

研究データの管理・利活用に関する 現状と課題について



令和4年4月28日

内閣府 科学技術・イノベーション推進事務局

第2章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

(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）

①信頼性のある研究データの適切な管理・利活用促進のための環境整備

研究データの管理・利活用、研究データのメタデータを検索可能とする体制の構築

②研究D Xを支えるインフラ整備と高付加価値な研究の加速

SINET、スパコン計算資源、研究設備・機器の共有、データ駆動型研究の推進

③研究D Xが開拓する新しい研究コミュニティ・環境の醸成

シチズンサイエンス、科学者とそれ以外の者の間の信頼感の醸成

新たな研究システムの
構築を目指す

公的資金による研究データの管理・利活用について

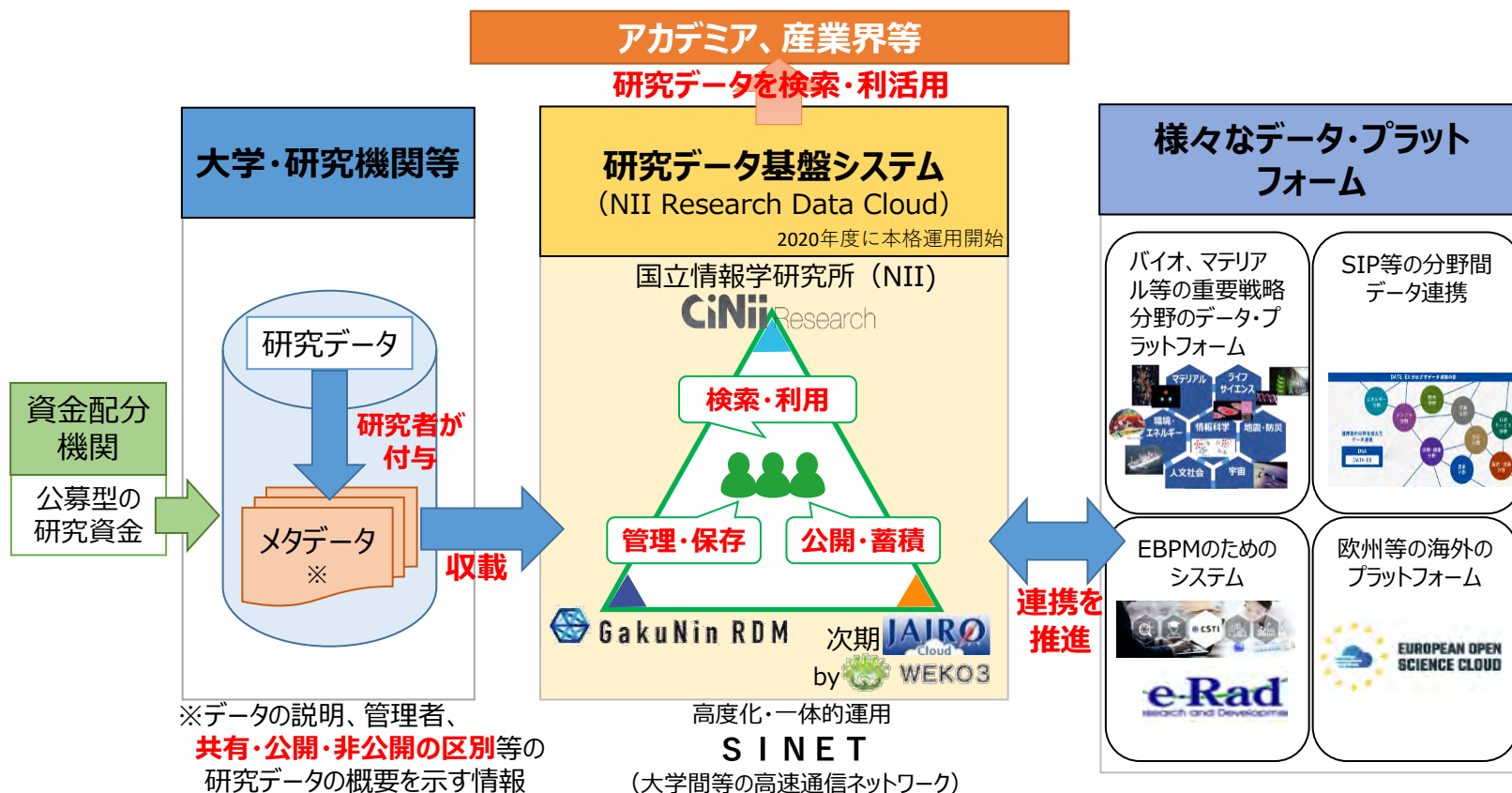
【背景】

- 知識をオープンにし、研究の加速化や新たな知識の創造などを促す **オープンサイエンスの動き** が活発化
- **世界的な出版社やIT企業** が、研究成果や研究データを **ビジネスの対象** として焦点を当てつつある

【政策文書】

- 第6期科学技術・イノベーション基本計画（2021年3月）
- 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）

研究データ基盤システムを中核としたデータ・プラットフォームの構築
→研究データの公開・共有を推進、産学官のユーザが **データを検索可能**



公的資金による研究データマネジメントの実現のための3つのアプローチ

公的資金による研究データマネジメントに求められること（先進的データマネジメント）

- 管理対象データの範囲の特定
- メタデータの付与（課題番号・課題名称、管理者、公開・共有の区分など）
- 機関リポジトリ等への管理対象データの収載
- 研究データの管理・利活用の実施状況に関する評価体系への導入

研究開発を行う機関

・データポリシーの策定
（国立大学・大学共同利用機関法人・国立研究開発法人においては2025年まで）

公募型の研究資金

・ムーンショット、次期SIP等の横展開
・全ての公募型の研究費の新規公募分に導入（2023年度まで）

研究分野

・マテリアル、バイオ、防災等の研究分野からの横展開

中核的なプラットフォームである研究データ基盤システム（NII-RDC）においてメタデータを検索可能な体制を構築

産学官のユーザーが研究データの所在を検索可能とし、管理者より入手できる体制を構築（第6期基本計画の計画期間2025年度までに）

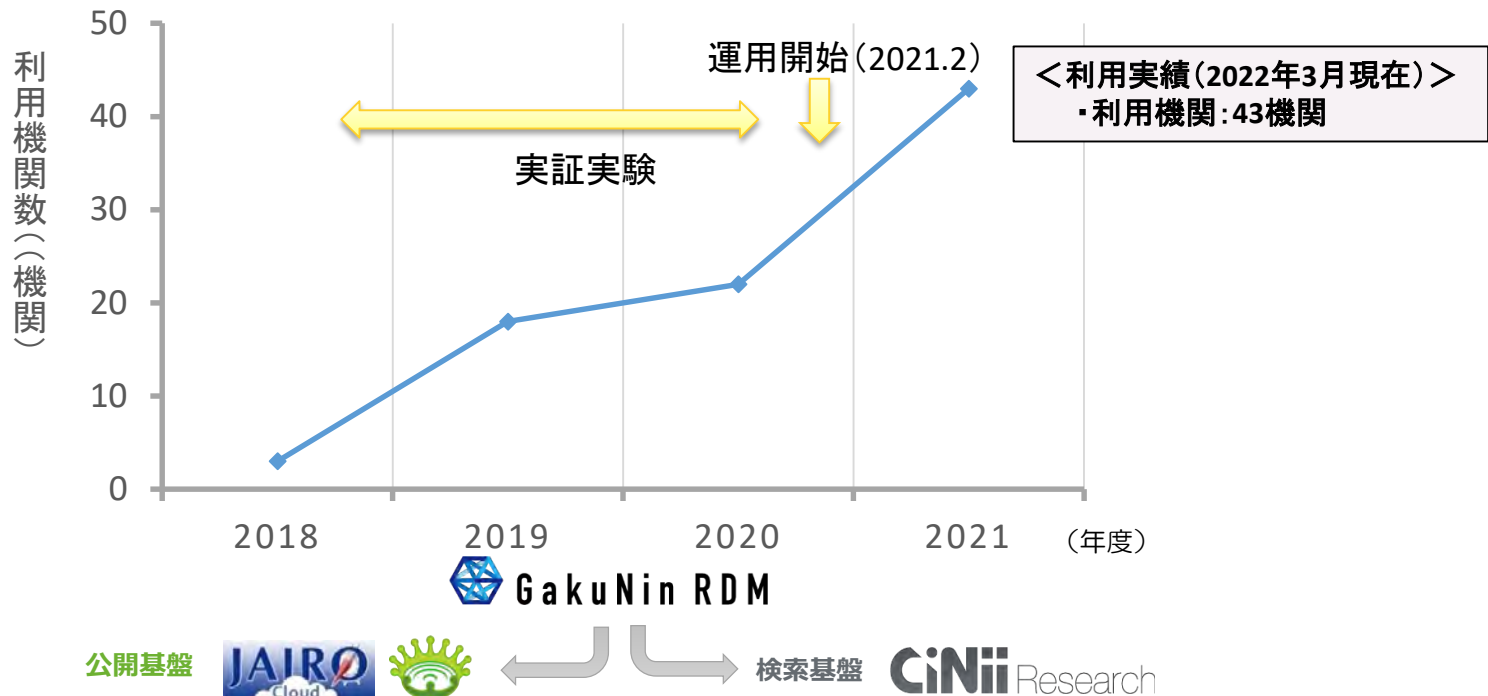
* オンラインでダウンロードできることが望ましいが、別の手段でも可。

研究データ基盤システムの利用実績(1)

GakuNin RDM (Research Data Management)

国立情報学研究所が構築し、2020年度(2021年2月本格運用開始)より運用・公開を開始した新サービス。研究プロジェクト実施中に、個人の研究者あるいは研究グループが研究データや関連の資料を管理するための研究データ管理基盤であり、既存のストレージや研究ソフトウェアと連携し、クローズドな空間で、研究プロジェクトに関わるファイルのバージョン管理や、メンバー内でのアクセスコントロールが可能。

<GakuNin RDMの利用状況（利用機関数）>



研究データ基盤システムの利用実績(2)

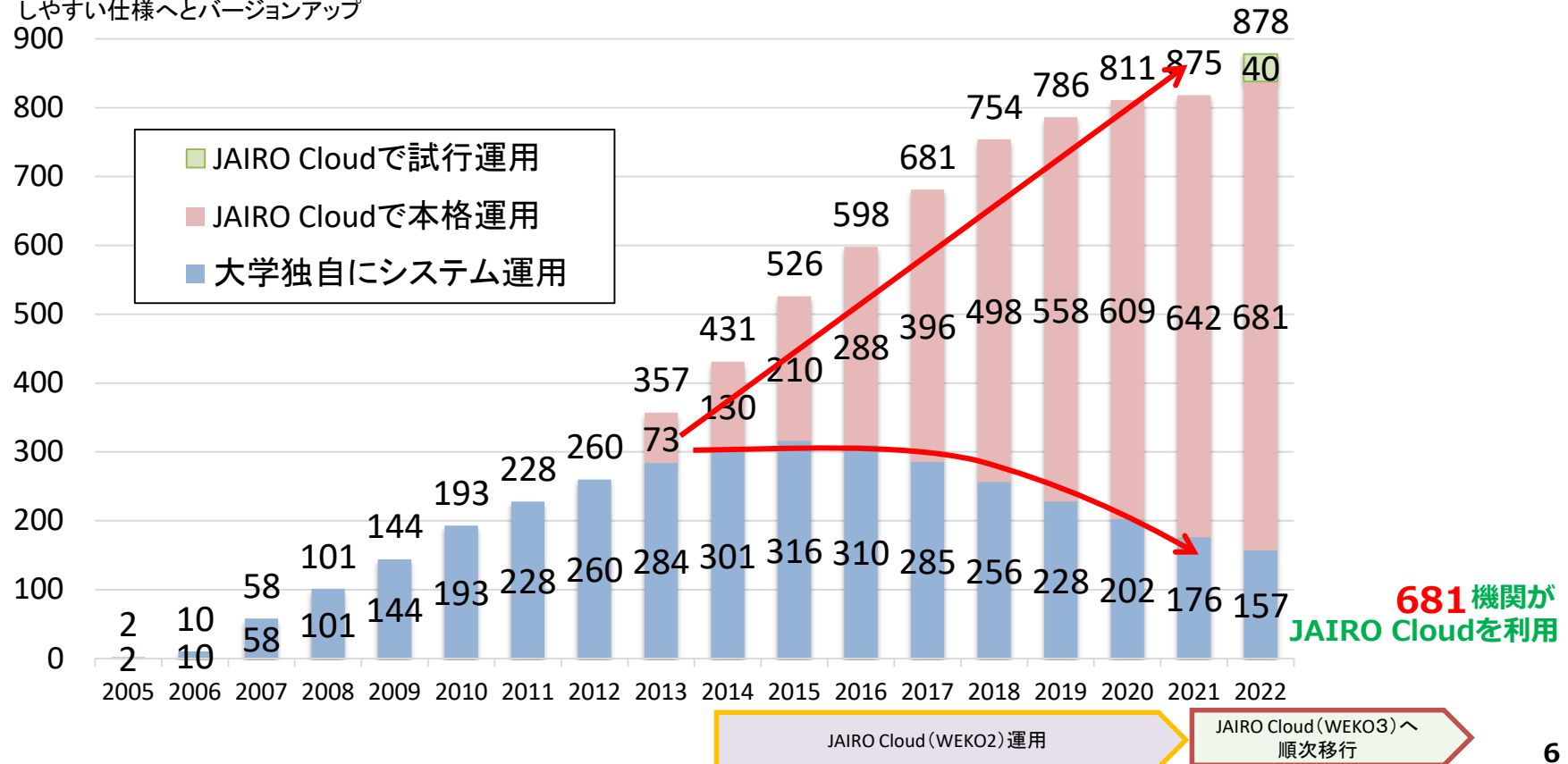
JAIRO Cloud

国立情報学研究所が構築し、2012年より運用・公開を開始したクラウド型の機関リポジトリ環境を提供するサービス。

<利用機関数推移>

- 2012年度の73機関から、1年後の2013年度には、130機関に増大。その後、コンスタントに利用機関は増大し、近年では、独自構築からJAIRO Cloudに移行する機関も増大傾向
- 研究データ基盤システムとして、2020年度にWEKO3(アプリケーション)を開発※し、バージョンアップ。2022年度現在、利用機関はWEKO2からWEKO3へ順次移行途中で、

※WEKO3では、バージョン管理機能や研究データに対応したメタデータ・スキーマであるJPCOARスキーマへ対応するなど研究データをコントロールしやすい仕様へとバージョンアップ

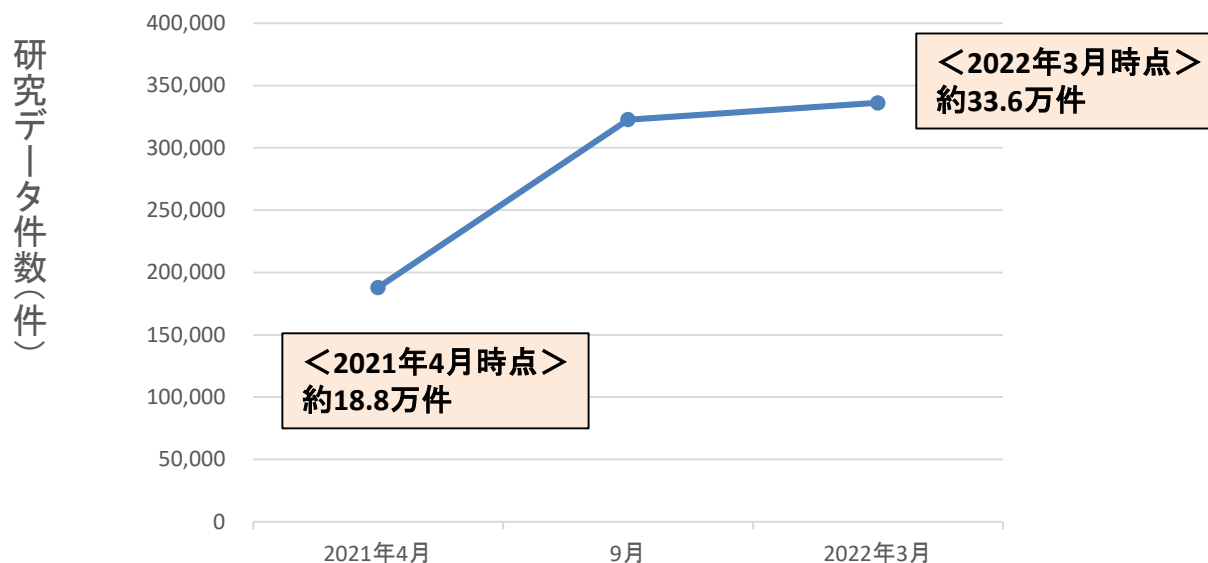


研究データ基盤システムの利用実績(3)

CiNii Research

国立情報学研究所が構築し、2021年より運用を開始した学術情報検索サービス。公開基盤(JAIRO-Cloud)等に登録された研究成果や論文情報のみならず、図書、研究データ、それらの成果を生み出した研究者、研究プロジェクトの情報などを包括して探索することが可能。

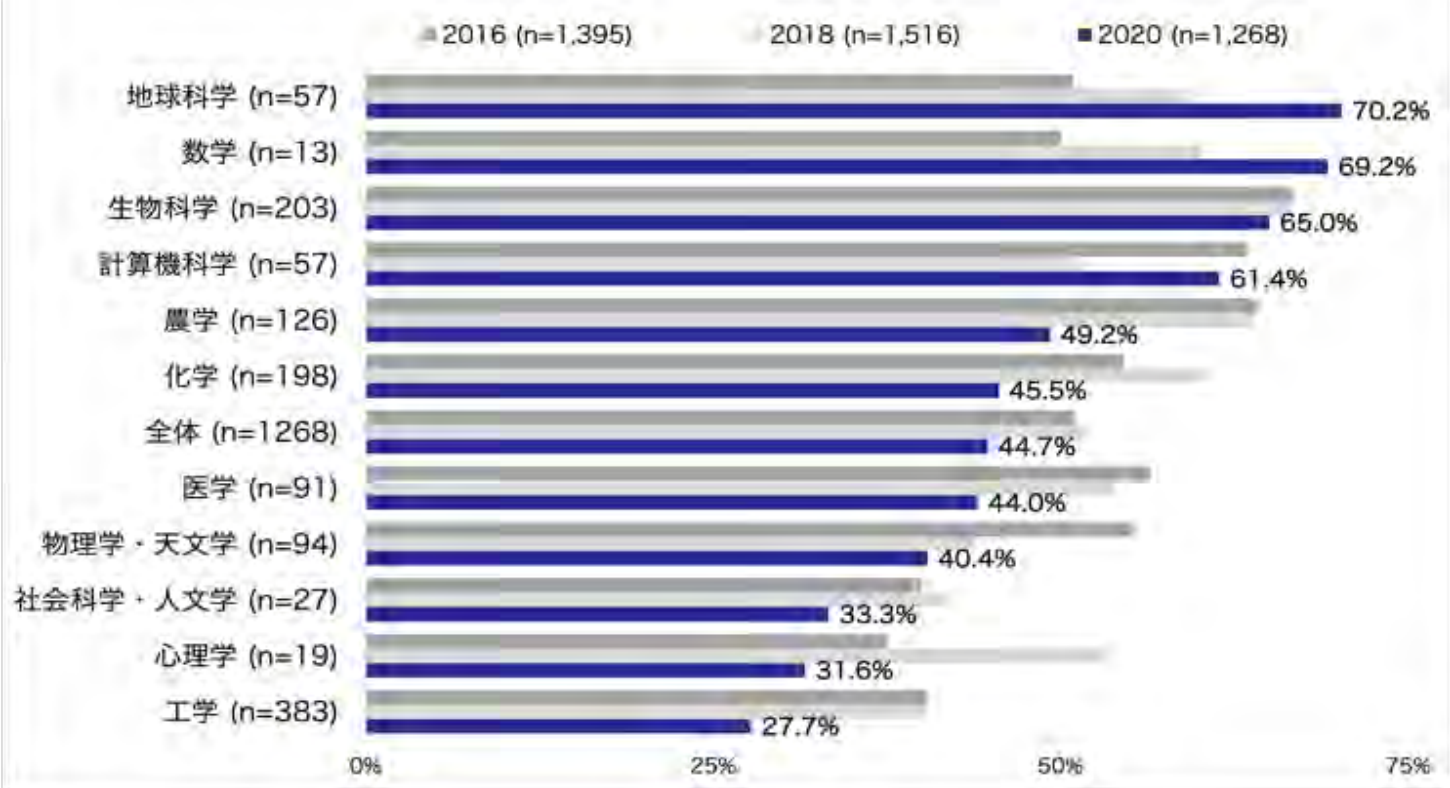
＜CiNii Research で検索可能な研究データ収載数＞



※CiNii Researchでは、これまでNIIが整備してきたサービス(CiNii ArticlesやCiNii Books等)で収載したデータを 内部で統合して、横断検索を可能にしている。

※研究データ以外の情報(論文、図書、その他の成果物等)についても2022年3月にかけて増加している。

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

出典： 科学技術・学術政策研究所「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」,
調査資料 (Research Material) 316, NISTEP, 2021-11(<https://doi.org/10.15108/rm316>)

ムーンショット研究開発制度における先進的データマネジメント

研究推進法人/PD: メタデータの保存・共有・公開

データカタログ

研究データ基盤システム (NII Research Data Cloud)
又はこれに相当するシステム

ムーン ショット 目標	データ No	データの 名称	データの 説明	管理者	連絡先	データの 所在場所	データ管理方針		その他
							公開・共有の方針	利活用・提供の方針	
	1								
	2								

集約・構成

メタデータ

メタデータ

メタデータ

PM/研究者:
管理対象データの保存・共有・公開

研究データ基盤システム等 (外部ストレージ、レポジトリ等可)

PM/研究者がメタデータを付与

メタデータ

メタデータ

メタデータ

管理対象
データ

管理対象
データ

管理対象
データ

PM/研究者が範囲を定める

非管理データ

プロジェクト① (PM/研究者)

集約・提出

研究データ基盤システム等
(外部ストレージ、レポジトリ等可)

メタデータ

メタデータ

管理対象
データ

管理対象
データ

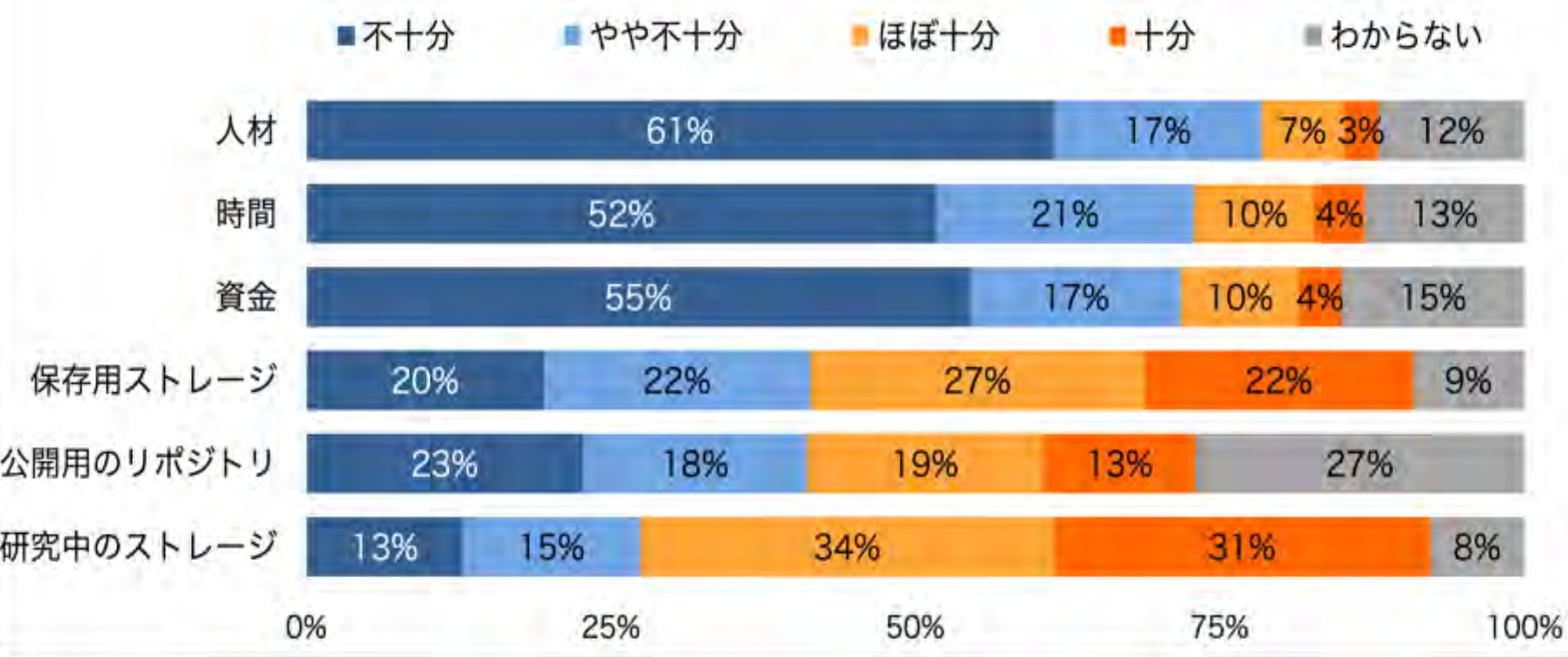
非管理データ

プロジェクト② (PM/研究者)

※PMは、管理対象データの管理・利活用の考え方を示した計画書（データ・マネジメント・プラン (DMP)）の原案を研究者に作成させ、取りまとめ、研究推進法人に提出する。

データの整備・公開資源の充足度

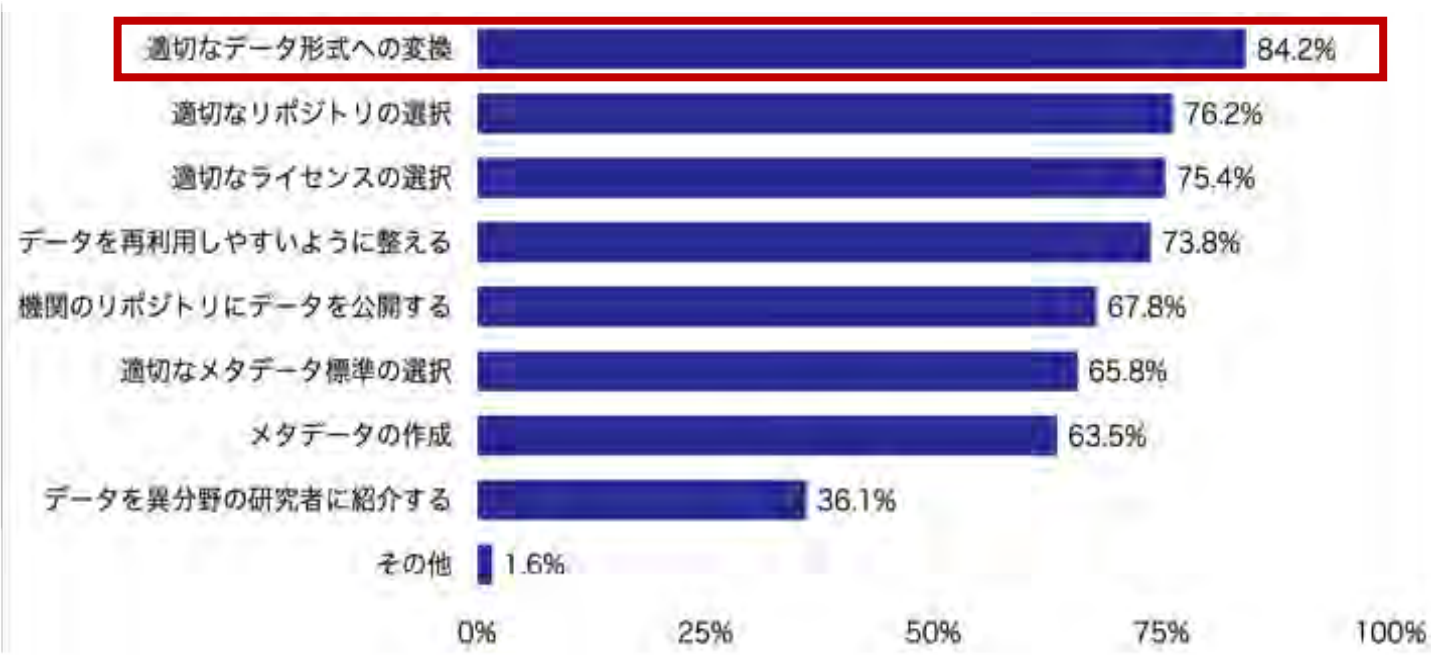
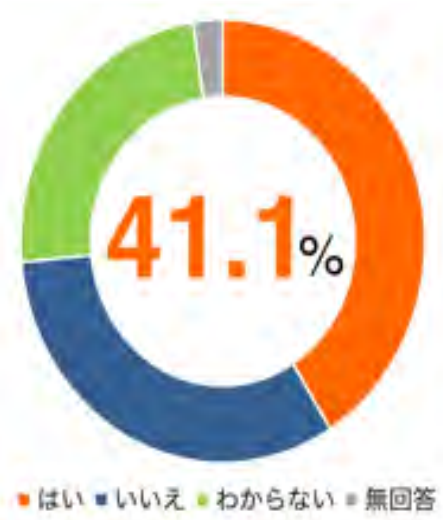
(n=1,188)



- 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いえない
- 公開用のリポジトリは「わからない」の比率が高い

出典： 科学技術・学術政策研究所「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」, 調査資料 (Research Material) 316, NISTEP, 2021-11(<https://doi.org/10.15108/rm316>)

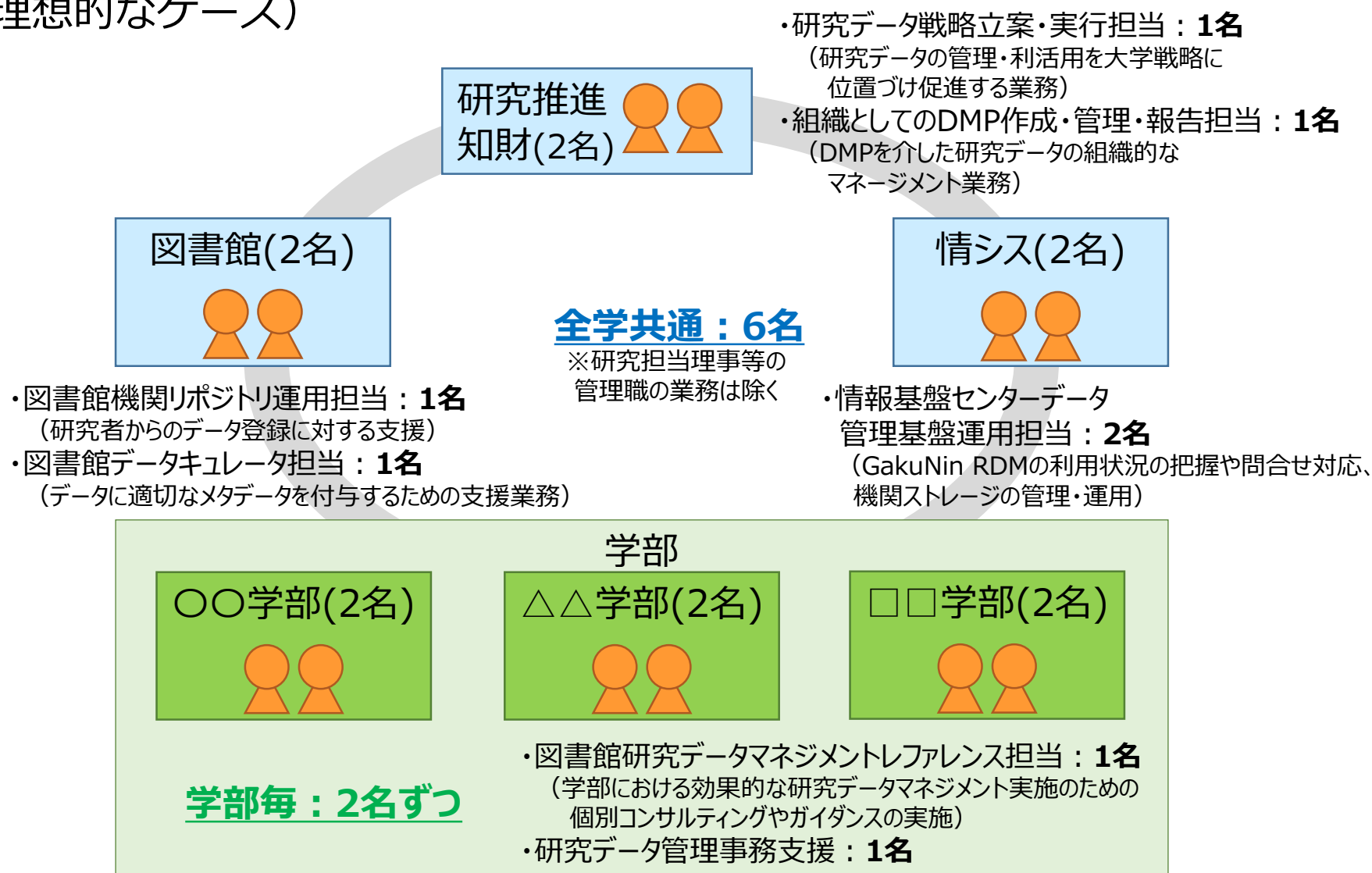
データの整備・公開・保存の依頼意思



- 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

研究データマネジメント業務の規模感（イメージ）

5,000人程度の大学の場合
(理想的なケース)



研究データの管理・利活用に向けた論点

- 公募型研究費（ムーンショット、次期SIPなど）からの横展開及び研究分野（マテリアル、バイオ等）からの横展開を進めつつあるが、大学等の組織からのアプローチが弱い面がある。
 - 形だけのデータポリシー策定と中身の無い機関リポジトリの構築が進む可能性がある。
- 大学研究力強化・大学支援施策と連動して、研究データの管理・利活用を進めるべきではないか。特に、研究データの共有・公開は、ハードとしての設備・機器の共有と密接な関係にあり、両者を連携して進めるべきではないか。
- 研究データの管理・利活用を担う人材（※）も含め、大学等における研究支援体制が脆弱であることが、研究者のインセンティブを削いでいるのではないか。
 - （※）データキュレーター・図書館職員（デジタル転換の取組等）、URA、シニア人材等。
- なお、日本学術会議に対し、アカデミアの立場から見た研究データ管理・利活用の課題やデータ駆動型研究の推進方策について、並行して審議依頼を行っている。

参考資料

データサイエンスに係る各国の取組

令和4年2月10日
情報委員会（第23回）
資料3-1より

【EU】

■ 欧州オープンサイエンスクラウド（EOSC）

効果的なオープンサイエンス移行の加速、支援を目的とし、サービス・システムへの信頼度の高いアクセスや、分野、社会、地理的境界を越えて共有される研究データの再利用を可能とするための基盤を構築。

■ 欧州データ戦略（A European strategy for data）

「データの単一市場」である「欧州データ空間」構築を目指す。

独仏が主導する産業界の欧州統合データ基盤プロジェクト「GAIA-X」とのシナジーも重視

【ドイツ】

■ ドイツ研究データインフラストラクチャ（NFDI）構築

研究データを共通基盤に集積する構想に毎年9,000万€（2020-2029）の投資を計画しており、研究データの収集、作成、分析等の学際的な研究データ・ライフサイクルのすべてのをサポートする「NFDIデータサイエンス・人工知能」の支援を決定。

■ データ統合シミュレーションの研究開発を実施

第1期は工学的アプローチからのモデル化などを実施（5年x2 合計約6,500万€）、第2期ではさらに大容量データやセンサー計測の取り扱い、データ解析手法の開発などを進め、統合システム科学に発展させる（1期7年 600万€/年＝予定）応用分野は、環境、健康、製造業など

【英国】

- 英国研究・イノベーション機構（UKRI）のオープンアクセス・ポリシーを策定し、研究データを含む研究成果のオープン化を推進

【中国】

■ デジタルインフラ整備への大規模投資

5G、IoT、AI、科学イノベーション施設等が対象で、追加投資額は2025年までに約10兆元（約150兆円、約1兆ドル規模）。

【フランス】

■ 新たな組織、基盤等の体制を整備

国立研究機構（ANR）が研究データ公開、共有等の推進のためデジタル戦略・データ部門を創設。また、研究データの共有、公開の促進のため国家研究データプラットフォーム（Recherche Data Gouv）を2022年第1四半期までに構築することを計画。

■ データサイエンスに係る投資

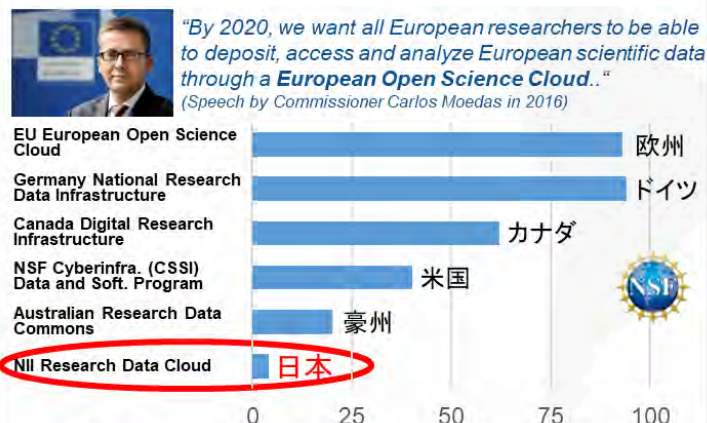
研究とイノベーションのための包括的で全国規模のeインフラの構築に2億2400万€を配分。また、大学のデジタル化に合計5600万€を投資

【米国】

- 7500万\$を投じ、5つのデータ駆動型科学の研究拠点を設立

研究データ基盤の現状

日本の研究データ基盤は2020年度に運用を開始したが、研究開発経費は世界に比して極めて低い



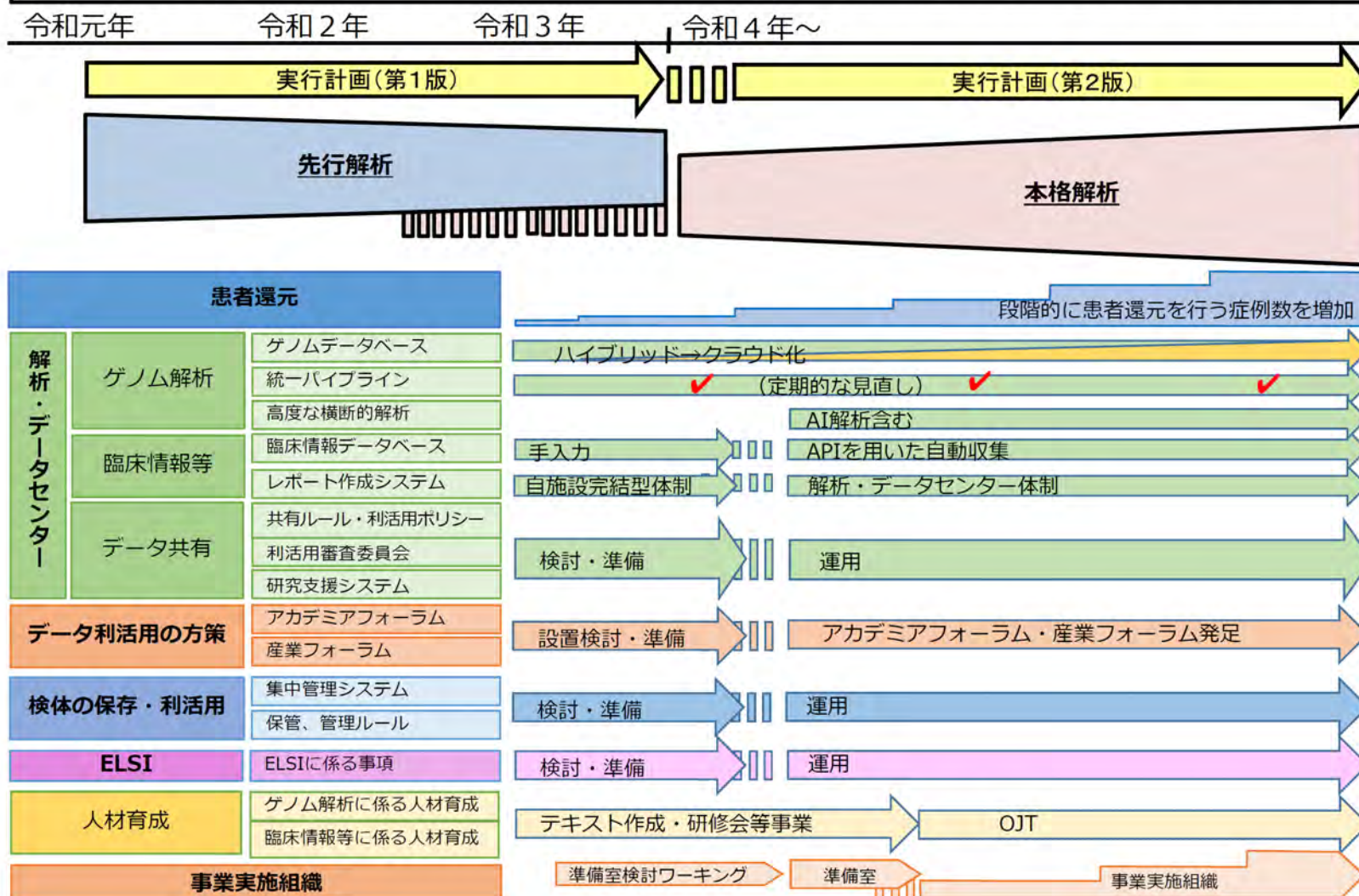
（各国の研究データ基盤関連開発経費より算出）[million USD/year]

- ・世界各国は2010年代よりデータ駆動型研究基盤開発に投資し、最近ではオープンサイエンス実現に向けた基盤を整備

関係府省の事例（厚生労働省）

「全ゲノム解析等実行計画」（第2版）について（案）

○「全ゲノム解析等実行計画」に基づき、全ゲノム解析等により明らかとなった当該疾患の治療等のために有益な情報等を患者に還元するとともに、研究・創薬などに向けた利活用を進め、新たな個別化医療等を患者に届けることを目指す。



関係府省の事例（厚生労働省）

「全ゲノム解析等実行計画」に係るこれまでの解析実績

「全ゲノム解析等実行計画（第1版）」では、がん領域においては、5年生存率が相対的に低い難治性のがんや、稀な遺伝子変化が原因となることが多い希少がん（小児がん含む）、遺伝性のがん（小児がん含む）について、全ゲノム解析等を行うこととされた。また、難病領域においては、単一遺伝子性疾患、多因子疾患、診断困難な疾患を優先して全ゲノム解析等を行うこととされた。それぞれの解析実績と解析予定は以下の通り。

がん領域	解析実績（※） （2019年12月以降）	解析予定 （2021年度）
○難治性のがん （白血病、食道がん、肝臓がん、 胆道/膵臓がん、肺がん、卵巣がん）	約500症例 ・膵臓がん ・白血病 ・肉腫	原則生存しており、解析結果を 還元可能な、保存検体を有する 約9,900症例 ・血液領域 ・消化器領域 ・婦人科領域 ・呼吸器他領域 ・希少がん領域 ・小児がん領域 この内、新規患者600症例
○希少がん（小児がん含む）		
○遺伝性のがん（小児がん含む）		
約3,250症例		
難病領域	解析実績 （2019年12月以降）	解析予定 （2021年度）
○単一遺伝子性疾患 ○多因子疾患 ○診断困難な疾患	約2,500症例	約3,000症例