

第6期科学技術・イノベーション基本計画を踏まえた
情報分野における研究開発等の進め方について 議論の中間まとめ（案）

令和3年 月 日
科学技術・学術審議会
情報委員会

第6期科学技術・イノベーション基本計画（令和3年3月26日閣議決定。以下、「基本計画」という。）では、Society 5.0の実現に向けた研究力の強化において、「研究活動のDX（デジタルトランスフォーメーション：研究DX）」の世界的な加速が指摘されるとともに、研究データの管理・共有・利活用を進める環境の整備の重要性が指摘されている。そこで、情報委員会においては、基本計画を踏まえた情報分野の研究開発等の進め方を検討するにあたり、まず、「研究DXとそれを支える学術情報基盤のあり方」に関して議論を行った。ここではその議論結果をまとめる。今後は、情報分野における研究開発課題等を議論しまとめる見込みである。

<議論のまとめ>

研究DXとそれを支える学術情報基盤のあり方

【研究DXとそれを支える研究データ基盤の重要性】

- 基本計画においては、研究DXを通じてより付加価値の高い研究成果の創出を目指す、としている。このために、研究データの戦略性を持った適切な管理及び共有と利活用、ネットワーク、データインフラ、計算資源等の研究基盤の形成・維持、さらに、リモート研究やスマートラボの普及の推進により、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、新たな仮説の発見等、個人の経験や勤からの研究のプロセスの変革を目指している。
- 特に、研究データの管理及び共有と利活用は、それによりインパクトが高い研究成果を創出する研究基盤の実現が求められるとしており、まず、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、研究データの管理・共有・利活用を進める環境を整備するとしている。
- 大学においては、目まぐるしく変わるテクノロジーやユーザー利用環境への対応、人材・予算の確保等の課題を抱えつつ、研究DXを含む大学DXの推進が求められている。大学ICT推進協議会では、今後の大学の情報環境整備のあり方に関する提言を出し、2030年の大学の情報環境を想定し、共通基盤開発体制の強化や、最先端設備の産学官共同開発の実現、大学間の情報環境格差の是正等の重要性を指摘している。
- これらを踏まえると、分野・機関を越えて研究データを管理・共有・利活用する全国的な研究データ基盤の構築を行うとともに、それが安定的に運用される仕組み・体制等を国全体として整備していくことが重要であると考えられる。

【研究データ等管理・共有・利活用のシステム・ツール】

- 研究データ基盤においては、利用者のニーズに応えるべく、研究データの管理・蓄積・利活用・流通、セキュリティ・トラスト、人材育成等の必要な機能の強化が重要である。
- 研究者が研究を行う際に、研究サイクル（テーマ決定、文献・データ調査、研究実施、評価、論文化等、公開）の中で、大きな負荷がかかることなく、研究データの管理・共有・利活用が進むような機能やシステム上の工夫を研究データ基盤に備えることが重要である。
- 研究と教育という観点のみならず、産学連携・社会での活用も含めて、研究データが共有・利活用されるシステムができることが望ましい。特に、学術研究で集めた個人のデータの利活用が民間でも進むような仕組み作りが促進されることが望ましい。
- マテリアル等の分野において、研究データ基盤を活用した、データ駆動型研究の成功事例を作り出していくことが重要である。

【研究データ等管理・共有・利活用のための制度・ルール】

- 研究データの管理・共有・利活用に関するルールの整備が重要である。大規模な大学等の研究機関は責任をもって研究データを維持管理するとともに、中小規模の大学などから出る研究データは、各機関での維持管理が困難な場合、研究データ基盤を整備する中核機関（群）などが中心になってデータを維持管理することが重要である。
- 研究データを共有して利活用するためには、研究分野ごとの代表的な機関がデータフォーマットやメタデータ等の標準化を推進することが望ましい。また、分野間データ連携基盤等との連携や国際的な通用性も想定した、標準化、ルール設定等が重要である。
- 自然現象に由来するデータ、個人のデータそれぞれの性質を踏まえ、個人情報保護法の改正等の法的・制度的観点に加えて、機密性、完全性の確保、社会受容性の向上も想定したルール・ガイドライン・マネジメントポリシーの整備、社会的コンセンサスを得る活動等、研究データ利活用促進のための活動が重要である。
- 多くの研究者が実験データ等の研究データを共有・利活用できるよう、データ提供者も評価される等のメリットがある仕組みづくりが重要である。

【研究データ等管理・共有・利活用にあたる人材】

- データエンジニア等、育成・確保に向けて、多くの研究機関等に共通するキャリアパス創出・待遇向上、産学の間での流動性の確保（特に、学から産への流動性の向上）、多様な人材育成メニューの提供等に取り組むべきである。

【研究 DX を進める上での留意点等】

- DX の効果を最大化するためには、研究、教育、事務作業の全体を含めて総合的に DX を進めていくことが重要である。
- 研究データ基盤の活用においては、SINET 等のネットワークや基盤センターの運用等、学術情報基盤全体の連携が重要であり、引き続き安定的な運用が求められる。

<付録：その他、関連する点、留意すべき点>

1. 研究 DX とそれを支える研究データ基盤の重要性

- ・大学 ICT 推進協議会 (AXIES) の提言「多様な教育研究活動の高度化を支える大学 ICT 基盤の集約化・共通化・協働化～コロナ時代における大学のデジタルトランスフォーメーションに向けて～」(令和2年12月10日)では、2030年の大学においては、①運営母体としての大学間の共同事業体、②大学経営における情報戦略、③ICT人材のキャリアパスの多層化、④国際通用性の担保が、より重要になると述べられている。
- ・研究 DX には、デジタル化による効率性向上と、新たな価値の創出がある。効率性向上を実現しつつ、新たな価値の創出までを想定すべきであり、その成功事例を作り出していくことが重要。新たな価値創出の研究として、①コンピューティング環境を使った DX 的研究(マテリアル、ゲノム合成の研究等)、②研究 DX を実現するための研究(ロボティクスバイオロジー等)が進展し始めてきている。これらの研究について、バラバラに支援するよりも、ある程度まとまった形(研究拠点等)で支援する必要があるのではないか。

2. 研究データ等管理・共有・利活用のシステム・ツール

- ・利用者のニーズに応えるべく、管理、蓄積、活用、流通、信頼、育成という6つの機能(データガバナンス機能、セキュア解析機能、データプロビナンス機能、キュレーション機能、セキュア蓄積環境、人材育成基盤)を強化するとともに、分野での利用の促進や研究機関内での意義の強化を図り、利用拡大に取り組んでいく必要がある。
- ・研究者が研究サイクル(テーマ決定⇒文献・データ調査⇒研究実施⇒評価⇒論文化等⇒公開)を回す際に、研究データの蓄積・管理・利活用が進むような機能やシステム上の工夫が、研究データ基盤に備わることが重要である。
- ・研究データ基盤は新たなシステム・サービスであり、利用者側に発信・説明していく必要がある。データエンジニアに加えて、データ営業、データヘルプデスク等のサービス・サポート体制を整備していくことが重要である。
- ・研究データは、研究と教育という観点のみならず、産学連携・社会での活用も含めて共有・利活用されるシステムができればよい。特に、大学等も関与してヒューマンデータの利活用が民間でも進むような仕組みが作られると、日本はもっと強くなるのではないかと。
- ・CO2排出量の側面として、情報基盤の開発・運用時の排出量の削減方針を提言することも検討しても良いのではないかと。

3. 研究データ等管理・共有・利活用のための制度・ルール

- ・研究データの蓄積は、大規模な大学、国研等の研究機関は責任をもってデータを維持管理するとともに、中小規模の大学などから出るデータはNIIなどが中心になって引き受けることが必要ではないか。
- ・地方創成の担い手となる地方大学への配慮が有意義。研究データ・リポジトリは各大学で設置するだけでなく、研究分野によってはNIIや学術団体等に設置する方が使いやすいこともある。
- ・蓄積したデータを使えるようにするため、研究分野ごとに中核機関を設定し、データフォーマットやメタデータ等の標準化を推進し、その機関が日本を代表して国際的にも連携するような構造とすることが良いのではないか。中核機関はシチズンサイエンスの推進にも取り組むべきである。
- ・分野別リポジトリや機関別リポジトリも含め、分散させつつ共有できるシステムとすべきである。
- ・研究データ基盤については、ネットワークやリポジトリと一体的に整備することが重要である。また、その実現に向けて、技術面（ネットワーク・クラウド・セキュリティ等）と非技術面（ガイドライン、人材育成、データ管理の手法、個人情報管理等）の両面で取り組むべきである。
- ・研究データ共有の大きな壁は、個人情報の問題が一番大きい。個人情報の取扱いに関するグレーゾーンが大きく、自己規制で厳しい方に倒れてしまい、法律上許される取扱いまで制限してしまうことがある。研究データの共有・利活用促進のために個人情報の取扱いについて議論することが必要である。
- ・データのインテグリティ（完全性）やコンフィデンシャルティ（機密性）が管理されないと研究成果がひっくり返ってしまうため、共有するデータこそインテグリティ管理が重要である。
- ・合法的なデータ利活用でも批判を受けることがあるため、データを使ったビジネスによって暮らしがより豊かになるという社会的コンセンサスも必要であろう。
- ・日本企業はオープンデータを利用していない、選択肢に入っていない、という傾向があり、行政・学術のデータが企業に認知されていない、利活用の検討対象になっていない、と推測される。

- ・研究データの蓄積・管理・利活用に関するルールの整理、整備が重要。特に、多くの研究者が実験データ等を共有・流通させていけるよう、データ提供者も評価されるというメリットがあるような仕組みづくりができるとういのではないか。
- ・研究 DX を進めるには、研究者の研究スタイルの変革が重要であり、先導的なプロトタイプの研究を進め、成功事例を公表するなど研究者全体を引っ張っていくことが必要。また、インセンティブ・呼び水となる研究支援も必要ではないか。

4. 研究データ等管理・共有・利活用にあたる人材

- ・データエンジニアが大変重要。育成・確保に向けて、多くの研究機関等に共通するキャリアパス創出・待遇向上、多様な人材育成メニューの提供等に取り組むべきである。
- ・民間で情報系人材は流動性が高いが、民から学は増えているものの学から民は変わっていない。産学の間での流動化も含めた情報系のキャリアパスを描いて成功例を作っていくことが重要である。
- ・全国的な研究データ基盤を支えるために、それを管理・運営していく人員を手当てできるか、また、データエンジニアのレベル（研究者クラスの知識と技能を持った上で、かつ、ある程度サービスに徹した仕事をする人材）をどのように確保していくかが課題である。
- ・人材の確保に関して、経済安全保障の観点から海外人材をどう考えるのか、早めに作戦を練っておくことが必要である。

5. 研究 DX を進める上での留意点等

（総合的 DX の推進）

- ・DX の効果を最大化するためには、研究、教育、事務作業の全体を含めて総合的に DX を進めていくことが重要である。
- ・コロナ禍によるネットワーク授業の結果、教育データや教育コンテンツの蓄積が進展している。これらを基盤として、データ駆動型の教育や教育に関するデータ駆動型の研究を振興し、その後、さらに広範囲にわたる研究へ拡張していくことができるのではないか。
- ・人文・社会科学系の実験や調査の DX も重要。また、研究・教育組織の経営や評価に当たってもデータを活用するのが重要ではないか。これらデータの蓄積に当たっては、ログデータ等だけでなく、そのメタデータも同時に集めていく必要がある。
- ・研究データ基盤の高度化に資する情報科学技術研究の実績に、グリーン関連で、低炭素、排出量削減、省エネ等の要素が入っていた方がよいのではないか。

(安定的運用、国全体での取組)

- ・日本の大学を含む研究・教育機関は SINET と NII のデータ基盤に頼っており、この基盤を長期にわたって支え続ける安定的な財源の確保と、利用者側のアプリケーションをつなぐ基盤センターの運用が重要である。
- ・現在、NII (SINET (ネットワーク) と研究データ基盤)、理研 (富岳) と大学の情報基盤センター等 (HPCI) が日本のアカデミアの情報基盤の開発・運用を支えているが、研究 DX の推進に向けて、これらの体制の強化 (枠組み、人員体制、資金等) も視野に入れて議論すべきではないか。特に、研究データ基盤についてはこれからの新たな取組であり、既存の枠組み・運用を超えたりソースの増強が必要である。
- ・データ駆動型研究を進める上では、データを蓄積することに加えて活用することが重要である。そのためには、データ基盤とともに計算基盤の整備が必要であり、かつ、柔軟に構成可能でセキュアなネットワーク環境の構築が必要である。さらに、これらは分野を越えた国全体の共通基盤であることも重要である。
- ・産学官で最先端設備の共同開発・整備を実施できる枠組みを実現するとともに、予算の安定的確保や柔軟な執行、情報環境の格差の是正に取り組むべきではないか。
- ・高品質なデータを構築・維持するためには、コストの見える化が重要。互換フォーマットへの変換やデータの維持等、見落とされがちなコストがある。
- ・各省庁で似たような取組をしているが、省庁をまたいだ戦略や統合化を考えることが重要である。