

AI for Science を支える研究データの管理・利活用と流通の在り方（審議まとめ）概要（案）

資料 3

背景

- AI技術の進展とデータ駆動型科学の深化による研究パラダイムの転換と研究データの重要性が飛躍的に向上
- 分野・機関別にデータ管理が進む中、AI for Science の進展により、分野横断的なデータ利活用に向けて、データ流通・基盤の高度化が鍵
- 「AI for Science の推進に向けた基本的な戦略方針(R8.3)」等においても、研究設備等、情報基盤、人材・運用体制の相互連携が重要と指摘
- AI for Science に即応するべく、実験基盤、AI統合型の研究データ基盤、計算資源、情報流通基盤の一体的な接続・連携等に向けた次世代情報基盤の在り方を整理

現状と課題

	現状	ニーズ・課題
情報流通基盤	・全国1000以上の機関への超高速・大容量の安定的通信提供	✓ APN等の最先端技術を活用した大容量・低消費電力等対応 ✓ 研究プロセスを同時並行的につなぐ基盤への進化
研究データ基盤	・我が国の中核的な研究データ基盤として位置づけ ・ルール等のデータエコシステム構築に向けた環境整備	✓ 国際動向を踏まえた我が国の研究ポータル構築 ✓ AIエージェントなどAI活用を安全に実現できるデータ空間 ✓ 分野を横断したデータ基盤の整備 ✓ オープン・アンド・クローズ戦略を技術的に可能とする基盤
セキュリティ	・国立大学に対するSINET上の挙動通信検知、人材育成(NII-SOCS)	✓ 支援対象拡大に伴うコスト・運用リソース確保 ✓ 認証基盤の強化、計算資源等との接続

国内外の動向

【欧米諸国】

- 研究拠点、大型研究施設、計算資源等を結ぶ高信頼性・大容量の研究ネットワークを整備
- 国内外の研究機関間での連携や、国際共同研究における基盤として活用
- EOSC(欧州)による分野、国境を越えた研究データ共有・利活用のための共有基盤を整備。各国ストレージや計算資源アクセスのための結節点として機能
- 米国ではDOEやNSFにおいて、国主導の研究基盤整備が進展 (AmSC、CloudBANK)

【国内】

- 共用実験基盤、研究データ基盤・分野データベース、富岳を含むHPCI等の計算基盤、SINETの各基盤の整備が我が国の強みである一方、各基盤間におけるデータ連携・集約、統合的な活用は課題

次世代情報基盤の方向性

- 付加価値の最大化を実現するネットワークの整備
 - ✓ 大容量化・低遅延化・低消費電力・高信頼化、AI時代における通信量増加にも耐え得る構成に高度化及び国際ネットワークとの接続強化により国際共同研究を支える基盤としての機能拡充
 - ✓ APNやNTN等の技術動向も踏まえ、ユーザーニーズに応える社会インフラとしての付加価値向上
 - ✓ 公私立大学へのNII-SOCS展開等、学術分野全体におけるサイバーセキュリティ対策を底上げ
- 研究プロセスを支える新たな研究データ基盤の構築
 - ✓ AI for Scienceに対応した分野横断的なデータ活用を可能にする研究者の研究プロセスを総合的に支えるAI統合型の研究データ基盤による研究データ空間の確立及び分野データベースの高度化
 - ✓ AI学習や新たな知の創出に資するダークデータやネガティブデータなどのデータセットを構築
- 計算資源等との一体的な接続・連携、認証強化
 - ✓ 実験基盤、研究データ基盤、計算資源を一体的に活用できる環境構築に向け、共通認証基盤の構築による研究者・研究機関、研究設備等、データ、計算資源を信頼性高く結節するパイプラインを整備

➔ 研究データの流通そのものが新たな付加価値を生み出す次世代情報基盤として刷新し、研究のスピードと質を向上

➔ 分野にとらわれずすべての研究者がAI for Science に対応した環境下での研究を可能に

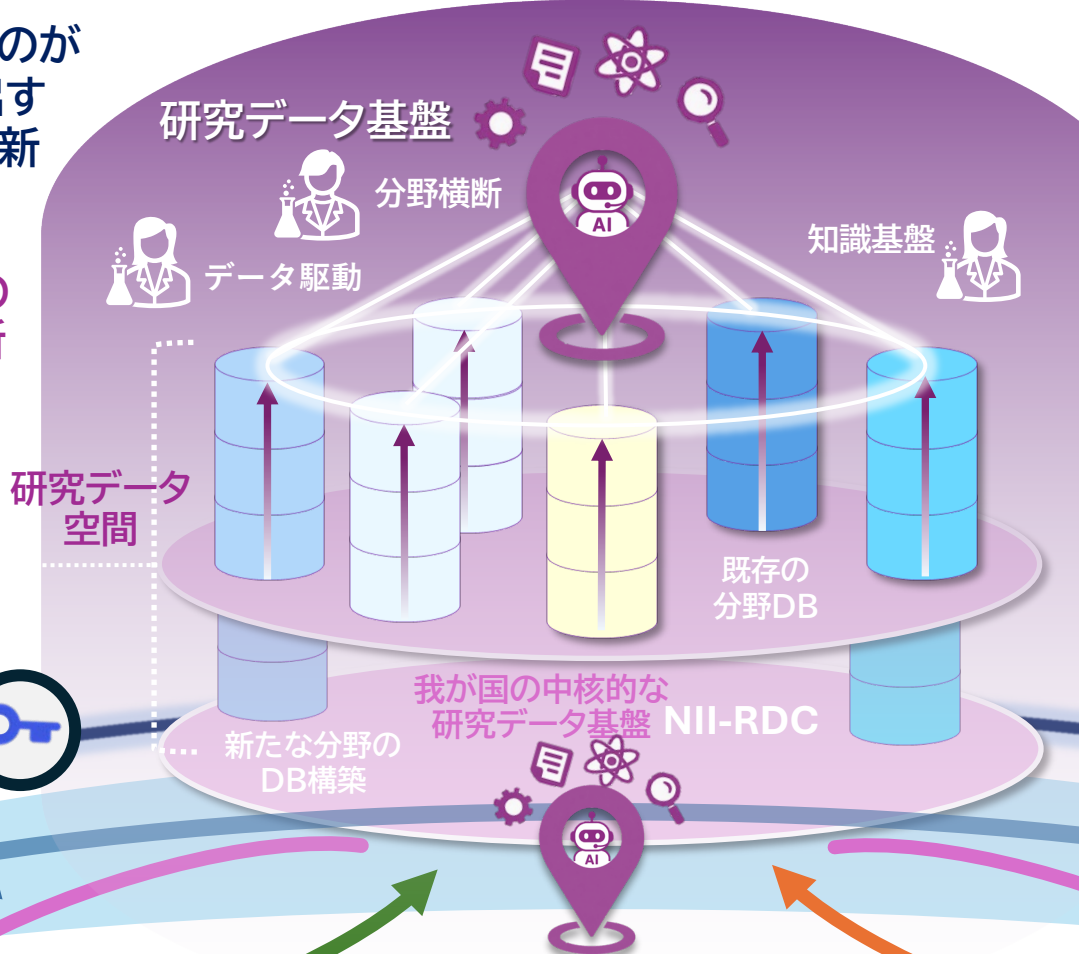
➔ AI for Science に対応した研究データの管理・利活用を実現し、研究者の研究プロセスを総合的に支援

AI for Scienceを支える研究データの管理・利活用と流通のイメージ(案)

研究データの流通そのものが
新たな付加価値を生み出す
次世代情報基盤として刷新

AI for Science時代の『研究データ基盤』へ刷新

- ・共有可能で機械可読可能な研究データ空間へ
- ・分野特性に応じたDBの最適化とAI利活用を前提とした形式化



- ✓すべての研究者がAIfS即応の研究環境下で研究が可能に
- ✓AIfS時代の研究者の研究プロセスを一元化
- ・研究データの管理・利活用を中核に国としてパイプラインを構築
 - ➔データ創出実験基盤、計算資源、研究データ基盤の連携強化(共通認証基盤の整備等)
 - ➔研究目的別のパイプライン構築が可能

GakuNin
パイプライン
(共通認証基盤等)

データ創出実験基盤
大学の共用研究設備、大型共用研究施設、大規模集積装置...

計算資源
HPCI、mdx...

AI for Science時代の『情報流通基盤』としてSINETの高度化(大容量化・高信頼化)