

ライフサイエンス分野のデータ基盤 を巡る国内外の動向

2026年3月24日

研究振興局ライフサイエンス課

AI for Science の推進に向けた基本的な戦略方針（概要イメージ）

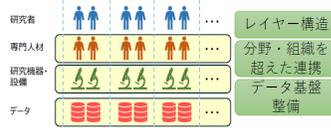
第7期科学技術・イノベーション基本計画の方向性

➤ 科学技術推進システムを刷新し、科学技術政策を大転換

・現状認識の一つとして… **A I と科学の融合**
による研究開発パラダイムの転換

2章 知の基盤としての「科学の再興」

4.AI for Scienceによる科学研究の革新



AI基本計画（R7.12閣議決定）

➤イノベーション促進とリスク対応の両立

➤信頼できるA I の追求

➤世界で最もA I を開発・活用しやすい国を目指す

海外動向

- ・各国は**AI for Science** を国家的ミッションと位置づけ
- ・研究投資、計算基盤整備、人材育成等を強化
- ・科学とビジネスが近接化し「勝者総取り」構造
- ・科学的発見のプロセス自体が、学際的、データ駆動型、計算集約型へと大きく変化

日本の強み

- **情報基盤**：世界最高水準の情報流通基盤（SINET）・研究データ基盤（NII RDC）・計算基盤（富岳・富岳NEXT・HPCI等）
- **研究基盤**：世界トップレベルの基礎科学力と多様な研究者層、世界最先端の研究装置群と大型研究施設、信頼性の高い実験・観測データの蓄積
- **社会基盤**：世界有数の経済規模、精密な製造・計測技術・ロボティクス、すり合わせや暗黙知を含む現場知、AIに対する社会的・産業的な需要

目的 I . 科学研究の革新と科学的発見の加速・質の変革、II . 研究力の抜本的強化と科学の再興、III . 国際的優位性・戦略的自律性の確保

中長期的な取組目標 **科学基盤モデル/エージェントやAI駆動ラボの活用により重要技術領域の先端的成果創出及び研究開発期間を1/10に**

今後5年間の目標 **AI for Scienceの推進により、日本の科学研究における国際優位性の確保**

（ターゲット例）

- **3年後までに、新素材開発速度10倍の潜在力を有するAI駆動ラボシステムを開発。**
将来は、AI駆動ラボシステムを用いて、我が国企業が国際的サプライチェーン上不可欠なマテリアルを量産する。
- **3年後までに、大規模なデータ取得を通じて、高機能なバイオ製品の高効率設計を実現するバイオ生成基盤モデルを開発。**
将来は、仮想細胞・生体モデルや、植物、動物、ヒト・臓器等の“デジタルツインモデル”を実現し、高精度かつ高効率なバイオ製品開発や創薬等に貢献する。
- **3年後までに、AIEージェント群による、最先端大型研究施設・研究装置からの大量・高品質データ産出や、仮説検証・実験の自動化・自律化を実現。**
新規性の高い研究を探索的に行うシステムの開発を通じて、科学研究の新しい方法論を示す。

戦略的な国際連携
(米国・英国など)

世界を先導する
科学研究成果の創出

AI for Science の波及・振興
による科学研究力の底上げ

AI for Science を支える
研究基盤の構築

(具体的な取組内容)

① **研究力・人材**
AI研究者等の育成
×
AI利活用の促進

② **計算資源**
戦略的増強
×
利便性向上

③ **研究データ**
高品質データの創出
×
データの一元化

- ・ AI for Science のあらゆる分野での波及・振興と日本の強みを生かした重点領域の設定・投資を両輪で推進、世界トップ層との戦略的国際共同研究を推進
- ・ AIの基礎研究含むAIそのものの研究の強化（リスク対応含む）
- ・ 国際連携・産学連携を通じ、AI・計算資源・データに精通した人材の参画・育成、技術専門職の育成・確保、評価や処遇の見直し
- ・ 世界最高水準の次世代AI・HPC融合プラットフォーム「富岳NEXT」の開発
- ・ AI共用計算資源の戦略的な増強と利便性（機動性、アクセシビリティ、相互運用性）の向上
- ・ 産業界との連携及び国際連携を通じた計算資源の有効活用
- ・ 戦略的価値の高いデータセットの特定・構築
- ・ 自動化・自律化した研究設備等の整備と研究データ創出プロセスの標準化
- ・ AI時代に即した次世代情報基盤の構築・活用、データの一元的運用

AI for Scienceにおける研究データの取扱い等に関する考え方（案）

【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）】

- 基本的な考え方として、**オープン・アンド・クローズ戦略の下で、研究データの管理・利活用を推進**する。
- また、研究データは、**国の安全保障を確保し、我が国の産業競争力や科学技術・学術上の優位性を確保するために重要な情報**を含む可能性があるため、以下の観点から留意すべき研究データは**非公開**とすることが求められる。
 - ✓ 輸出管理や個人情報保護等に関する国内関係法令やガイドライン等で取扱いに制限あるもの
 - ✓ 企業の秘密性、研究の新規性、研究セキュリティ等の観点から非公開とすべきもの

■ AI for Scienceにおいて留意すべき事項

- 加えて、AI for Scienceの推進においては、**日本のもつ研究データがAIの活用等により意図せず流出し、日本のもつ優位性が損なわれる恐れ**があることを踏まえ、AI技術の進展や、研究分野・研究データの特長・状況等に留意しつつ、日本の強みとなるAIモデルや研究データ等については、**国益保護の観点から適切に管理**する必要。

■ AI for Scienceにおける研究データの管理に関する具体策（案）

- **日本の強みを有する研究データを適切に管理**するため、研究データの国外移転の可否、学習利用の可否、物理的なサーバーの場所等について、国として指針を明らかにした上で、以下の方策を講じる。
 - ① 国が実施する「**AI for Science による科学研究革新プログラム**」等については、その実施手順※において、上記指針も含め、**データマネジメント、デュー・ディリジェンスに係る対応を徹底**。政府において研究機関のリスク軽減措置の内容を確認し、**必要に応じて追加措置を要請する仕組み**を検討する。
 - ② 今後、国内で創出される研究データについて、当該データの流出等により日本の優位性が損なわれることがないよう、**国全体として、セキュアかつアクセス制御可能な研究データ基盤を構築**した上で、**保存・管理**されることを目指す。

※その際、**研究セキュリティの確保に関する取組のための手順書も参照**（令和7年12月研究セキュリティと研究インテグリティの確保に関する有識者会議）

- 関係機関において、以下の【**基本的な枠組み**】を通じて、**研究セキュリティ確保に関する取組を実施**。

【基本的な枠組み】

政府/資金配分機関	研究機関の リスクマネジメント結果を確認 し、十分か判断、必要に応じて追加のリスク軽減 措置を要請 。
研究機関	研究参画者の リスク評価を行う とともに、研究に関するデータマネジメントプランや、研究機関その他個人・機関との間で締結する共同研究契約その他 契約・協定について 、協力の内容、研究データ等へのアクセス、発明・特許等の知的財産の取扱い、守秘義務の 内容の適切性について確認 。
研究者	自身の研究活動に関する情報の申告、 データマネジメントプランの作成 、研究データ・情報システム等の 適切な管理

※技術流出防止に努めることは重要である一方で、**リスクゼロを求めすぎると国際共同研究等に抑制的になるなど研究現場に悪影響が生じる恐れ**があることを踏まえ、**リスクの程度に応じた合理的な対処**を求める。

※データの取扱いについては、内閣府等における検討とも整合性をとりながら取組を進める。

■ 欧州分子生物研究所(EMBL)の Science AI Strategy (2025年2月)の 3 つの柱

● AI手法と理論

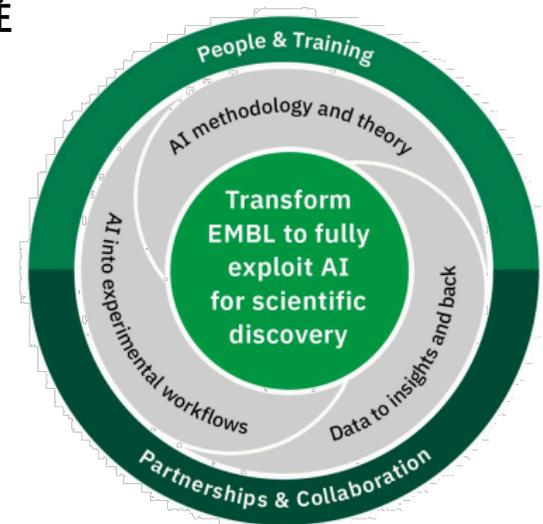
- ✓ 生物学的ドメインに特化した基礎モデル(foundation models)
- ✓ 多階層モデリング、因果推論(Causal representation learning)、AutoML、ベンチマーク指標、オープンサイエンス推進

● データから洞察へ、そしてその逆も (Data to insights and back)

- ✓ EMBLが保有する **大規模バイオデータをAI訓練・評価用に整備**
- ✓ モデル・リポジトリやAI予測の成果の共有基盤の整備

● 実験ワークフローにおけるAI活用 (AI for experimental workflows)

- ✓ AIによるサンプル準備、自動化された実験デザイン、フィードバック駆動型データ取得、仮説生成へのAI活用
- ✓ 高速処理やバリデーション統合による実験スループット向上



https://www.emblaustralia.org/wp-content/uploads/2025/05/EMBL_AI-Strategy_Feb2025_Accessible.pdf

■ 欧州バイオインフォマティクス研究所(EBI)

<https://www.ebi.ac.uk/>

- EMBLの6拠点の1つとして英国ケンブリッジに位置する約850人超の職員数から成る研究所。
- 欧州のライフサイエンス研究リソース統合組織であるELIXIRと協働。
- EMBLのデータサイエンスセンター(DSC)が、EMBL内の横断的なデータ品質管理・構造化・標準化の機能を担い、EBIはそのデータの活用のための標準ツール開発やAI用データセットの提供、AIモデル共有インフラの運用等を担う
- AIモデル開発に伴う計算需要を満たすため、民間クラウドやGPUメーカー等との連携のほか、**欧州のAI FactoriesやGigafactories等の計算資源とも連携しながら、簡素化されたユーザーフレンドリーなインターフェースの開発**や生命科学特有のデータ処理環境の整備等も担っている。

■ EBI - BioAlrepo

<https://www.ebi.ac.uk/biostudies/BioAlrepo>

- BioStudiesデータベースの中に設けられた**AIモデルのレポジトリ**の試行版。
- 開発した**AIモデルをオープンに共有**し、他者の再利用を可能にしている。これには**メタデータ、モデルの重み(計算パラメータ)、その他のデータファイル、更にモデル構築、テスト、検証に使用されるデータセットへのリンク**が含まれる。
- 初回リリースには、顕微鏡、スプライシング予測、タンパク質構造決定、オミクス解析等の分野のモデルセットが含まれている。
- 相互運用性や効率性の確保のため、**DOME (Data, Optimization, Model and Evaluation)ガイドライン**等を標準として導入。



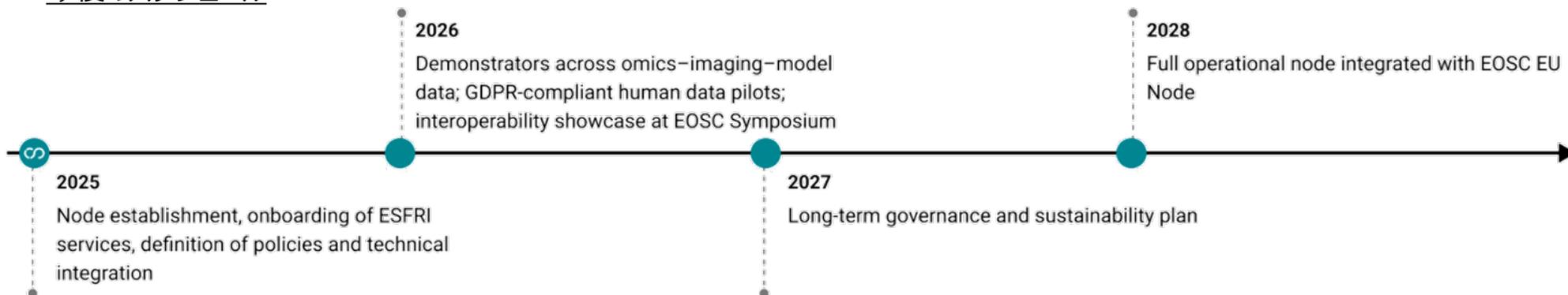
主な目的： EOSC内のバイオデータの専門知識を強化し、ヨーロッパ全域で生命科学のデータ、サービス、計算資源への相互運用可能で公正かつ安全かつ倫理的に準拠したアクセスを確保すること。

対象分野： ゲノミクス、プロテオミクス、メタボロミクス、イメージング、神経科学、臨床研究、システム生物学、バイオインフォマティクス等の生命科学分野

ユースケース： 「科学分野にわたるFAIR画像解析」、「生物多様性観測の調和」、「構造生物学データの連合データ保存」等

体制： ELIXIR、EMBL、Euro-BioImaging、Instruct-ERICの4組織が提携

今後のスケジュール



(出典) <https://elixir-europe.org/news/EOSC-LSC-MoU>
<https://eosc.eu/building-the-eosc-federation/eosc-node-life-sciences-connect>

5 BER Resources Collaborating to Prototype a Unified Data Access Layer



Unifying Access Layer Components
BERtron - Global Search supported by common APIs - **find and reuse** data
Data Transfer Service - **maintain provenance, propagate credit**



Sample metadata,
standardized data
products

Biogeochemical
measurements,
sensor data

Proteomics,
metabolomics, imaging

(meta)genomics,
-transcriptomics,
metadata

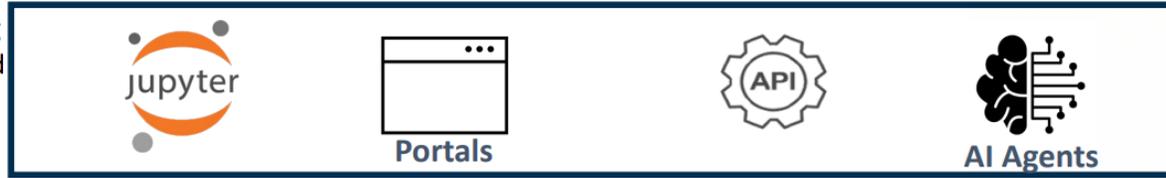


The BER Data Lakehouse Architecture to Enable Humans and AI



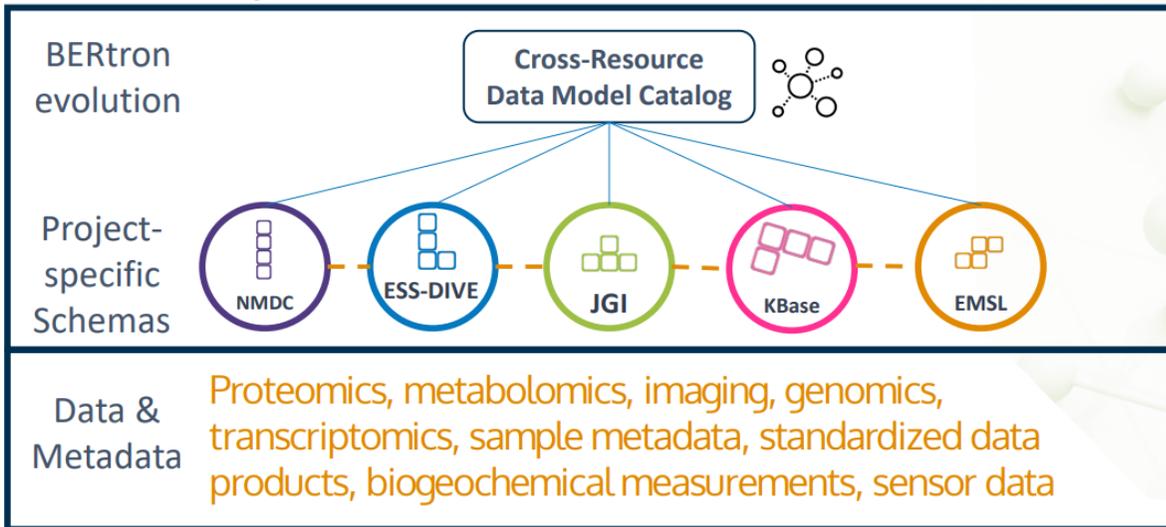
User Interfaces for Data Science and Analytics

New and existing tools can be used to access BER data

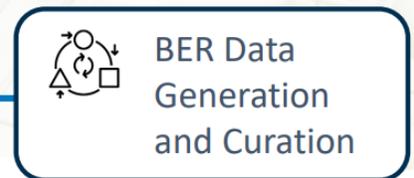


Metadata and Governance Layer

Users can access data using either the unified catalog or project-specific schemas



Queryable information resides in an object store



AI活用の促進

■STRIDESイニシアティブ

(Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability)



商用クラウドサービスの利用に伴う経済的・手続き的な障壁を減らし、生命医学研究を加速させることを目的に、NIH研究者・NIH支援研究者に、**主要商用クラウドの割引料金**のほか、専門家コンサルティングや、**高性能計算資源(HPC)**や**大規模データストレージへのアクセス**を提供

■NIH Cloud Labによる実践トレーニング

研究者のスキル向上支援のため、研究者が隔離された安全な環境で、AIやバイオインフォマティクスのチュートリアルを学習したり、概念実証の構築、コスト見積もり等を行うことができる**最大\$500のクレジット(最大90日間)を無償提供**(これまでの利用者は1,000人以上)

■安全なコード管理 NIH GitHub Resource Center

NIH所内 (Intramural) の職員 (研究者・開発者等) のみを対象に、研究者や開発者がコードを安全に保存・管理できるよう、高度なデータ損失防止機能やアクセス制御機能など、特別なセキュリティ機能を整備。また、個人のGitHubアカウントでは得られない多くのNIH特有の機能やサポートも提供。

■AI・クラウド環境の活用事例

- ・TOPMed (Trans-Omics for Precision Medicine)
STRIDESを通じた大規模な全ゲノムシーケンスデータのクラウドへの移行により大人数での同時アクセスを可能にし、共同研究を加速
- ・Advanced Imaging and Microscopy (AIM) Resource
ローカル環境のインフラ容量の制約を、クラウドの拡張性で補うことで、データ分析の高速化や複数の並行研究が可能に。

■AI活用促進の取組

- ・AIM-AHEAD 電子カルテを含む大規模データを基盤にAI/MLの研究・基盤整備・人材育成を推進し、健康格差の是正と医療革新を目指す全米規模のイニシアチブ
- ・Bridge2AI AIに適した高品質なデータの整備や、データの使い方や評価方法の共通ルール、解析ツールの提供、AI人材の育成体制の構築等を推進

リスクへの対応(研究の公正性の確保、データ管理)

■助成金申請におけるAI利用の制限(2025年9月～)

AIツールを用いて大量の申請が行われる懸念に対応するため、2025年9月以降、AIで実質的に作成された申請は審査対象外に。また、PIが1年間に提出できる申請数は最大6件に制限。

■査読(ピアレビュー)プロセスでのAI利用の使用禁止

未公開の研究アイデアの漏洩を防ぐため、査読プロセスにおける生成AIの使用は禁止。

■制限データ(ヒトゲノムデータ等)を使用する際の制限

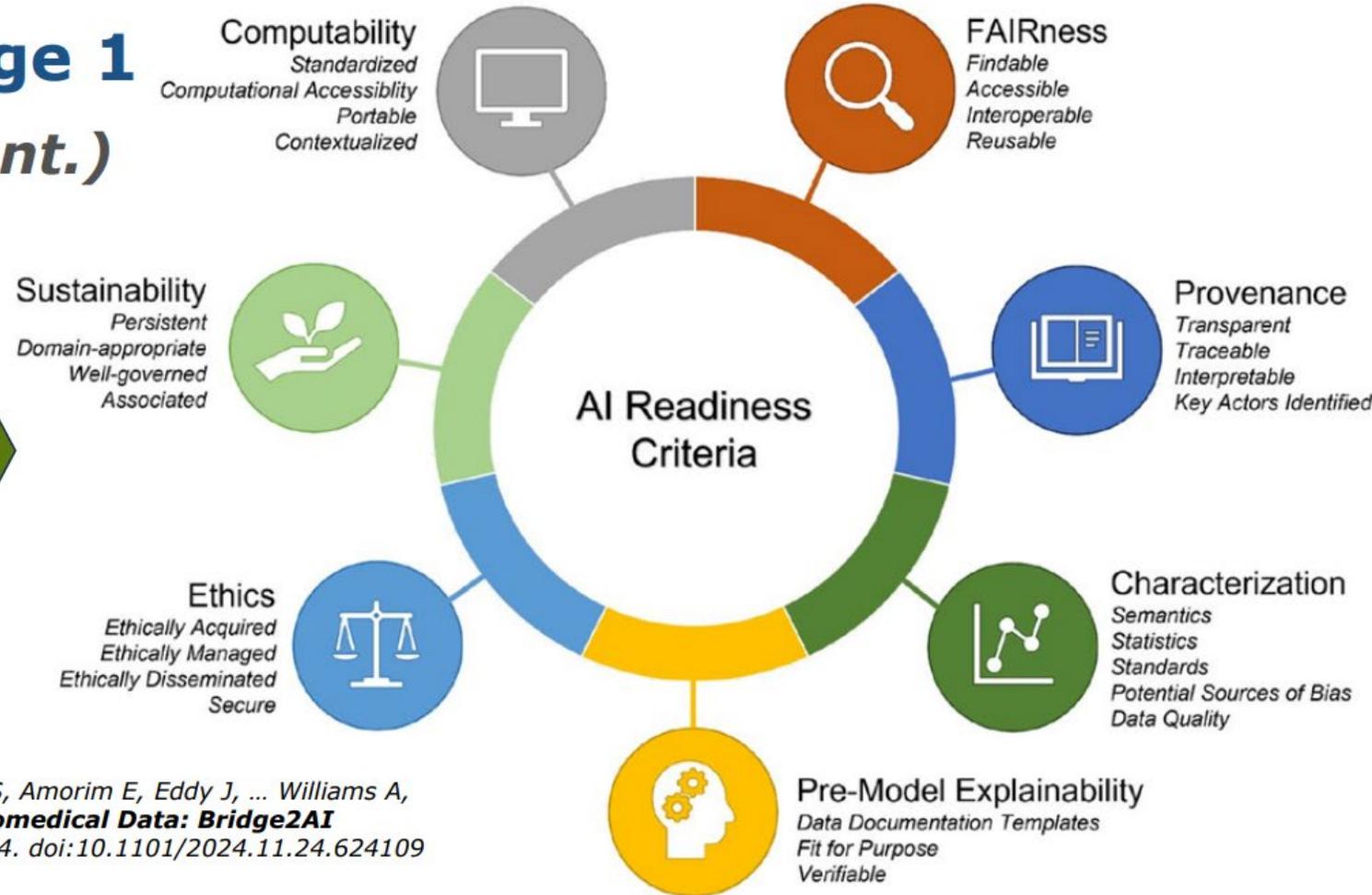
- オープンなAIモデルへの入力制限
- 開発したAIモデルの共有制限
 - ✓ プライバシーの懸念がない一般的なデータを用いて開発されたAIツールやソフトウェアは、NIHの**オープンサイエンス原則**に基づき、無料でオープンな形式で共有することが原則。
 - ✓ 他方、**ヒトのゲノムデータなどのアクセス制限データを学習させたAIモデルのパラメータ**は「データ派生物」と定義され、漏洩リスクを防ぐために、**プロジェクト終了後の保持や第三者への共有を、現在、一時的に禁止**

■プライバシー強化技術(PETs)等の検討

イノベーションを促進しつつ、制限データ漏洩リスクを軽減するための、プライバシー強化技術 (Privacy Enhancing Technologies) やその他の方策について、研究コミュニティや一般に向けて広く情報提供依頼 (RFI) を発出

Bridge2AI Stage 1 Deliverables (cont.)

Core Principles & Tools for AI-Readiness



Clark T, Caufield H, Parker J A, Al Manir S, Amorim E, Eddy J, ... Williams A, Muñoz-Torres M C. **AI-readiness for Biomedical Data: Bridge2AI Recommendations.** bioRxiv, 2024-11-24. doi:10.1101/2024.11.24.624109 / PMID:39484409

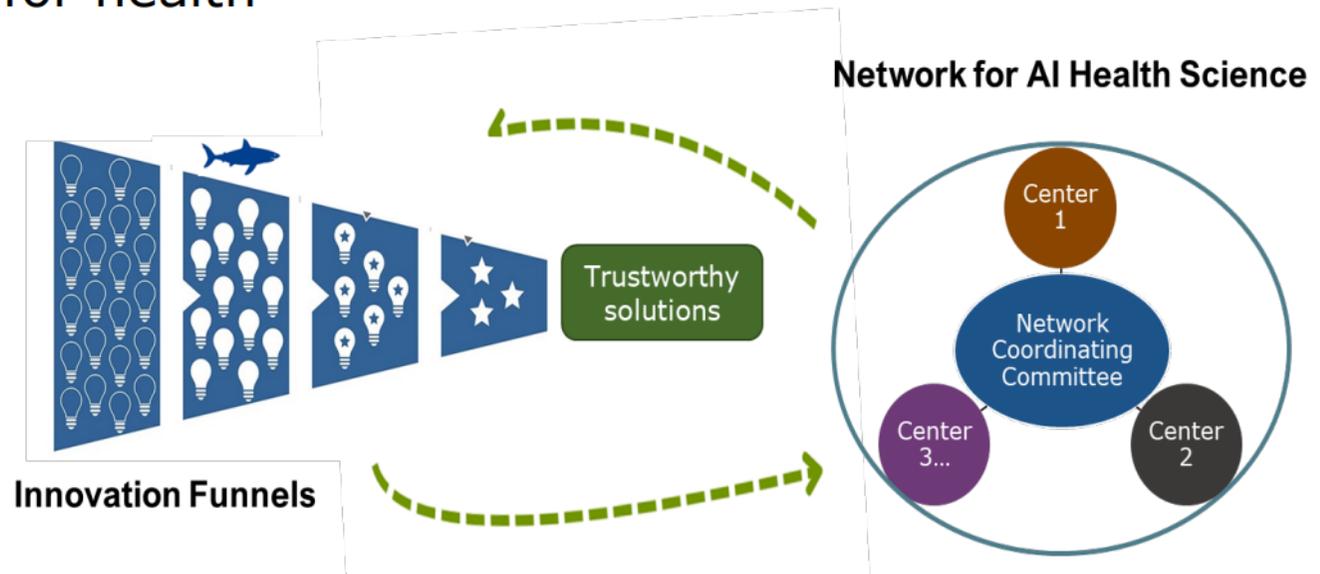
Program Goal and Initiatives

Goal:

Deliver AI-based solutions to biomedical Grand Challenges and support AI-enabled discovery for health

Two Initiatives:

1. Innovation Funnel
for safe, rigorously developed AI-enabled solutions
2. Network for AI Health Science
build the framework for AI health science of the future



米国議会・国家安全保障新興バイオテクノロジー委員会報告書(2025年4月) (National Security Commission on Emerging Biotechnology, NSCEB)



Current State of Biodata

01

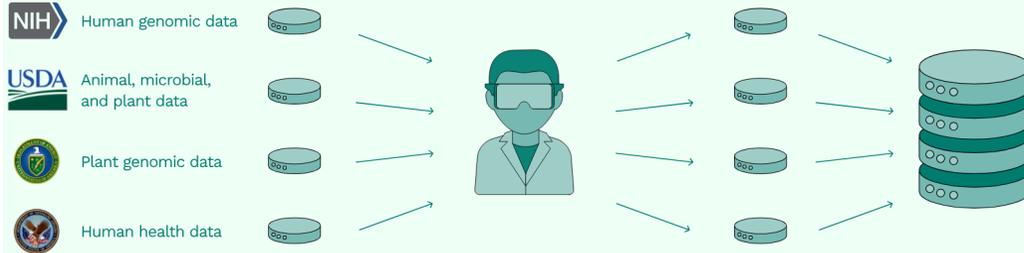
Researcher manually searches and recalls data

02

Researcher manually cleans each dataset

03

Researcher manually combines datasets



Recommendation

AI-Ready Biodata + Web of Biodata

01

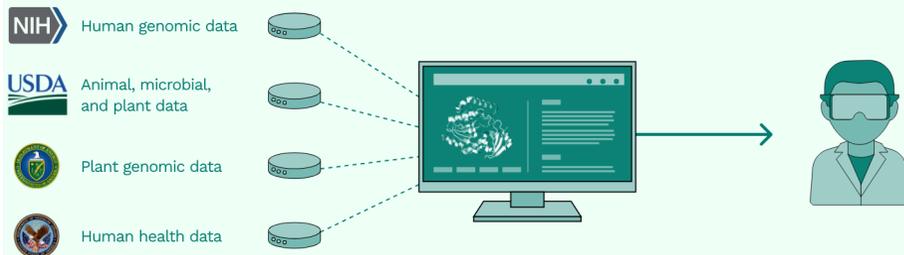
Agencies generate data in-line with NIST standards for AI-ready biodata

02

Agencies connect data to DOE Web of Biodata

03

Researcher pulls AI-ready data from Web of Biodata



※本報告書での生物学データ (biological data) の定義

“the information, including associated descriptors, derived from the structure, function, or process of a biological system(s) that is either measured, collected, or aggregated for analysis”

本報告書の第4-1章は、生物学的データを戦略的資源として扱うべきとして、議会に対して、下記4点を提言：

4.1A : Congress must authorize the Department of Energy (DOE) to create a Web of Biological Data (WOBD), a single point of entry for researchers to access high-quality data.

(議会は、エネルギー省(DOE)に対し、研究者が高品質なデータにアクセスできる一括窓口としての生物学データウェブ(WOBD)の作成を認可すべき)

4.1B : Congress should authorize the National Institute of Standards and Technology (NIST) to create standards that researchers must meet to ensure that U.S. biological data is ready for use in AI models.

(議会は、国立標準技術研究所(NIST)に対し、米国の生物学データがAIモデルに活用できるよう、研究者が満たすべき基準の作成を認可すべき)

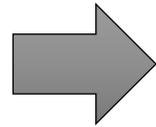
4.1C : Congress should authorize and fund the Department of Interior (DOI) to create a Sequencing Public Lands Initiative to collect new data from U.S. public lands that researchers can use to drive innovation.

4.1D : Congress should authorize the National Science Foundation (NSF) to establish a network of “cloud labs,” giving researchers state-of-the-art tools to make data generation easier.

< 欧米における傾向 >

従来型のデータベース

- データの保管・検索を主な目的とした整備・設計



AI時代のデータ基盤

- AI開発用の活用も見据えたAI-readyデータの整備
- インフラ上でAIモデルを開発・活用可能な計算資源と連動した設計

< 共通的な課題 >

- AI-readyデータの標準化
 - 持続可能な運用
 - データセットのオープン/クローズ
 - AIモデルのオープン/クローズ
 - セキュリティ・セーフティの確保
- 等

➤本報告書は、合成生物学とAIとの融合研究（SynBioxAI）についての、OECDの新興技術予測ガバナンス枠組みにもとづく戦略的インテリジェンス分析とともに、技術の現状・見通しについて、具体的な事例も含めて紹介。

➤その中で、データセットやそれを活用したAIモデルのアルゴリズムのオープン/クローズの政策オプションについても分析・提言。



Table 3.3. Policy options for data

Sub-topic	Policy option	Opportunities	Consideration
Aggregation	Support efforts and develop incentives for data generation, aggregation and use, including common data sharing infrastructures (e.g. data from publicly funded projects to be shared into a government-run, accessible repository)	Targeting data silos is key to obtain enough high-quality data to train AI models. Learn from successful experience of the Protein Data Bank which enabled AlphaFold.	Obligations can be set on publicly funded projects but sharing of privately funded data may require incentives. Engage researchers to identify which subsets of data would be most useful and focusing standardization of those may be most efficient path forward.
Standards and benchmarks	Encourage international standards (e.g. ISO) for data formats and benchmarks for evaluation of AI models in synthetic biology (e.g. protein structure prediction, sequence design...).	Internationally agreed data standards would support aggregation and sharing, whilst quantitative benchmarks would facilitate comparison between models to support the most successful ones.	Learn from successful examples in other fields (e.g. DICOM for medical imaging) ⁴⁴ . Standards can also lead to restriction towards certain innovation paths, so consider (e.g. who is included) when developing.
Open vs closed models	Landscape examples of open and closed datasets and models to compare differing incentives and outcomes	Mapping drivers could help policymakers understand incentives and policy levers available to them.	Factors affecting decision over open vs closed models (e.g. IP, biosecurity, trust...) will be different for different models.

Note: Additional suggestions to policymakers on promoting the sharing, (re)use, and governance of data can be found in several OECD Council Recommendations, such as: Recommendation of the Council on Enhancing Access to and Sharing of Data [[OECD/LEGAL/0463](https://www.oecd.org/legal/0463)]; Recommendation of the Council concerning Access to Research Data from Public Funding [[OECD/LEGAL/0347](https://www.oecd.org/legal/0347)]; Recommendation of the Council on Health Data Governance [[OECD/LEGAL/0433](https://www.oecd.org/legal/0433)], etc.

Source: OECD Research.