

研究DXの推進について

令和4年6月17日

文部科学省 研究振興局 参事官（情報担当） 付

研究デジタルトランスフォーメーション（研究DX）について

- 新型コロナウイルス感染症を契機として、世界的に社会のデジタル変革が一気に進展。
- 研究分野においても、人工知能（AI）技術やビッグデータ解析の発展、研究機器等の遠隔化・自動化、スパコン・ネットワーク等のデジタルインフラ利用拡大が推進。また、世界的な知の共有を目指した研究成果のオープン化（オープンサイエンス）が進行。**研究の在り方そのものに大きな変革期**が到来。
- **新たな価値創造を目指し、デジタル技術とデータ活用によって研究活動を変革（研究デジタルトランスフォーメーション（研究DX））**することで、社会や経済、科学を大きく動かす可能性。

AI・データ駆動型研究

AI技術やビッグデータ解析により、人間の認知の限界・バイアスを超えて、科学的知見や社会にとっての新たな価値の発見の可能性。

➡ **人の能力を超えた新たな発見・理解が拡大・加速**

オープンサイエンス

世界的なオープンサイエンスの潮流により、研究論文や研究データの公開・共有や市民の研究活動への参画などが促進。

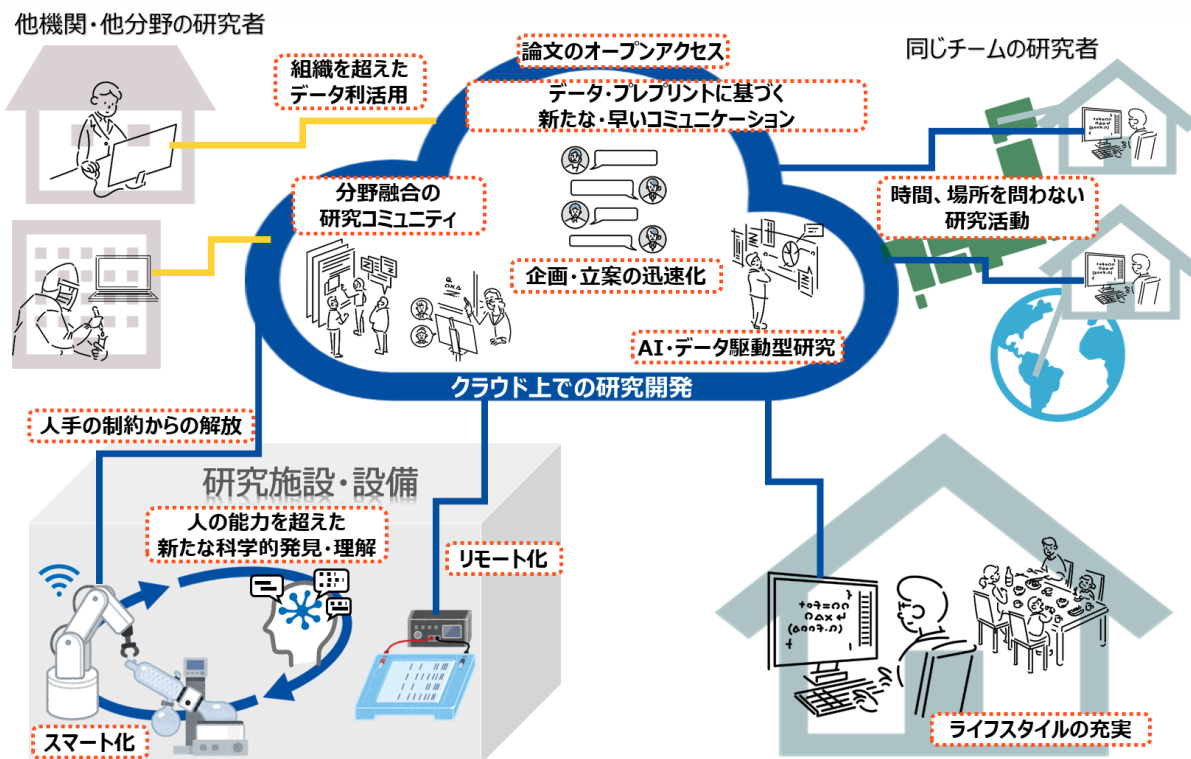
➡ **分野や組織を越えた新たな協働による総合知の創出**

研究環境のデジタル化

研究装置等のリモート化・スマート化や、クラウド、ネットワーク、スパコンなど研究デジタルインフラの整備が進み、サイバー空間とフィジカル空間の融合が加速。

➡ **時間や場所、人手の制約から解放され、研究生産性が爆発的に向上**

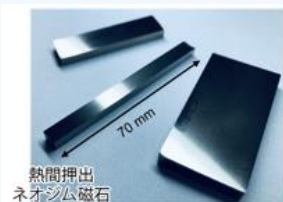
「AI」×「データ」×「リモート化・スマート化」⇒ 価値創造 研究デジタルトランスフォーメーション（研究DX）



研究DXの先行的な成果事例

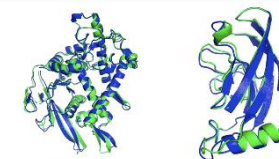
AIが約6600万通りから最適な作製条件を探索 → AI解析前後で材料性能が1.5倍に向上

- 世界の電力消費量の約50%を占めるモーターの効率化・省エネ化のために、永久磁石の高性能化が期待。
- ネオジム磁石の作製プロセス条件は約**6600万通り**であり、網羅的な探索が不可能。AI解析により提案された**最適な作製条件に基づいて実験計画を立てる**ことで、**わずか40回程度の実験でネオジム磁石の強さをAI解析前の約1.5倍向上**させることに成功。
(2021.11.15 プレスリリース)



AIによるタンパク質の立体構造解析ツール → 数年かかっていた作業を圧倒的に短縮

- 2014年にGoogleが約670億円で買収したAI開発企業DeepMindが無償公開した**遺伝子配列情報からタンパク質の立体構造を解析するAI「AlphaFold2」**。
- 従来、タンパク質のアミノ酸配列の構造を特定するには**数か月～数年**かかり、多大な時間と費用が掛かっていたが、**AlphaFold2では、限られた情報から構造と機能を推定することが可能に**。



T1037 / 6vrr4
90.7 GDI
(RNA polymerase domain)

T1049 / 6y4f
93.3 GDI
(adhesin tip)

実験結果とシミュレーションの結果がほぼ類似

引用：新しい資本主義実現会議

スパコン「富岳」を活用したゲリラ豪雨予測 → 30秒ごとに更新するリアルタイム性を実現

- ゲリラ豪雨は**5～10分**といった短時間で状況が急激に変化するため**予測が困難**。理研等の研究グループは、マルチパラメータ・フェーズドレイ氣象レーダによる雨雲の詳細な観測データと、「富岳」による大規模計算により、高頻度かつ高精度な予測を実現。
- 2021年夏に、**首都圏において30秒ごとに更新する30分後までのリアルタイム降水予報を世界で初めて実施**。
(2021.7.13 プレスリリース)



実証実験で表示される「3D 雨雲ウォッチ」アプリイメージ

仮説・実験・検証・修正のサイクルを自動化 → これまで着手が困難だった希少疾患等の創薬開発が推進

- ケンブリッジ大学とマンチェスター大学により開発されたロボットサイエンティスト「イブ」は、一日当たり**1万化合物のスクリーニング**を行うとともに、**実験結果を統計学・機械学習により解析し、より高い活性をもつ新たな化合物の構造を予測し、実験・検証・修正のサイクルを繰り返す、仮説主導のハイスループット研究を自動化**した。
- 製薬企業が手を出しにくい**希少疾患等に対する創薬開発が推進**されることが期待。



引用：JST リサーチトランスフォーメーション (RX) ポスト/with コロナ時代、これからの研究開発の姿へ向けて

Robot scientist Eve

「AI」、「データ」、「リモート化・スマート化」による圧倒的な生産性の向上とハイインパクトな成果が生まれ始めている。しかし、まだ一部の研究者・研究領域のみ。この変革の動きを、日本全体に発展させることが必要。

研究DXの課題と必要なアプローチ

- AI技術の活用や研究データの共有など研究DXの取組を進めるに当たっては、具体的に何をすればよいのかわからない、メリットが不明確である、研究データの性質に応じた公開・非公開のバランス、データ形式の共通化にかかるコストなど様々な課題が存在。
- これらを解決し研究DXを全国的な動きにするためには、「ユースケースの形成、普及」、「データ共有・利活用の促進」、「研究デジタルインフラ等の効果的活用」を一体的に進めることが必要。

【研究DXを進める上での主な課題】

研究DXに対する理解が浸透していない

そもそも研究DXとは何か、何をすればよいのか、といった研究DXに対する認識や理解が浸透していない。

研究DXのメリットが不明確

研究DXはまだ萌芽的なフェーズであり、具体的なケースが乏しく、そのメリットや意義が理解されにくい。

オープンアンドクローズの難しさ

分野やデータの特性によって、公開・非公開の戦略が求められるが、その考え方やノウハウが蓄積されておらず、効果的な研究データの共有や利活用が進みにくい。

研究DXの実施にかかるコスト

メタデータの付与やデータの構造化など、研究DXに必要となる人手や時間といった追加的負担をどのようにカバーするか。

計算資源やストレージの確保

研究DXが進めば進むほど、取り扱うデータの量が増加し、必要となる計算資源やストレージの規模が大きくなってしまふ。

分野と基盤（情報、計測等）の研究者の連携

各分野における研究DXを進めるためには、分野の研究とAI等の情報科学、計測等の実験装置開発の連携が不可欠だが進みにくい。

【課題解決に必要なアプローチ】

1. 価値創造を目指したユースケースの形成、普及

- 具体的な研究DXの成功事例（産業界の参画、参加者のインセンティブ形成、海外連携含む。）を創出し、その成果や過程で得られる仕組みや経験・ノウハウを普及することが必要ではないか。

2. データ共有・利活用を促進する基盤的機能の強化

- 研究データの共有・利活用に取り組むよう研究者や大学、研究機関の活動を促すとともに、それらを支援する機能の整備や共通的な課題への対応が必要ではないか。
- また、分野・基盤の研究コミュニティの融合による革新的な研究手法の開発を促進することが必要ではないか。

3. 研究デジタルインフラ等の効果的活用

- 我が国には先端的な研究インフラがすでに数多く整備されている。これらを効果的に活用する方策を進めることが必要ではないか。

研究DX実現に向けた取組（研究DXプラットフォーム概念図）

社会課題解決や新たなサービスなどの価値創造

超効率化による生産性向上と研究者の働き方の変革

医療

健康

環境

エネルギー

交通

都市

安全・安心

1. 価値創造を目指したユースケースの形成・普及 （AI・データ駆動型研究開発の推進、データプラットフォーム形成など）

民間企業

海外機関

参画

人文社会

気候変動・レジリエンス



ライフサイエンス

NBDC



DATA-EX

国プロ

デジタル人材育成



マテリアル

データ管理・利活用に関する
制度の改善（評価等）



大学・研究機関

分野・基盤の融合
（AI×データ×スパコンなど）



CiNii Research

GakuNin RDM 次期 LAIRO by WEKO3

2. データ共有・利活用を促進する基盤的機能の強化 （横断的なデータ検索、制度改善、分野融合、人材育成など）



クラウド



3. 研究デジタルインフラ等の効果的活用 （スパコン、SINET、ストレージや大型研究施設等の高度化や利便性の向上など）

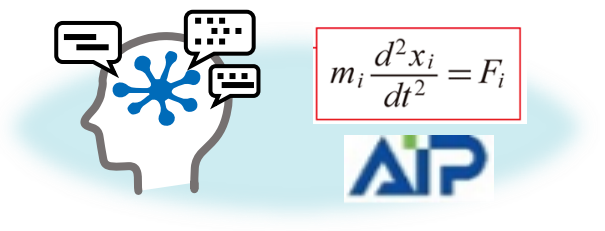
AI×スパコン×研究ビッグデータによる研究DXの推進

高品質な研究データの共有・利活用を促進



- 大型研究施設等から創出される大規模の研究データや各機関が所有する研究データが繋がり、多様なデータが利活用できる環境を整備

我が国の人工知能 (AI) 研究の英知を結集

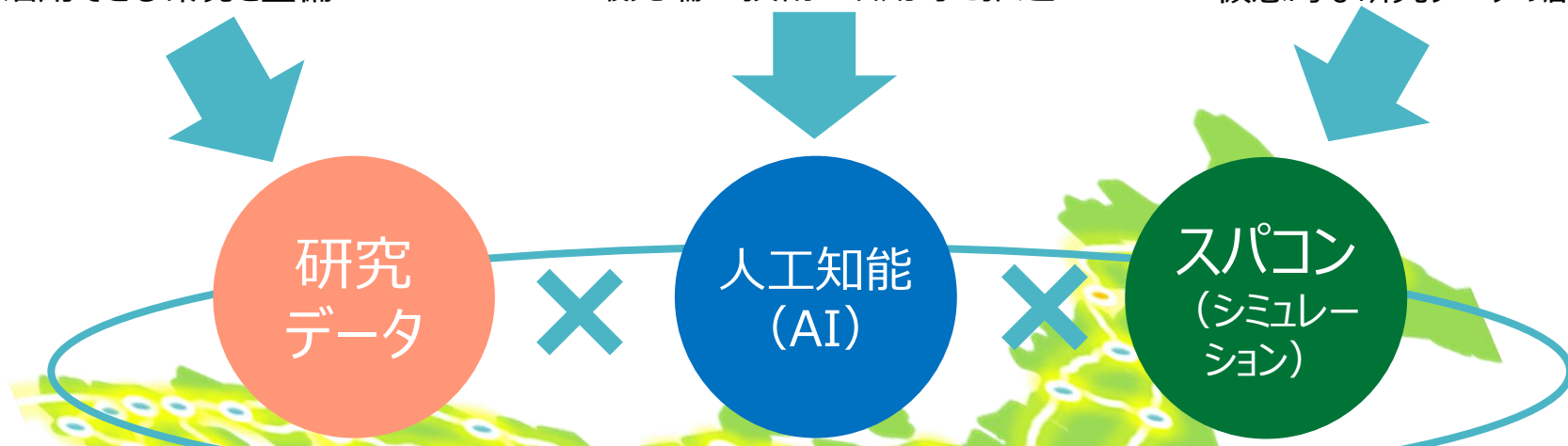


- 機械学習の原理の解明等による信頼できるAIの開発
- 日本が強い分野、社会課題への最先端AI技術の活用等を推進

世界最高水準の計算基盤の活用



- 世界トップレベルの「富岳」などスパコン群の活用し、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



多次元の研究データを、人工知能をはじめとする数理・情報科学を結集し、世界最高性能の計算基盤で解析

⇒ 我が国の研究力を結集したAI・データ駆動型研究開発を推進

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



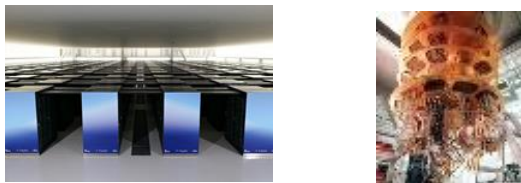
【ライフ分野のデータアセットの例】

- ◆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

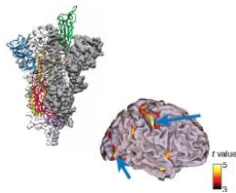
シミュレーションによりリアル空間のデータを補完

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



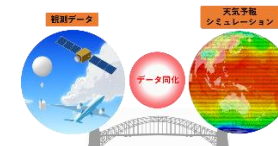
解析環境の提供

シミュレーションの高度化

多次元データ

データ統合

人工知能をはじめとする数理・情報科学の結集



データ同化

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

データ⇒数理モデル

- ・多次元データを活用し、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

解決すべき課題

- 各分野データプラットフォームや、各機関リポジトリの構築等が進められているが、これらをつなぎ、分野・機関を越えてデータを共有・利活用するための全国的な研究データ基盤が未実装であり、国際的にも遅れ。
- 政府全体の方針に基づき、公的資金による研究データの取扱いに当たり、研究者に求められる責務が増大（データマネジメントプランの作成、メタデータ付与等）しており、対応が必要。
- 研究データの取扱いルール等の制度の整備や普及が追いついておらず、データサイエンスに不可欠であるデータマネジメント人材も不足。
- DXによる研究手法の変革が一部にとどまっており、デジタル基盤を徹底的に活用したAI・データ駆動型研究の進展が不十分。

実施内容

事業期間：R4年度～R8年度

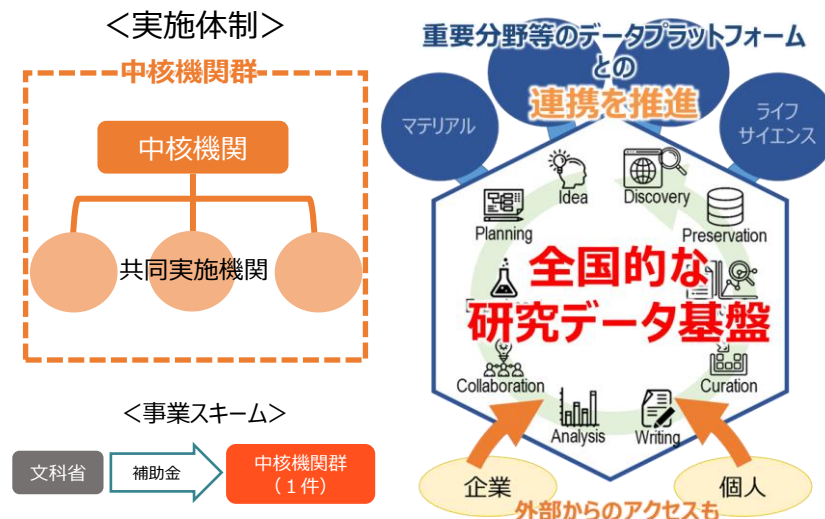
- 我が国の研究力の飛躍的発展を図るため、各分野・機関の研究データをつなぐ**全国的な研究データ基盤の構築・高度化・実装**と、**AI解析等の研究データ基盤の活用**に資する環境の整備を行う、**研究DXの中核機関群を支援**する。

● 全国的な研究データ基盤の構築・高度化・実装

- ユーザーニーズを踏まえながら、研究データの管理・蓄積・利活用・流通といった点で適切かつ実用的な機能を確保した全国的な研究データ基盤を整備。
- 構築が進む各機関・各分野のリポジトリやデータプラットフォームとの連携・接続。

● 研究データ基盤の活用に係る環境の整備

- 効率的なAI活用のための、機械可読データの統一化や標準化等を含めたルール・ガイドライン整備、データマネジメント人材育成支援等、ユーザー視点に立って研究データ基盤を最大限に活用するための環境整備。





① 研究データ基盤高度化チーム

【求められる機能例】

- DMPに基づく研究データの取扱い（管理対象データの取捨選択やメタデータの付与等）
- 研究データの出所や修正履歴等を管理し、対外的にその真正性の説明
- 研究データを秘匿したまま安全に解析可能とするための環境構築

② プラットフォーム連携チーム

- ・ 分野・機関リポジトリやデータプラットフォームにおける研究データの連携・接続

③ 融合・活用開拓チーム

- ・ 異分野間でのデータ連携を前提としたAI・データ駆動型研究のシーズ・ユースケースの創出

④ ルール・ガイドライン整備チーム

- ・ 個人情報等の留意事項を含む研究データの取扱いに関するルール・ガイドラインの整備

⑤ 人材育成チーム

- ・ データマネジメント人材の育成に向けた教材開発やコンテンツ整備等