

# オープンサイエンスと研究データ基盤整備

---



2022年11月14日

内閣府 科学技術・イノベーション推進事務局 参事官(統合戦略担当)／  
文部科学省 科学技術・学術政策研究所 上席フェロー  
赤池 伸一

# 科学技術・イノベーション政策

## 国内外における情勢変化

(先端技術の覇権争い、  
災害の激甚化、  
情報独占と富の偏在化…)

## 我が国の科学技術・ イノベーションを取り巻く現状

(目的化したデジタル化、  
研究力の低下…)

## COVID-19による変化の加速

(リモート化・スマート化、データ利活用、  
I T 投資、メタバース…)

## 目指す社会 (Society 5.0)

**国民の安全と安心を確保する**  
持続可能で強靱な社会



**一人ひとりの多様な幸せ(well-being)**  
が実現できる社会

# Society 5.0（2016年に提唱）

- サイバー空間とフィジカル（現実）空間を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会（Society）
- 第5期科学技術基本計画（2016年1月閣議決定）で初めて打ち出した概念

新たな社会  
"Society 5.0"



Society 1.0 狩猟



Society 2.0 農耕



Society 4.0 情報



Society 3.0 工業



# サイバー空間とフィジカル空間の高度な融合

フィジカル（現実）空間から**センサー**と**IoT**を通じてあらゆる情報が集積（**ビッグデータ**）  
**人工知能（AI）**がビッグデータを解析し、高付加価値を**現実空間にフィードバック**

## これまでの情報社会(4.0)



# オープンサイエンスについて

ICTの活用により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

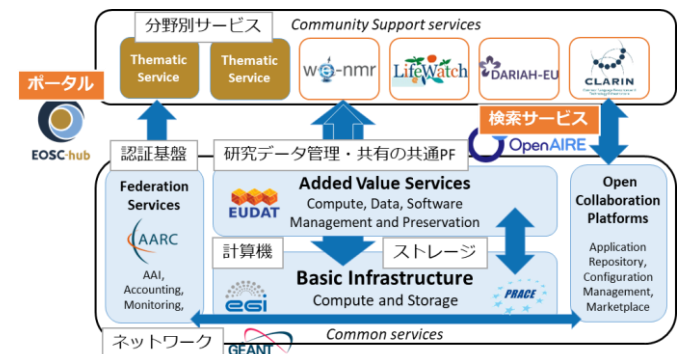
オープンサイエンス＝オープンアクセス＋オープンリサーチデータ

**オープンアクセス（OA）：（研究論文をオープンに）**

- インターネットの特性を生かして論文を誰でも自由に利活用できるように
- 商業出版社の寡占に端を発する学術誌高騰問題への対処
- Green OA：（米国、日本）出版者版論文の代替物（著者最終稿）を機関リポジトリ等に掲載して公開
- Gold OA：（英国、欧州）オープンアクセス掲載料（APC:Article Processing Charge）を支払うことで出版者版論文をオープンに

**オープン（リサーチ）データ：（研究データをよりオープンに）**

- 論文の根拠データを皮切りに、研究データを共有・公開することで新しい科学的価値をイノベーションを効率よく生み出す基盤づくりを推進。（論文で起きた問題の根本的解決を目指す）
- 政策は研究データ基盤整備とインセンティブ設計に注目



出典：国立情報学研究所 山地一禎教授 提供資料

**進む国際イニシアチブでの検討**

**G7：**

- オープンサイエンスWGを設置（EUと日本が共同議長）。2016年より毎年WS等を開催。

**OECD：**

- 公的資金による研究データアクセスに関する理事会勧告（2021年1月）

**UNESCO：**

- オープンサイエンス勧告（2021年12月）



出典：文部科学省科学技術・学術政策研究所 林和弘データ解析政策研究室長 提供資料を基に内閣府作成 5



## 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

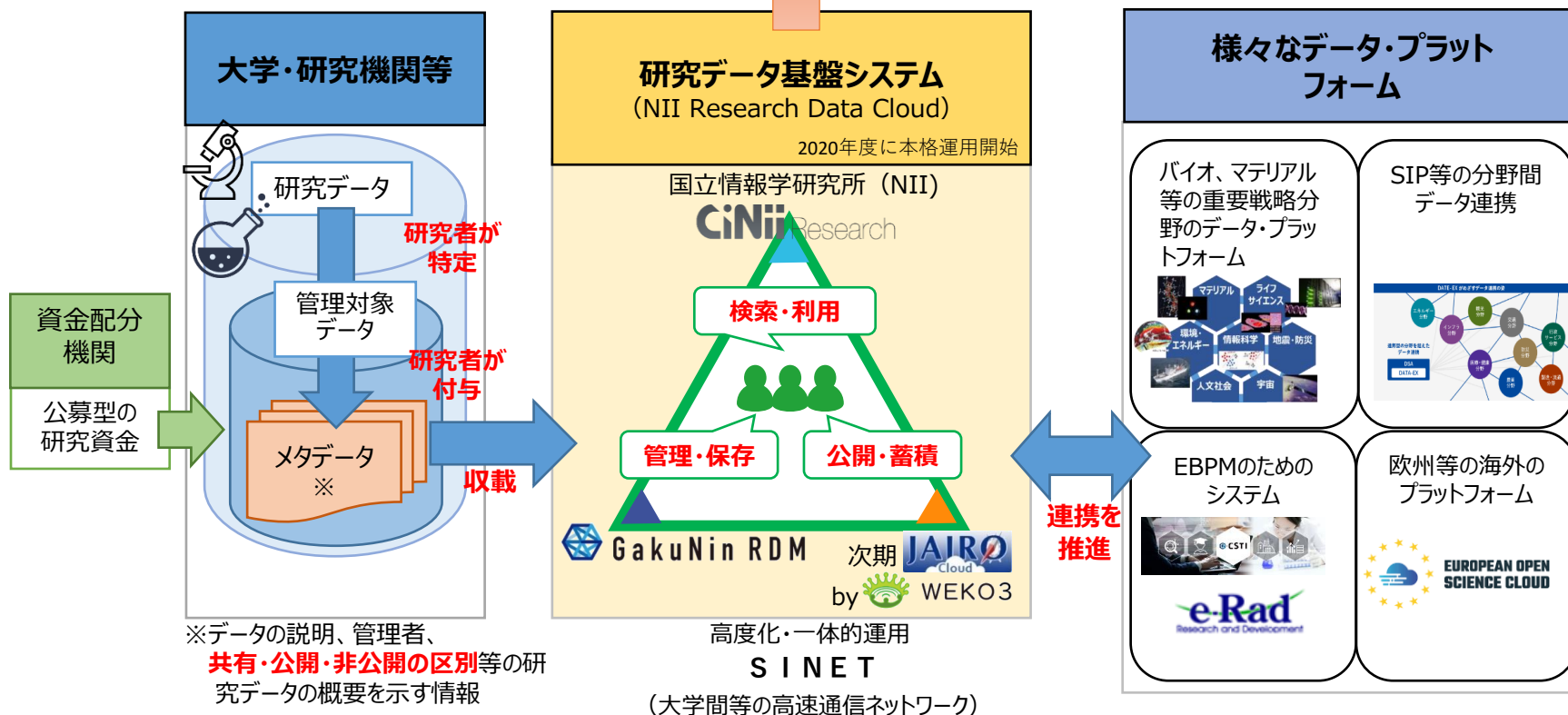
## 【政策文書】

- 第6期科学技術・イノベーション基本計画（2021年3月）
- 統合イノベーション戦略2022（2022年6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）

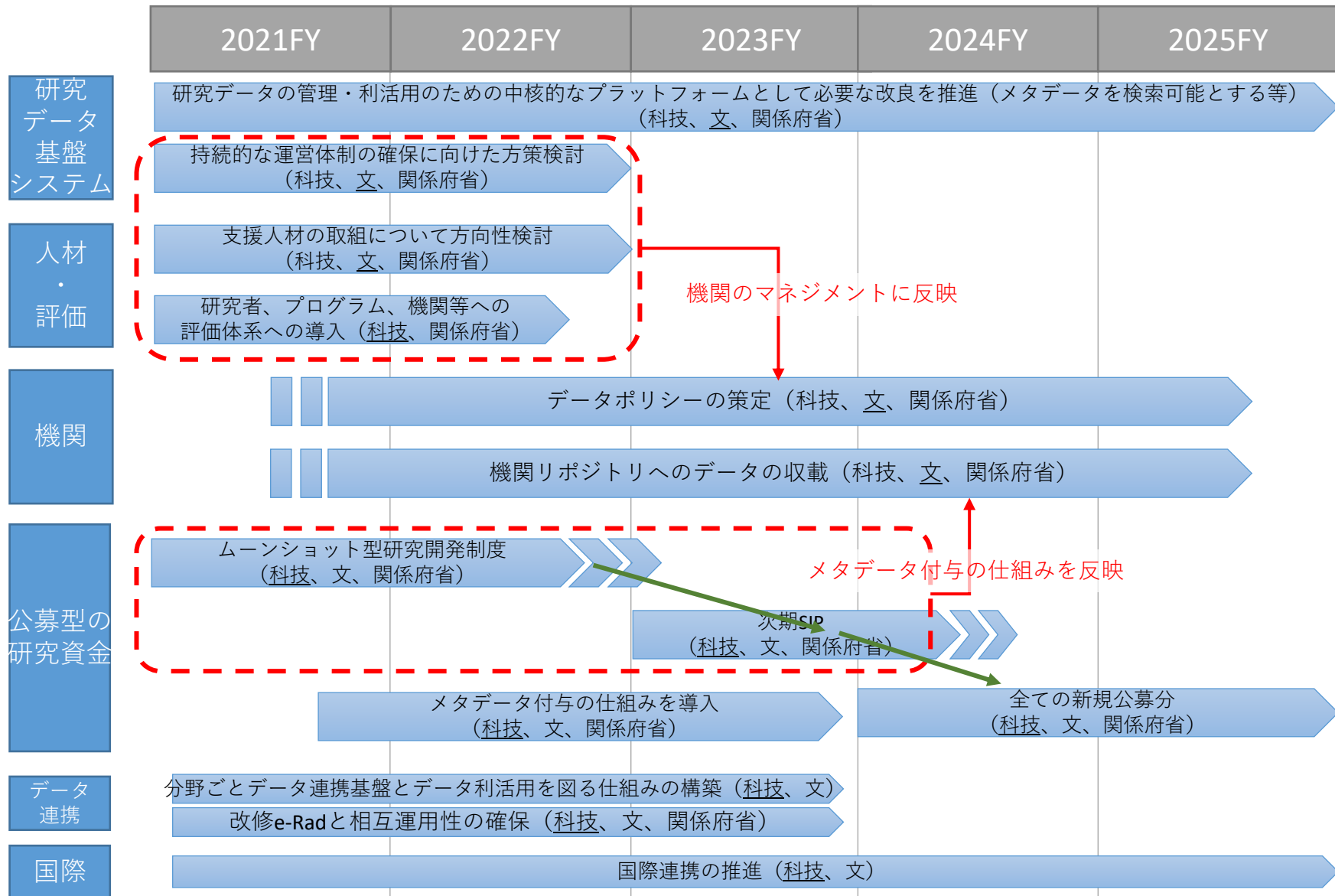
研究データ基盤システムを中核としたデータ・プラットフォームの構築  
→研究データの公開・共有を推進、産学官のユーザが **データを検索可能**

アカデミア、産業界等

研究データを検索・利活用



# 研究データ管理・利活用の取組に関するロードマップ





# 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  
（国立大学法人・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

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で、電磁的な形態により管理可能なものをいう。研究ノートやメモ、実験や観測、シミュレーション等から直接得られたデータやそれを加工したデータ、論文のエビデンスとなるデータ等が含まれる。

## 管理対象データ

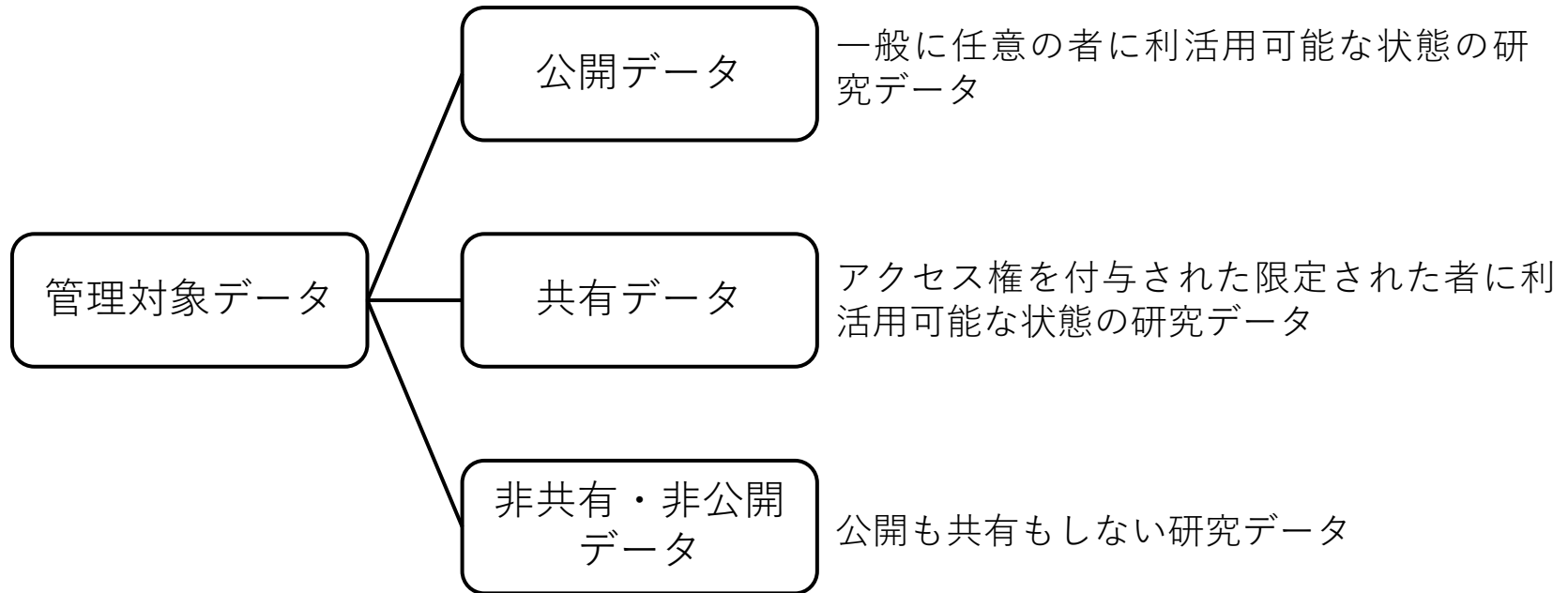
研究データのうち、研究者の所属する研究開発を行う機関や資金配分機関の基準等に基づいて、管理・利活用の対象として、研究者がその範囲を定めるものをいう。

対応

## メタデータ

管理対象データを説明するための情報から構成されるデータをいう。研究データの名称、研究データの説明、研究データの管理者及びその連絡先、研究データの所在場所、研究データの保存・公開・共有の方針等の情報を含む

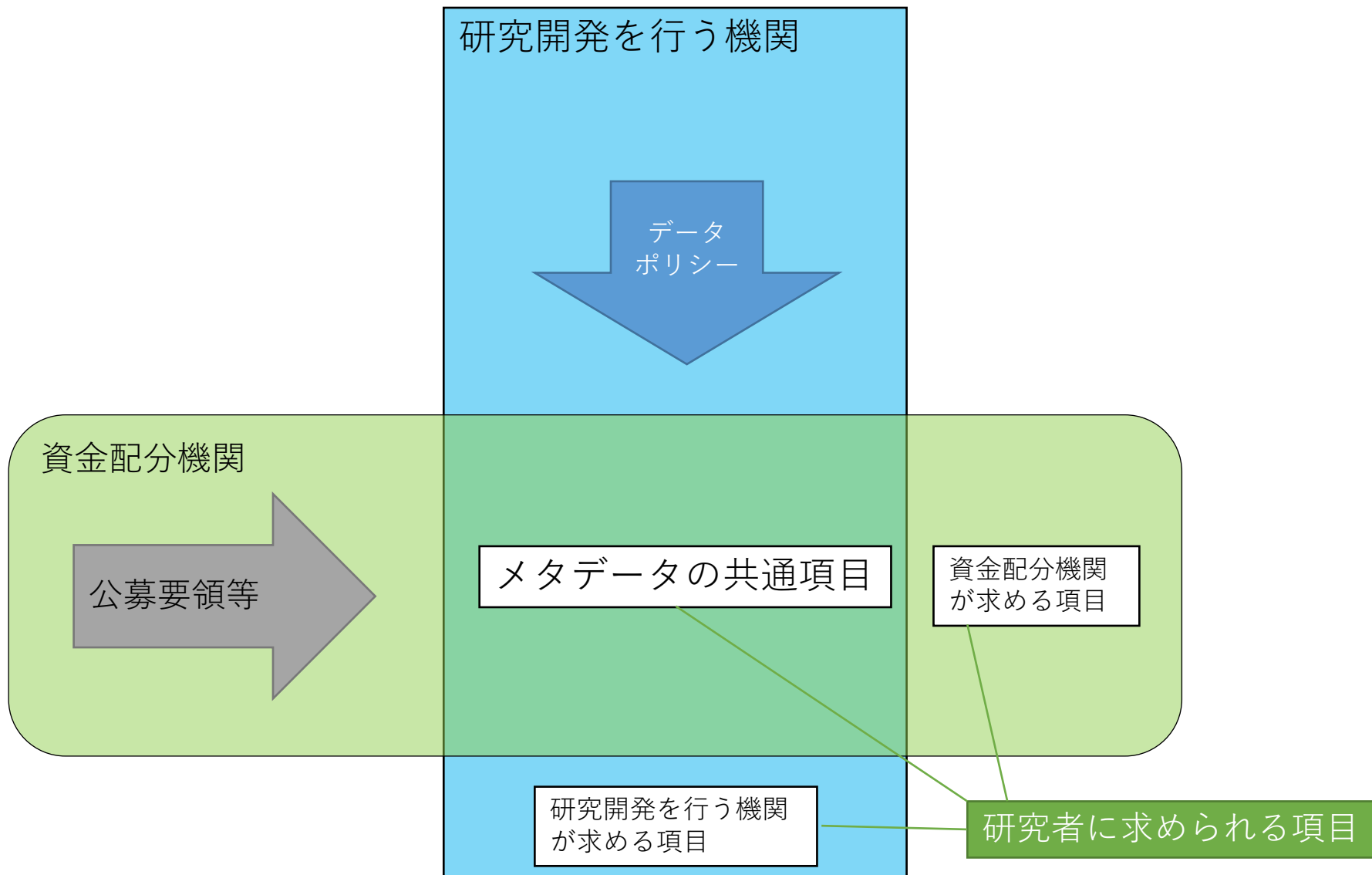
# 管理対象データの公開及び共有の区分



※) 「公的資金による研究データに関する基本的な考え方」から要約

- ✓ 研究分野等の特性や、大学、大学共同利用機関法人、国立研究開発法人等のデータを管理する組織の特性に配慮して、公開、共有、又は非共有・非公開の判断が行われる必要がある
- ✓ 我が国の産業競争力や科学技術・学術上の優位性を確保するために重要な情報を含む可能性があるため、個人情報、企業の秘密情報、研究の新規性、我が国の安全保障等の観点から留意すべき研究データは非公開とすることが求められる
- ✓ 産業競争力や科学技術・学術的な優位性を確保するために、公開による利活用の促進とのバランスを考慮しつつ、適切なエンバーゴ（時限非公開）期間を設定することも想定される
- ✓ 関係諸法令に従うとともに、データの取り扱いに関する各国の国内法及びEU規則並びにデータ管理の原則であるFAIR原則等の国際的な規則や慣行等との整合性に十分留意する必要がある

# メタデータに関する機関と公募型の研究資金の関係



# 基本的な考え方におけるメタデータの共通項目

(2021年11月30日時点)

|    | 項目                | 必須／任意 | 備考                                                                               |
|----|-------------------|-------|----------------------------------------------------------------------------------|
| 1  | 資金配分機関情報          | 必須※   | 公募型の研究資金を配分した資金配分機関（府省含む）の英語略称<br>公募型の研究資金以外の場合は記入不要                             |
| 2  | e-Radの課題番号        | 必須※   | 公募型の研究資金の場合は、e-Radに登録した課題番号<br>公募型の研究資金以外の場合は記入不要                                |
| 3  | プロジェクト名           | 必須※   | プロジェクトの研究代表者が統括する研究開発の範囲の名称<br>公式な名称がない場合は、研究者の所属機関のルールに従って入力                    |
| 4  | データNo.            | 必須    | 管理対象データを一意に特定するための番号<br>公募型の研究資金を配分した資金配分機関が付与<br>公募型の研究資金以外の場合はデータ管理機関（項目14）が付与 |
| 5  | データの名称            | 必須    | 学会資料、報告資料、測定結果などの中身の分からない名称は避ける                                                  |
| 6  | 掲載日・掲載更新日         | 必須    | メタデータの掲載日・掲載更新日                                                                  |
| 7  | データの説明            | 必須    | 端的かつ中身の分かる内容を記載                                                                  |
| 8  | データの分野            | 必須※   | e-Radの研究分野（主分野）。e-Radとの連携により、自動入力される予定。                                          |
| 9  | データ種別             | 必須    | 研究データ基盤システム上では、通常は「データセット」を標準とするが、データの特性に応じて「データセット」以外の種別を選択可能                   |
| 10 | 概略データ量            | 任意    | 1GB未満、1GB以上10GB未満、10GB以上100GB未満、100GB以上等の区分により記載                                 |
| 11 | 管理対象データの利活用・提供方針  | 必須    | 無償／有償、ライセンス情報、その他条件（引用の仕方等）等を記載                                                  |
|    | アクセス権             | 必須    | 公開／共有／非共有・非公開／公開期間猶予から選択                                                         |
|    | 公開予定日             | 必須    | 公開期間猶予を選択した場合、公開予定日を記載                                                           |
| 12 | リポジトリ情報           | 必須    | 現在のリポジトリ情報、あるいはプロジェクト後のリポジトリ情報                                                   |
|    | リポジトリURL・DOIリンク   | 任意    | 情報があれば記載                                                                         |
| 13 | データ作成者            | 任意    | 管理対象データを生み出した研究者の名前                                                              |
|    | データ作成者のe-Rad研究者番号 | 任意    | 管理対象データ作成者のe-Radの研究者番号                                                           |
| 14 | データ管理機関           | 必須    | 各データを管理する研究開発を行う機関のe-Radに登録された法人名                                                |
|    | データ管理者            | 必須    | データ管理組織において各管理対象データを管理する担当者の名前                                                   |
|    | データ管理者のe-Rad研究者番号 | 任意    | 管理者のe-Radの研究者番号<br>e-Rad研究者番号がない管理者は記入不要、ある場合は必須                                 |
|    | データ管理者の連絡先        | 必須    | データ管理者の所属機関の住所や電話番号、メールアドレス等                                                     |
| 15 | 備考                | 任意    |                                                                                  |

※公募型の研究資金による研究活動の場合

# 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<br>ショット<br>目標 | データ<br>No | データの<br>名称 | データの<br>説明 | 管理者 | 連絡先 | データの<br>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

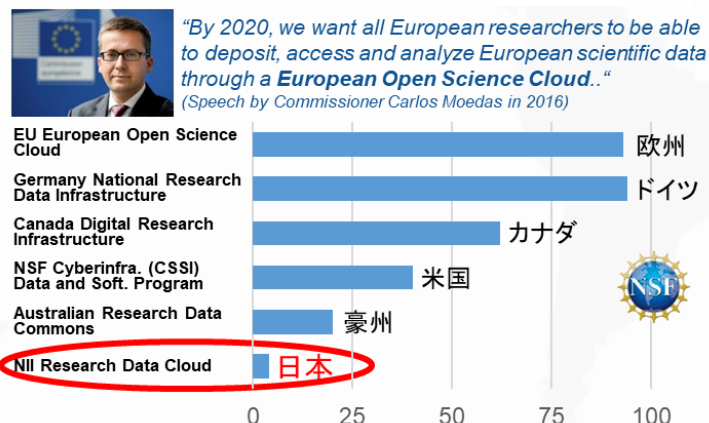
研究とイノベーションのための包括的で全国規模のeインフラの構築に2億2400万€を配分。また、大学のデジタル化に合計5600万€を投資

## 【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

## G7 Research Compact 「研究協約」

2021年6月 G7コーンウォール（英）・サミット 首脳コミュニケ附属文書

<主として以下の点を確認>

- 「開放性(openness)」、「相互主義(reciprocity)」及び「協力(cooperation)」が G7共通の価値であること
- 可能な限りオープンで、安全かつ効果的な国際協力を支える原則を、堅持し守るため協働すること
- Covid19の経験を教訓に、将来の様々な未曾有の危機への備えとして、また人類共通の課題解決に資するものとして、国境を越えたオープンで、迅速かつ機動的な研究協力、データ共有の重要性



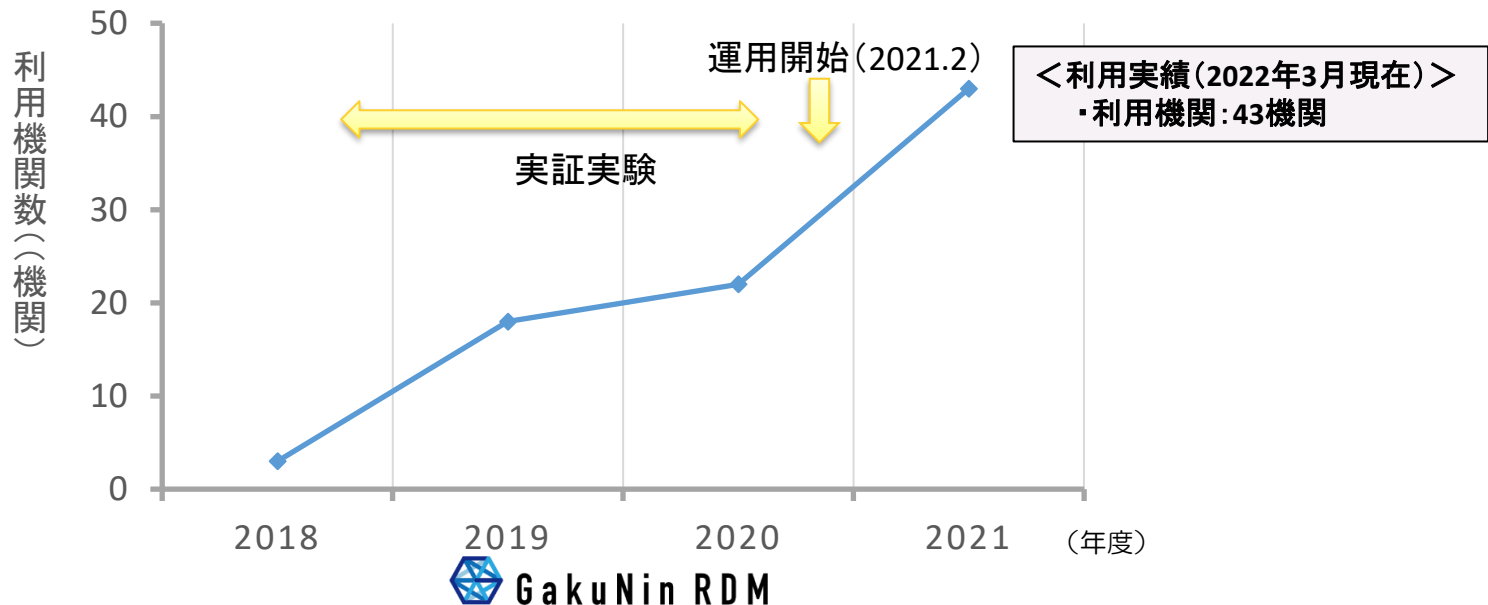
- 2016年発足の**G7オープンサイエンスWG**（共同議長：日本及びEU）により議論を継続

# 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)

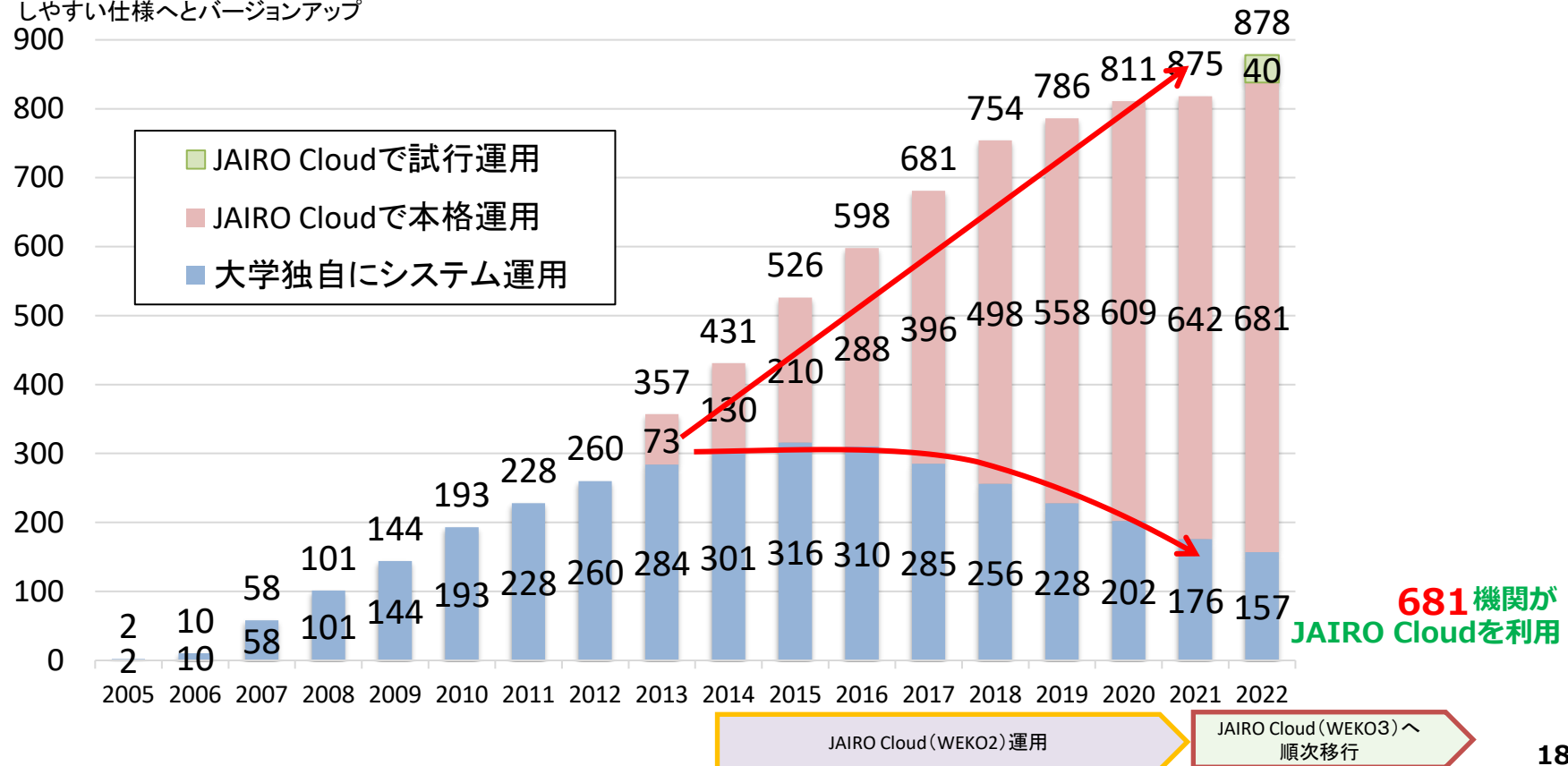
## JAIRO Cloud

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

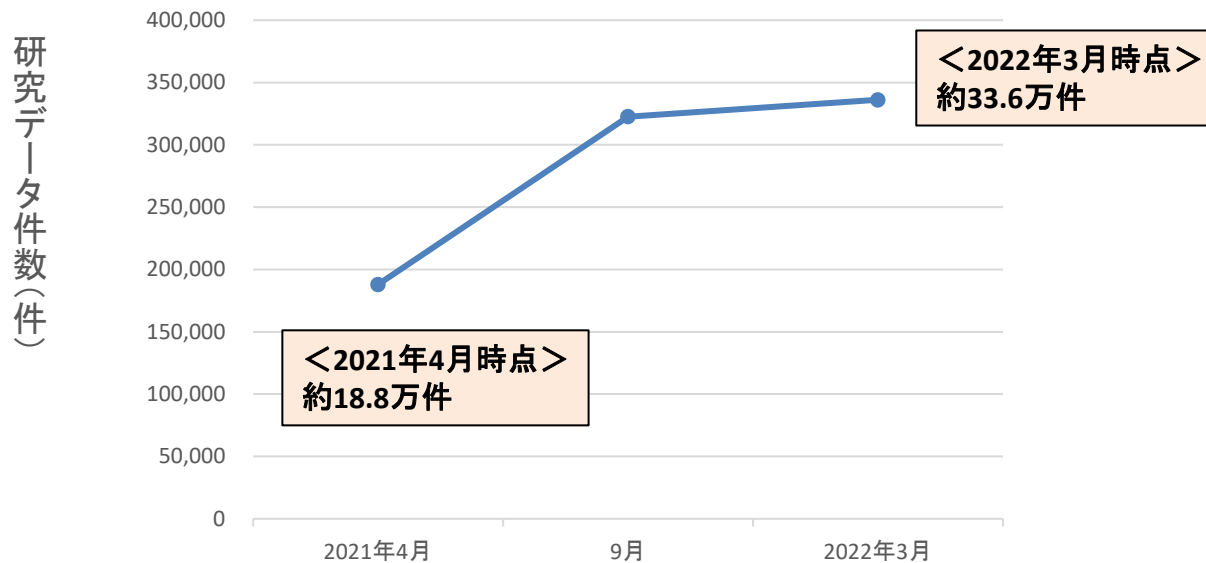


# 研究データ基盤システムの利用実績(3)

## 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>

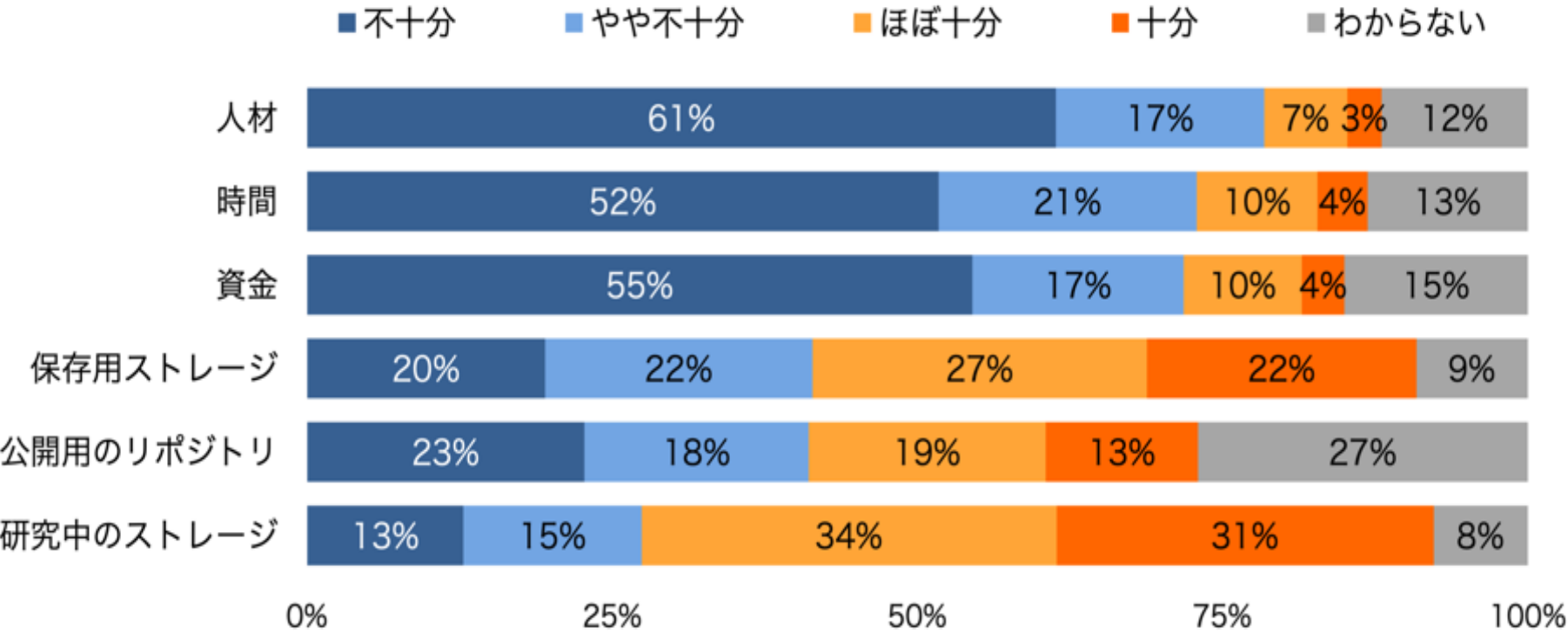


※CiNii Researchでは、これまでNIIが整備してきたサービス(CiNii ArticlesやCiNii Books等)で収載したデータを 内部で統合して、横断検索を可能にしている。

※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。

# データの整備・公開資源の充足度

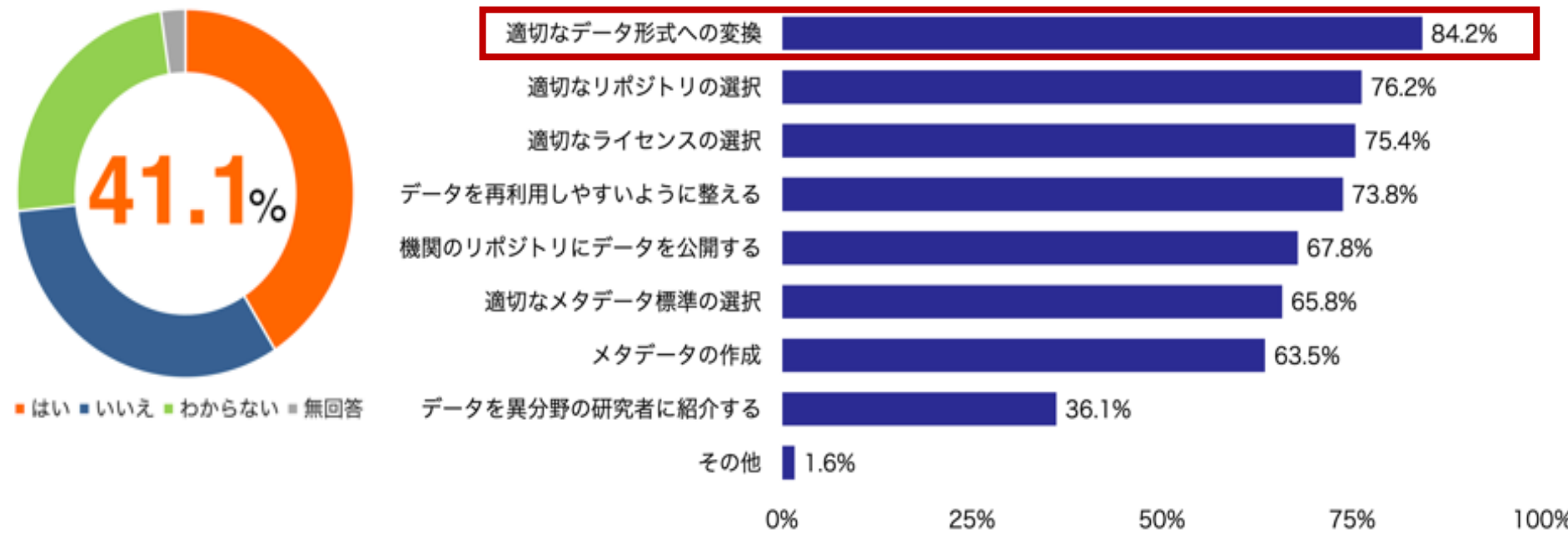
(n=1,188)



- 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

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> )

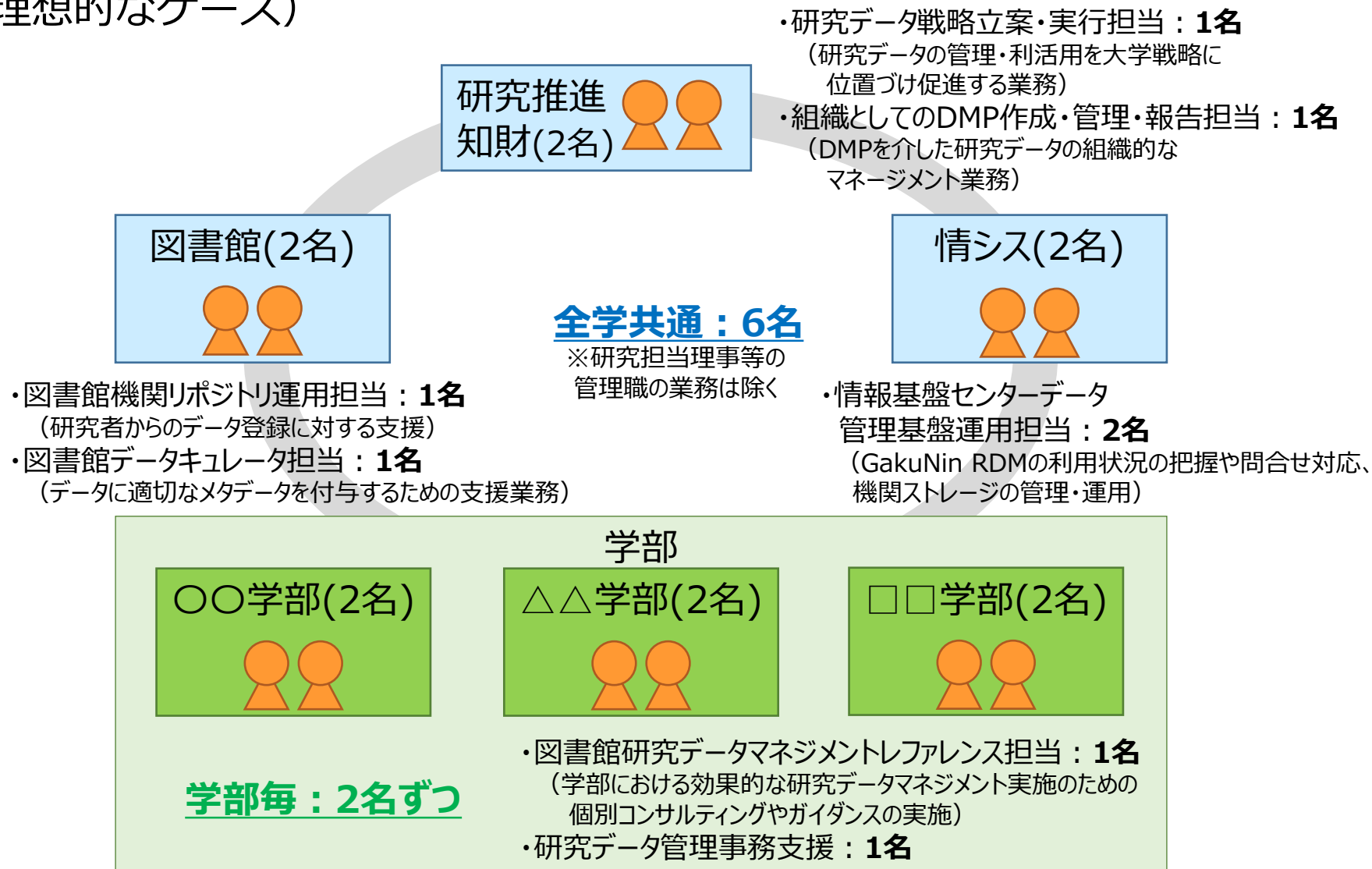


# 研究に専念できる時間の確保に向けた取組



# 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）



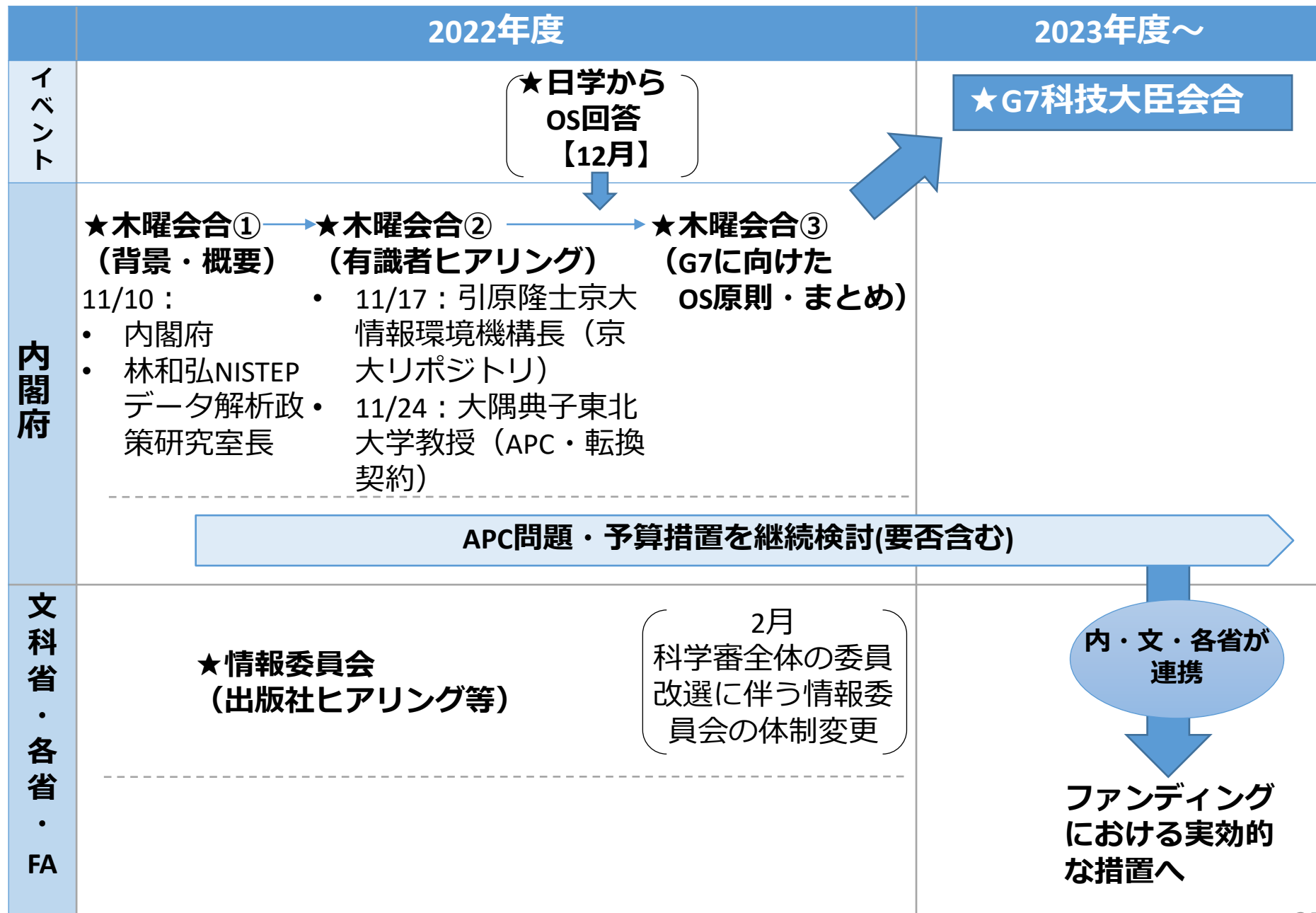
# 米国OSTP 即時OA方針

2022年8月25日 OSTPより政府機関に対するメモランダム

- 米国の連邦機関は、1億ドル超の年間研究開発費の連邦機関については180日以内、その他の連邦機関については360日以内に、**即時OA方針の実行計画をOSTP及び行政管理予算局（OMB）に提出**する。
- 各連邦機関の即時OA方針は、2024年末までに確定および公開され、公開の1年後までに施行されることとする。
- 即時OA方針の対象は、公的資金から生み出された査読付き学術論文および研究データ。
  - 論文：出版後、即時に助成機関の指定するリポジトリを通じ、無償で誰でもアクセス可能とする。
  - 研究データ：論文の根拠データは原則論文の公開と同時に公開。
- 研究インテグリティの確保に向けた研究助成機関の取るべきアクションについても言及。

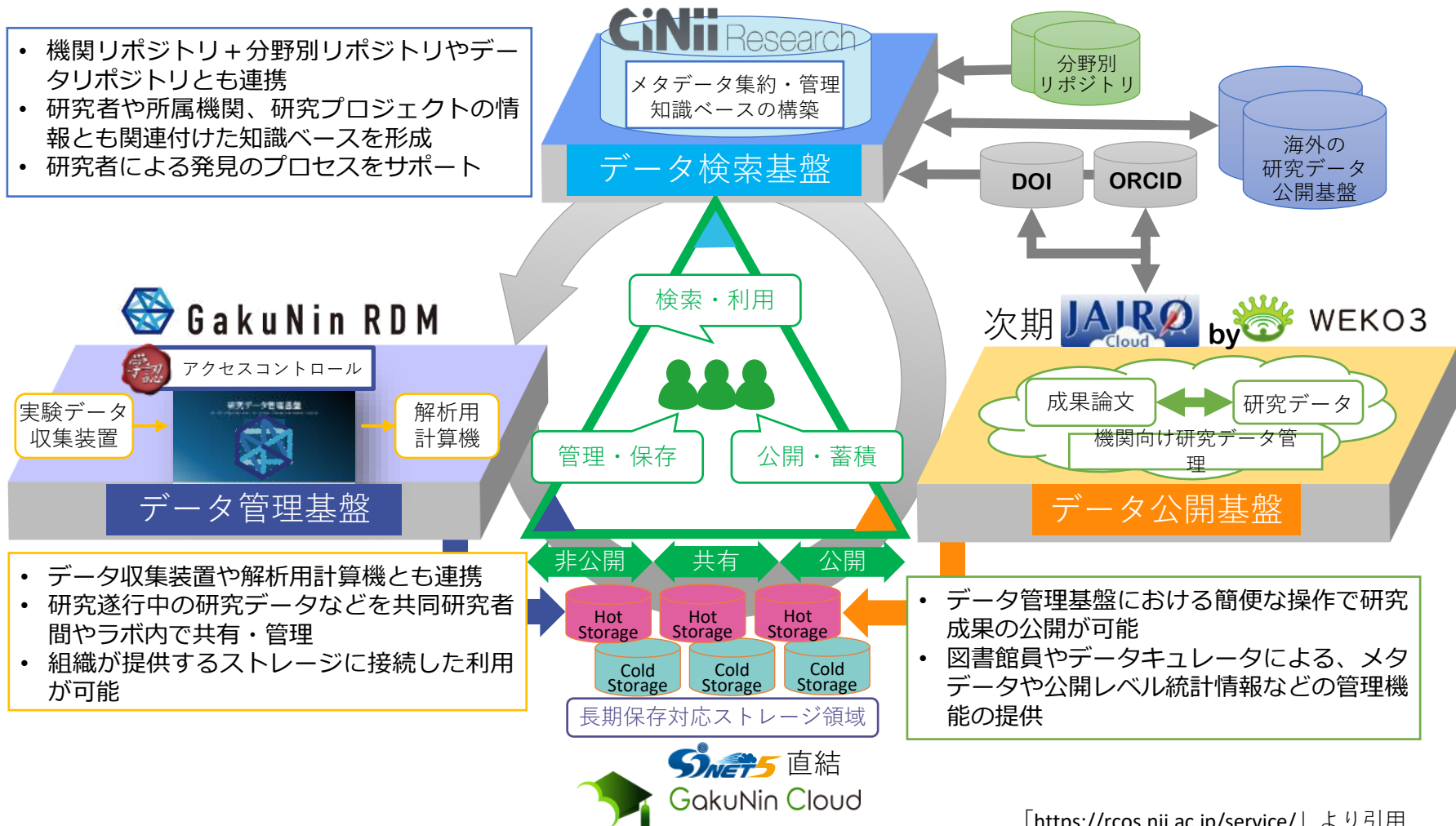
<https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>

# オープンアクセス問題に関する今後の進め方



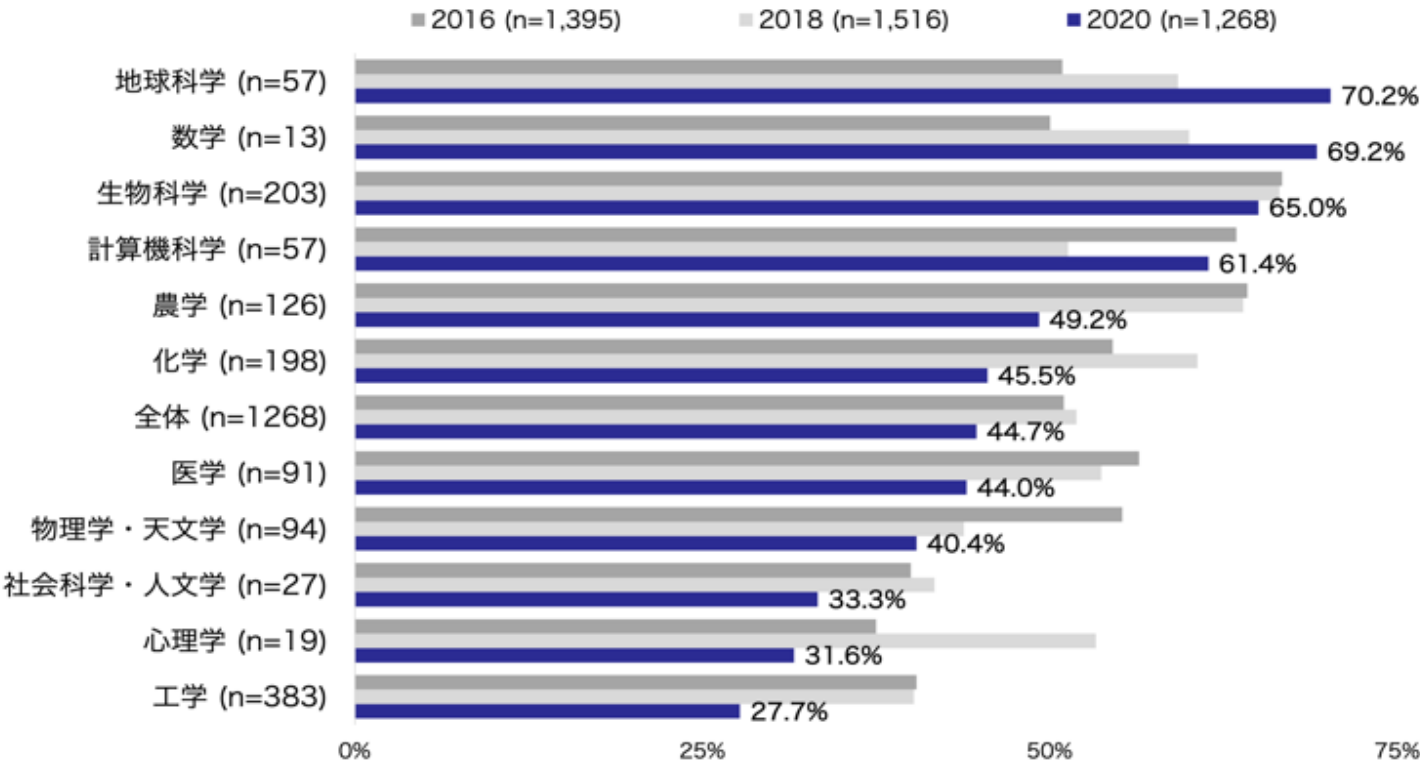
# 参考資料

# 研究データ基盤システム (NII Research Data Cloud)



「<https://rcos.nii.ac.jp/service/>」より引用

# 分野別データ公開経験



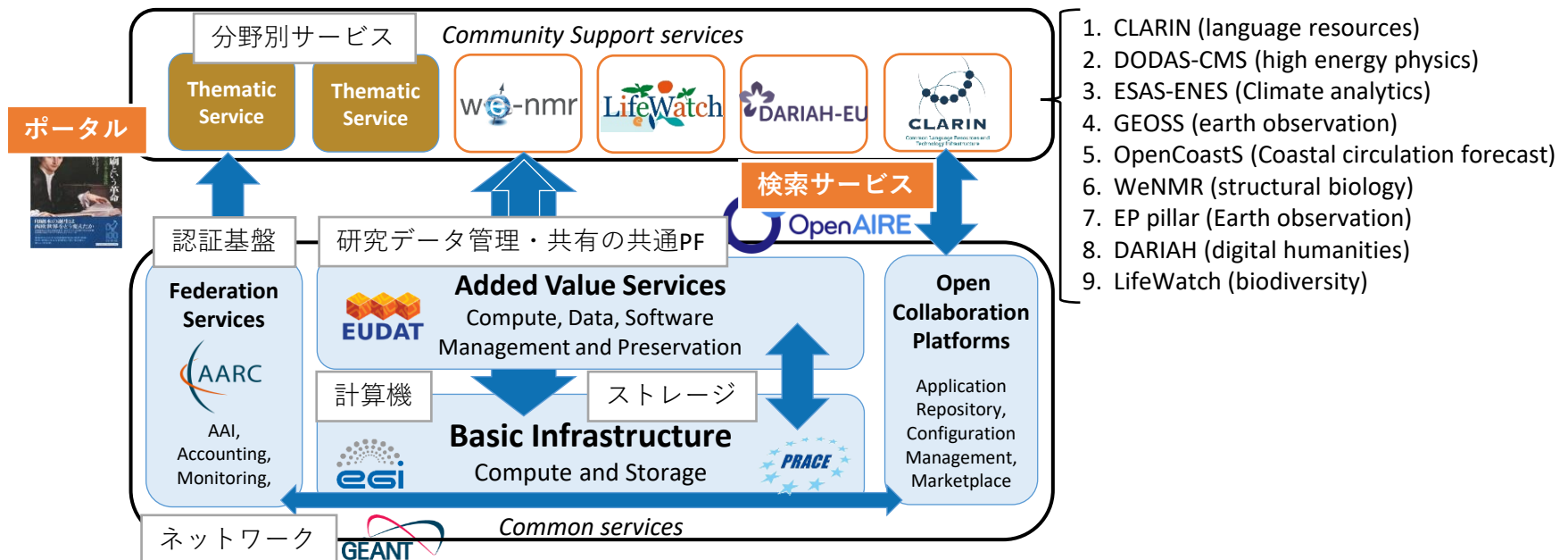
- 分野による差が大きい
- 地球科学（1位）と数学（2位）は前回調査から増加。ただし回答者数が比較的少ないため少数の回答が影響している可能性あり

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> )



# European Open Science Cloud (EOSC)

- **従来**：個々の大学や研究機関でEインフラを構築
- **今後**：個々のインフラを繋ぎEU全体でEインフラを構築
  - 既存のEインフラをベースにネットワークから分野別DBまでを有機的に結合
  - ポータル（EOSC-hub） 検索サービス（OpenAIRE） から具体化
  - 人社系へのサポートなどロングテールにおける活用も考慮
  - 産業界との連携も考慮



出典：国立情報学研究所 山地一禎 教授