

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

資料 2-1
情報委員会 (第25回)
令和 4 年 5 月 17 日

令和4年度予算額 991百万円 (新規)



文部科学省

背景

新型コロナウイルス感染症の猛威により、我が国のデジタル化への遅れが顕著になったことから、次の成長の原動力として「デジタル」「AI」が最重要視されている。そのような中、AI・データ駆動型研究の重要性が高まるなど、研究手法が大きく変化しており、研究DXにより生産性を飛躍的に向上させるためには、膨大な量の高品質なデータの利活用を推進していくことが鍵である。このため、我が国における研究データの管理・利活用を促進するための中核的な研究データ基盤の構築・高度化・実装を行い、各分野等で構築が進められているデータプラットフォーム等と連携した、オープン・アンド・クローズ戦略に基づく研究データの管理・利活用を促進し、AI等の活用を推進することが求められている。

【経済財政運営と改革の基本方針2021】(令和3年6月18日閣議決定)

研究の生産性を高めるため、研究DXを推進するとともに、研究を支える専門職人材の配置を促進する。

【成長戦略フォローアップ2021】(令和3年6月18日閣議決定)

・研究のDXの実現に向け、AI・データ駆動型研究を推進するため、全国の先端共用設備や大型研究施設も効果的・効率的に活用し、2022年度からマテリアル、ライフサイエンス等多様な分野の研究データを戦略的に収集・共有・活用する取組を強化する。

未解決の課題

- 各分野におけるデータプラットフォームや、各機関におけるリポジトリの構築等が進められている。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の実装が未実施であり、国際的にも遅れをとっている。
- 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らず、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ぎ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。

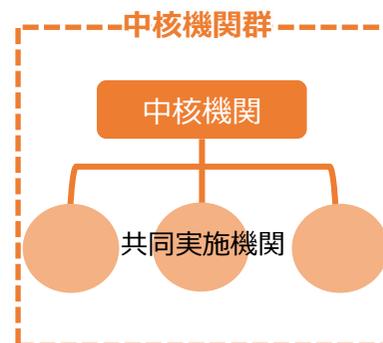
●全国的な研究データ基盤の構築・高度化・実装

- ・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・流通といった面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し、AI・データ駆動型研究を推進。
- ・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。

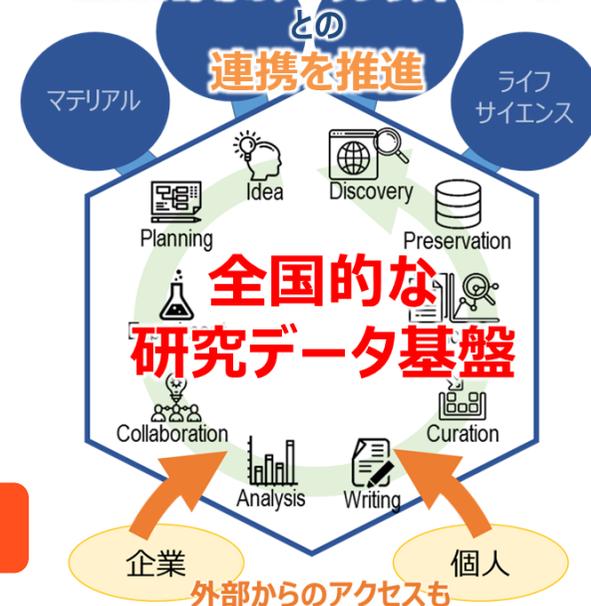
<実施体制>



<事業スキーム>



重要分野等のデータプラットフォームとの連携を推進



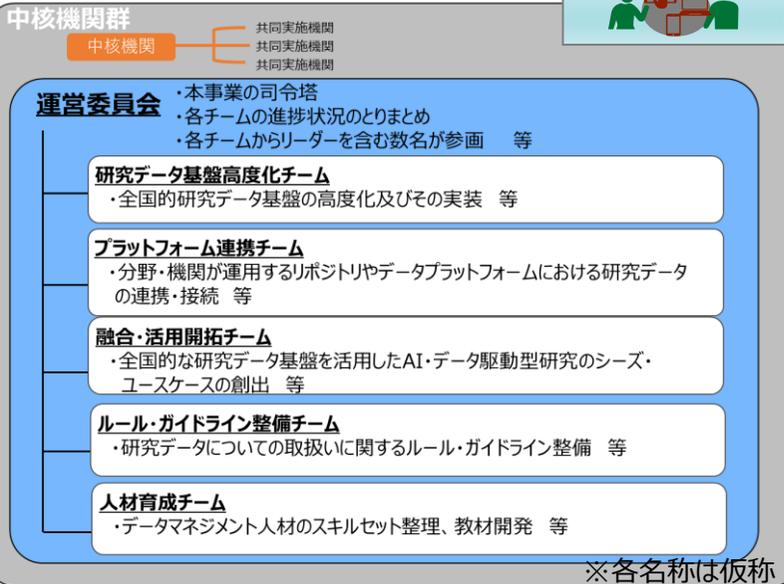
研究データエコシステムの構築に向けた具体的な取組イメージ

■ 全国的な研究データ基盤

全国的な研究データ基盤を中心に、オープン・アンド・クローズ戦略に基づいた研究データの管理・利活用が持続的に行われる仕組み（研究データエコシステム）の構築を目指す。そのために、全国的な研究データ基盤の構築・高度化・実装等や研究データ基盤の活用に係る環境整備を実施



■ 事業実施体制



中核機関群

- ・研究データ基盤の構築・高度化・実装の中心的役割を担う中核機関及び複数の共同実施機関により有機的な連携を行う連合体
- ・中核機関群には、本事業の司令塔となる運営委員会及びそれぞれ推進すべき事業の内容に基づき中核機関群による機関から構成したチームを構成して事業推進

<各チームのミッション>

① 研究データ基盤高度化チーム

- ・全国的な研究データ基盤の高度化及びその実装

【最低限求められる機能】

- データマネジメントプラン（DMP）の内容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）を支援する機能
- 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を説明可能とすることを支援する機能
- 研究者が管理する研究データを共有・公開した際、その研究データの利用状況や公開後の再利用の状況をモニタリングし、研究者にフィードバックする機能
- 研究者が管理する研究データ・プログラム・解析環境等の情報を、再現性の検証や再利用可能な形でパッケージ化し公開することを支援する機能
- 秘匿すべき情報が含まれる研究データについて、その管理をシステム上で適切に行うとともに、これらの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を支援する機能

② プラットフォーム連携チーム

- ・分野・機関が運用する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
- ・全国的な研究データ基盤上で統合的なメタデータ検索の実現
- ・全国的な研究データ基盤との接続を希望する産学官の機関等に提示する要件等の整備

③ 融合・活用開拓チーム

- ・全国的な研究データ基盤を活用し、異なる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
- ・全国的な研究データ基盤のユーザー拡大に向けた周知・広報活動

④ ルール・ガイドライン整備チーム

- ・研究データの効果的・効率的な活用のための機械可読データの統一化や標準化等や個人情報等の研究データを取り扱う上で留意すべきデータを含む研究データについて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

- ・研究データ管理に必要なデータマネジメント人材も必要なスキルセット等の要件の整理
- ・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

■ スケジュール

公募開始：4月21日

締め切り：5月31日

事業開始：7月下旬～8月上旬予定

研究データエコシステムの構築に係る事業実施体制イメージ

