

資料2

科学技術・学術審議会学術分科会
学術情報委員会(第12回)
平成30年9月18日(火)

ライフサイエンス研究における オープンサイエンスの推進

国立研究開発法人 理化学研究所

理事 小安 重夫

✓ データベースの統合、データ統一化による オープンサイエンスの推進

- 国際的な取組み
 - ヒトセルアトラス (Human Cell Atlas) プロジェクト
 - ヒト常在菌メタゲノム解析
 - 脳科学研究分野における国際協力
 - その他の事例
- バイオサイエンスデータベースセンターの取組み (JST)

✓ オープンサイエンスのための基盤構築

- データ基盤の構築と、活用に向けた基盤の推進
- 統合のための技術開発



背景・目的・期待される成果

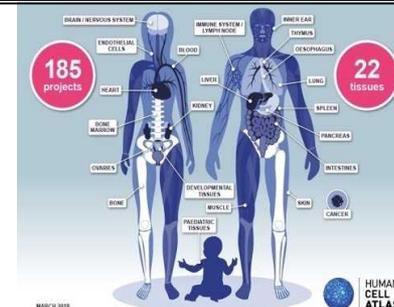
- ▶ ヒトゲノム情報の多くは細胞の集合（バルク）に由来する平均値的な情報であり、生命現象を本質的に理解し将来の医療をさらに発展させるためには、生体の究極の基本単位での数百種類にのぼる細胞種を「1細胞レベル」で解析することが必要とされる。
- ▶ こうした背景をふまえ、**ヒトの全細胞（約37兆個）について、種類・状態・系統などを最新の1細胞解析手法によって分類し、カタログ化することを目的とした国際プログラム「セルアトラス（ヒト細胞百科事典、HCA）プロジェクト」が欧米の主要な研究機関（欧州：サンガー研究所・カロリンスカ研究所、米国：ブロード研究所、等）を中心に提案され、参画機関の研究者間の会議で議論が進んでいる。**
- ▶ 本プロジェクトの推進のためには、系統的かつ高精度でヒトを構成する全ての細胞について一つ一つの細胞固有の遺伝子発現状態を網羅的解析を可能とするための技術として、**日本の持つ1細胞ゲノミクス解析技術が重要な役割を果たすことが欧米の研究機関から強く期待されている。**

各国で HCA Meeting が開催され議論が進んでいる

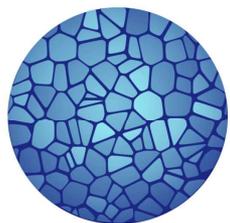


米国では、NIHがこの二年間でセルアトラス関連のプロジェクトに約200億円の予算をを投じている。欧州でも、特定の臓器等を対象としたパイロットスタディが進行しており、わが国の参画が強く求められている。

研究が進められている解析対象



- ▶ 免疫系
- ▶ 肝臓
- ▶ 腎臓
- ▶ 肺
- ▶ 皮膚
- ▶ 腸
- ▶ 心臓
- ▶ 神経系（等）



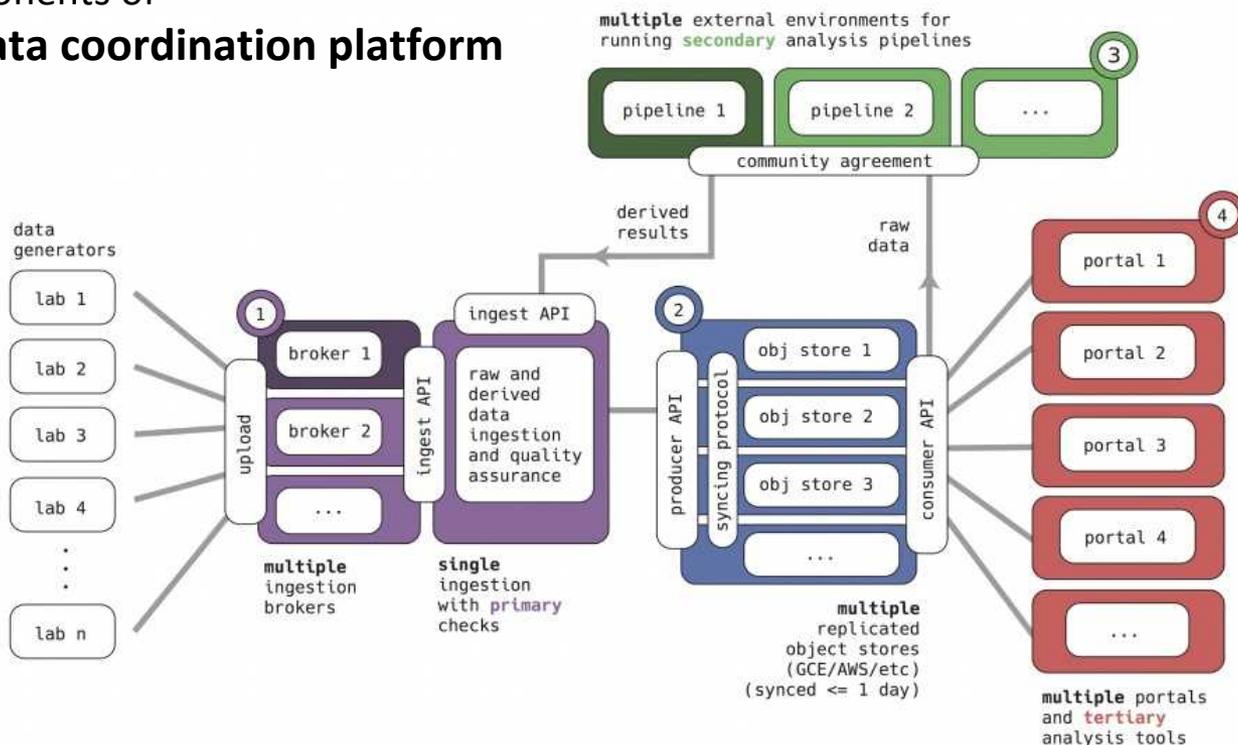
HUMAN CELL ATLAS

AREAS OF IMPACT

Where do we start?

Studying all the cells in the human body is an enormous endeavor—current estimates suggest that an average human being is made of at least 37.2 trillion cells. To take on this bold task, we are conducting preliminary pilot projects that will not only reveal interesting biology, but also inform us about efficient and effective sampling and analysis strategies for a full-scale cell atlas effort. These pilot projects will also begin to build an international network that is essential for the cell atlas's success. A few examples are:

Diagram of key components of the open-source data coordination platform



ヒトエピゲノム解析の国際動向



International Human Epigenome Consortium (IHEC)
(12カ国・地域参画)

欧州 :

BLUEPRINT project



(39.4 million€/4.5 年)

BLUEPRINT
epigenome

2017年-
個別疾患プロジェクト
(**SYSCID**)

(14.4million€/5years)

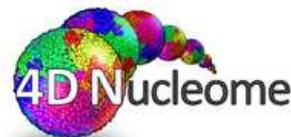
The European Union's Horizon 2020 research and innovation programme



米国NIH :

Roadmap Epigenomics Project

29 innovative areas of research supported by the NIH Common Fund



2018年-
Common Fund's
4D Nucleome Program

日本 :

AMED-CREST (H26年度までJST-CREST+さきがけ)

「エピゲノム研究に基づく診断・治療へ向けた新技術の創出」

【研究開発目標】 疾患の予防・診断・治療や再生医療の実現等に向けたエピゲノム比較による疾患解析や幹細胞の分化機構の解明等の基盤技術の創出 (H23-30)



ヒト常在菌メタゲノム解析の国際動向

METAGENOMICS

Massive Microbial Sequence Project Proposed

Science 315, 1781 (2007) Mar. 30

Understanding the microbiome-human, animal, and 'environmental' is as important as the human genome.

ヒト、動物、環境中の細菌叢を理解することはヒトゲノムを理解することと同じように重要である。

2007年12月：国際コンソーシアム (IHMC) 設立

日本からはHuman MetaGenome Consortium Japan (HMGJ)

欧州：

MeTaHIT

(Metagenomics of Human Intestinal Tract)



21.2 million€/4.5年

含 公的資金
(European Commission)
11.4million€

13 partners from
academia and
industry, a total of
8 countries.

米国NIH：

HMP

(Human Microbiome Project)

173 million \$ /2008-2012

2013年-
各疾患プロジェクト予算に振り分け、
メタゲノム解析予算総額は増加



NIH HUMAN
MICROBIOME
PROJECT

欧米人の非常に大きな微生物層ゲノムリソースが構築され、約500万個の遺伝子が「カタログ化」された。

日本：AMED-CREST+PRIME

「微生物叢と宿主の相互作用・共生の理解と、それに基づく疾患発症のメカニズム解明」

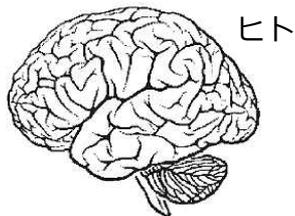
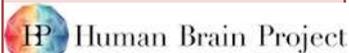
(2016-最長7年半)

【研究開発目標】 宿主と微生物叢間クロストーク・共生の解明と健康・医療への応用)

世界の脳科学研究の国際動向

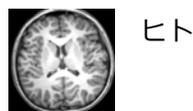
欧州：Human Brain Project

1.2 billion€/10年



- ・情報通信技術と生物学の収束によるヒト脳理解
- ・ヒト全脳のモデル化によるシミュレーション

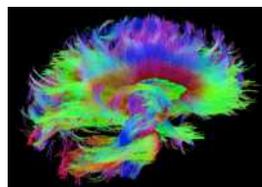
日本：革新脳



- ・霊長類の構造・機能マップ作成
- ・霊長類の構造・機能マップ作成に寄与する革新的技術開発

米国：Brain Initiative

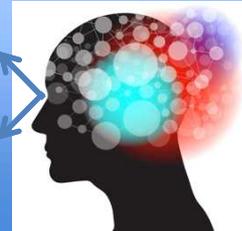
4.5 billion\$/10年



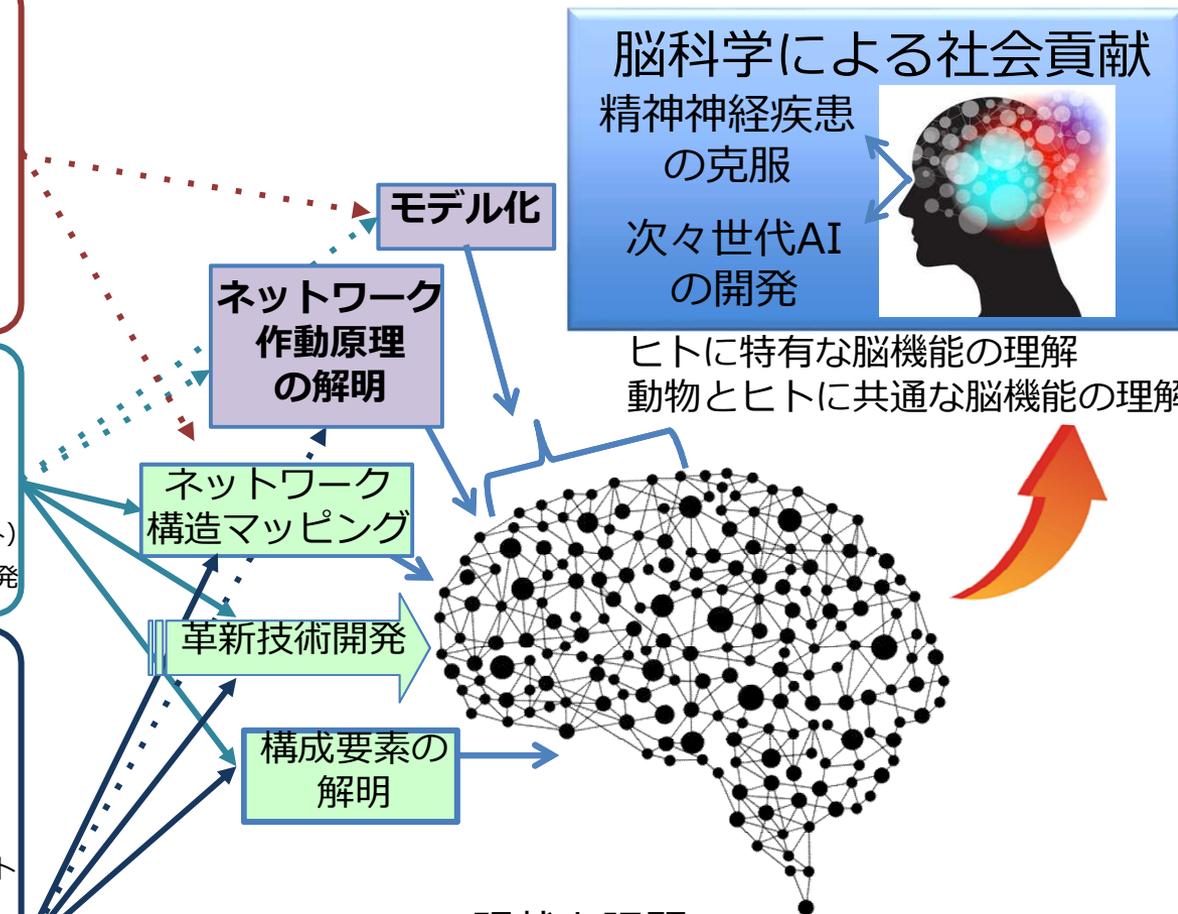
- ・革新的技術開発
- ・シナプスから全脳レベルの構造的結合マップ作成
- ・大規模脳活動計測による脳活動マップ作成
- ・脳活動の階層をつなぐ統合的理解
- ・革新的論理・データ解析技術を用いた精神機能の生物学的基盤解明

脳科学による社会貢献

精神神経疾患の克服
次々世代AIの開発



ヒトに特有な脳機能の理解
動物とヒトに共通な脳機能の理解



現状と課題

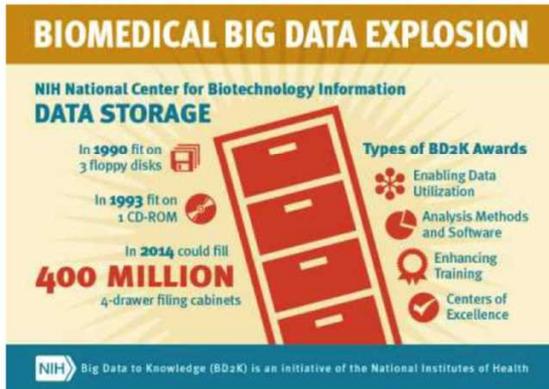
- ・脳構造マップ構築は大型国家プロジェクトで進みつつある
- ・しかし、同定した脳構造のネットワークがどのように機能しているか（**作動原理**）の解明は進んでいない
- ・ネットワークの機能の理解には、数理モデル化が必須
- ・**とくにヒトに特有な高次機能に関しては**、数理モデル化によるネットワーク作動原理の解明に照準を合わせた取り組みが**世界的にもほとんどない**
- ・**数理脳科学と実験脳科学の密な連携**が必要

海外のデータ共有化の動向

2017年7月19日 ライフサイエンス委員会資料より

○ Big Data to Knowledge (BD2K)

- ・米国NIHのCommon Fund 26件のうちの1件、**NIH横断的な取組み**を実施(FY2017予算は\$74M)
- ・第二フェーズ(FY2018~2021)では、**FAIRデータの集積、共有、活用**のための”Data Commons Pilot Phase”が計画されている



<https://www.nih.gov/news-events/news-releases/nih-invests-almost-32-million-increase-utility-biomedical-research-data>

○ Global Alliance for Genomics and Health (GA4GH)

- ・ヒトゲノム情報や臨床情報に関する**データ共有プロジェクトを推進**
- ・**44カ国、487機関が参画**
(NIH, EMBL-EBI, Wellcome Trust, 中国BGI, Pfizer, Merck, Philips, Google, Intelなど。日本からは、NBDC, DBCLS, AMED、国立がん研究センター、DDBJ、理研、ToMMoなど15機関)
- ・**Beacon Project**では、世界の200以上のヒトゲノムデータセットを検索可。検索の方法と結果情報を限定することで、**データ利用審査なしに迅速な情報入手が可能**
- ・Matchmaker Exchange(希少疾患)、BRCA Challenge(がん)など、他にも様々なデータ共有の取組み



○ ELIXIR

- ・**20カ国の180機関が参画**
(EMBL、英国、ドイツ、オランダ、ベルギー、イタリア、フランス、スペイン、イスラエルなど)
- ・FAIR principlesに基づいた**ELIXIR Interoperability Platform**などのデータ基盤を提供
- ・EXCELERATE (2015-2018, €19M)、CORBEL (2015-2018, €14.5M)、ENVRIPlus (2015-2019)などのHorizon2020プロジェクトを実施

Platforms

ELIXIR's activities are divided into five areas called 'Platforms'. These are Data, Tools, Interoperability, Compute and Training. The Platforms are managed by Platform leaders and the work is carried out by groups within the Platforms.

Data Platform
Aims to identify key data resources across Europe and support the linkages between data and literature

Tools Platform
Helps researchers find the best software tools to analyse their data.

Compute Platform
Develops services to make it easier to store, share and analyse large datasets.

Interoperability Platform
Develops and encourages the adoption of standards to describe life science data.

Training Platform
Helps scientists and developers find the training they need, and also provides that training.

<https://www.elixir-europe.org/platforms>

ライフサイエンス分野におけるデータの統合について

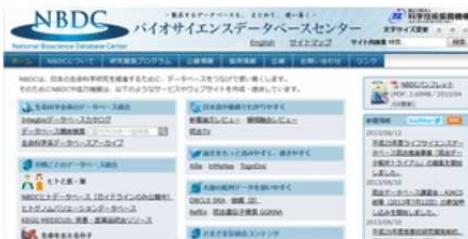


NBDC（バイオサイエンスデータベースセンター）におけるデータベース統合化の取り組み

- 当初、各省で独自に進められていたが、2006(平成18)年頃から内閣府の主導のもと、省庁連携によるオールジャパンの協力体制を構築して推進。
- 2009(平成 21)年 5 月には、「統合データベースタスクフォース報告書」(総合 科学技術会議 ライフサイエンス PT 統合データベースタスクフォース)の中で、DBCLS と BIRD の一体的な運用を、JST に新たに設置する組織で行うことが提言。
- 2011(平成 23)年 4 月、JST に NBDC が発足し、ライフサイエンスデータベース統合推進事業が開始。



文部科学省：バイオサイエンスデータベースセンター(NBDC)ポータルサイト



NBDCは、データベース整備・統合の様々な実務に携わるとともに、ファンディングを介して、統合のための基盤技術開発や、生命科学の各分野のデータベースの統合を進めています。平成22年度までの「文部科学省データベース統合プロジェクト」と平成23年度までの「バイオインフォマティクス推進事業」の流れを継承しています。

URL: <http://biosciencedbc.jp/>

ライフサイエンス分野におけるデータの統合について

バイオサイエンスデータベースセンターの今後のあり方について(提言)

平成29年3月 国立研究開発法人科学技術振興機構

バイオサイエンスデータベースセンター運営委員会 より

～ 今後 5 年をめどに NBDC の活動を拡大・強化して取り組むべき内容 ～

1. 支援対象を、公開済みデータだけでなく 未公開データ まで拡大する

【新たな対象】

- ・我が国のライフサイエンス研究データを網羅的に把握・可視化し、大型プロジェクト 研究のデータが関係者間でいち早く最大限に利活用されるように支援する

2. データベース構築者の視点から、利用者の視点に転換 する

【新たな利用】

- ・ヒト由来データの共有と利活用は喫緊の課題であり、ヒトとモデル動物の研究データをゲノムから表現型まで統合し、利用者の要請に応えられるデータ基盤を確立する(当面の連携先として、東北メディカル・メガバンク計画、臨床ゲノム情報統合データベース整備 事業、ゲノム医療実現推進プラットフォーム事業、AMED 微生物叢関連プロジェクト群、COI STREAM 健康・医療データ連携事業、などが想定される)

3. 新たな知識やイノベーションを生み出すデータベースを構築 する

【新たな価値】

- ・世界に伍する研究開発が実現するように、大規模データ解析や人工知能を含む多様な分野の研究者と共に、データベースの戦略的統合と利活用に取り組む

ライフサイエンス分野におけるデータの統合について

バイオサイエンスデータベースセンターの今後のあり方について(提言)

平成29年3月 国立研究開発法人科学技術振興機構

バイオサイエンスデータベースセンター運営委員会 より

～ NBDC が関係する府省や機関と共に解決に当たるべき課題 ～

4. **データ整備・統合の継続性**を保障する体制の整備

【研究基盤には継続性が不可欠】

- ・ 恒久的な財政措置(国の研究開発予算のうち一定割合を確保するなど)
- ・ バイオインフォマティクス人材の育成とキャリアパス設計

5. 我が国の顔としての**求心力・発信力の強化**

【プレゼンス強化には一元化が不可欠】

- ・ 継続性のある一元的な統合データベース機関の設立
- ・ 我が国を代表する統合データベースとしての国際活動

分野別統合と分野を超えた統合のための技術開発の展望

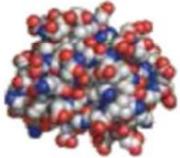
2017年7月19日 ライフサイエンス委員会資料より

ファンディングによって分野別(生物種やオミクス別)のデータベース統合を推進するとともに、分野を超えた統合のための技術開発を実施している。

分野別統合データベース



プロテオーム
蛋白質立体構造



SSBD Database
Systems Science of Biological Dynamics

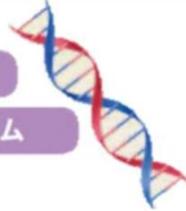
バイオイメージ
生命動態



植物



ゲノム
エピゲノム



多階層オミクス



医薬品

KEGG
MEDICUS



糖鎖



Microbe DB^{JP}

微生物



分野を超えた統合のための技術開発

- RDF化等による高度な検索の実現
- データベース関連ツール開発
- データベース統合化の国際的標準化など

探す



整理する



つなげる



抽出する



解析する



ライフサイエンスにおける
ビッグデータ基盤の構築

オープンデータサイエンス基盤の政策的位置づけ

第5期科学技術基本計画

第4章 科学技術イノベーションの基盤的な力の強化

(2) 知の基盤の強化

① イノベーションの源泉としての学術研究と基礎研究の推進

企業のみでは十分に組み込まれない未踏の分野への挑戦や、分野間連携・異分野融合等の更なる推進といった観点から、国の政策的な戦略・要請に基づく基礎研究は、学術研究と共に、イノベーションの源泉として重要である。このため、国は、政策的な戦略・要請に基づく基礎研究の充実強化を図る。(中略)

また、学際的・分野融合的な研究の充実を図る。

② 研究開発活動を支える共通基盤技術、施設・設備、情報基盤の戦略的強化

ii) 産学官が利用する研究施設・設備及び知的基盤の整備・共用、ネットワーク化

加えて、幅広い研究開発活動や経済・社会活動を安定的かつ効果的に促進するために不可欠なデータベースや計量標準、生物遺伝資源等の知的基盤について、公的研究機関を実施機関として戦略的・体系的に整備する。

iii) 大学等の施設・設備の整備と情報基盤の強化

(前略)国は、資金配分機関、大学等の研究機関、研究者等の関係者と連携し、オープンサイエンスの推進体制を構築する。公的資金による研究成果については、その利活用を可能な限り拡大することを、我が国のオープンサイエンス推進の基本姿勢とする。その他の研究成果としての研究二次データについても、分野により研究データの保存と共有方法が異なることを念頭に置いた上で可能な範囲で公開する。ただし、研究成果のうち、国家安全保障等に係るデータ、商業目的で収集されたデータなどは公開適用対象外とする。また、データへのアクセスやデータの利用には、個人のプライバシー保護、財産的価値のある成果物の保護の観点から制限事項を設ける。なお、研究分野によって研究データの保存と共有の方法に違いがあることを認識するとともに、国益等を意識したオープン・アンド・クローズ戦略及び知的財産の実施等に留意することが重要である。

また、国は、科学研究活動の効率化と生産性の向上を目指し、オープンサイエンスの推進のルールに基づき、適切な国際連携により、研究成果・データを共有するプラットフォームを構築する。

統合イノベーション戦略

第2章 知の源泉

(2) オープンサイエンスのためのデータ基盤の整備 ③ 今後の方向性及び具体的に講ずる主要施策

研究データを我が国のデータインフラから公開できるよう、主として機関リポジトリを対象としたシステム開発や、国際認証基準等を参考にしたりポジトリの整備・運用を進める。その上で、研究分野の特性等を踏まえたオープン・アンド・クローズ戦略を考慮したデータポリシーやデータマネジメントプランの策定を促進し、これらに基づく研究データの管理・公開等を促進するとともに、公的資金による研究成果としての研究データについては、データインフラを通して機械判読可能化を促進する。

第6章 特に取組を強化すべき主要分野

(1) AI技術 ③ 今後の方向性及び具体的に講ずる主要施策

i) 桁違いな規模による人材の育成・活用

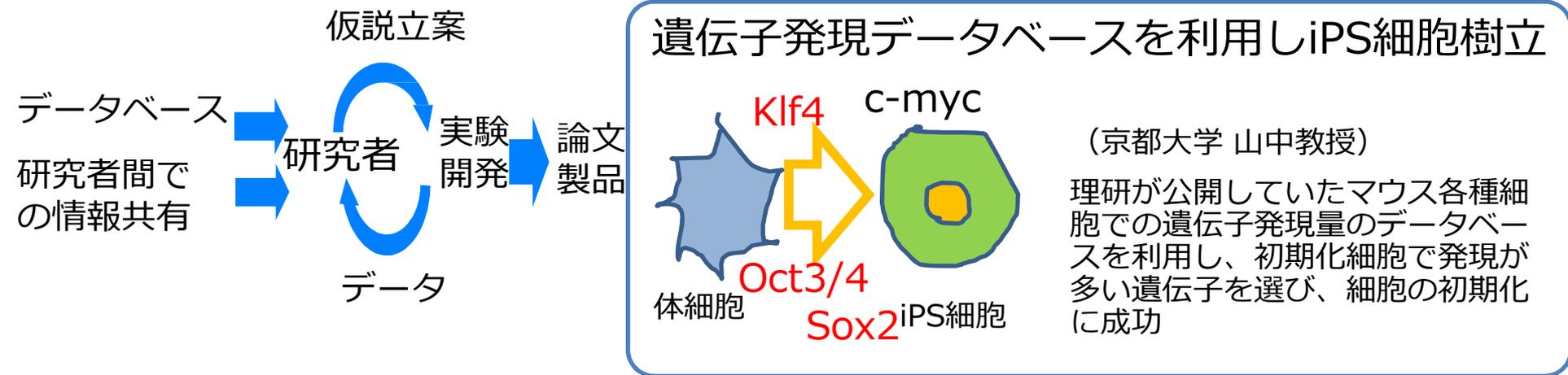
データ連携基盤を活用したAI技術の社会実装、ロボット技術等と組み合わせた応用開発、現在のAI技術の弱みを克服する基礎・基盤的な研究開発を産学官が一体となって強力に推進する。

(2) バイオテクノロジー ③ 今後の方向性及び具体的に講ずる主要施策

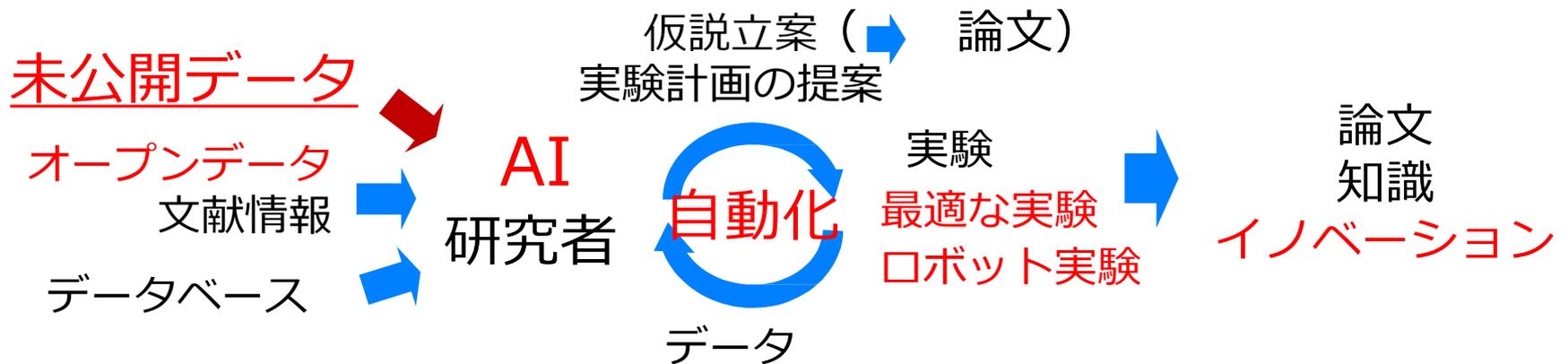
今後は、過去の戦略等をより詳しく分析した上で、医療・非医療分野が一体となった新たなバイオ戦略について、2019年夏を目指して策定する。

他方、「データ駆動型」の技術開発・社会実装を早急に加速するため、当面の取組として、以下のとおり環境整備を行いつつ、我が国の強みを最大限に発揮できる分野をターゲットに、「帰納的」な発想での解析等による技術開発・社会実装を行う。

オープンサイエンス基盤推進の意義



データ利用研究の展望



- AIは、インプットデータが多いほど精度が向上し、分野横断分野融合の新しい発見をもたらすことが期待される。

オープンサイエンスのためのデータ基盤の構築

- 機関の強みを生かした連携研究による成果の創出を目指す世界的な潮流の中で、国際連携の活性化を目指し、セキュアかつ参画者が自由に使えるデータ基盤が必要。
- その基盤構築にはデータ産出の現場とユーザー及び情報科学の研究者の共同構築で作られることが望ましい。
- 情報科学者のみによるフレームワーク構築でもなく、ビッグデータによるデータベースを作るでもなく、研究者にとって一番効果のある基盤を構築し、オープンサイエンスとして活用することを目指すべき。

分野ごとのデータ基盤構築と分野横断集約化した公開基盤構築、その利用研究の実証

1. ポリシー策定とデータ管理の基盤の構築
2. 研究分野ごとのデータ基盤を構築
3. ユーザーニーズからのクラウド解析環境の整備
4. 人材養成

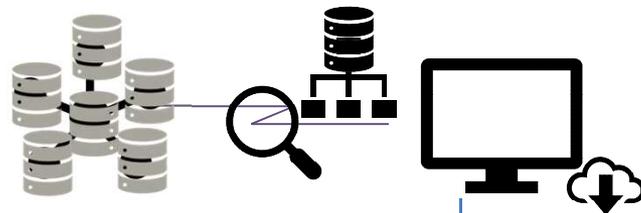


**データ公開のみならず
データ科学の拠点構築**

**真のオープンサイエンス
実施可能な環境**

理化学研究所におけるオープンサイエンスに向けた取組み

「自然科学知の統合」を先導するオープンサイエンス基盤の構築・強化



★研究データを統合データベースに集約・再統合することにより、組織横断的な管理・共有が可能

情報科学 熱力学、力学系、宇宙科学、生態学、化学、物理学、医科学、社会科学、経済学、...

総合自然科学研究所としてのポテンシャルを最大限に活かす統合データベース

★統合データベースを通じて、検索ワードと関連付けられた様々な実験データや論文を組織横断的に表示・可視化

- 分野横断的な数理的アプローチの開発、共通の数理・統計モデル等の開拓
- 分野で知られた知見やノウハウの、他分野への展開における適用・活用

- 既成分野・領域にブレークスルーをもたらす画期的な研究手法の開発
- 時代を先導する真にハイ・インパクトな新興・融合領域の開拓
- 頭脳循環拠点としてのポテンシャル強化

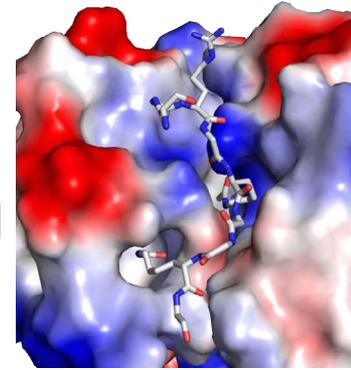
異分野データからの研究事例

観測結果からの
恒星間重力相互作用の
シミュレーション

宇宙創成の瞬間の状態
をシミュレーション

タンパク質の
立体構造解析

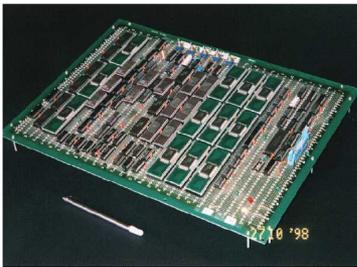
創薬に向けた生体内
タンパク質相互作用
シミュレーション



点と点が力を及ぼし合うというデータの構造が一致

重力多体問題のシミュレーション
のための専用計算機

GRAPE計算機開発



1989年にGRAPE1を開発

GRAPEの計算原理を
応用し、生命科学の
ための計算機を開発

分子動力学シミュレーションのための
専用計算機

MD-GRAPE計算機開発



2006年にMD-GRAPE3を開発
世界初の1ペタflopsの計算速度
を達成
(ゴードン・ベル賞授賞)

現在では創薬における計算化合物スクリー
ニングの標準技術になっている。

(MD-GRAPE4を企業等と開発中)

理化学研究所におけるオープンサイエンスに向けた取組み

オープンサイエンス基盤の構築と その実践によるイノベーション創出戦略



イノベーションにつなげる 機関戦略



データ戦略ポリシー
産業連携スキーム
(イノベーション事業化支援のための
法人設置による共創機能の強化)

研究データの集約化、AI可読化が必須



企業連携の敷居を下げ、民間資金の呼び込みと
イノベーションの共創を実現