共同利用施設等におけるデジタル化推進 *
4. 既存のデータベース、研究データに対するリファレンス支援 *
5. 機密性の高いデータ保存・解析環境 *
6. 学内外研究者との協業を支援するデータ管理環境 *
7. データの分析・可視化基盤の提供 *
8. メタデータの設計と付与、データ整理支援 *
9. データの品質チェック *
10. 永続的なデータ公開基盤 *
11. データの収集、利用に関する法的、倫理的課題の解決 *

大学のRDMシステムへの興味・期待

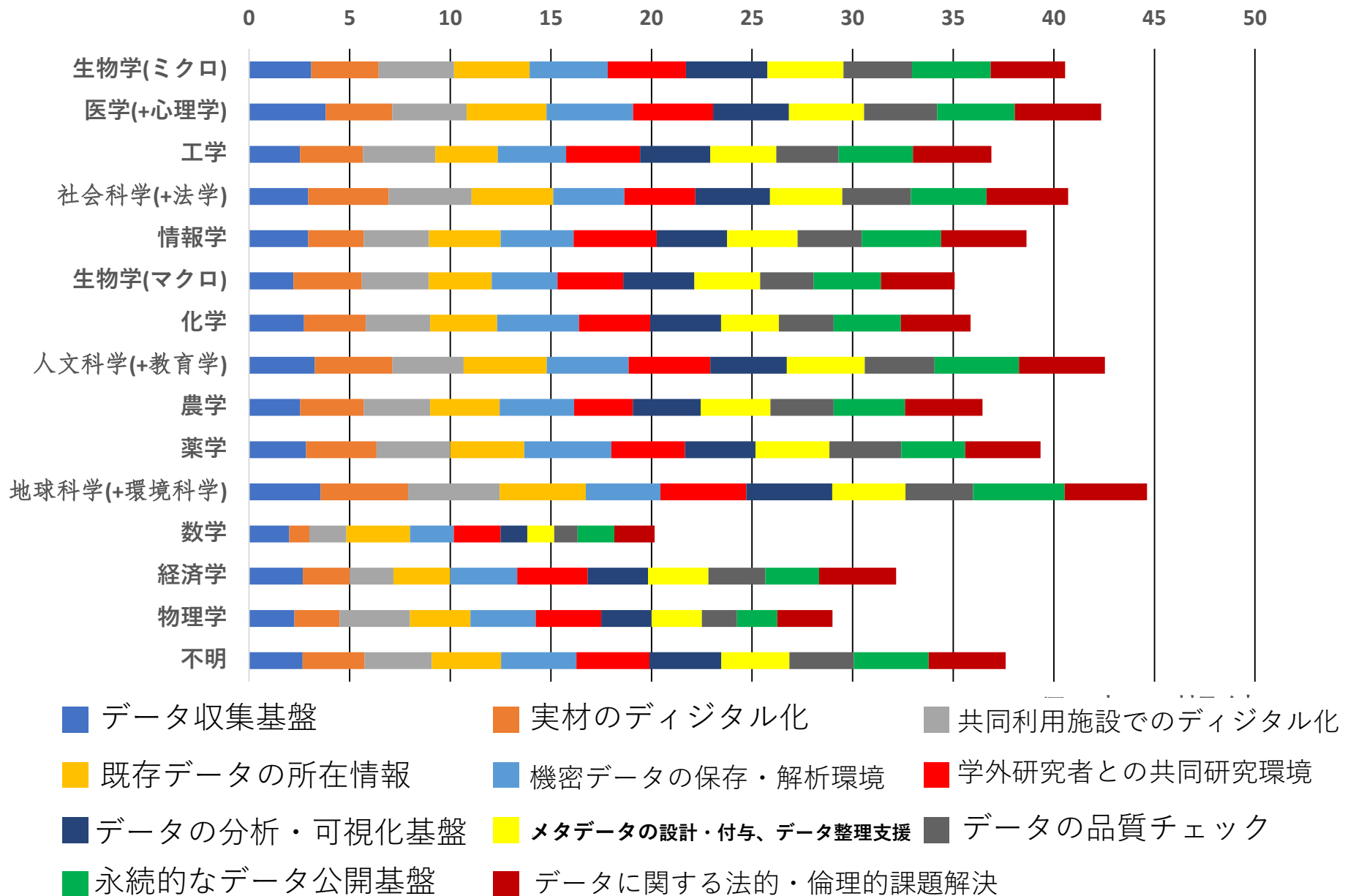
大いに有り 有り 普通 あまり無し 無し

提供項目



0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

学内RDMシステムの機能についての興味・期待の強さ(1～5段階)の平均値



アンケート調査結果から

1. 生命科学分野の関心が高い (分野による関心度の差異が大)
 ← 学術誌によるデータソース公開あるいは明示の要求の影響?
2. 自身のデータを公開することに関して否定的な研究者が多い
 - ・ 守秘義務
 - ・ 優位性の確保
3. 他方、50件以上でデータを公開
 - ・ ただし、学外システムの利用が過半数(生命科学関係が多い)
 - ・ 公開に否定的でなくても、時間がない、データ公開の動機が不十分である等の理由をあげる人が多い
4. RDMシステム構築への期待: 分野および項目により多少の差異

当面の需要・必要

1. 研究グループ内でオープンに使えるデータマネジメントシステム
2. 論文に使用したデータの公開・保管システム
3. RDMに関する教育 (大雑把に分野分けが必要? 例えば、「学域」別)
 - ➔ 研究の効率化
 - ➔ 公開を困難にしている要因の一部解決
 - ➔ 研究公正