

## (2) 新たな研究システムの構築（オープンサイエンスとデータ駆動型研究等の推進）

### (a) 現状認識

ビッグデータ等の多様なデータの収集や分析が容易となる中で、計算機を活用したシミュレーションやAIを活用した研究のインパクトがより一層大きくなっている。さらに、新型コロナウイルス感染症を契機として、研究交流のリモート化や、研究設備・機器への遠隔からの接続、データ駆動型研究の拡大など、世界的に研究活動のDX（研究DX）の流れが加速している。

これに伴い、論文のオープンアクセス化や研究成果の迅速な公開の場の一つとしてのプレプリントの活用も一層加速しており、研究データの公開・共有を含め、オープンサイエンス等の世界的な知の共有を目指した研究成果のオープン化が進みつつある。その一方で、信頼性のない研究データを利用した論文が撤回される事例や、世界的な出版社やIT企業がビジネスの対象として研究成果や研究プロセス全体で得られたデータを囲い込む動きも見られる。このような状況を踏まえ、各国政府、国際機関、産業界、アカデミア等において、研究活動における自由と多様性を尊重しつつ、国際的な貢献と各主体の利害の双方を考慮に入れた、オープン・アンド・クローズ戦略に基づく研究プロセスのマネジメントを実行することが求められている。

我が国においては、ネットワーク、機関リポジトリ、データプラットフォーム、計算資源等の研究基盤や各種ガイドライン等の制度環境の整備が行われてきた。

特に公的資金により得られた研究データの管理・利活用については、大半の大学において機関リポジトリが整備されてきたが、その一方で研究データの収載が進んでいないことや、データポリシーが未整備であるといった課題がある。一部の先行的なプログラムでは、研究データの幅広い利活用を促進するため、そのメタデータを集約し、検索・閲覧可能とするための取組が行われているが、研究データの管理・利活用に向けた取組は道半ばである。さらに、研究データの帰属や、個人情報の取扱いなど、研究データに関する法的・制度的な諸問題もある。

また、我が国のデジタル研究基盤については、2020年、4つのスーパーコンピュータランキングにおいて、世界1位となった「富岳」や学術情報ネットワーク（SINET）などのインフラの整備が進み、質の高い研究・教育に貢献している。他方、コロナ禍において、大学等の共用施設・設備の多くが古いシステムを活用していたため、外部ネットワークへの接続が困難となり、その遠隔利用について課題が顕在化している。さらに、研究DXは研究活動そのものの変容・発展を伴うものであるが、研究設備・機器の整備と活用についても、依然として囲い込みと自前主義的な文化は残っており、改善が求められる。

#### 【現状データ】（参考指標）

- ・ 国立研究開発法人における研究データポリシーの策定法人数：11法人（2020年9月）
- ・ 競争的研究費制度におけるデータマネジメントプラン（DMP）の導入済み府省・機関数：8省・機関（2020年9月）
- ・ 国内における機関リポジトリの構築数：811個（2019年度）
- ・ 研究データ公開の経験のある研究者割合：51.9%（2018年度）
- ・ プレプリント公開の経験のある研究者割合：20.4%（2020年度）
- ・ HPCI提供可能資源量：年間25ペタflops（2019年度）
- ・ 研究設備・機器の共用化の割合：大学等における研究機器のうち相当程度の市場規模のある10機器（2012～2016年度購入）について、競争的資金で購入したもののうち9割は研究者個人や研究室単位での利用にとどまる。（2017年度）

## (b) あるべき姿とその実現に向けた方向性

社会全体のデジタル化や世界的なオープンサイエンスの潮流を捉えた研究そのもののDXを通じて、より付加価値の高い研究成果を創出し、我が国が存在感を発揮することを目指す。特に新型コロナウイルス感染症の研究においても、論文のオープンアクセス化やプレプリントの活用が更に拡大する中、研究プロセス全般で生まれるデータについて、戦略性を持って適切な共有と利活用を図るとともに、それによりインパクトの高い研究成果を創出していくための研究基盤の実現が求められる。

このため、まず、データの共有・利活用については、研究の現場において、高品質な研究データが取得され、これら研究データの横断的検索を可能にするプラットフォームの下で、自由な研究と多様性を尊重しつつ、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・利活用を進める環境を整備する。特にデータの信頼性が確保される仕組みが不可欠となる。また、これらに基づく、最先端のデータ駆動型研究、AI駆動型研究の実施を促進するとともに、これらの新たな研究手法を支える情報科学技術の研究を進める。

同時に、ネットワーク、データインフラや計算資源について、世界最高水準の研究基盤の形成・維持を図り、産学を問わず広く利活用を進める。また、大型研究施設や大学、国立研究開発法人等の共用施設・設備について、遠隔から活用するリモート研究や、実験の自動化等を実現するスマートラボの普及を推進する。これにより、時間や距離の制約を超えて、研究を遂行できるようになることから、研究者の負担を大きく低減することが期待される。また、これらの研究インフラについて、データ利活用の仕組みの整備を含め、全ての研究者に開かれた研究設備・機器等の活用を実現し、研究者が一層自由に最先端の研究に打ち込める環境が実現する。

以上の質の高い研究データの適切な管理・利活用や、AIを含めた積極的なデータサイエンスの活用、そして先進的なインフラ環境の整備は、単に研究プロセスの効率化だけではなく、研究の探索範囲の劇的な拡大、新たな仮説の発見や提示といった研究者の知的活動そのものにも踏み込んだプロセスを変革し、従前、個人の勘や経験に頼っていた活動の一部が代替されていくことになる。これにより、データを用いたインパクトの高い研究成果の創出につなげるほか、研究者の貴重な時間を、研究ビジョンの構想や仮説の設定など、より付加価値の高い知的活動へと充当させていく。同時に、グローバルな視点からも、オープンサイエンスの発展に貢献する。

さらに、このような研究活動の変革や我が国全体の雇用慣行の変化によって、研究者の在り方も変わる面があり、既に世界各地では見られる、シチズンサイエンスとしての市民の研究参加や研究者のフリーランス化など、多様な主体が研究活動に参画し活躍できる環境が我が国でも実現し、研究者とそれ以外の者が、信頼感を醸成しながら、知の共有と融合を進め、新たな形での価値創造を実現する環境整備を図っていく。

### 【目標】

- ・ 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、世界最高水準のネットワーク・計算資源の整備、設備・機器の共用・スマート化等により、研究者が必要な知識や研究資源に効果的にアクセスすることが可能となり、データ駆動型研究等の高付加価値な研究が加速されるとともに、市民等の多様な主体が参画した研究活動が行われる。

### 【科学技術・イノベーション政策において目指す主要な数値目標】（主要指標）

- ・ 機関リポジトリを有する全ての大学・大学共同利用機関法人・国立研究開発法人において、2025年ま

で、データポリシーの策定率が100%になる。公募型の研究資金<sup>162</sup>の新規公募分において、2023年度までに、データマネジメントプラン（DMP）及びこれと連動したメタデータの付与を行う仕組みの導入率が100%になる。

### (c) 具体的な取組

#### ① 信頼性のある研究データの適切な管理・利活用促進のための環境整備

- 研究データの管理・利活用のための我が国の中核的なプラットフォームとして2020年度に本格運用を開始した研究データ基盤システム（NII Research Data Cloud）の普及・広報と必要な改良を引き続き進める。また、公的資金により得られた研究データについて、産学官における幅広い利活用を図るため、2023年度までに体系的なメタデータ<sup>163</sup>の付与を進め、同年度以降、研究データ基盤システム上でこれらのメタデータを検索可能な体制を構築する。さらに、メタデータをE B P M<sup>164</sup>に活用するため、e-Radの改修に合わせて、相互運用性を確保する。研究データ基盤システムについて、持続的な運営体制の確保に向け2022年度までに方策を検討する。【科技、文、関係府省】
- 公的資金により得られた研究データの機関における管理・利活用を図るため、大学、大学共同利用機関法人、国立研究開発法人等の研究開発を行う機関は、データポリシーの策定を行うとともに、機関リポジトリへの研究データの収載を進める。あわせて、研究データ基盤システム上で検索可能とするため、研究データへのメタデータの付与を進める。【科技、文、関係府省】
- 公募型の研究資金の全ての新規公募分について、研究データの管理・利活用を図るため、データマネジメントプラン（DMP）及びこれと連動したメタデータの付与を行う仕組みを2023年度までに導入する。次期S I Pにおいても同様に、DMPの策定とメタデータの付与を実施することとする。【科技、文、関係府省】
- 研究データ基盤システムと内閣府が実施する研究開発課題（S I P等）で構築する分野ごとデータ連携基盤との間で、相互にデータの利活用を図るための仕組みを2023年度中に構築する。【科技、文】
- 研究者の研究データ管理・利活用を促進するため、例えば、データ・キュレーター、図書館職員、U R A、研究の第一線から退いたシニア人材、企業等において研究関連業務に携わってきた人材、自らの研究活動に資する場合にはポストドク等の参画や、図書館のデジタル転換等の取組について、2022年度までにその方向性を定める。【科技、文、関係府省】
- 自由で開かれた研究活動を尊重し、我が国と価値観を共有する国・地域・国際機関等（E U、G 7、O E C D等）との間で、研究データの管理・利活用に関する連携を進める。我が国の研究データ基盤システムとこれに相当する取組との国際連携を図り、研究データの管理・利活用に関する国際的な相互運用性を高めることにより、本計画期間中に、グローバルプラットフォームの構築を目指す。【科技、文】
- 研究データの管理・利活用に関する取組を更に促す観点から、2022年までに、これらの取組の状況を、研究者、プログラム、機関等の評価体系に導入する。【科技、関係府省】

<sup>162</sup> 「府省共通研究開発管理システム（e-Rad）について」（[https://www.e-rad.go.jp/dl\\_file/particulars\\_e-rad.pdf](https://www.e-rad.go.jp/dl_file/particulars_e-rad.pdf)）において、システムの対象として規定される公募型の研究資金。

<sup>163</sup> 体系的なメタデータとは、統一した様式により研究データの概要を示したデータであり、研究データの名称や説明、管理者、保管場所、共有・公開の有無等の情報を含む。「統合イノベーション戦略2020」（2020年7月17日閣議決定）において、ナショナルレベルでのデータポリシーを定めることとしている。

<sup>164</sup> E B P M：Evidence-based Policy Making。エビデンスに基づく政策立案。

## ② 研究DXを支えるインフラ整備と高付加価値な研究の加速

- 2022年度に、我が国の大学、研究機関等の学術情報基盤として、全国をつなぐ超高速・大容量ネットワーク（SINET）を増強し、これを研究データ基盤システムと一体的に運用することで、最先端の研究教育環境を提供する。また、引き続きこれらの学術情報基盤を支える技術の研究開発を推進する。さらに、2021年度までに、学術情報基盤としての役割のみならず、大学等の知を生かせる我が国の社会基盤インフラとして、民間と連携しつつ利活用できる環境整備の方策を検討する。【科技、文】
- スパコン計算資源については、2021年よりスーパーコンピュータ「富岳」の本格的な共用を進めるとともに、国内の大学、国立研究開発法人等のスパコン計算資源について、全国の研究者の多様なニーズに応える安定的な計算基盤として増強する。加えて、次世代の計算資源について、我が国が強みを有する技術に留意しつつ、産学官で検討を行い、2021年度までに、その方向性を定める。この検討の結果を踏まえ、必要な取組を実施する。【文、関係府省】
- 研究設備・機器については、2021年度までに、国が研究設備・機器の共用化のためのガイドライン等を策定する。なお、汎用性があり、一定規模以上の研究設備・機器については原則共用とする。また、2022年度から、大学等が、研究設備・機器の組織内外への共用方針を策定・公表する。また、研究機関は、各研究費の申請に際し、組織全体の最適なマネジメントの観点から非効率な研究設備・機器の整備が行われていないか精査する。これらにより、組織的な研究設備の導入・更新・活用の仕組み（コアファシリティ化）を確立する。既に整備済みの国内有数の研究施設・設備については、施設・設備間の連携を促進するとともに、2021年度中に、全国各地からの利用ニーズや問合せにワンストップで対応する体制の構築に着手し、2025年度までに完了する。さらに、現在、官民共同の仕組みで建設が進められている次世代放射光施設の着実な整備や活用を推進するとともに、大型研究施設や大学、国立研究開発法人等の共用施設・設備について、リモート化・スマート化を含めた計画的整備を行う。【科技、文、関係府省】
- データ駆動型の研究を進めるため、2023年度までに、マテリアル分野において、良質なデータが創出・共用化されるプラットフォームを整備し、試験運用を開始する。また同様に、ライフサイエンス分野においても、データ駆動型研究の基盤となるゲノム・データをはじめとした情報基盤や生物遺伝資源等の戦略的・体系的な整備を推進する。さらに、環境・エネルギー分野、海洋・防災分野等についてもデータ駆動型研究の振興に向けた環境整備を図る。加えて、プレプリントを含む文献など、研究成果に係る情報を広く利用できる環境の整備を推進するとともに、これらを支える基盤分野（OS、プログラミング、セキュリティ、データベース等）を含めた数理・情報科学技術に係る研究を加速する。【文、経】
- 2020年度に実施した試行的取組をベースとして、DXによる研究活動の変化等に関する新たな分析手法・指標の開発を行い、2021年度以降、その高度化とモニタリングを実施する。【文】

## ③ 研究DXが開拓する新しい研究コミュニティ・環境の醸成

- 地方公共団体、NPOやNGO、中小・スタートアップ、フリーランス型の研究者、更には市民参加など、多様な主体と共創しながら、知の創出・融合といった研究活動を促進する。また、例えば、研究者単独では実現できない、多くのサンプルの収集や、科学実験の実施など多くの市民の参画（1万人規模、2022年度までの着手を想定）を見込むシチズンサイエンスの研究プロジェクトの立ち上げなど、産学官の関係者のボトムアップ型の取組として、多様な主体の参画を促す環境整備を、新たな科学技術・イノベーション政策形成プロセスとして実践する。【科技、文】