

# 研究データ管理を支える学術情報基盤の紹介

国立情報学研究所  
オープンサイエンス基盤研究センター

谷藤幹子

# 研究データとは.

## 世界での研究データ(research data) :

- **研究データとは、研究の結論を裏付けるあらゆるデータであり、メタデータとスキーマ（記述規則）によって機械可読に構造化し、PIDで識別・引用可能にすることで、FAIRな共有・再利用を可能にする対象のこと。**
  - 活用するデータ
    - 研究データへのDOI登録ガイドライン：メタデータの定義、メタデータ・スキーマ、DOI 付与の考え方、分野横断スキーマ設計、schema.org + JSON-LDの利用例

## 日本で「研究基盤」の対象としている研究データとは :

- **科学・人文社会を含む研究過程で生成・取得・加工され、研究の結論や再現性を裏付ける一次・二次のデータ全般**
  - 数値表、画像・音声、コードと設定、計測ログ、シミュレーション出力、調査回答、試料情報など、形式は多様
  - 本質は「いつ・誰が・どの環境で・どう作ったか」という来歴と文脈が伴うこと。
  - この「文脈」をメタデータとして構造化し、検索・共有・引用・再利用を可能にすることが、国際的な研究データ管理の前提※となっている。

## AI Readyとは :

- 研究データ基盤における「AI-Ready」とは、単にデータが保存されている（Findable/Accessible）だけでなく、機械学習モデルが即座に、かつ効果的に学習・解析を行える状態にあることを指し、メタデータの標準化※、クレンジングされた研究データが利用できることが期待されている。

※国際的な研究データ基盤では、FAIR原則（Findable/Accessible/Interoperable/Reusable）を満たすため、分野横断で通用するメタデータ項目と記述ルール（スキーマ）のが進む。

# 研究データをめぐる世界動向（欧州） -AI4EOSC

## 欧州の政策・エコシステムとしてのEOSC（European Open Science Cloud）

- └ AI4EOSC：EOSCにAIを統合するプロジェクト。AI活用を本格化
  - └ AI4OS：AI/ML/DLの実行・共有・再利用のための技術基盤を開発
    - FAIR原則によりデータの再利用や分散的な利用を可能年、研究者が簡単に使える形で提供することを目指す



目的	EOSCにおける統合AI基盤提供プラットフォームの整備
提供形態	PaaSに近い（ダッシュボード、一体化ツール群）
主な機能	<ul style="list-style-type: none"> <li>• 開発（Jupyter等）、実験管理（MLflow）、API化（DEEPaaS）、サーバーレス推論PF（OSCAR）といった各環境へのワンクリック展開</li> <li>• Nextcloud統合、Zenodo連携、DOI付与での公開などのデータ/成果連携</li> <li>• Federated Learning（Flower/NVFlare等）機能による分散データの活用</li> <li>• FAIR/プロベナンス機能による実行メタデータやモデル証跡の自動記録</li> </ul>
ユースケース	都市熱解析：熱画像×DLで建物の熱損失箇所検出 気象：雷雨・雹など局地事象のAI予測で早期警戒 植物病害：衛星/観測×AIで発生リスク予測、農薬削減
開発予算	約500万ユーロ（2022-2025） ※今日のレートで約9億2千万円

Ref. AI4EOSC報告書「CORDIS – EU research results」<https://doi.org/10.3030/101058593> によるとEUの Horizon Europe 助成により総予算約 €4.997M（約5百万ユーロ）で実施

Home About Powered by AI4EOOSC Documentation **Code**

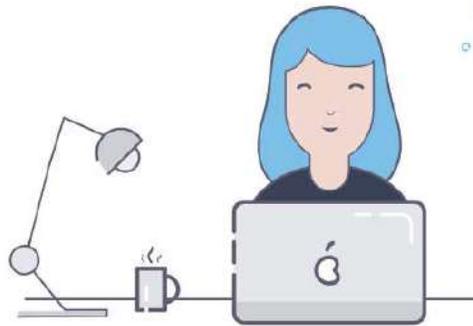
GitHubでソース公開 AI4OS Hub · GitHub

# 欧州統合AI基盤プラットフォーム「AI4OS」

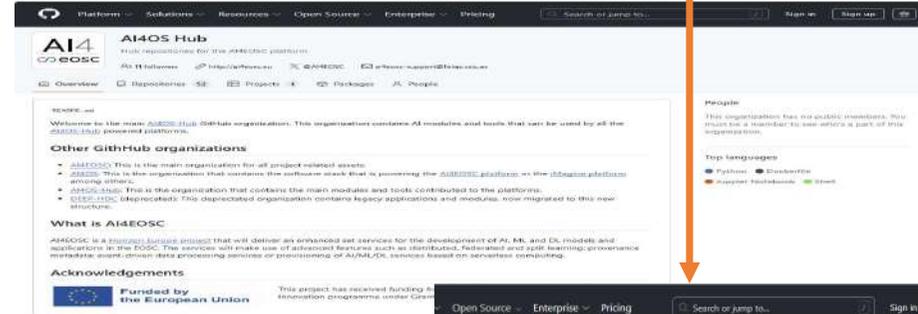
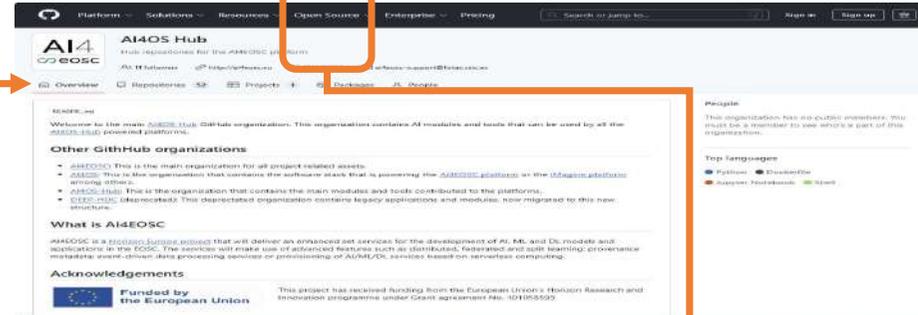
## AI4OS: Build distributed AI platforms.

AI4OS allows to create customizable AI platforms over distributed clouds.

Get started



They trust us



## AI基盤の利用ガイド

Welcome to AI4OS/AI4EOOSC documentation!

The **AI4EOOSC** platform and the **AI4OS** stack empowers scientist by lowering the barrier to adopt the latest AI models and tools. The platform provides a user-friendly workbench to easily train, deploy, share and monitor AI models. The stack provides a set of tools and services to easily create and deploy AI models.

It covers the full ML cycle: from model creation, to training, deployment and monitoring in production. Following the FAIR principles for science, both our software (platform and models) are fully open-source and easily portable to any other platform (no vendor lock-in).

Current supported platforms:

- AI4EOOSC: AI for the European Open Science Cloud
- iMagine: Imaging data and services for aquatic science
- AI4LIFE: AI models and methods for the life sciences

### Getting started

Start here to get an overview of the platform and how to use it.

### How-Toe

How to accomplish common tasks, such as trying a model, developing a new one, training, deploying, etc.

Ref.

- AI4OS : 欧州の統合AI基盤プラットフォーム <https://ai4os.eu/>
- iMagine : 水域科学のための画像データとサービス。テーマはHealthy Oceans, Seas, Coastal and Inland Waters <https://ai4os.eu/powered-by/>

# 研究データをめぐる今日の日米欧

- **科学の再興**：AIにより研究プロセス（仮説・実験・解析）を自律化・高速化し、研究力を強化
- **AI for Scienceの推進\***：
  - ・ データ創出・活用基盤の整備
  - ・ 次世代情報基盤の構築
  - ・ AI駆動型研究の実装、チャレンジ型支援

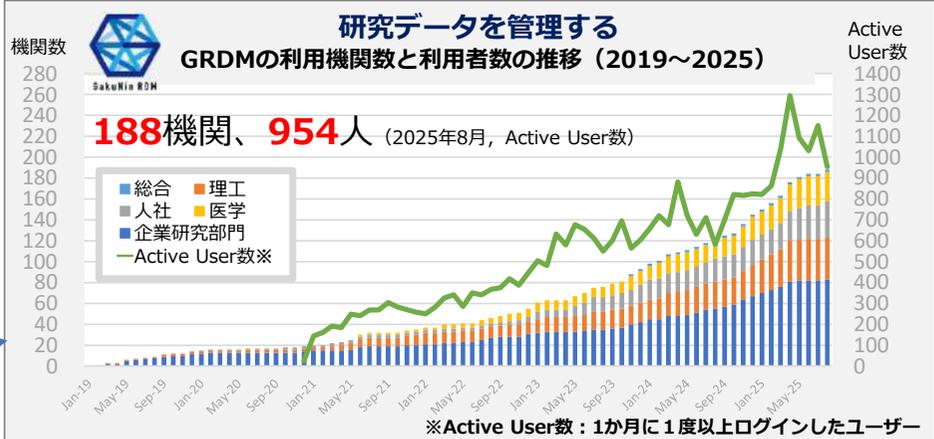
研究基盤	米国 (US)	欧州 (EU/UK)	日本 (Japan)
基本姿勢	<p><b>「全方位・覇権」</b> 安全保障・エネルギー・創薬など全領域でAI主導</p>	<p><b>「統合・連携」</b> AI大陸として統合、運用（RAISE構想）</p> <ul style="list-style-type: none"> <li>・ 計算資源、データ</li> <li>・ 人材・資金</li> </ul> <p>スタートアップ支援としてAI専用のスーパーコンピュータ拠点（AI Factories）</p>	<p><b>「選択と集中・再興」</b> 重点領域に特化し、研究生産性を向上</p> <ul style="list-style-type: none"> <li>・ 「科学の再興」を掲げ*、日本が強みを持つ「良質な実験データ」と「計算基盤（富岳NEXT等）」をつなぎライフ・マテリアル等の分野で世界をリード</li> </ul>
機能的な優位性	<ul style="list-style-type: none"> <li>・ <b>NAIRR</b>：AI計算資源とデータを全米の研究利用</li> <li>・ <b>FASST (DOE)</b>：エネルギー省等の国立研究所が保有する膨大なデータを「AI-Ready」化、プラットフォーム提供</li> </ul>	<ul style="list-style-type: none"> <li>・ HPCやAI Factories：研究者やスタートアップがRAISE経由でAIギガファクトリー等の計算資源利用</li> <li>・ 欧州戦略やAI規則と整合する「AI科学基盤」（対商業主導、他国家主導）</li> </ul>	<ul style="list-style-type: none"> <li>・ <b>学術情報ネットワーク</b>：SINETによる全国直結</li> <li>・ <b>世界最高水準の計算基盤</b>：富岳NEXT / HPCI</li> <li>・ <b>研究データ基盤</b>：NII RDCによるデータ標準化、AIメタデータ生成等で「AI-Ready」化</li> <li>・ <b>コアファシリティ化と自動化ラボ</b></li> </ul>

\*AI for Science の推進に向けた基本的な考え方について、2025年10月、[https://www.mext.go.jp/content/20251006-mxt\\_jyohoka01-000045188\\_04.pdf](https://www.mext.go.jp/content/20251006-mxt_jyohoka01-000045188_04.pdf)

# NIIが提供する学術研究プラットフォーム

最先端のネットワーク基盤「SINET」と研究データ基盤「NII RDC」を整備。  
我が国の基幹的ネットワークとして大学等の学術研究や教育活動全般とデータ駆動型研究に資する。

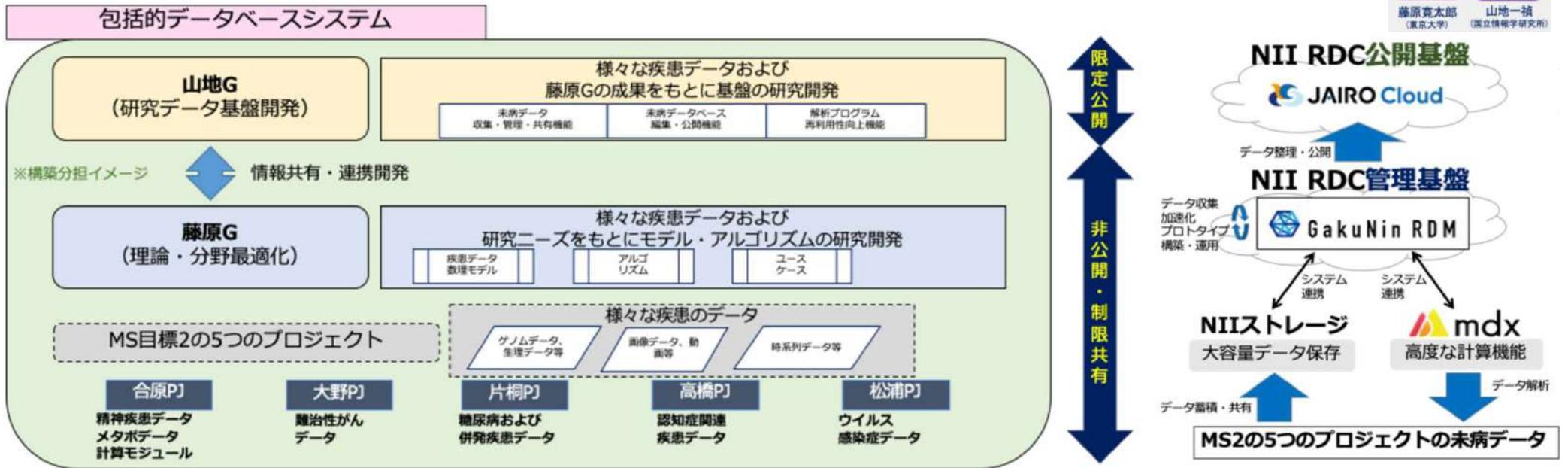
## 学術研究プラットフォーム (2022~2027年度)



1,000機関以上の340万人以上、大規模研究機関間では100Gbpsを超えるネットワーク利用

# NII RDC活用事例 1:

ムーンショット目標2で「医療と数理の融合研究を支援する包括的データベース」を構築\*



- データ管理GakuNin RDMで高度なメタデータを生成、管理。
- データ公開JAIRO Cloudでプロジェクトのデータカタログとして共同利用。

\*JST ムーンショット型研究開発事業 目標2【合原プロジェクト】『複雑臓器制御系の数理的包括理解と超早期精密医療への挑戦』および AI等の活用を推進する研究データエコシステム構築事業【融合・活用開拓 (東京大学) チーム】

# NII RDC活用事例 1:

## プロジェクトで共有する未病データ（メタデータ）カタログの利用イメージ



PJ研究者  
(数理)

メタデータを検索、  
目的のデータがあるGRDMプロジェクトへアクセス



目標2内で共有されて  
いるデータであれば即時アクセス



GakuNin RDM



PJ研究者  
(実験)

未共有データの場合、  
研究者確認の上、共有プロジェクトに招待

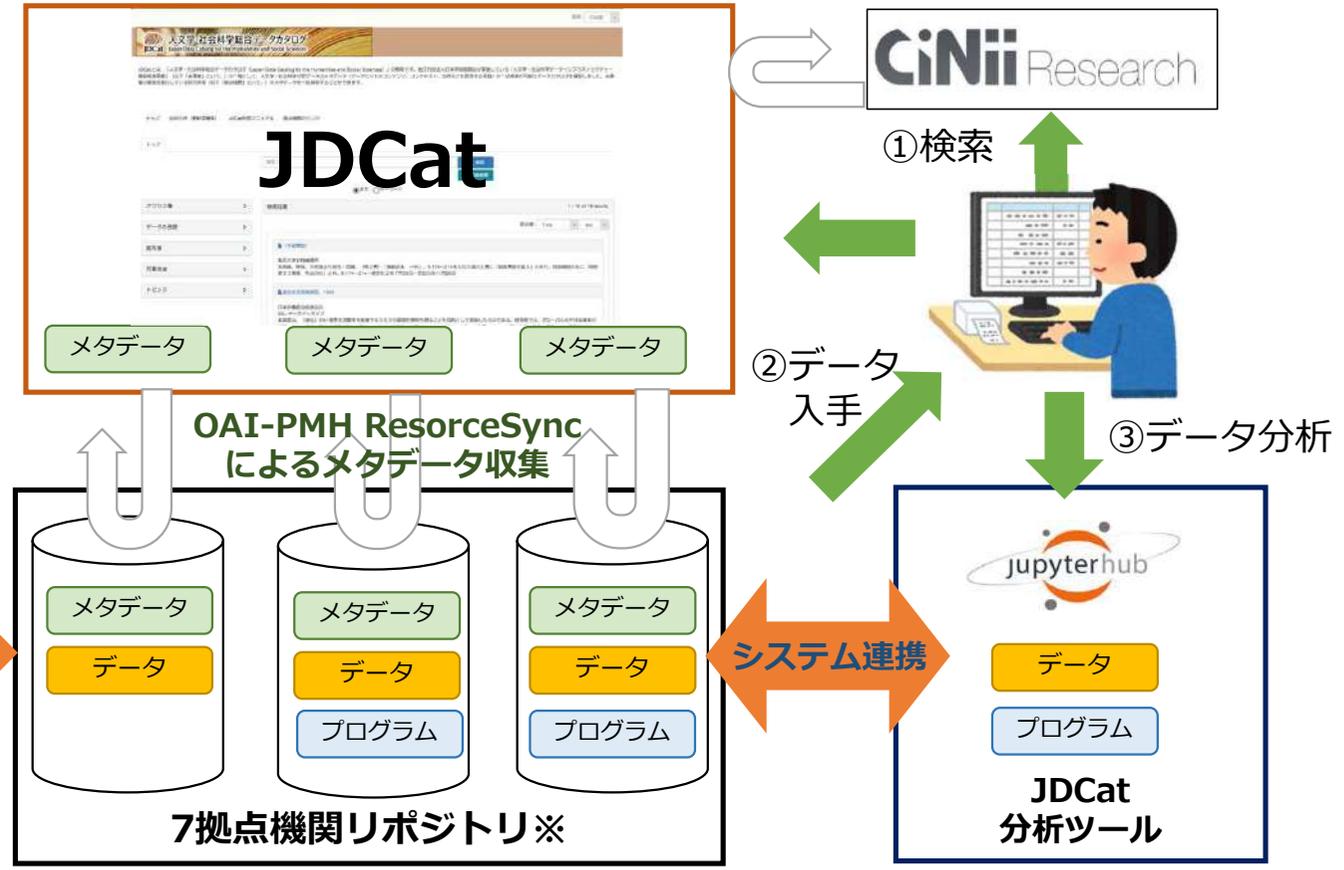
# NII RDC活用事例 2:

## 人文学社会科学の総合データカタログ「JDCat」※2026年1月23日時点



研究データを提供

- ※7拠点機関
- ・ 史料編纂所
  - ・ 東京大学社会科学研究所
  - ・ 神戸大学附属図書館
  - ・ 奈良文化財研究所
  - ・ 慶應義塾大学パネルデータ設計・解析センター
  - ・ 一橋大学 経済研究所
  - ・ 大阪商業大学 JGSS研究センター



JSPS「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」および後継事業の支援を受けて構築。7拠点機関整備の73,988件※の研究データのメタデータをカタログ化

# NIIにおける新たな取り組み-2026

## AI for Scienceを支える情報基盤の高度化

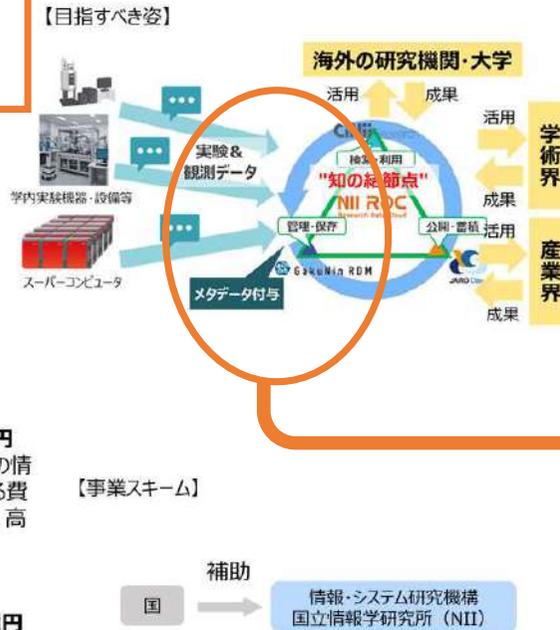
令和7年度補正予算額（案） 5億円 文部科学省

### 現状・課題

- 世界的なAI for Scienceの流れ等により、研究データの重要性がこれまで以上に増しており、我が国の資産として有効に活用していくためには、AI readyな形で研究データを整えて保存・管理することが求められている。しかし、現状はメタデータ付与に多くの負担がかかっている状況であり、AI for Scienceの促進のためには、AI時代に即した研究データ基盤の構築が不可欠。
- また、「学術論文等の即時オープンアクセスの実現に向けた基本方針」（令和6年2月16日統合イノベーション戦略推進会議決定）では、2025年度から新たに公募を行う科研費等の特定の競争的研究費に対して論文及びその根拠データの即時オープンアクセス（OA）化が義務化されており、今後、研究データ基盤に膨大な研究データが集約されてくる見込み。
- 研究データは我が国の貴重な財産であり、それらを死蔵させないためにも、研究者の負担を軽減できる研究データ基盤の高度化が急務。

### 事業内容

- **先行事例やニーズの調査研究 0.5億円**  
AI時代に即した研究データ基盤の構築のために必要な海外における研究データ基盤等の情報基盤の状況やAIへの対応状況等の先行事例及び情報基盤の構築・整備・管理に係る費用・工数や研究現場のAI活用に関する情報基盤へのニーズ等に係る調査研究を実施し、高度化する情報基盤の概念設計を行う。
- **次世代研究データ基盤の構築 4.9億円**  
全国の研究者がAIやデータを最大限に活用できるよう、研究者の負担となっているメタデータ付与について、AIによるメタデータ付与支援機能を先行して開発を進めることで、いち早く研究データ基盤の高度化を実施し、AI for Scienceの促進及び我が国の研究力強化・産業競争力強化を目指す。



実験や観測データ、他の研究情報に対して「AIによるメタデータ付与」により：

- データ整理の省力化
- 共同研究の全体を把握
- データ来歴情報の管理など

信頼性と再現性が担保された  
研究データの生成と蓄積が  
AI-Readyな研究基盤の新たな課題か。

（担当：研究振興局参事官（情報担当） 付） 65