

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

令和8年 7 月

科学技術・学術審議会 情報委員会
AI for Science を支える研究データの管理・利活用
と流通の在り方ワーキンググループ

-目次-

1. はじめに.....	3
2. 研究データの管理・利活用と流通を支える次世代情報基盤の必要性.....	3
(1)背景(AI for Science 推進とオープンサイエンス進展)	3
(2)次世代情報基盤を巡る政策動向.....	4
3. 学術研究プラットフォームの構築・実施状況	6
(1)情報流通基盤(SINET 等)の現状.....	6
(2)研究データ基盤(NII RDC 等)の現状	7
4. 国内外の動向	10
(1)海外の情報通信ネットワークの動向.....	10
(2)海外の研究データ基盤の動向	10
(3)国内における現状と国際連携.....	11
5. ユーザーニーズ、社会からの要請.....	12
(1)情報流通基盤.....	12
(2)研究データ基盤.....	13
(3)知識創出のための新たな AI 統合型の研究データ基盤の構築	14
(4)研究分野からのニーズ	14
6. AI for Science の推進に向けた基本的な戦略方針への対応.....	16
7. 次世代情報基盤整備の方向性	17
(1)付加価値の最大化を実現するネットワークの整備	17
(2)研究プロセスを総合的に支える新たな研究データ基盤の構築.....	18
(3)計算資源との一体的な接続・連携、認証強化	20
8. おわりに	20
参考資料.....	22
(1)科学技術・学術審議会 情報委員会 における下部組織の設置について	23
(2)科学技術・学術審議会 情報委員会 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ワーキンググループ 委員名簿	24
(3)審議経過	25
審議まとめ概要.....	27
<別冊> 基礎資料(関係資料等抜粋)	

1. はじめに

近年、AI 技術の飛躍的な進展とデータ駆動型科学(Data-driven Science)の深化を背景として、研究の在り方は大きな転換期を迎えている。観測・実験機器やシミュレーションの高度化により、研究分野横断で膨大かつ多様な研究データが日常的に生成・蓄積される一方、それらを効果的に管理・共有し、AI をはじめとする先端技術と組み合わせて新たな知を創出するための情報基盤の重要性が急速に高まっている。特に「AI for Science¹」は、研究プロセスそのものを変革し、従来では到達困難であった科学的知見の創出を可能とする中核的な手段として位置付けられている。

こうした状況を踏まえ、我が国では第7期科学技術・イノベーション基本計画²において、研究データの戦略的利活用とオープンサイエンスの推進、これらを支える情報基盤の整備・高度化が重要課題として掲げられている。また、情報委員会による「次世代の科学技術イノベーションを支える情報基盤の在り方について(中間とりまとめ)³(以下、「中間とりまとめ」とする。)」や「AI for Science の推進に向けた基本的な戦略方針⁴(以下、「戦略方針」とする。)」においても、研究データの円滑な流通を支える基盤、分野横断的なデータ連携や AI が出力する情報の信頼性を担保する新たな情報基盤の構築、計算資源・ネットワーク・認証基盤の一体的整備の必要性が示されている。

今回、「科学技術・学術審議会 情報委員会 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ワーキンググループ(以下、「WG」とする。)」では、これらの政策文書に示された方向性を踏まえつつ、AI for Science を支える 研究データの管理・利活用および流通を支える次世代情報基盤の必要性について整理するとともに、我が国の学術研究プラットフォームの現状、国内外の動向、ユーザーニーズや社会的要請等を俯瞰し、議論を重ねた。その上で、AI for Science の本格的な推進に資する情報基盤の在り方と、今後重点的に取り組むべき次世代情報基盤整備の方向性について検討し、政策的示唆を提示するため、結果をとりまとめたところである。

2. 研究データの管理・利活用と流通を支える次世代情報基盤の必要性

(1)背景(AI for Science 推進とオープンサイエンス進展)

¹ 研究プロセスのあらゆる段階において AI を利活用することにより、研究の効率性・生産性を向上させ、科学研究の在り方を革新させる取組

² 令和 8 年 4 月 27 日 閣議決定

³ 令和 7 年 5 月 30 日 科学技術・学術審議会 情報委員会 とりまとめ

⁴ 令和 7 年 3 月 31 日 文部科学省 決定

オープンアクセスと研究データのオープン化を含む「オープンサイエンス」の進展、新型コロナウイルスを一つの契機とした研究 DX(デジタル・トランスフォーメーション)により、世界は知識集約型社会へと大きな変革を遂げている最中であったが、生成 AI をはじめとする AI 技術の飛躍的な進展は、データ駆動型科学を深化させ、データの価値・概念を一新させた。観測・実験・シミュレーション・解析の各研究段階においてデジタル技術が深く組み込まれ、研究活動全体を通じて膨大かつ多様な研究データが生成・蓄積されるようになった。特に、AI を活用して研究仮説の創出や実験計画の最適化、解析の高度化等を行う「AI for Science」は、従来の研究手法を補完・拡張するのみならず、研究プロセスそのものを変革し得るものとして国際的にも注目されている。

このような研究環境の変化の中で、研究データは単なる研究成果の副産物ではなく、新たな知を生み出すための基盤的資源としてその重要性が一層高まっている。一方で、研究データの管理方法や記述形式、公開・共有・非公開・非共有の在り方(オープン・アンド・クローズ戦略)は分野や機関ごとに大きく異なり、データの再利用や分野横断的な利活用を阻害する要因となっていることが、大学等研究機関等から指摘されている。

また、社会課題の複雑化・高度化に伴い、単一分野や単一組織による研究だけでは十分な対応が困難となっており、異分野連携や産学官連携を通じた研究データの共有・流通の重要性はこれまで以上に増していると言える。感染症対策、気候変動、エネルギー、材料開発などの分野では、データを迅速かつ安全に流通させ、AI を活用して知見を統合することが、研究成果の創出を左右する鍵となっているとも言われている。

こうした背景の下、オープンサイエンスの理念に基づき、研究データを適切に管理し、AI 技術を取り入れつつ、FAIR 原則⁵を踏まえて、共有・再利用を促進することは G7、UNESCO 等をはじめ、国際的な潮流となっている。他方で、知的財産、個人情報、経済安全保障の観点から、必ずしも全てのデータを一律に公開することは適切ではなく、オープン・アンド・クローズ戦略を踏まえた戦略的なデータ管理・利活用とデータ流通を促進するための設計が求められている。

このため、研究のライフサイクル全体を見据えた次世代情報基盤の整備が不可欠となっている。

(2)次世代情報基盤を巡る政策動向

第7期科学技術・イノベーション基本計画をはじめとする関連政策文書において、研究データの戦略的利活用と、これを支える情報基盤の整備・高度化は重要な柱として

⁵ データを共有するための基準となる国際的な原則

Findable(見つけられる)、Accessible(アクセスできる)、Interoperable(相互運用できる)、Reusable(再利用できる)

位置付けられている。特に、AI for Science を推進するためには、高品質かつ多様な研究データを分野横断的に活用できる環境の構築が不可欠であり、研究データ基盤、流通基盤、計算基盤を一体的・シームレスに連携させた情報基盤の実現が求められている。

中間とりまとめでは、研究データの保存・管理、流通、活用を担う情報基盤が、AI 時代における新たな科学技術・イノベーションを切り開く中核的インフラであることが示されている。同時に、国立情報学研究所(NII)が中心となって構築を進めてきた、我が国の研究データの管理・利活用のための中核的なプラットフォーム⁶として位置づけられた研究データ基盤システムである NII RDC⁷について、機能の高度化や利活用の拡大、研究者の視点で改善の必要性が指摘されている。

加えて、研究データを AI の学習データとして活用し、AI の性能を高度化させ、高度化された AI を用いた研究で取得したデータを情報基盤に還元し、さらなる AI の高度化に活用していくという好循環を創出することが、我が国全体の研究力強化や産業競争力の向上につながるとも指摘されている。

また、AI for Science の本格的な推進に当たっては、研究データ、計算資源、人材を一体的に支える情報基盤の高度化が不可欠である。研究プロセス全体に AI が深く関与する中で、戦略方針においては、日本の取るべき基本戦略は、日本の資産とリソースを最大限に活用し、勝ち筋になり得る分野等の研究力を世界のトップ水準に引き上げることにありとし、AI for Science を推進するためには、それを支える研究インフラ(情報基盤、データ創出基盤等)の構築・整備が不可欠であると指摘している。

さらに、第 7 期科学技術・イノベーション基本計画においては、AI for Science の進展により、研究データの量・多様性が飛躍的に拡大することを前提に、研究データ基盤や情報流通基盤を含む情報基盤の高度化が不可欠であるとされている。特に、学術情報ネットワーク SINET⁸は、全国の研究大学等を結ぶ高信頼で大容量なネットワークとして、研究設備の遠隔利用や研究データの集約・流通を支える中核的インフラと位置付けられている。

これらのことを踏まえると、次世代情報基盤の整備に当たっては、個別システムの短期的な更新や機能拡張にとどまるのではなく、AI for Science の本格的な推進を見据え、中長期的な視点から研究エコシステム全体を支える基盤として設計・運用していくことが極めて重要であるといえる。すなわち、研究設備・機器等、研究データ、計算

⁶ 「公的資金による研究データの管理・利活用に関する基本的な考え方」(令和 3 年 4 月 27 日 統合イノベーション戦略推進会議) 参照

⁷ NII Research Data Cloud:研究データのライフサイクルに即した「管理基盤(GakuNin RDM)」、「公開基盤(JAIRO Cloud)」、「検索基盤(CiNii Research)」の3つの基盤から構成

⁸ Science Information Network :昭和 62 年に学術情報ネットワークパケット交換網として運用が開始され、令和 4 年度より現行の SINET6 を NII において運用中

資源、ネットワーク、人材・運用体制が相互に関連し、持続的に価値を創出できる情報基盤として高度化を図ることが、今後の政策上の重要な課題であると認識される。

3. 学術研究プラットフォームの構築・実施状況

(1) 情報流通基盤(SINET 等)の現状

我が国の学術研究を支える SINET は、研究機関間の通信確保を目的とする共同利用のための基盤として整備が始まった。初期段階においては、主として学術情報の共有や基礎的な研究連携を支えるためのネットワーク環境が重視されており、通信の安定性や基本的な接続性の確保が中心的な課題であり、この段階では、ネットワークは研究活動を「補助するインフラ」として位置付けられていた。

その後、計算機性能の向上、大型研究施設の整備、分野横断型研究の拡大などを背景に、学術ネットワークに求められる役割は大きく変化した。研究データの大規模化・通信の高速化が進む中で、情報流通基盤は単なる接続のための手段ではなく、研究そのものを成立させる基盤として位置付けが変化していった。

現在運用されている SINET は、コミュニティと対話しながら、こうした段階を経て構築されたものであり、今や全国 1000 を超える組織かつ 300 万人以上のユーザーに対して、超高速・大容量通信を安定的に提供することで、スーパーコンピュータ、研究データ基盤、クラウドサービス、研究設備等を分断なく接続することを可能としている。

本 WG においては、SINET が「研究成果を単に流すインフラ」ではなく、「研究プロセスを同時並行的につなぐ研究基盤」へと進化していると指摘している。AI for Science の本格化により、情報流通基盤は新たな段階に入っている。研究は、データ収集・解析・学習・検証が高速に循環するプロセスへと変化しており、研究拠点間、研究設備間、計算資源間を超高速で、セキュアに結ぶネットワークが無くては成立しない構造であるといっても過言ではない。こうした中、SINET は全国の大学・研究機関、コアファシリティ、大型研究施設・設備等を結び、研究設備の遠隔操作、分散拠点における協調実験、国際共同研究の基盤として、より一層機能していくことが重要なポイントである。SINET は、研究の「効率化」を超え、研究の「可能性そのもの」を規定する基盤として進化することが求められている。

一方で、AI for Science がもたらす今後の通信量急増、国際接続の更なる高度化、運用人材の確保、セキュリティ確保と利便性の両立など、中長期的課題が顕在化しているため、短期的な更新対応として捉えるのではなく、研究構造の変化を前提とした次世代のアーキテクチャーを構成することが鍵となる。

また、NII が運用する情報セキュリティ運用連携サービス(NII-SOCS)についても、サイバーセキュリティの脅威が叫ばれる中、その価値は高まっている。NII-SOCS は、SINET 上の通信内容を確認することなく観測・解析し、不審通信を自動検知して参加機関等へ通知する仕組みであり、運用に際しては通信の秘密保護を徹底し、NII が通信の中身を任意に閲覧することは行わず、暗号化や通信の振る舞いを踏まえた自動化された解析手法により検知・通知を実施している。最大のミッションは、重大なサイバー攻撃の検知・情報提供、人材育成(CISO⁹や CSIRT¹⁰要員向け研修)、および観測データの統計化・研究利用支援の三本柱であり、観測体制は SINET とインターネット接続点における対外観測を中心に、サンプリングと深掘りの二段階観測を組み合わせる形態を採用している。本取組を通じて、平成 29 年の試行運用開始以降、国立大学等の即応力は著しく向上してきた。現状の課題は、観測・支援の対象が主に国立大学等に限定されていることであり、サービス拡大に伴うコスト負担と運用リソースの確保が障害となっている。

このほか、今後のセキュリティへの対応として、SINET 上の DDoS 攻撃¹¹対策等によるセキュリティ対策と一体的にクラウド型ファイアウォールの共同利用や認証基盤のホスティングなど、共同防御の仕組み導入の検討も進められている。今後は、NII-SOCS の公私立大学への対象拡大と拡張に伴う財政的・組織的課題の解決、および認証基盤等他の基盤との連携強化が主要な論点である。

(2) 研究データ基盤(NII RDC 等)の現状

研究データの管理については、長らく研究者個人や研究グループ単位での対応が中心であり、保存や共有の方法は分野・機関ごとに大きく異なっていた。この段階では、研究データは主として成果論文の裏付けとして扱われ、長期的利活用や再利用は限定的であった。オープンサイエンスや FAIR 原則など国際的な潮流、研究の公正性確保への要請を背景に、研究データを組織的に管理し、共有可能な形で保存する必要性が高まった。これにより、研究データは「個別管理」から「組織的管理」へと段階的に移行している最中である。

こうした流れの中、NII は、研究データ基盤「NII RDC」を構築してきた。NII RDC は、機関リポジトリ、研究データ管理、検索基盤等を統合し、研究データを共通的に取り扱うための基盤として令和 3 年より本格的に運用を開始している。

オープンサイエンスの潮流、研究 DX の進展と呼応するように、NII RDC は研究データの管理・利活用における我が国の中核的な研究データ基盤として位置付けられ、これまで以上に研究者の研究プロセスを牽引する基盤として進化するため、令和 4 年

⁹ Chief Information Security Officer(最高情報セキュリティ責任者)

¹⁰ Computer Security Incident Response Team(情報システムに対するサイバー攻撃等の情報セキュリティインシデントが発生した際の情報セキュリティに関する統一的な窓口)

¹¹ Distributed Denial of Service attack(分散型サービス拒否攻撃)

度から「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」(以下、本事業という。)により、研究者に寄り添ったNII RDCの機能強化とユースケース創出、他分野のデータプラットフォームとの接続等実証のための分野連携、データ管理・利活用を支える人材の育成、そしてルール・ガイドラインの整備といった研究データを取り巻く課題の解決と向き合ってきた。

第6期科学技術・イノベーション基本計画期間中の5年間における主要KPIの一つとして、大学、大学共同利用機関法人、国立研究開発法人においては研究データポリシーの策定が求められた。特に大学等では、組織のマネジメントの下、大学図書館、情報基盤・メディアセンター、研究推進部、URA組織等が一体となってデータポリシーの実効的な運用に取り組みつつある。本事業では、ルール・ガイドライン整備を担当する名古屋大学を中心に、地域コンソーシアムの構築を段階的に進め、北海道から九州・沖縄地域まで全国の地域で研究データ管理・利活用を支えるためのコミュニティが誕生した。今後は、各地域コミュニティが自律的かつ継続的に活動を発展・高度化させながら、地域間の結節をより強固なものとし、自機関のみならず全国規模での支援を実現できるような環境へと深化していくことが課題として挙げられる。

また、研究データは研究分野ごとに異なる顔を見せる。データ利活用が先導するマテリアル分野やライフサイエンス分野をはじめとして、データの種類・形態・形式等が大きく異なることや、技術進展に伴いデータの解像度が高まり、一つ一つのデータ量が極めて大きくなってきているほか、大型実験装置から創出されるデータは1実験あたり数ペタバイトを超えるものも存在するなど、NII RDCのみで、解決できるものではないことにも留意が必要である。特に先端研究基盤刷新事業「EPOCH¹²」や大規模集積研究システム形成先導プログラム¹³などの国家プロジェクトが推進される中、各方面との連携・協働を強め、研究データの管理・利活用の促進を進めていくためにも、NII RDCの進化及び体制整備の強化が鍵を握る。

さらに、AI for Scienceの核心はデータの多様性と質にあるとされ、分野により状況は異なるが、非公開データや失敗事例を含む非競争領域での高品質なデータは、科学基盤モデルの構築にも重要な資源となるため、学術界限のみならず、産学間においても秘匿性を保ちつつ、共同活用・共有するためのプラットフォームが形成されることにより、新たな競争力の源泉となり得る。

本WGでは、AI for Scienceの進展により、研究データが「保存対象」から「解析・学習・再利用を前提とした研究資源」へと位置付けが変化している点が指摘され、

¹² 令和7年度補正予算より文部科学省が推進する事業:

我が国の研究基盤を刷新し、若手を含めた全国の研究者が挑戦できる魅力的な研究環境を実現するため、全国の研究大学等において、地域性や組織の強み・特色等も踏まえ、技術職員やURA等の人材を含めたコアファシリティを戦略的な整備等を推進

¹³ 令和7年度補正予算より文部科学省が推進する事業:

我が国が有する強みを活かした、オートメーション/クラウドラボの形成により、AI時代にふさわしい研究システム改革を先導

この段階では、研究データ基盤は研究活動の後工程ではなく、全国の研究者にとって簡便に使いこなすことができる AI 時代に適した基盤として、研究プロセスの中核要素へと進化することが必要である。

<研究分野における取組状況例>

i) ライフサイエンス分野(ライフサイエンス統合データベースプロジェクト)

我が国ではライフサイエンスデータベース統合推進事業(NBDC)を通じて基盤整備を進めてきたが、その取組を発展的に再編したのが NLDP(ナショナルライフサイエンスデータベースプロジェクト)である。NLDPは、分散し、形式の異なる多様な研究データベースを連携・統合し、統合ポータル整備や横断検索機能の提供を通じて、研究者の利活用環境を向上させてきた。

一方で、ライフサイエンス分野のデータは、ゲノム・臨床・画像など多様かつ大規模・高次元であり、分散性や形式の非統一性に加え、実験条件や解析手法に依存して再現性の確保が難しく、倫理的配慮も求められるという特徴を有する。こうした中、AI 利活用の観点からは、機械学習への対応や、セキュリティ・プライバシー・データ主権の確保といった課題も含め、これらの分野特有のデータの特徴に配慮した適切なデータ管理・利活用の枠組み整備が求められている。

ii) マテリアル分野(マテリアル DX プラットフォーム事業)

マテリアル分野においては、論文情報を基にしたキュレーションデータ、系統的に整備されたリファレンスデータ、日常的な実験から生成されるワーキングデータの三層構造でデータが蓄積されており、特に RDE 等の仕組みにより実験データの構造化・収集が進められている。また、データ基盤上での機械学習やデータ解析環境の整備、エンバーゴ期間を経たデータの広域共有、さらにデータの有償提供といった新たな流通モデルの取組も進展している。こうした環境のもと、データ駆動型研究から AI 駆動型のデータ利活用への移行が進みつつある。

iii) 大型研究施設(放射光施設・SPring-8)

放射光施設・SPring-8 においては、実験中の迅速な意思決定を支援するため、データのリアルタイム解析や研究者間でのデータ流通を重視した基盤整備が進められており、データセンターと HPCI 計算資源を連携させた大規模解析環境が構築されている。また、GakuNin RDM との連携により研究室レベルのデータ管理との接続も進展しているほか、高速測定装置の導入に伴う大量データ生成に対応するため、リアルタイム処理や圧縮技術などの高度化も進められている。

4. 国内外の動向

(1) 海外の情報通信ネットワークの動向

研究データ基盤や情報流通基盤の整備を巡っては、近年、AI 技術の急速な進展を背景として、国際的な競争環境の変化が一層顕著となってきた。特に、AI for Science の進展に伴い、AI を活用したデータ駆動型研究や分野横断型研究によるデータ量の肥大化、高速な通信の必要性が高まっており、これらの研究活動を支える情報通信ネットワークの在り方は世界規模で大きく変化している。

欧米諸国では、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が整備されており、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されている。これらの情報通信ネットワークは、国境を越えた相互接続を前提とする形で運用されており、研究データや解析結果を広域かつ円滑に流通させる役割を担っている。また、従来以上に大容量通信、低遅延、安定性及びセキュリティの確保が求められる状況となっており、本 WG では、こうした要件を背景として、情報流通基盤が研究活動の付加的要素ではなく、研究プロセスを成立させる基盤の一つとして位置付けられつつある点が確認されている。

(2) 海外の研究データ基盤の動向

研究データ基盤についても、海外では国家レベルでの整備が進展している。

欧州においては、EOSC (European Open Science Cloud) の構築が進められており、分野や国境を越えた研究データの共有・利活用を可能とする共通基盤の整備が段階的に進められている。EOSC は、研究データの保存や公開に加え、AI 活用や分野横断研究を見据えたデータ流通の仕組みを備える点に特徴がある。FAIR 原則に基づく分散・フェデレーション型の研究基盤として、認証や機能統合を実現したことで、分散された各国の研究データに対してシームレスなアクセスを可能としている。これは、EOSC が各国のストレージや計算資源へのアクセスのための最初の結節点・入口として機能することを意味する。

米国においても、連邦政府機関を中心に、研究データの管理や共有に関する方針と、それを支える技術的基盤の整備が進められている。研究データの公開や再利用に関する取組は、研究の透明性や再現性の確保に加え、AI 学習用データとしての活用を視野に入れたものとなっている。

特に DOE がジェネシス・ミッション¹⁴の実現のための主要な研究基盤として整備する The American Science Cloud (AmSC) や NSF が ID 連携を通じた CloudBank といった仕組みを通じた商用クラウド利用まで先端計算資源へのアクセスを統合するために整備する NSF ACCESS など、AI 時代に向けたデータ利活用に係る基盤の投資、整備が進行している。

これら海外の取組に共通する点として、研究データ基盤が単なるデータの保管場所ではなく、研究データの流通・利活用を前提とした共用基盤として設計・運用されている点や認証を通じた資源統合などが挙げられる。また、基盤の整備と併せて、データガバナンス、セキュリティ、国際標準への対応といった制度的・運用的側面が重視されている点も特筆すべき点である。

(3) 国内における現状と国際連携

国内においては、SINET、NII RDC、富岳を含む HPCI¹⁵等の情報基盤が整備されていることが我が国の強みとして位置付けられている一方、それらをどのように統合的に活用していくかが、課題として認識されている。

本 WG では、今後の情報基盤整備が「単なる基盤整備」ではなく、AI for Science を前提とした研究システム全体の変革に対応する必要があることが示されている他、研究データの創出・収集・利活用のプロセスが高度化・自動化していく中で、基盤側もそれに対応した機能を備える必要があるとの課題が共有されたところである。

また、最先端の研究設備を集積し、研究設備の自律化・自動化・遠隔化による大規模オートメーション/クラウドラボの構築を実現する拠点の形成や全国の研究大学等においてコアファシリティの戦略的な整備と先端的な研究設備・機器の整備・共用・高度化の推進等により、AI 利活用可能なデータセットの構築・共用化が促進されていく中、AI 駆動型研究を支えるデータ創出基盤と情報基盤との接続環境の構築は、欧米の状況も踏まえると、早急に取り組むことが必要である。

さらに、国際頭脳循環、国際共同研究の拡大に伴い、研究データの越境移転やセキュリティ、データガバナンスの在り方についても、研究活動の円滑な実施や研究環境の魅力度に影響を与える要素となっている。戦略方針では、今日の国際環境において、科学研究は一国のみで完結するものではなく、国際的なネットワークの中でいかなる価値を提供できるかが国家の研究力を規定していると指摘する。

加えて、AI for Science の推進に当たって、世界の科学研究にとって不可欠な高品質なデータ、高度な研究基盤、革新的な研究プロセスを提供する立場を確立するこ

¹⁴ 米国における AI による科学研究と技術革新の抜本的改革を目指す国家プロジェクト。令和 8 年6月4日、文部科学省、経済産業省及び米国エネルギー省(DOE)は、日本の AI for Science の取組と米国のジェネシス・ミッションとの連携に向けた日米戦略的パートナーシップを締結。

¹⁵ 革新的ハイパフォーマンス・コンピューティング・インフラ：国内の大学や研究機関の計算機システムやストレージを高速ネットワークで結んだ共用計算環境基盤

とが、日本が取るべき戦略的かつ現実的な選択であるとする中、こうした観点からも、経済安全保障にも留意した上で、国内の情報基盤を国際的な情報通信ネットワークやデータ基盤と接続可能な形で整備していくことも重要である。

5. ユーザーニーズ、社会からの要請

(1) 情報流通基盤

今後の AI for Science の進展に伴い、研究活動におけるデータの通信量や通信頻度は飛躍的に増大することが予想される。情報流通基盤には従来とは量的・質的に異なる役割が求められ、特に、大型研究施設・設備や国際共同研究においては、数百 Gbps 規模の通信がすでに発生しており、今後はテラビット級への拡張が必要とも言われている。

研究設備、計算資源、クラウドを組み合わせた研究プロセスが一般化していく中で、即時にデータを取得・解析・共有でき、分散した研究資源を一体的に運用する基盤として確立することを通じて、研究プロセスそのものの変革へとつながっていく。AI 技術の導入により、研究活動において発生する通信量はこれまで以上に増大する可能性があるが、AI エージェントによる反復的な解析やモデル呼び出しが常時行われる状況では、従来型の単純な帯域拡張では対応が困難となる。これらのことから、次世代の情報流通基盤はオール光ネットワーク (APN¹⁶) 等の最先端技術を活用した大容量・低遅延・低消費電力・高信頼性を同時に満たす高度な基盤である必要がある。

あわせて、オープンサイエンスによる研究データの国際的な流通が常態化する中で、国際ネットワークとの接続性及び相互運用性の確保が重要となる。FAIR 原則に基づき、研究データや解析処理ツールなどを国外の研究機関と共有するためには、性能面に加えて、セキュリティやアクセス制御に係る基盤機能の強化が不可欠である。

社会的要請の観点からは、機微データや重要な研究データの取扱いに関する懸念が高まっており、安全にデータを「流せる」環境の整備が強く求められている。研究データの国外移転や不適切な利用に対するリスクを低減するため、情報流通基盤自体が信頼性を担保する役割を果たすことも必要である。

また、研究環境の公平性の観点から、地方大学や小規模機関においても同等の通信環境を利用可能とする全国的な均質性の確保が求められている。教育研究機会の確保という観点からも、情報流通基盤は共用インフラとして高度化すると同時に、その利用格差を解消する方向で整備する必要がある。NTN¹⁷ (非地上系ネットワーク) などのモバイル環境の充実に関しても関心が寄せられている。加えて、初等中等教育を含

¹⁶ All-Photonics Network

¹⁷ Non-Terrestrial Network

む教育現場等における AI 活用のニーズを踏まえると、ネットワークの付加価値を高めることは、SINET が社会インフラとしてさらなる発展を遂げるうえで重要な要素であるといえる。そのため、これまで培ってきた SINET 自身の価値を、あらためて強く認識していく必要がある。

(2) 研究データ基盤

研究データ基盤に求められる機能は、これまでの「研究データを適切に管理・公開する機能」から「研究データを糧に AI を活用した新たな研究を駆動する原動力」へと大きく変化している。

現状では研究データが研究プロセスの途中で散逸・停滞し、十分に利活用されていない実態に関しても指摘がなされているが、データの収集・保存・解析・共有そして公開までを一体的にマネジメントすることにより、日々の研究の流れの中で自然に研究データが管理・利活用される基盤の実現が必要となっている。

また、多くのデータは機械学習での利用を前提として整備されていないほか、データセキュリティやプライバシー保護、データ主権の確保といった観点への対応も重要であり、適切なデータ管理・利活用の枠組みの整備が求められる一方で、そもそも研究データを集めること自体が容易ではないとの認識も指摘されている。AI 利活用を前提としたデータ形式やデジタル化が大きく遅れている分野特有のデータや、個人情報等により他の分野以上に共有・公開が困難な医療データなど、価値が高いにもかかわらず十分に活用されていないデータが多数存在している。これらの課題に対して、データの創出・収集・AI 利活用を前提としたデータ形式化そのものを支援する仕組み及びそれを担う人材の育成に関しても強く求められているところである。

さらに、AI 利活用が前提となっていくと仮定した場合、研究成果の創出において、データの量だけでなく質が重要となるため、メタデータ整備や構造化、標準化の要求は、コミュニティから強く求められていることに加え、分野横断研究では、異なる形式のデータを結びつけるための機能が不可欠であり、各研究コミュニティで求める共通のセキュリティ意識を実現する認証基盤の強化が重要な論点となる。

一方で、医療・個人情報・産業データなど、外部公開が困難なデータの利活用も重要な論点となっており、機微データを安全に扱うための秘密計算技術やアクセス制御、管理などオープン・アンド・クローズ戦略を技術的に可能とする基盤であることが重要であるという指摘もなされている。

また、研究の再現性や信頼性の観点から、データの来歴や処理過程を追跡できる仕組みが不可欠であるとの認識も共有されている。これは AI 利活用の進展に伴い、さらに重要性を増している。

加えて、産学連携の進展に伴い、民間データの活用や企業との共同研究を前提としたデータ基盤へのニーズも高まっている。研究データは組織内に閉じるのではなく、多

様なプレイヤーが参加可能な形で流通・活用されることが求められ、確実な身元保証を行う認証の確保が期待されている。

(3) 知識創出のための新たな AI 統合型の研究データ基盤の構築

上述の通り、AI for Science が加速度的に進展していく中で、研究データ基盤に対する大学等研究機関のステークホルダーからの期待は刻々と変化している。研究者のニーズはデータの蓄積・共有から、知識の生成・活用そのものを支援する基盤へと拡大しており、AI が単なる解析ツールにとどまらず、実験設計や自動化、仮説生成にまで関与し、研究プロセスそのものを変革するという可能性も指摘されている。

このような状況に対応するためには、データ解析だけでなく、実験・シミュレーション・知識生成を一体的に扱う環境が必要となる。従来のように個別ツールを組み合わせる形ではなく、データ創出基盤や計算基盤からの様々なデータを一元的に管理・活用する場の提供など研究プロセス全体を支える統合環境としての AI 統合型の研究データ基盤へのニーズが高まっている。

一方で、現実には AI 利活用には大きな障壁が存在し、GPU 環境やモデル運用に関する専門知識を持つ研究者・支援組織は限られており、誰もが利用できる共通的な AI 環境の整備が不可欠であるとの問題意識が共有されているところである。

また、研究の再現性の観点から、クローズドな AI サービスへの依存に対する懸念も示されている。研究者が利用するモデルや処理を把握・制御できる、再現可能で透明性のある AI 環境が求められている。

さらに、分野横断研究の進展に伴い、AI を活用して研究テーマや共同研究者を見つけるなど、知識の接続や研究コミュニティの形成を支援する機能へのニーズも顕在化している。

加えて、AI の利活用によって研究作業が効率化される中、その時間を新たな発見や創造的活動に振り向けることが期待されており、単なる効率化を超えて研究活動全体の質を高める基盤としての役割が求められている。

社会的観点からは、AI の利活用に伴う判断過程の透明性や信頼性の確保が重要な課題となっており、透明性・信頼性のあるオープンな学術特化型 AI モデルによる共通的な AI 利用環境を備えた知識基盤¹⁸が構築され、AI の利活用のコンサルテーションや情報共有の場が提供されることにより、AI for Science を加速していくことが不可欠であることも指摘されている。

(4) 研究分野からのニーズ

i) ライフサイエンス分野

¹⁸ AI を活用し、データの解釈、知識への汎化、知識の関係付け・体系化を自動化し、分野を横断する新たな知の創造を支援するもの

ライフサイエンス分野固有のデータ基盤の整備は引き続き重要である一方で、分野横断的な知識基盤の構築も不可欠である。このため、ライフサイエンス分野にとどまらず、あらゆる分野のデータを統合的に利活用できる共通基盤に対しては、国際的に通用する機能を備えた新たな共有認証機構の提供により、安全かつ円滑なデータアクセスを実現することが期待される。

加えて、分野間の壁を越えたデータの連携・検索機能を提供することで、異分野融合による新たな知見創出を促進するとともに、増大し続けるデータを安定的に蓄積・活用できる大規模ストレージの提供が求められる。これらを通じて、分野横断的なデータ利活用を支える高度な知識基盤の実現が期待される。

ii) マテリアル分野

データは蓄積されているものの利活用に十分つながっていないこと、AI 向けのデータセット提供への転換が遅れていること、ワーキングデータの構造化に大きな負担があること、さらには探索空間に対してデータ量が不足していることが課題となっている。これらに対し、データ提供から知識・AI モデル提供への転換、AI による自動構造化の導入、分野横断的データ連携と AI 基盤との統合を進めることが必要である。

iii) 大型実験施設

データが装置・施設・計算資源に分散していることにより利用が複雑化していること、データ量の急増に既存基盤が追従できていないこと、共同研究単位にとどまるデータ共有、さらには産業利用の増加に伴うセキュリティ要件への対応が課題となっている。これに対し、データを一元的に扱える環境の整備、データ圧縮・高速転送・解析の一体化、計算資源との近接化によるデータ移動の最小化、セキュリティと利活用を両立するデータ管理基盤の構築が求められる。

iv) その他

地球環境分野においては、DIAS¹⁹をはじめとするデータ基盤を通じ、研究者に加え地方公共団体や企業等の多様な主体によるデータ利活用が進展している。一方で、データ基盤ごとに地球観測データ等の提供形式、検索方法や API の整備状況が異なること、観測・モデル・計算資源がワンパッケージとなっていないこと、AI 基盤が十分でないことが、ユーザーが複数基盤を横断的にアクセスし活用する上での課題となっている。

また、情報の創出主体と利活用主体の間を仲介する科学・データサイエンス・AI を横断的に理解した人材が不足しており、体系的な活用が進みにくい状況にある。この

¹⁹ データ統合・解析システム(Data Integration and Analysis System)。地球観測データや気候予測データを蓄積・統合・解析・提供するプラットフォームとして運用されている

ため、自然言語を通じてデータを理解・分析可能とする分野特化型 AI モデルや、データ解析からレポート生成までを一体的に支援する AI エージェントの整備、さらにデータ公開・検索・解析・知識生成・意思決定支援を統合した総合的な AI 支援型基盤への高度化が求められるとともに、分野横断での連携・知見共有や人材育成を促進する仕組みの構築が重要である。

また、HPC コミュニティにおいても、計算性能の向上に対してデータ転送や連携処理が追従せず、ネットワークがボトルネックとなっているほか、複数の計算資源やデータ基盤を利用する際に認証が分断され、利用手続きが煩雑化していることが大きな課題となっている。また、学術機関外（産業界等）との連携においては認証制度の違いが障壁となっているほか、実験装置や IoT データ、外部データベースとの連携が不十分であり、リアルタイム処理や統合的なデータ利活用が困難な状況にあるため、単一 ID によるシームレスな資源利用を可能とする統合認証基盤の整備、外部 ID との連携を含む柔軟な認証体系の構築、実験装置・データ・計算資源を統合するオーケストレーション機能の実装が求められる。

6. AI for Science の推進に向けた基本的な戦略方針への対応

戦略方針においては、次世代情報基盤の構築として、計算資源の戦略的な増強、研究データ基盤の強化、情報流通基盤の強化が基本的な取組の考え方として整理されている。本 WG では、こうした方向性も踏まえつつ、特に AI for Science の推進に不可欠となる研究データの管理・利活用および流通に焦点を当てて議論を行ってきた。

研究大学等におけるコアファシリティや大規模集積拠点、大学共同利用機関、共同利用・共同研究拠点等に整備された最先端の研究設備・機器からは、今後も、高品質かつ大規模な研究データが継続的に創出される。これらに AI 利活用が加わることで、データ生成量およびそれに伴う通信量は、従来の想定を大きく上回る速度で増大していくことが見込まれる。

一方、これらの研究データは、全国に展開された SINET を通じて流通し、研究データ基盤に蓄積・管理されるとともに、計算資源へとシームレスに接続されることにより、分野横断的なデータ活用を可能とする。特に、AI for Science 時代にはこれまで高性能な計算機を利用してこなかった多くの研究者が、日々の研究活動の中で計算資源を用いて蓄積されたデータの高度な解析が容易にできる環境の整備も不可欠である。このような環境の実現は、各分野における研究サイクルそのものの高度化・加速化をもたらすものと期待される。

このように、情報流通基盤は研究活動を支える基盤であり、その役割は、停止が許されないこと、安全性が確保されていること、さらには大容量データを高速に伝送でき

ることといった要件を満たす「研究活動の動脈」として位置付けられる。このため、SINET の大容量化および高信頼化は最重要課題である。

さらに、研究者・研究組織と研究設備・機器、研究データ基盤、計算資源が相互に結節するためには、共通認証基盤の果たす役割は極めて重要である。このためには、単一 ID での利用を可能とし、かつ認証のセキュリティ強度を大幅に強化する共通認証基盤の高度化も、早急に取り組むべき重要な論点である。さらに重要なことは、高度化された共通認証基盤をより広い研究者層に提供できる基盤運用のための技術の開発が視野に入らなければならない。

加えて、データ利活用が進んでいるマテリアル分野、ライフサイエンス分野等におけるデータベースとの相互連携に関して、メタデータ付与やキュレーション体制の確立、幅広い分野の研究者に対するシームレスなデータと計算資源の活用基盤を含めて一体的に取り組み、AI 利活用による新たな知識創出環境(知識基盤)の構築など対応を進めていく必要がある。

戦略方針に規定された具体的なアクション項目を踏まえて、令和8年から令和12年の集中改革期間と呼応し、これらの課題に早急に対応していくことで、我が国の資産とリソースを最大限に活用し、勝ち筋となり得る分野等の研究力を世界トップ水準に引き上げることに貢献することを期待する。

7. 次世代情報基盤整備の方向性

(1)付加価値の最大化を実現するネットワークの整備

研究活動で生成・流通するデータ量の急増と高度化に対応するため、学術ネットワークには単なる通信手段を超えた価値が求められる。特に、AI の利活用の進展により、研究データの生成・分析・再利用が連続的かつ高度に行われるようになることで、通信量は加速度的に増大するとともに、高速通信や低遅延性に対する要求も飛躍的に高まるが見込まれる。こうした状況を踏まえ、大容量データを高速かつ安定的に伝送するだけでなく、データ流通の効率化、分散環境における共同研究の促進、リアルタイム処理の高度化といった付加価値を創出する基盤としての抜本的な機能強化が必要である。

このため、次期 SINET では、さらなる大容量化・低遅延化・低消費電力・高信頼化を図るとともに、AI 時代における爆発的な通信量増加にも耐え得るスケーラブルな構成への高度化を進める。また、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を拡充するとともに、ネットワーク上でのデータ流通の最適化やサービス高度化を通じて、研究データの迅速かつ安全な共有を実現する。

これにより、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基盤となり、研究成果の創出スピードと質の向上を図る。

戦略的に投資が進む APN や NTN など最先端の情報通信技術の動向も踏まえつつ、ユーザーニーズに最大限応えるアーキテクチャーとし、教育現場も含め、社会インフラとしての付加価値向上に取り組むこととする。

さらに、NII-SOCSについては、国立大学や大学共同利用機関にとどまらず公私立大学への展開を推進し、学術分野全体におけるサイバーセキュリティ対策の底上げを図る必要がある。この取組に併せて、各機関における通信の安全性確保の観点からセキュリティ機能の高度化を促進することで、認証基盤の強化・連携を図ることが可能となる。このように、ネットワーク全体の信頼性を一層向上させることが求められる。

(2) 研究プロセスを総合的に支える新たな研究データ基盤の構築

AI for Science の進展においては、研究データの質と量の確保に加え、その円滑な管理・利活用が不可欠である。さらに、今後、コアファシリティや大規模集積拠点等からは、AI 利活用が進み、高品質かつ大規模なデータが大量に創出される見込みである。世界各国で AI を中核とした研究の高度化・高速化が急速に進む中、我が国における研究環境を俯瞰し、国際競争に伍していくためにも、AI for Science に対応した研究データの管理・利活用を中核とする研究者の研究プロセスを総合的に支える新たな研究データ基盤によってデータ空間を確立することが求められる。

具体的には、各分野のデータベースとの連携強化や共有・利活用されやすくなるためのメタデータの整備により、分野横断的なデータ活用を可能とするとともに、データの蓄積・管理・共有・再利用を一体的に支援する環境や仕組みを構築する。また、メタデータ自動付与等、AI をデータの管理・共有・再利用の仕組み等自体に活用することが重要となるため、FAIR 原則に則っており、かつ AI による解析・再利用を前提としたデータ整備を推進し、研究データの価値を最大化する環境を整える。

先導する分野におけるデータベースについては、各分野の特性を踏まえつつ、分野内での統合基盤の高度化、データ品質の確保、分野内の体系的整理を担う情報処理機能の強化及び AI 利活用を前提とした形への整備のみならず、あらゆる分野との連携が重要であるため、我が国の中核的な研究データ基盤の立場から各分野におけるデータベースのハブとして連携強化の中心となって分野を牽引する取組を支援することも必要である。

これにより、研究データの利活用を一層活性化し、分野融合や新たな研究領域の創出を促進するとともに、研究の再現性・透明性の向上にも資することが期待される。

さらに、AI for Science を組織内に浸透するためには、研究者のみならず、図書館、情報基盤・メディアセンター、研究推進部、URA 等の研究支援・マネジメント組織、そして経営層を巻き込んだタテの糸を強靱にした環境を構築することは重要な論点で

ある。加えて、ヨコの糸を太くしていく視点に立てば、地域間の協働体制を強化することも重要であり、地域コンソーシアムの質的拡大を通じて、大学等研究機関の成長を支える全国規模のコミュニティへと発展していく必要があるため、機関内の連携のみならず、ノウハウ共有や人材育成支援、キュレーション機能の確立等組織間の協働を加速することが求められる。

また、AI 学習等に不可欠な学習データの枯渇に関しては、AI 基盤モデルの構築に当たって大きな論点となる。このため、AI 利活用を前提とした形式やデジタル化に至らず可視化されていないようなダークデータや、実験における失敗等のネガティブデータをいかに活用できるかが科学研究力の向上に直結すると考えられる。さらに、大学等研究機関が所蔵する図書資料等については、本文データのデジタル化・公開が十分ではなく、AI 活用も見据えた機械可読なデータ整備の推進を通じた有効活用にも対応が必要である。

あわせて、医療・個人情報・産業データなど、外部公開が困難なデータの利活用も引き続き重要な論点であり、機微データを安全に扱うためのアクセス制御や管理などオープン・アンド・クローズ戦略を意識した基盤として成長するためのセキュリティ体制強化についても留意しなければならない。

加えて、AI for Science による研究活動・研究成果の最大化を図るには、単にデータを蓄積・共有するだけでなく、それらを構造化・意味付けし、信頼性の高い体系的な知識を知識グラフ²⁰として構築することが重要である。知識グラフや機械学習モデル等を活用し、分野横断的な知識の発見や仮説生成を支援する環境を構築することで、AI によるデータ解析、知識抽出、関係性の把握・可視化等が可能となる。知識グラフと信頼性・透明性のある AI モデルが一体となった知識基盤が構築されることによって、データの創出から保存・管理、利活用(計算・解析等)に至る様々なインターフェースにおける一体的な連携を実現させることにより、AI for Science の実効性を高め、研究プロセスの高度化・効率化を図るとともに、分野の枠を越えた新たな科学的知見の創出を促進することが期待される。

これらの取組を一体的に推進することは、AI がもたらす新たな研究プロセスへの転換に対応することであり、我が国の科学研究の発展を支える鍵となる。AI 及び研究データを中心とした科学研究の変革が目前に迫る中、AI を活用して新たな知識を創出し、その知識を研究データやこれまでの知見と結び付けながら、次の知識創出へと連続的に発展させていくことが不可欠である。このため、こうした取組を支援する大学等研究機関における人材の育成・確保についても、早急に取り組むべき重要な課題である。

²⁰ 学術における「知識グラフ」とは、論文・著者・トピックスなどの関係を構造的に表現したもので、学術知識の表現の一つ

(3) 計算資源との一体的な接続・連携、認証強化

AI を活用した高度な研究を推進するためには、情報基盤を一体的に活用できる環境の構築が不可欠である。このため、計算資源と情報流通基盤、研究データ基盤との連携を強化し、日々の研究プロセスの中で、研究者が意識することなくシームレスに利用できる環境の実現と、その成功・失敗の過程を記録して研究の来歴管理だけでなく AI で研究を加速する基盤を構築する必要がある。

あわせて、これらの多様な資源を安全かつ円滑に利用するためには、認証基盤の高度化が重要となる。共通認証基盤の構築による研究者・研究機関、研究設備・機器、データ、計算資源を信頼性高く結節するパイプラインの整備を進め、利便性とセキュリティを両立した利用環境を確立する。さらに、認証基盤の高度化は、国内のみならず、EOSC 等の海外プラットフォームとの連携にも大きく寄与する。AI for Science の推進に当たって、世界の科学研究にとって不可欠な高品質なデータ、高度な研究基盤、革新的な研究プロセスを提供する立場を確立するためにも、経済安全保障にも留意した上で、情報基盤を国際的な研究ネットワークやデータ基盤と接続可能な形で整備する必要があることにも留意が必要である。

8. おわりに

本報告書では、AI for Science の進展を背景とした研究環境の変革を踏まえ、研究データの管理・利活用および流通を支える次世代情報基盤の在り方について整理するとともに、今後重点的に取り組むべき方向性を示した。

AI の利活用の急速な進展により、研究データの生成・流通・活用はこれまでにない規模と速度で拡大しつつあり、研究活動そのものの在り方も大きく変化している。このような状況においては、情報流通基盤、研究データ基盤、計算資源、共通認証基盤といった各要素を個別に整備するだけでなく、それらを連携させ、一体的に機能させることが不可欠である。

特に、情報流通基盤については、研究活動を支える基盤インフラとして、その大容量化・高信頼化に加え、AI 時代における加速度的な通信量増加への対応や、安全性の確保が求められている。また、研究データについては、分野横断的な利活用を可能とする基盤の構築とともに、AI 利活用による高度な分析・知識創出を支える研究者の研究プロセス全体を総合的に支える新たな研究データ基盤によって、研究データ空間を確立することが必要である。

さらに、これらの基盤を安全かつ円滑に利用するためには、単一 ID での利用やセキュリティ強度の強化等、共通認証基盤の高度化及び安定した運用を通じた信頼性の確保が不可欠である。

これらの取組を通じて、研究資源の統合的な活用が実現されることにより、研究プロセスの高度化・効率化が進展し、分野の枠を越えた新たな知の創出が加速することが期待される。ひいては、我が国の研究力の強化及び国際競争力の向上にも大きく寄与するものである。

AI for Science はこれまでにない速度で進展していくことが想定される。その影響は研究者にとどまらず、研究支援を担うセクターをはじめ、様々なセクターに及ぶものであり、それぞれの立場において急速な変化を常に意識した上で対応する必要があることから、多様な人材が一体となって AI for Science を支えることが重要である。また、この変革を次世代の若者へと継承していく観点からも、持続的な支援環境を構築する必要がある。そのため、これらの支援環境構築に当たっては、様々な課題に対して柔軟な対応が求められることから、基盤の利用環境の提供・利用支援、コンサルティング能力等を含め、組織の体制強化、人材の育成・確保、そして、これらに必要な財源の確保についても、中長期的な観点から検討すべき課題として認識する必要がある。

今後は、本報告書で示した方向性に基づき、関係組織・ステークホルダーが連携・協働しつつ、次世代情報基盤の整備を着実に推進していくことが求められる。その際には、技術の進展や研究現場のニーズを的確に捉え、柔軟かつ継続的に、基盤整備をはじめとする取組の強化・高度化を図っていくことが重要である。

(参考資料)

科学技術・学術審議会 情報委員会における下部組織の設置について

令和 7 年 10 月 6 日
科学技術・学術審議会
情 報 委 員 会

情報委員会運営規則第2条の規定に基づき、科学技術・学術審議会 情報委員会に以下の下部組織を設置する。

名称	調査審議事項
AI for Science を支える 研究データの管理・利活用 と流通の在り方ワーキング グループ	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に係る事項について

以上

科学技術・学術審議会 情報委員会
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
ワーキンググループ
委員名簿

(50音順)

臨時委員

◎ 尾上 孝雄 大阪大学理事・副学長

専門委員

石田 栄美	九州大学データ駆動イノベーション推進本部教授
○ 江村 克己	福島国際研究教育機構理事
工藤 郁子	大阪大学社会技術共創研究センター特任准教授
千葉 滋	東京大学情報基盤センター長
林 理恵	NTT株式会社ネットワークサービスシステム研究所主幹 研究員
宮田 なつき	産業技術総合研究所人工知能研究センター研究チーム 長
矢守 恭子	朝日大学情報教育研究センター長
吉田 亮	情報・システム研究機構統計数理研究所副所長
若目田 光生	株式会社日本総合研究所創発戦略センターシニアスペ シャリスト

◎:主査、○主査代理

(令和6年6月16日現在)

※上記の他、

情報・システム研究機構 国立情報学研究所 がオブザーバーとして参加

科学技術・学術審議会 情報委員会
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
ワーキンググループ
審議経過

※科学技術・学術審議会 情報委員会(第44回:令和7年10月6日)において、本ワーキンググループの設置を決定

第1回 R7.12.24(水)

○ワーキンググループの運営規則等について審議

○AI for Science を支える研究データの管理・利活用及び流通の在り方について
議論

-現状・背景等について意見交換

-学術研究プラットフォームの現状と今後について意見交換

第2回 R8.2.13(金)

○AI for Science を支える研究データの管理・利活用及び流通の在り方について
議論

-SINET の現状と今後およびセキュリティについて意見交換

-「AI for Science の推進に向けた基本的な戦略方針について」の検討状況について意見交換

第3回 R8.3.26(木)

○AI for Science を支える研究データの管理・利活用及び流通の在り方について
議論

-「AI for Science の推進に向けた基本的な戦略方針について」素案、AI for Science における研究データの取り扱いに関する考え方について意見交換

-研究データ基盤の展開について意見交換

第4回 R8.4.24(金)

○AI for Science を支える研究データの管理・利活用及び流通の在り方ワーキンググループとりまとめに向けた整理について議論

○AI for Science を支える研究データの管理・利活用及び流通の在り方について
議論

-知識基盤の構築について意見交換

第5回 R8.5.22(金)

○AI for Science を支える研究データの管理・利活用及び流通の在り方について
議論

-AI for Science を支える研究データの管理・利活用(主要分野の状況)について
意見交換

<ゲストスピーカー>

- ・五斗 進 情報システム研究機構 国立遺伝学研究所 バイオデータ研究拠点
(BSI)副拠点長
- ・出村 雅彦 国立研究開発法人物質材料研究機構 技術開発・共用部門長
- ・初井 宇記 国立研究開発法人理化学研究所 放射光科学研究センター
制御情報・データ創出基盤グループ グループディレクター

○AI for Science を支える研究データの管理・利活用及び流通の在り方ワーキング
グループとりまとめに向けた整理について審議

第6回 R8.6.16(木)

○AI for Science を支える研究データの管理・利活用及び流通の在り方(審議まと
め)案について審議

(審議まとめ)概要

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