

研究DXの推進について

令和5年 5月25日 基礎研究振興部会

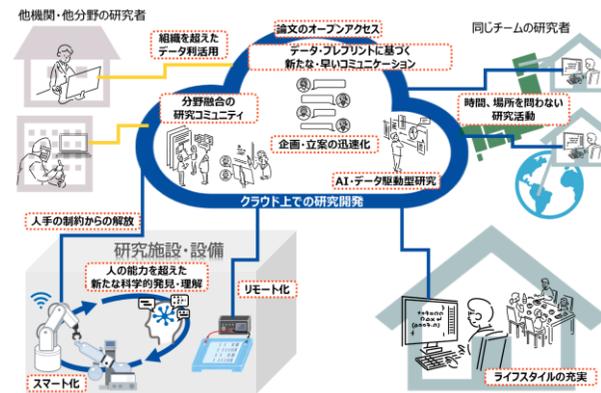
研究振興局 参事官（情報担当）付

研究のデジタルトランスフォーメーション（研究DX）の推進 — 研究DXプラットフォーム開発の加速・高度化 —



令和5年度予算額 441億円
(前年度予算額 414億円)

令和4年度第2次補正予算額 185億円



「AI」×「データ」×「リモート化・スマート化」による研究DXのイメージ図

背景・課題

- 研究のデジタルトランスフォーメーション（研究DX）は、AI技術やビッグデータ解析、研究機器等のリモート化・スマート化等により、**人の能力を超えた新たな発見・理解の拡大、時間や場所、人手の制約からの解放、分野融合による総合知の創出**、といった研究活動に大きな転換をもたらし、**研究の生産性や成果のインパクトを飛躍的に高める成果**が生まれ始めている。
- これらの動きを**全国規模に発展**させ、世界に先駆けてAI・データ駆動型研究開発による成果創出を推進することが必要。

方向性

世界に誇る我が国の研究リソースを結集し、
研究DXプラットフォームの開発を加速・高度化する。

取組内容

価値創造を目指したユースケースの形成・普及

138億円 (135億円)

- 気候変動・レジリエンス、マテリアル、ライフサイエンス、人文社会の分野におけるAI・データ駆動型研究開発によるユースケース形成を推進するため、分野の中核機関が全国の大学等と連携し、研究データの創出から管理、利活用までを一気通貫で行うプラットフォーム構築を進める。また、その成果を発信・普及する。

全国的なデータ共有・利活用を促進する基盤的機能の強化

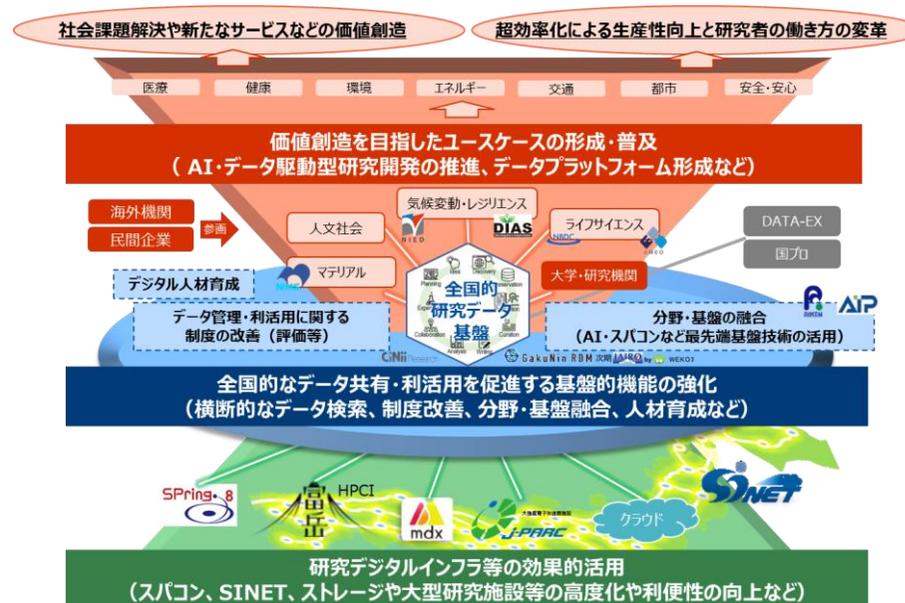
122億円 (98億円)

- 全国の大学・研究機関におけるAI・データ駆動型研究開発を推進するため、分野・機関を越えた研究データの共有・利活用を促進する全国的な研究データ基盤の整備を進める。
- 信頼性の高い最先端AI基盤技術の開発を進めるとともに、量子や数理の活用による先駆的な分野・基盤融合の研究DXプラットフォームを形成し、次世代を見据えたAI・データ駆動型研究開発を先導する。

研究デジタルインフラ等の効果的活用

181億円 (181億円)

- SPRING-8をはじめとしたHPCIや学術情報ネットワーク「SINET」の高度化・活用や、将来必要な計算資源の確保に向けたポスト「富岳」等の性能・機能等を調査検討を進める。
- SPRING-8とともに、産学に幅広い利用ニーズがあるJ-PARCや次世代放射光施設等の他の大型研究施設についても、ユーザーニーズを踏まえつつ、ハード面・ソフト面の整備等を進める。



研究DXプラットフォーム イメージ図

※ () 内は前年度予算額

理論・実験・計算・データ（4つの科学）

演繹

理論・原理から予測

帰納

データを解釈・探索

人間・頭脳
デジタル空間

理論科学

- ・理論構築
- ・頭脳戦

$$\frac{1}{2} \hat{P}(ij|ab) \sum_{k,l} \sum_{c,d} \langle kl || cd \rangle t_{ik}^{ac} t_{jl}^{bd}$$

実験科学

- ・経験から解釈
- ・人海戦術



コンピュータ
サイバー空間

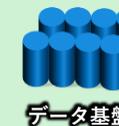
計算科学 (シミュレーション)

- ・モデル化
- ・スパコン勝負



データ科学 (インフォマティクス)

- ・AI・データ解析
- ・質と量が重要



研究DXプラットフォームの形成と展開のイメージ（AI×スパコン×研究ビッグデータ）

データの計測・収集

多様かつ優れた解析基盤



大量の高品質な研究データ

膨大な研究データの蓄積・統合

- 大規模データアセット群を高速転送、蓄積、体系化（メタデータ付与、標準化等）し、セキュアな環境下に保護。
- データ間のつながりを含む多様な分野の膨大データアセットによる知識空間を構築



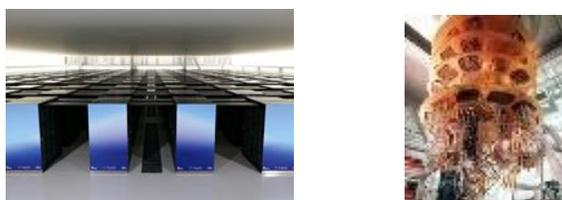
【ライフ分野のデータアセットの例】

- 生体分子の構造データ、細胞の動態データ
- ヒトや動物、植物と共生する微生物
- 環境中（土壌や海など）の微生物

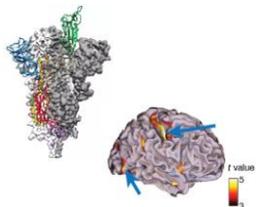
シミュレーションによりリアル空間のデータを補完

世界最高性能の計算資源

スパコン・量子コンピュータ



- シミュレーションによるリアル空間では取得が難しい仮想的な研究データの創出

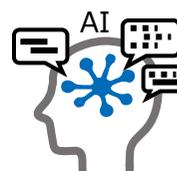


解析環境の提供

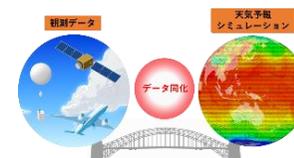
研究データ

データ解析

人工知能をはじめとする数理・情報科学の結集



弱教師付機械学習
因果探索



データ同化

$$m_i \frac{d^2 x_i}{dt^2} = F_i$$

データ⇒数理モデル

- データを活用し、新たな数理モデルや予測手法の開発、高付加価値化、新機能付与等を実現

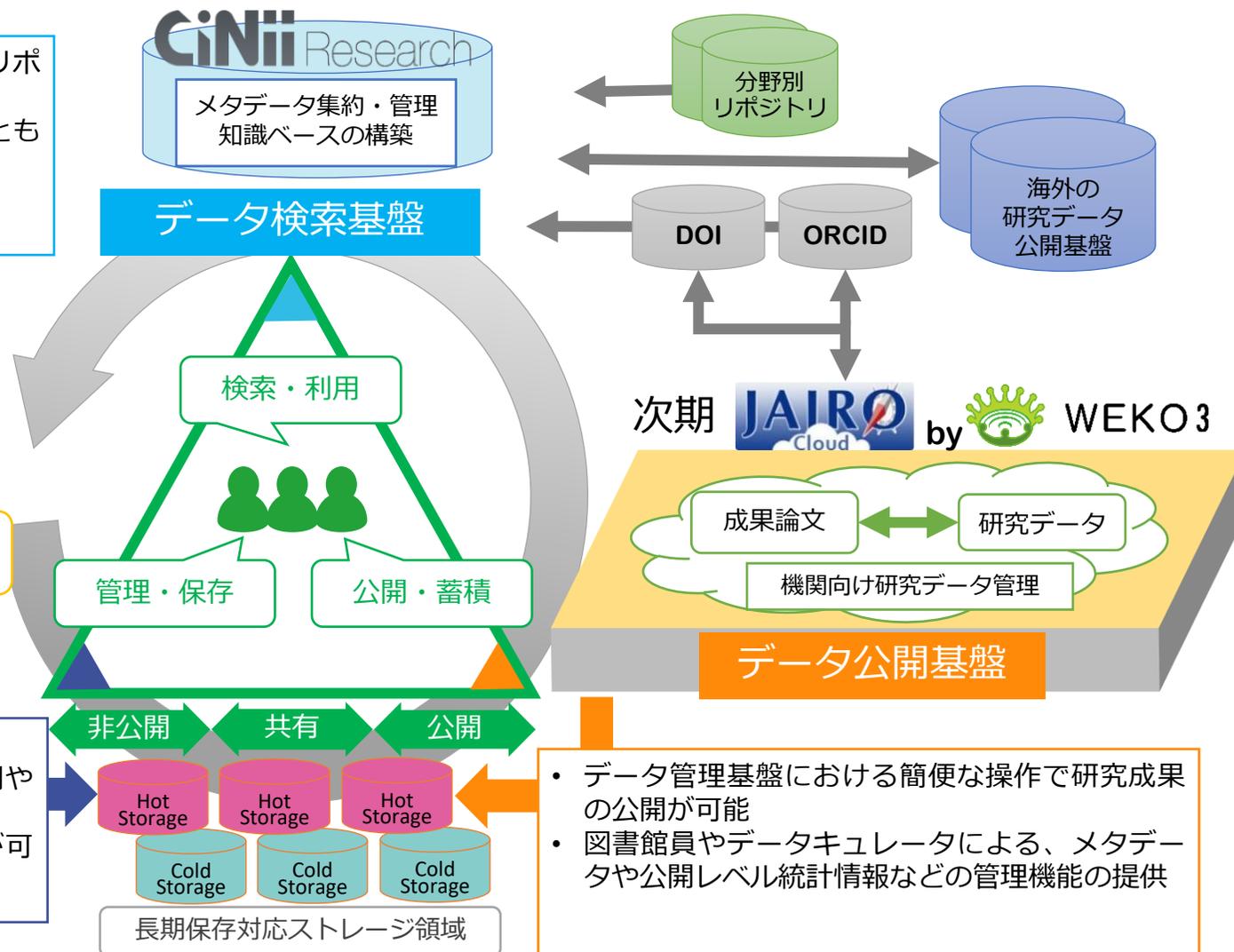
シミュレーションの高度化

研究データ基盤の構築 (NII-RDC)

- 機関リポジトリ+分野別リポジトリやデータリポジトリとも連携
- 研究者や所属機関、研究プロジェクトの情報とも関連付けた知識ベースを形成
- 研究者による発見のプロセスをサポート



- データ収集装置や解析用計算機とも連携
- 研究遂行中の研究データなどを共同研究者間やラボ内で共有・管理
- 組織が提供するストレージに接続した利用が可能



- データ管理基盤における簡便な操作で研究成果の公開が可能
- 図書館員やデータキュレータによる、メタデータや公開レベル統計情報などの管理機能の提供

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

研究データ基盤高度化チーム
NII Research Data Cloudを
7つの側面から機能拡張

研究データ基盤の機能実装

活用

コード付帯機能

データ・プログラム・解析環境のパッケージ化と流通機能を提供し、研究成果の再現性を飛躍的に向上

信頼

データプロビナンス機能

データの来歴情報の管理から利用状況を把握でき、データ公開へのインセンティブモデルを提供

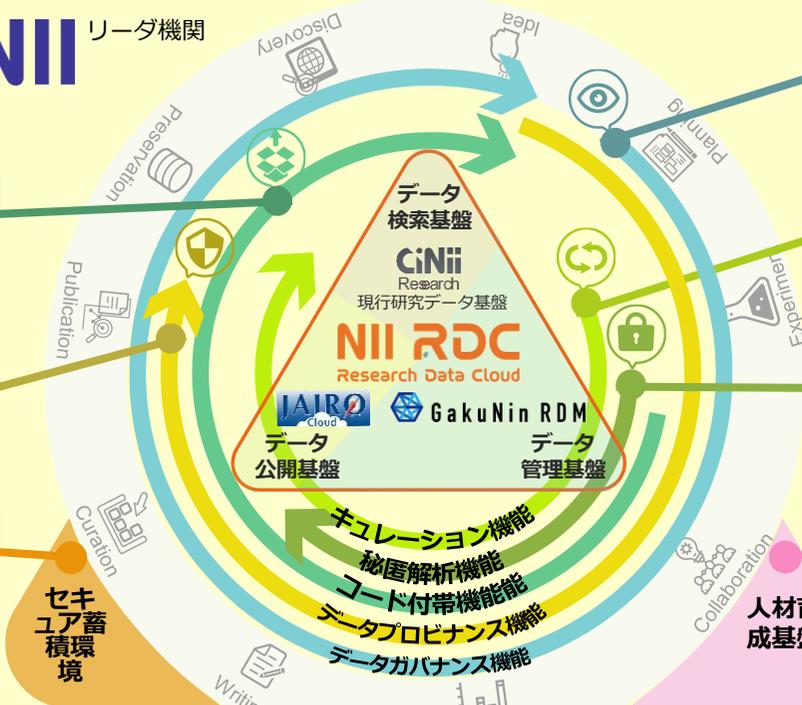
蓄積

セキュア蓄積環境

安全で強固なデータの保存・保護機能を有する超鉄壁ストレージを提供し、機微な情報も安心して保全

NII

リーダ機関



データガバナンス機能

管理

計画に基づきデータ管理等を機械的に支援し、DMPをプロジェクト管理に不可欠な仕組みへと変革

キュレーション機能

流通

専門的なキュレーションを実践できるエコシステムを構築し、データ再利用の促進に寄与

秘匿解析機能

保護

秘密計算技術で機微な情報も安心して解析できる環境の提供で、新しいデータ駆動型研究の世界を開拓

人材育成基盤

育成

RDMに必要なスキルを学ぶ環境を提供し、全ての研究者を新しい科学の実践者へと育成

プラットフォーム連携チーム



理化学研究所

リーダ機関

- ・ 機関内サービス等とNII RDCの連携機能の整理と設計
- ・ 計測機器等からの大量データを効果的に管理するための要件整理と機能開発
- ・ 管理対象となるメタデータの設計と実証
- ・ 関連する高度化機能との仕様調整と共同開発

融合・活用開拓チーム



東京大学
THE UNIVERSITY OF TOKYO

リーダ機関

- ・ 異なる分野間でのデータ活用やデータ連携に発展する取り組みを精査
- ・ 異なる分野間でのデータ活用やデータ連携に関する具体的なユースケースを創出
- ・ ユースケースをまとめたツールキットの作成とそれを用いた広報活動

ルール・ガイドライン整備チーム



名古屋大学
NAGOYA UNIVERSITY

リーダ機関

- ・ 研究データの活用に適した機械可読データの統一的な記述ルール設計
- ・ 研究データの公開に必要な要項や作業フローの整備
- ・ 研究データを適切に取扱うための指針のまとめ
- ・ 学内整備のための事例形成

人材育成チーム



大阪大学
OSAKA UNIVERSITY

リーダ機関

- ・ 人材育成を主とした研究データ管理体制の構築を推進する学内組織構築の事例形成
- ・ 研究データ管理人材に求められる標準スキルに関する検討
- ・ 研究データ管理人材育成のためのカリキュラムの作成、オンライン学習コースの整備

基盤の活用に係る環境整備

中核機関群の代表からなる運営委員会が全体を統括し研究データエコシステムの全国展開に向けて共同実施機関を随時拡大

海外では、既に基盤モデル活用のためのプラットフォームや、
それを活用したサービス・アプリケーション開発が展開されている

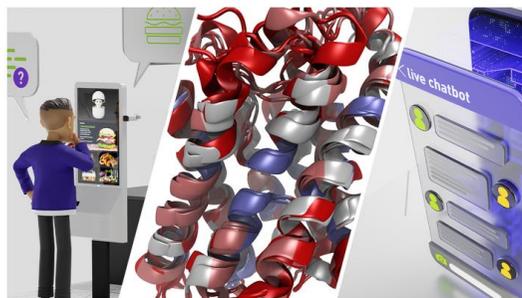
我が国においても、様々な分野での基盤モデルの活用（生成系AIのサービス等）を進めることが必要と考えられる

(例) 基盤モデルを活用したサービスのうち、研究に活用した海外の事例

NVIDIA Launches Large Language Model Cloud Services to Advance AI and Digital Biology

NVIDIA NeMo LLM Service Helps Developers Customize Massive Language Models; NVIDIA BioNeMo Service Helps Researchers Generate and Predict Molecules, Proteins, DNA

Tuesday, September 20, 2022



GTC—NVIDIA today announced two new large language model cloud AI services — the **NVIDIA NeMo Large Language Model Service** and the **NVIDIA BioNeMo LLM Service** — that enable developers to easily adapt LLMs and deploy customized AI applications for content generation, text summarization, chatbots, code development, as well as protein structure and biomolecular property predictions, and more.

The NeMo LLM Service allows developers to rapidly tailor a number of pretrained foundation models using a training method called **prompt learning** on NVIDIA-managed infrastructure. The NVIDIA BioNeMo Service is a cloud application programming interface (API) that expands LLM use cases beyond language and into scientific applications to accelerate drug discovery for pharma and biotech companies.

"Large language models hold the potential to transform every industry," said Jensen Huang, founder and CEO of NVIDIA. "The ability to tune foundation models puts the power of LLMs within reach of millions of developers who can now create language services and power scientific discoveries without needing to build a massive model from scratch."

AI向け半導体を開発しているNVIDIA社において、
MIT・ハーバード大学ブロード研究所と連携し、
2022年9月に**生体分子データによるAI基盤モデルを
数万人規模の研究者に提供することを発表。**

既に、特定の学術分野におけるAI基盤モデル
の利用を開始している形。