

AI for Science 時代に対応した HPCI 全体でのを中心とした我が国の計算資源を
持続的に維持・発展できるエコシステム構築に向けた戦略的な整備や制度の共通化に向けた方針(案)

令和 8 年 9 月
HPCI 計画推進委員会

1. 背景

AI 技術、特に大規模基盤モデル、生成 AI、AI エージェント、科学基盤モデル等の急速な発展により、科学研究自体が大きく変化しつつある。材料科学、生命科学、気象・気候、創薬、ロボティクス、宇宙、ものづくり等の幅広い分野において、データ、モデル、シミュレーション、実験・観測装置、フィールド、エッジデバイス、自動化システム等を統合的に活用する「AI for Science」の重要性が急速に高まっており、従来の計算科学・計算工学の研究者のみならず、幅広い分野のドメイン研究者がシミュレーション、生成 AI、推論サービス等を統合的に利活用して研究を推進する状況が広がりつつある。

こうした中で、AI for Science を支える共用計算資源についても、例えば、従来型の利用者が個々の計算機にコードやデータを持ち込むバッチジョブ中心の利用形態にとどまらず、Web UI、API、コンテナ、データ共有、基盤モデル活用、AI エージェントからの利用、外部サービスとの連携を利用環境整備の前提とする等といった、よりサービス指向の計算基盤へと発展させる必要性が求められており、特に生成 AI や科学基盤モデルの活用においては、従来型のバッチ処理だけではなく、推論サービス、対話型 AI サービス、研究ログや電子実験ノートの解析、AI エージェントによる実験・ワークフロー支援等、常時稼働・長期稼働するサービス型の利用が増加すると考えられる。

また、HPCI はこれまで、「富岳」をはじめとするフラッグシップシステムと、大学等の情報基盤センターが整備・運用する資源が相互に補完しながら、我が国の計算科学・計算工学を支えてきた。今後も、全国の研究者が高度な計算資源を利用できる環境を維持・発展させることは重要である。その一方で、GPU 等の AI 計算資源の高額化、電力・冷却・設置スペースの制約、ベンダーとの交渉、運用人材の不足、ソフトウェア・サービス開発体制等を踏まえると、各機関が個別に計算資源を整備・運用する従来型の延長のみでは、十分な規模・機能・利便性を確保することが困難ともなりつつある。特に、AI 計算資源が一部の研究者のためのみではなく幅広い研究者が継続的に利用する研究インフラとなることを踏まえ、米国 Genesis Mission や欧州 AI Giga Factories をはじめとして、整備費及び運営費の全額を公的機関が措置し研究者が無償利用する従来型ではなく、民間投資を組み合わせた大規模投資と研究費等による利用料負担を組み合わせた、新たな整備・運営の考え方が示されてきている。

このため、これらのことから、HPCI を中心とした我が国の計算資源を持続的に維持・発展させ、AI for Science 時代にふさわしい次世代の共用計算基盤を構築するため、HPCI についても、AI for Science 時代にふさわしい次世代の共用計算基盤として、その整備・運用・利用制度の在り方を見直す必要がある。

2. HPCIに関するこれまでの課題

現行の HPCI は、主として利用者が個別の資源提供機関にアカウントを取得し、SSH 等によりログインし、バッチジョブを投入する利用形態を前提として発展してきた。しかし、AI for Science 時代には、Web ブラウザ上の統合環境、API を通じた利用、AI エージェントによるワークフロー自動生成、複数の計算資源・データ基盤・EPOCH 等により整備される共用研究設備や実験・観測装置の組合せ利用、コンテナ化されたアプリケーションやモデルの配布・実行等が拡大すると考えられるが、長期稼働するサービス、外部システムとの連携、対話型 AI 推論、AI エージェントによる継続的なワークフロー実行等を、利用者やプロジェクト単位で柔軟に構築・運用する仕組みは十分ではない。また HPCI 全体としても、各資源の利用状況、ジョブ情報、キュー状態等を統合的に把握し、最適な資源へ適切なロードバランシングをする機能は存在せず、HPCI は複数機関の計算資源を接続する仕組みとしては機能してきた一方で、利用者から見て資源を最適に利用できる環境には至っていない。

また、AI 計算資源は大規模化・高額化しており、個別機関ごとの調達では、価格、納期、保守、運用人材、電力・冷却設備等の面で限界も指摘されており、加えて、産業利用、政策対応、国際連携含めたプロジェクト利用等についても、AI for Science 時代の多様な利用形態に対応できるよう、より柔軟な仕組みが求められる。複数の計算資源、データ基盤、Web サービス、API 等を横断して利用する場合、利用者が資源提供機関ごとに個別にアカウントを取得し、利用申請やアクセス権付与を受ける仕組みでは、利用開始までの手続や運用負担が大きくなる。また、利用者が意識しない形で複数の資源を横断的に利用し、実際の利用実績に応じて事後的に精算するような利用形態も想定されるが、このため、利用実績の可視化、共通ポイント・クレジット、後払い型の精算、政策的な資源配分等を可能とする制度が求められる。

加えて、GPU や NPU 等の AI 計算資源は技術進歩が速く、数年程度のサイクルで大規模な更新投資を行うことが必要となっていく一方、現行の国立大学等における計算基盤については、初期システムの整備費、運用費、さらに次期システムへの更新の原資等を一体として確保する仕組みとはなっておらず、AI 計算資源の急速な技術進歩に対応した機動的な設備更新や需要の増大に応じた継続的な規模拡大に対応する、官民による投資、運営費及び次期システムへの再投資までを含めた利用料負担、持続的に計算資源を維持・更新できる仕組みの構築も課題となっている。

さらに、今後の HPCI には、単なる計算機管理にとどまらず、クラウド的なサービス運用、推論サービス運用、セキュリティ監視、経済安全保障対応等、多様な専門性が必要となり、これらを大学教員や研究者の兼務に依存して維持することには限界があり、専門職員、リサーチエンジニア含めた、人材の確保・育成・キャリアパス形成も重要な課題である。

3. 次世代 HPCI 環境の基本的方向性の考え方

HPCI は、我が国の計算科学・計算工学を支える中核的な研究基盤として、これまで大きな役割を果たしてきた。今後も、全国の研究者が高度な計算資源を利用できる環境を維持・発展させることは極めて重要である一方で、とりわけ AI for Science の急速な進展により、計算資源に求められる機能、規模、スピード、利用形態は大きく変化している。今後の HPCI は、単なる高性能計算機の共同利用にとどまらず、EPOCH 等により大学等に整備される共用研究設備とも接続し、計算資源と研究データを一体的に利活用できる次世代の基盤としてさらに発展していく必要がある。その際には、米国 DOE、欧州等において進む AI for Science 基盤の国際的な潮流との整合性を確保し、日本独自の閉じた仕組みに陥らないよう留意する必要がある。

文部科学省として、関係機関と連携しながら、AI for Science 時代にふさわしい次世代 HPCI 環境をさらに発展させていくための検討を、以下の方向で進めていく。その際には、公的研究費等による継続的な研究需要を基礎として民間投資を誘発しつつ、官民が適切に役割を分担しながら、計算資源を持続的に維持・更新・拡充できる「成長可能なエコシステム」を構築することを基本的な方向性とする。

(1) AI for Science 時代の研究ワークフローを支える

次世代 HPCI は、単に高性能計算機を共同利用する仕組みではなく、エッジ・クラウド・HPC の連続体も意識しつつ、データ、モデル、ソフトウェア、計算資源、EPOCH 等により大学等に整備される共用研究設備を含む実験・観測装置、センサー、ロボット、AI エージェント等を結びつける研究ワークフロー基盤として再設計する。従来型のバッチジョブ利用を引き続き支えつつ、Web UI、API、コンテナ、ワークフロー、データフロー、データ・モデル共有、リアルタイム・対話的利用、AI エージェント利用等に対応したサービス基盤を整備する。

材料科学、ライフサイエンス・医療、環境・気候、ものづくり等の分野ごとに、データ、モデル、解析環境、推論サービス、可視化環境等を備えた分野別プラットフォームを、共通インフラ上に必要に応じて切り出して構築できるようにすることが重要である。

その際に、HPCI や JHPCN がこれまで担ってきた全国共同利用の理念は、AI for Science 時代においても引き続き重要である。高度な AI 計算資源、科学データ、学習済みモデル等について、所属機関や地域にかかわらず、必要な計算資源、データ、モデル、ソフトウェアにアクセスできる環境を整備する。その際、各拠点に同種の資源を個別に薄く広く整備することのみを目指すのではなく、利用者から見て統合的に利用可能なサービス基盤を構築することが重要である。

(2) 共通要件に基づく戦略的整備の推進

次世代 HPCI に参加する資源提供機関については、以下のような共通要件への対応を求めることを検討する。

学認等を活用した認証連携、認可・利用者管理
プロジェクト単位のアクセス権管理、環境隔離

コンテナ実行環境、コンテナ管理・統合環境

Web UI・ポータル機能

API 利用、エージェント利活用環境への対応、常駐サービス運用

データ・モデル共有基盤との連携

EPOCH 等により整備される共用研究設備を含む実験・観測装置、センサー、ロボット、

エッジデバイス等との連携

共用ストレージおよび外部データ基盤との連携

要求に応じて適切な資源へ割当てを行うメタスケジューラ、資源ブローカ等の機能

国際的に広く利用される技術への対応

利用実績の把握・可視化

課金・ポイント・クレジット等の統合的な制度改善

セキュリティ、ログ管理、脆弱性対応

ユーザー支援、AI ヘルプデスク

これらの要件に対応できる機関を次世代 HPCI に位置づけることにより、利用者にとって一貫性のあるサービスを提供する。

(3) HPCI 計算資源整備の方向性

国全体として、AI for Science 時代にふさわしい一体的・共通的な計算基盤へ移行していくためには、支援を重点的に行う拠点数を無秩序に増やすのではなく、整備すべき規模・機能、効率的な運用、効果的な調達等の全体的な対応の方向性を総合的に勘案し、我が国の計算基盤の中核的な拠点群が連携して機能する方向で進める。その際には各機関が個別の時期に個別仕様で調達するのではなく、中核的な拠点群の中で調達時期、仕様、運用方針を可能な限り調整し継続的に資源を増強・更新していくことを検討する。

その際には、複数のモデルを組み合わせることで検討することとし、具体的には、国として中核的な大規模 AI 計算資源を整備するナショナルセンター型、複数大学・情報基盤センターが共同調達・共同運用を行う大学連合型、mdx、NII Research Data Cloud 等の学術情報基盤と近接して整備するモデル、国内民間データセンター・クラウド事業者と緊密に連携するモデル、米国・欧州等の大規模計算資源との国際連携モデル等が考えられるが、これらは相互排他的ではなく、研究分野、利用形態、データ管理要件、セキュリティ要件、整備スピード、コスト等に応じて組み合わせることが重要である。

その際、特に、大規模な AI 計算資源については、国費による整備のみを前提とせず、国と民間が共同して初期投資を行い、その後のシステム更新については、利用料収入や民間投資を主体としつつ、必要に応じて公的支援を組み合わせるなど、官民の資金を活用した持続的な整備方式を状況を見つつ段階的に取り入れる。~~次世代 HPCI においては、物理的にどこに計算資源を設置するかと、誰がサービス設計・運用責任・利用者支援を担うかを分けて考える必要があり、例えばサービス責任を持つ機関と、物理的な設置場所は別にすることも考えられる。~~また、複数大学が共同で運用責任を持ち、NII 等の学術情報基盤と連携しながら、物理資源は複数拠点又は民間施設に分散することもあり得る。

／以上