

## 社会を支える生命医学データ基盤の確立

### ① ビジョンの概要

ゲノム解析技術が成熟し、あらゆる生体・環境からゲノムデータを安価に得られるようになった。ゲノム医学やヒトを含む各種生物のバイオバンクが今後普及し、地球生命全体の多様性を網羅するだろう。我々は環境資源や予算が有限であることを見据えた科学を構築する責務がある。科学者が社会と対話しながら持続可能なビッグサイエンスを継続するには政府主導によるデータ拠点、情報拠点が必要となる。

### ② ビジョンの内容

人新世とよばれる時代に入り生命医学を取り巻く環境は急速に進歩している。数十年先には、全国民の個人ゲノムをデータベース化して利活用する国が数多く存在するはずだ。また地球上の主だった生命はゲノムおよびその生息域が明らかにされるだろう。医療画像の AI 診断に代表されるバイオイメージング技術は、ヒト健診のみならず環境測定・保全にも利用される。これらのビッグデータはアーカイブに適切に保存し、適切な用途で利用されねばならない。未来の科学者はこの情報を駆使して、有限の資源を絶やさぬよう保護しつつ持続可能な形で利用するための科学を構築する必要に迫られる。また感染症やバイオテロなど新しく生じうる脅威に対しても科学的に適切な判断を迫られる。ここでもっとも重要なのは幅広い高度人材がデータへ自由にアクセスできることであろう。

国内には、大量データを発見・取得・分析し、個々の社会的問題に照らして応用するための「データ基盤」が欠けている。生命医学がヒトを扱う以上、倫理的・法的・社会的課題 (ELSI)こそ科学先進国として力を入れて解決すべき部分である。これを実現すべく、ゲノムを含む個人データを適切に保管し、利活用できるデータベースセンターを設置したい。ここにはヒトのみならず、あらゆるバイオリソースの情報が同様に蓄積される。そしてデータを処理するための大容量高速ネットワークおよび高性能のスーパーコンピュータ(以下スパコン)が全国から利用できるよう整備される。さらに研究者に限らない多様な人材がそのデータを駆使して様々な社会活動に参画可能とする。ビッグデータ・アーカイブを社会全体で利活用できる「データ基盤」の充実こそ、数十年後に日本が科学先進国として残るための必要条件である。

### ③ 学術研究構想の名称

社会を支える生命医学データ基盤の確立

### ④ 学術研究構想の概要

本提案に基づく持続可能な「データ基盤」には、データベース、スパコン、人材の3要素が含まれる。

データベース：複数ある国内のデータベース関連機関を統合あるいは連携させた、国際的な競争力と存在感を持つ国立バイオデータベースセンター(以下DBセンター)を構築する。DBセンターは収集すべきデータの策定からデータ算出・整備プロセスにも積極的に関与し、データ産出プロジェクトを立案する時点から産出側と連携する体制をとる。つまり国の生命医学研究を主導する中核機関として機能する。

スパコン：DBセンターと不可分の関係にあるのが情報処理用のスパコンである。計算機環境は今後もムーアの法則に従って大規模化する。そこでクラウドを含めた複数拠点の計算ノードとストレージを高速ネットワークで結んだ連携スパコンを構築し、必要に応じてバックアップ等を作成、提供する。

人材：上記の二項目は優れた人材があって初めて実現できる。DBセンターには国内主要機関の人材を結集させ、大学等教育機関とDBセンターが連携してキュレータ、エンジニア、研究者を育成するトレーニングコースやカリキュラムを立案する。トレーニングは短期のものからセンターにおける実地研修形式で実施するものまで含み、既に類似するコースを実施する国内主要機関とも密に連携する。

### ⑤ 学術的な意義

科学の究極目標はヒトのより良い未来を構築することであり、ヒトという生命体の動作原理を知ることこそ最重要である。ヒトを含む生命体について圧倒的な量の実データ(ゲノムデータ、オミックスデータ、画像データなど)を取得し、そこからのデータ駆動形式によって科学を発展させねばならない。それにはヒトゲノム解析に関するELSIを克服し、我々が処理しうる情報量や環境負荷も考慮しながら、我々はどこまで知るべきなのかという哲学的問題にも答えねばならない。

本構想のデータ基盤はヒト以外の生物にも及ぶ。我々にはデータを使い捨てるのではなく未来へのアーカイブ

イブに残す責務がある。まずは有する資源をデジタルデータとして残し、実体よりもデータを利活用すべきである。本提案は ELSI を踏まえたうえで地球に存在する生物情報を幅広く記録して利用するための「データ基盤」確立であり、科学者だけによる科学を脱して社会で取り組む科学（オープンサイエンス）を開始する。

#### ⑥ 国内外の研究動向と当該構想の位置付け

国内の学術情報ネットワーク SINET6、文部科学省バイオサイエンスデータベースセンター（NBDC）や、日本医療研究開発機構（AMED）等と連携し国立と呼べる中核データセンターを政府直轄で実施する。米国の NCBI や欧州連合の EBI、韓国の KOBIC と同等の組織とする。

#### ⑦ 社会的価値

データを社会と幅広く共有することはデータ利用者の時間短縮効果、生じる知的財産の価値、そして将来的な経済効果など複数段階の価値を有する。ただし現在のオープンサイエンスがはらむ問題は、データに付随するライセンスや用語問題、そして個人情報保護である。法整備や ELSI で社会からの理解を得るためヒト研究・医療のあり方について正面から向き合っ解決するための組織を構成する。

#### ⑧ 実施計画等について

**実施計画・スケジュール** 準備組織の設置（1～2年次）：国内研究拠点が競争的研究資金獲得や研究インパクト向上のために費やすエフォートを DB センター設立に振り向けるため、省庁等研究費配分機関や産業界を含めた準備組織を構成する。中核機関の公募（2年～3年次）：準備組織のもと維持管理すべき中核データおよび機関を公募する。恒久財源をうたう事業のため、公募機関や選定には十分な時間と透明性を確保して実施する。準備組織はそのまま DB センター執行部として拠点の連携や評価に携わる（任期制）。とりわけ海外との既存大型国際連携を引き継ぎ、データの場所やアクセス方法等に変更を及ぼさないように留意する。

**データ基盤の長期安定運営（4年次以降）**：DB センター参画機関は執行部による戦略やデータ共有政策のもと、データの産生から維持管理までを実施する。参画機関には5年を目処に定期的な評価も実施する。ユーザ目線による利用指標や研究者以外の人材も含めた外部委員会による評価を通じ、特にヒトゲノム情報の利活用や経済活動と環境維持のバランスについてデータを元に議論できる場を構築する。実際に安定運営を開始するまでには5年以上かかると考えられるが、設立後は上記の国際組織と連携しつつデータを維持管理できる。

**実施機関と実施体制** 世界に情報発信する国内のデータ重要拠点、学術情報ネットワーク SINET6 で結ばれるスパコン、更には商用クラウドサービスが公募に応じて参画するネットワーク型研究組織を構成する。公募を実施するため初期のみ任期制による準備組織を用意する。

技術基盤のパイロットスタディとして AMED が実施する臨床ゲノム情報統合データベースやデータ利活用プラットフォームを利用する。国内で将来的に維持管理すべき中核データおよび機関は公募形式で選定する。各機関には条件としてデータ共有政策の遵守、トレーニングコースの設置等を義務づける。DB センターは国公立大学が形成する拠点ネットワークに似た組織構造をとるが、大きく異なる点はクロスアポイントメント等を用いて DB センターへの人的エフォートを確保すること、そして国外からの利用や研究者以外からの利用も視野に入れたユーザ目線からの情報提供サービスを徹底すること、そして何より恒久財源に基づく事業とすることにある。

総経費 286 億円

所要経費「データ基盤」を実現するための DB センター構築にかかる準備費用

準備組織の構築・中核拠点の選定（初期3年、各2億円）

- ・ 準備組織執行部スタッフ クロスアポイント・兼務費用年間1億円（15名）
- ・ 準備組織事務局スタッフ、オフィスレンタル、旅費等雑費年間1億円

「データ基盤」10 拠点に配賦する計算機資源・人件費 40 億円/年（4年次以降）

- ・ 拠点に配賦するハードウェア経費（ストレージ、計算ノード、保守費） 10 拠点計 30 億円
- ・ 拠点に配賦する人件費、開発費等諸経費 10 拠点 10 億円

#### ⑨ 連絡先

有田 正規（国立遺伝学研究所）