基礎・横断研究戦略作業部会（第1回）	資料2-3
令和7年11月25日	

ライフサイエンス委員会
基礎・横断研究検討部会

NLDP説明資料

ＮＬＤＰプログラムディレクター
ＮＢＤＣライフサイエンスデータベース特別主監
高木 利久(富山国際大学　学長）

統合DBプロジェクトの目的とその推進方策

(2001-2010 BIRD) 2006- 統合DBプロジェクト　2007-DBCLS　2011-NBDC　2025-NLDP

目的：世界的に分散され、フォーマット、オントロジー、IFがバラバラなDBを連携・統合して格段に使い易くすることにより、生命研究・バイオ産業の大幅な効率化を図る

・その手段として、以下を推進
  ・オープンサイエンスの推進（データの流通、データの保全、ヒトデータ共有ポリシー）
  ・統合に資する様々な技術開発
  ・カタログ、横断検索、アーカイブ、など統合の成果や各種DBのポータルサービス事業
  ・ファンディングによる統合利用を意識した多様なDB構築（本格型、育成型）
  ・国際標準化、国際連携、国内連携、技術普及活動
  ・DB人材、バイオインフォ人材の育成

・ターゲット層
  ・ライフサイエンス研究者
  （基礎・応用生命研究者、医薬科学研究者、バイオインフォマティシャンなど）

ライフサイエンスデータベース系事業の事業実施主体の見直しについて

＜ライフサイエンスデータベース系事業のあらまし＞

- 個人のゲノムデータやタンパク質の立体構造データ、遺伝子発現データ、細胞レベルでの発現情報など、ライフサイエンス研究にデータベースの活用は必須であり、我が国のライフサイエンス研究全体を推進・加速させていくためには、**産出された大量のデータを生かすためのデータベースの整備が不可欠であり、また、データベースが効率的に活用されるための仕組みも必要**である。
- 平成20年に内閣府統合科学技術会議ライフサイエンスPTが「統合データベース　タスクフォース報告書」において、ライフサイエンス分野における我が国全体の恒久的かつ一元的な統合データベースの整備について方針を取りまとめるなど、**統合化されたデータベースによる新たな研究成果の創出や研究の効率化を我が国が一体となり推進**してきた。
- 上記の報告書における提言を踏まえ、ライフサイエンスデータベース系事業は、**令和6年度までJSTバイオサイエンスデータベースセンター（NBDC）が「ライフサイエンスデータベース統合推進事業」として実施**。データ解析ツール等の技術開発を通じ、網羅的に解析可能な世界最大級の知識グラフの構築、複数の統合ツールが開発されてきた。

＜見直しの背景及び方針＞

- 令和6年2月に、ライフサイエンス委員会において、「ライフサイエンスデータベースは研究基盤として重要であり、国が直轄で実施すべき」との方向性が承認され、**ライフサイエンスデータベースの基盤的な技術開発の実施主体を見直す必要**が生じた。
- 見直しの結果、令和７年度からNBDCのプロジェクトの一部を、**JSTから国に段階的に移行することが決定**。内局化したプロジェクトを**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（NLDP）と命名**。

ライフサイエンスデータベースの重要性等に関する政府文書

科学技術・イノベーション基本計画（令和３年３月26日閣議決定）

- ライフサイエンス分野においても、**データ駆動型研究の基盤となるゲノム・データをはじめとした情報基盤**や生物遺伝資源等の戦略的・体系的な整備を推進する。
- **研究データの管理・利活用を進める環境を整備**する。**最先端のデータ駆動型研究、ＡＩ駆動型研究の実施を促進**する。

生物多様性国家戦略2023-2030（令和５年３月31日閣議決定）

- 2002年度より開始された、ライフサイエンス研究の発展のために多様なバイオリソース整備を行う「ナショナルバイオリソースプロジェクト」において、時代の要請に応えたリソースの収集・保存・提供を推進するとともに、**利活用に向けたデータベースや付随情報の整備**に引き続き取り組む。

統合イノベーション戦略（令和７年６月６日閣議決定）

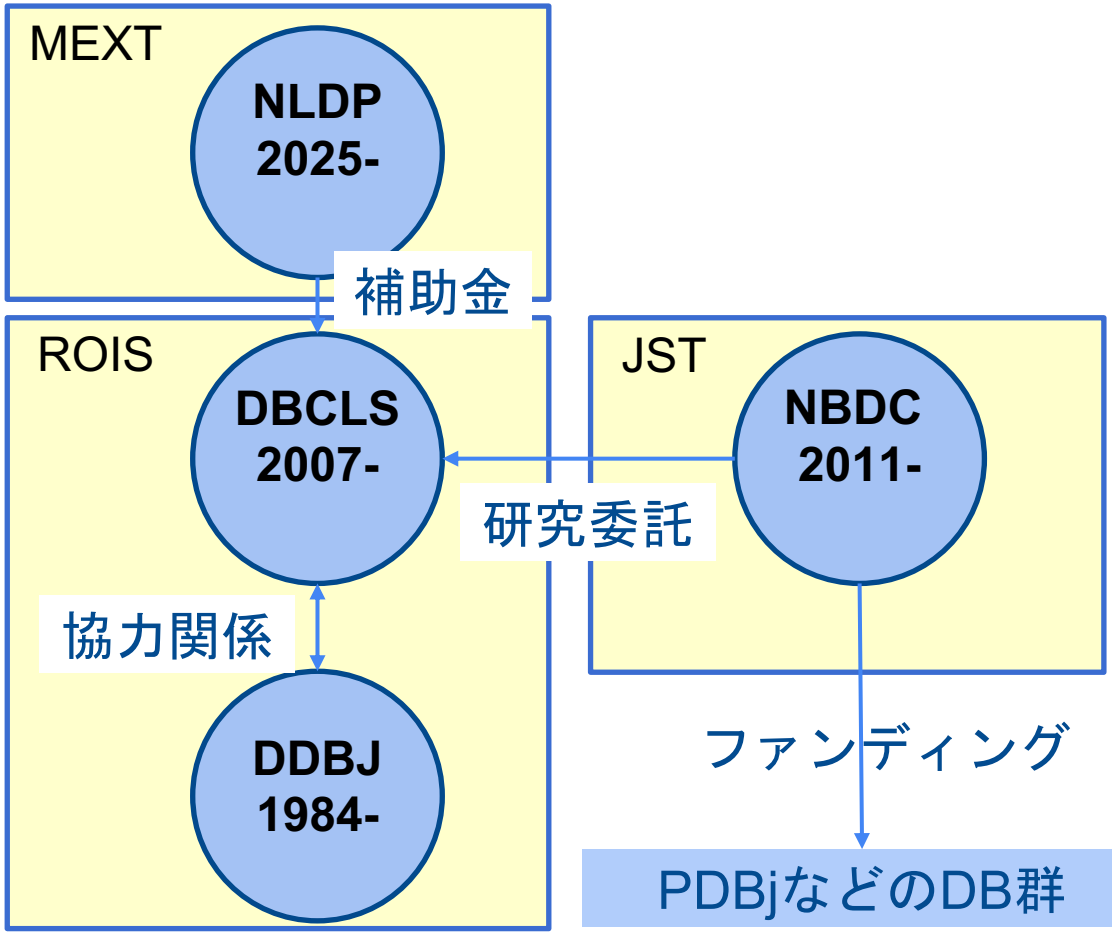
- バイオエコノミー拡大の源泉となる生命科学研究を支える人材育成、ライフコースに着目した研究等の基礎生命科学の振興、**データベース・バイオリソース・バイオバンク等の次世代情報研究基盤の整備・充実、それらを活用したデータ駆動型研究を推進**。

バイオエコノミー戦略（令和６年６月３日統合イノベーション戦略推進会議決定）

４．基盤的な施策　（１）基礎生命科学の研究力強化　　３）生命科学研究を下支えする研究基盤の強化

（前略）**ライフサイエンス系データベースの構築については、引き続き競争的なファンディングにより実施する**ことに加えて、大学共同利用機関法人情報・システム研究機構において、**重要なデータベースの安定的な維持・管理**や、**AIを用いた統合検索技術等のデータベース高度化のための技術開発**等の推進、ライフサイエンス系データベースの維持・管理・開発に必要な**バイオインフォマティクス人材の育成**に取り組む。

統合DBプロジェクトを推進するDBセンター群



JSTライフサイエンスデータベース統合推進事業（NBDC）

- ポータル事業（DBカタログ、DB横断検索、DBアーカイブ、RDFポータル、NBDCヒトデータベース、TogoVar、など）
- DB機能連携・統合化のための基盤技術開発（DBCLSに委託）
- 統合化推進プログラム（DICP）ファンディング事業
- 企画・広報（トーゴーの日、AJACS講習会など）

2023年度

- ポータル事業のうち研究要素の強いRDFポータル、NBDCヒトデータベース、TogoVarもDBCLSが担当

2025年度

- 基盤技術開発をNLDPに移管

2026年度

- RDFポータル、NBDCヒトデータベース、TogoVarをNLDPに移管

2027年度以降

- DBカタログなど他のポータル事業もNLDPに移管するか検討中
- ファンディングをどのような形で実施するか検討中

統合プロジェクト、DBセンターの成果(一部）

・**ライフサイエンス研究者向けAIツールなどの開発・ポータル事業**
アプリケーション開発：ヒトDB(提供申請約1300, 利用申請約600 – 半数以上が海外から)、
TogoVar(20DB, 9.3億バリアント, 25.8万人)、
PubCaseFinder(15DB, 200万データ)、TogoTV(2311動画, 300万回視聴, 1.1万人登録)、
DBカタログ(2,572DB)、DB横断検索(819DB)、アーカイブ(157DB)、など
データセット整備：TogoID(114DB, 52億IDペア)、RDFポータル(70DB, 1600億トリプル)、
PubAnnotation(1,700万文献, 650注釈プロジェクト)、TogoDX(20DB, 65属性)、など
ツール開発：SPARQL-proxy、SPARQList、Grasp、RDF-config、TogoStanza、TogoWS、など

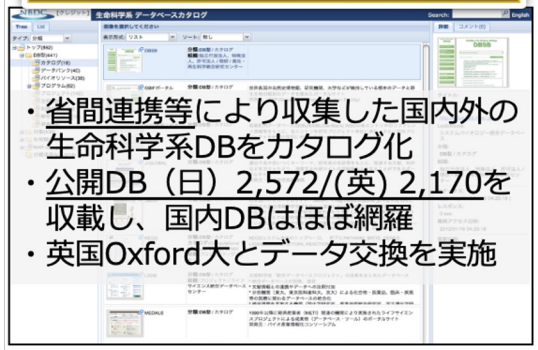
データセット整備やツール群開発は当初バイオインフォマティシャン用を想定していたが、その後AI用の基盤になることが判明。現在AI利活用に不可欠の存在となった

・**その他の活動**
国際連携(BioHackathon15回100人, BLAH9回50人, グラフサミット6回30人, 20ヶ国)、
国内連携(バイオハッカソン16回80人, Togothon157回60人)、
トーゴーの日シンポ(500名)、AJCAS(5回, 1,700名)、など

・**ファンディングにより構築されたDB群**
（本格型）PDBj, KEGG MEDICUS, jPOST, GlyCosmos, Shin-MassBank, SSBD, INTRARED, MicrobiomeDatahub
（育成型）ATTED-II, JoGo, DeepspaceDB, MIIB-AI, Cell IO, integMet, SSCV DB, PHi-C DB, Cura Toxii

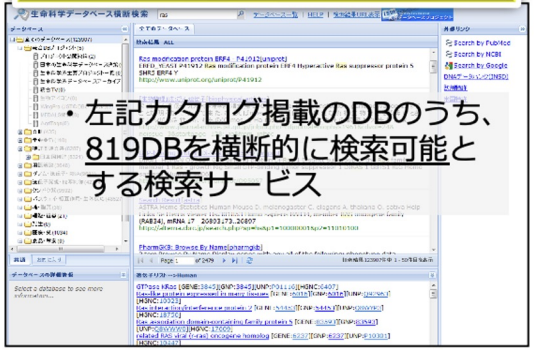
DBサービスの一例

生命科学系データベースカタログ



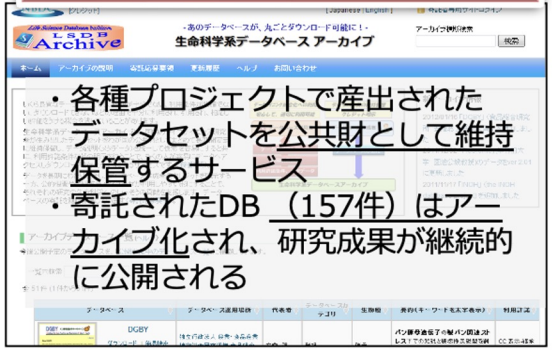
- 省間連携等により収集した国内外の生命科学系DBをカタログ化
- 公開DB（日）2,572/(英) 2,170を収載し、国内DBはほぼ網羅
- 英国Oxford大とデータ交換を実施

生命科学データベース横断検索



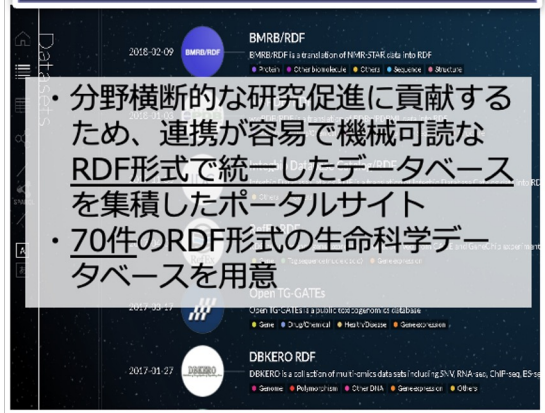
- 左記カタログ掲載のDBのうち、819DBを横断的に検索可能とする検索サービス

生命科学系データベースアーカイブ



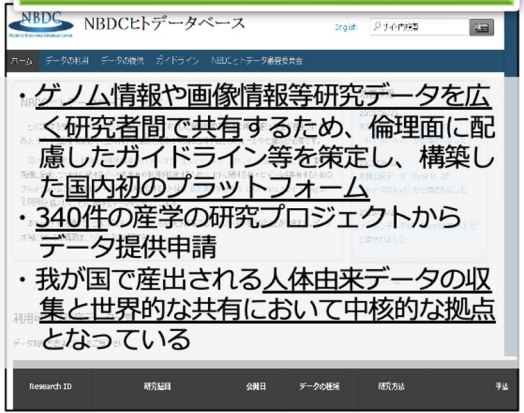
- 各種プロジェクトで産出されたデータセットを公共財として維持保管するサービス
- 寄託されたDB（157件）はアーカイブ化され、研究成果が継続的に公開される

RDFポータル



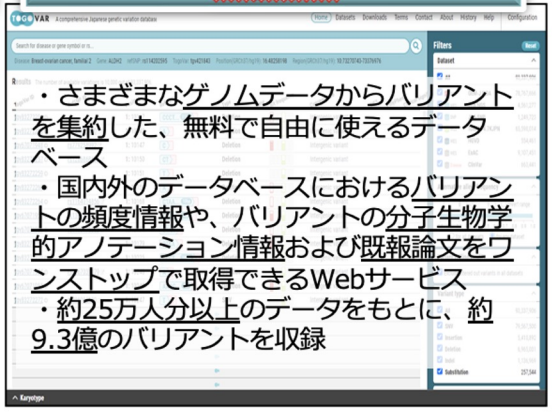
- 分野横断的な研究促進に貢献するため、連携が容易で機械可読なRDF形式で統一したデータベースを集積したポータルサイト
- 70件のRDF形式の生命科学データベースを用意

NBDCヒトデータベース



- ゲノム情報や画像情報等研究データを広く研究者間で共有するため、倫理面に配慮したガイドライン等を策定し、構築した国内初のプラットフォーム
- 340件の産学の研究プロジェクトからデータ提供申請
- 我が国で産出される人体由来データの収集と世界的な共有において中核的な拠点となっている

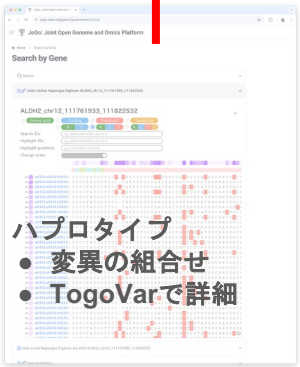
TogoVar



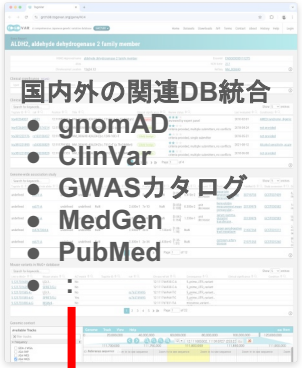
- さまざまなゲノムデータからバリアントを集約した、無料で自由に使えるデータベース
- 国内外のデータベースにおけるバリアントの頻度情報や、バリアントの分子生物学的アノテーション情報および既報論文をワンストップで取得できるWebサービス
- 約25万人分以上のデータをもとに、約9.3億のバリアントを収録

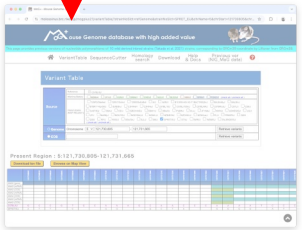
統合DBとそれに基づく外部DBとの連携の一例

TogoVar: 日本人ゲノムバリアント頻度の統合DB



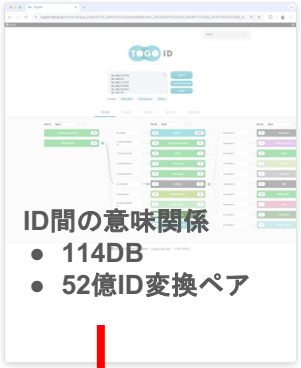
JoGo: ハプロタイプDB
(九州大学)



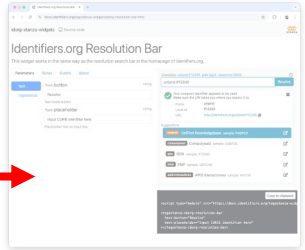


MoG+: モデルマウスDB
(RIKEN)

TogoID: DB間のID対応サービス

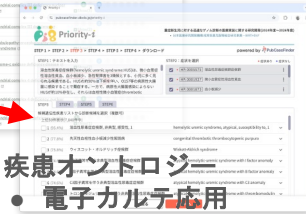


Id.org に変換機能提供
(欧州EBI)



DBCLS技術提供
(欧州EBI)

PubCaseFinder: 疾患検索統合DB



NanbyoData: 難病DB
Priority-i: 重症新生児ゲノム診断
(厚労省)

統合化事業により開発したツールの例

元データ

TogoDB (2007年〜)	**研究者のデータをデータベース化 - 小規模DB構築のお手本** 表計算ソフトのようなテーブル形式のデータを、誰でも容易に高度なデータベースとして公開できるサービス。データベースをRDFに変換することもできる。論文発表する新規DBの公開に使われたり、DBアーカイブで利用されてきた。データベースが基本的に備えるべき特質を追求したもの。
TogoWS (2008年〜)	**主要な公共データベースの利活用 - 外部DBのAPIによる利用を標準化** 国内外の主要な公共DBに対して、共通の検索・取得APIを提供しているサービス。結果を、テキストや、JSON、RDFなどに変換して取得できるほか、データの部分取得にも対応している。ウェブサービス化によるプログラミング言語非依存というパラダイムを打ち出したもの。
TogoGenome (2013年〜)	**RDFによる統合データベースの開発 - 大規模DB構築の実証実験** RDFのみで実装されたゲノムデータベース。全生物のゲノムという膨大な情報をRDFでも実用的なデータベースとして実装できることを実証し、RDFならではの機能として、遺伝子アノテーションから環境情報まで多様なデータの統合と、ファセット検索などの仕組みを開発した。
TogoStanza (2013年〜)	**分散化データの利活用を促進 - DB開発のお手本** ウェブブラウザで動作する汎用的かつ再利用可能な可視化モジュールを開発するためのツール。TogoGenomeを構築するにあたり、ウェブページ内の情報単位・可視化単位を他のウェブサイトでも容易に再利用するための汎用的な仕組みを提案したもの。同じようなモジュールを何度も再開発する必要がなくなり、DB開発コストを大幅に下げるとともに、統合化推進プログラム内の連携も創出した。各種外部サービスでも利用されている。
D2RQ Mapper (2014年〜) LODチャレンジ基盤技術部門最優秀賞 (2015年)	**RDB資産のRDF化を支援** 関係データベースに格納されているデータを、RDFのデータ構造にマッピングすることでRDF化およびSPARQLエンドポイント化するミドルウェア。
RDFポータル (2014年〜)	**全RDFデータを統合した知識グラフの構築 - 統合化推進から世界標準へ** 生命科学のRDFポータルサイトとしては、データの種類、データ量、ともに世界最大。DBCLSの各種サービスの共通データベースであり、あらゆるデータを共通の技術基盤に揃えて真に統合するためのプラットフォームを提供。

知識グラフ

知識グラフ

アプリ開発

SPARQL-proxy (2016年〜) LODチャレンジ基盤技術部門優秀賞 (2018年)	**SPARQLエンドポイントの安定運用 - 利用の増大に対応** RDFのデータベースであるSPARQLエンドポイントを安全に公開するためのミドルウェア。データベースの種類によらず共通の使いやすい検索インターフェイスを提供するほか、データの書き込みを拒否し読み出しだけ許可する、同じ検索にはキャッシュされた結果を高速に返す、大量の検索リクエストをジョブ管理しサーバの負荷を守る、などの機能をもつ。
SPARQList (2017年〜) LODチャレンジ基盤技術部門優秀賞 (2017年)	**SPARQLをアプリケーション用のAPIに転換 - 開発効率の最大化** SPARQLエンドポイントをバックエンドとして、再利用可能なウェブAPIを容易に構築するためのツール。複雑で長大なSPARQLクエリを引数を変えて呼び出したり、出力結果をJavaScriptでカスタマイズすることができる。以後のアプリケーションの開発効率を大幅に向上させたほか、ブラックボックスになりがちな検索クエリを公開することで透明性の確保にもつながった。Togoシリーズはじめ、様々なアプリケーションで利用されている基盤ツールの一つ。
TogoVar (2018年〜)	**日本人ゲノム変異データの統合 - 医科学データ統合への取り組み** 日本人のゲノム変異を網羅的に収集している。米国のgnomADの偏りを補完する存在で、自国のデータを自国内で責任を持って管理する意義もある。国内外のDBと連携し、バリアントの解釈に有用な約20のデータベースを統合。
RDF-config (2019年〜)	**複雑なRDFグラフを理解可能な表現に転換 - 認知の限界への挑戦** RDFのグラフごとに、コアとなるデータモデルを樹形のYAML形式で記述するという近似的な工夫を取り入れることにより、データ構造のスキーマ図、任意のSPARQLクエリ、ShExのバリデーションなどを自動生成するツール。これまで手作業で行っていた時間のかかる作図やクエリ開発が一瞬で可能になった。
MetaStanza (2020年〜) LODチャレンジデータ活用部門優秀賞 (2021年)	**典型的な可視化をノーコードで実現 - 可視化のコモディティ化** TogoStanzaを使って、特に汎用的に利用されうる可視化アプリを整備している。高機能なテーブルから棒グラフなどの各種チャートまで、コードを一切書くことなく、ウェブアプリを構築できる。WebComponentsの属性で細かなパラメータを調節できるほか、CSSのテーマをカスタマイズすることで埋め込み先のサイトのデザインと合わせることができる。今後はMetaStanzaを組み合わせた際に連動する仕組みを充実させることで、入力から出力までをカバーする高度なウェブアプリの開発を実現したい。
TogoID (2021年〜)	**古典的なID変換の課題を解決 - 新規DBペアもオープンソースで開発** 主要なDBのID間の関係を収集している。データベースを辿るにはIDの関係が必須であり、TogoDXでも活用されている。これまでのID変換サービスでは対応していないDBが多かったが、オープンソースで新規なDBを追加できる仕組みを提供した。今後コミュニティの要望に応えてより多くのID関係を取り込む予定。ウェブアプリだけでなくAPIも完備していること、ID変換の意味をオントロジーで整備したことも特徴といえる。

ファンディングによるDB開発
－ NBDC統合化推進プログラム（本格型）研究開発課題 －

研究代表者	所属・役職	研究開発課題	データベース名	研究開発期間
石濱　泰	京都大学　教授	jPOST prime：コミュニティ連携を基盤とするプロテオームデータベース環境の実現	jPOST	令和5年4月 ～ 令和10年3月
松田　史生	大阪大学　教授	次世代低分子マススペクトルデータベース シン・マスバンクの構築	Shin-MassBank	
大浪　修一	理化学研究所　チームリーダー	バイオイメージングデータのグローバルなデータ共有システムの構築	SSBD：database	令和4年4月 ～ 令和9年3月
粕川　雄也	理化学研究所　チームリーダー	細胞ごとの多様な活性やゲノム変異・多型との関連を探索できるシスエレメント・データベース	INTRARED （ChIP-Atlas, fata.bio）	
金久　實	京都大学　特任教授	ヒトゲノム・病原体ゲノムと疾患・医薬品をつなぐ統合データベース	KEGG MEDICUS	
木下　聖子	創価大学　副所長	異分野融合を志向した糖鎖科学ポータルのデータ拡充と品質向上	GlyCosmos Portal	
栗栖　源嗣	大阪大学　教授	蛋白質構造データバンクのデータ駆動型研究基盤への拡張	PDBj	
森　宙史	国立遺伝学研究所　准教授	マイクロバイオーム研究を先導するハブを目指した微生物統合データベースの特化型開発	MicrobiomeDatahub	

ファンディングによるDB開発
- NBDC統合化推進プログラム（育成型）研究開発課題 -

研究代表者	所属・役職	研究開発課題	データベース名	研究開発期間
白石　友一	国立がん研究センター　分野長	大規模言語モデルを活用した病的スプライシング変異データベースの自律的構築	SSCV DB	令和7年4月 ～ 令和10年3月
新海　創也	理化学研究所　上級研究員	4Dゲノム状態の理解と可視化を支援するデータベースの構築	PHi-C database（開発中）	
水野　忠快	東京大学　助教	化合物と個体をつなぐ毒性病理画像データベースCuraToxiiの開発	Cura Toxii（開発中）	
池田　和由	理化学研究所　上級研究員	AI駆動型データキュレーションによる持続可能な中分子相互作用統合データベースの開発	MIIB-AI（開発中）	令和6年4月 ～ 令和9年3月
尾崎 遼	理化学研究所 チームディレクター	細胞レベルの機能・表現型と遺伝子発現を関連付ける「Cell IO」データベースの開発	Cell IO（開発中）	
早川 英介	九州工業大学　准教授	創発的再解析のためのメタボローム統合データベース	integMET（開発中）	
大林　武	東北大学　教授	非モデル植物のための遺伝子ネットワーク情報活用基盤	ATTED-II	令和5年4月 ～ 令和8年3月
長﨑　正朗	九州大学　教授	日本人塩基配列情報の公開可能なゲノム・オミクス情報基盤による双方向型研究教育データベース開発と国際連携	Japanese Open Genome Omics Platform	
Vandenbon Alexis	京都大学　准教授	空間オミックスデータ解析用データベースの開発	DeepSpaceDB	

外国のDBセンターの状況

米国

- NCBI
  - 人員 341人 (2025年)
  - 予算 290億円 (2025年)
    - 推計：FY2017がNLM予算の約45%
  - ストレージ >100PB (SRAだけで70PB) (2025年)
- CZIやAllen Instituteなど
  - 多様な細胞種でのRNA-seqなど大規模データを集積、細胞シミュレーションを標榜

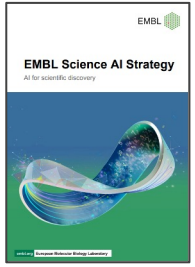
https://virtualcellmodels.cziscience.com/
https://alleninstitute.org/

- NAIRR
  - 生命科学外も含めた、大規模データの分散型の集積と研究者のためのAIインフラの提供

https://nairrpilot.org/

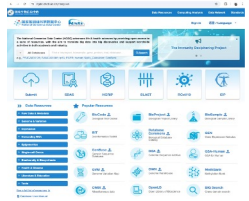
欧州

- EMBL-EBI
  - 人員 657人 (2024年)
  - 予算 174億円 (2024年支出実績)
  - ストレージ 390PB (2024年)
  - EMBLのウェット実験系・EBIのデータベースとともに、AIストラテジーとして、「AIの理論的な基盤構築」「AIレディなデータ整備」「自動実験ワークフローによる研究の加速」を3つの主要な柱として提唱



https://www.embl.org/documents/document/embl-ebi-highlights-report-2024/
https://www.embl.org/documents/document/a-european-strategy-for-ai-in-science/

中国



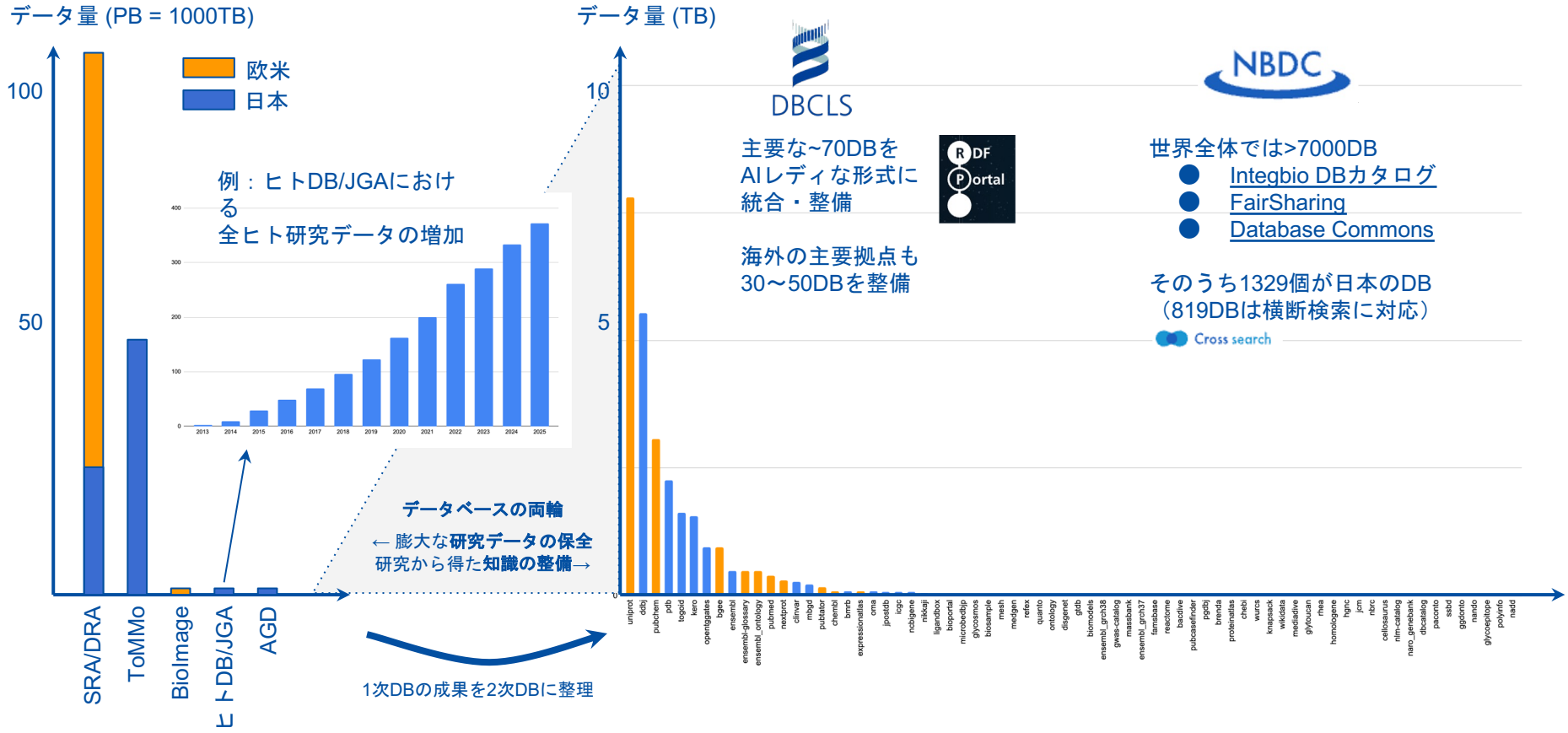
- NGDC (BIG)
  - 人員 215名 (2025年)
  - 予算 100億円 (2025年)
  - ストレージ 108PB (2025年)
  - マルチオミックスデータの集積に注力、国際的認知度向上を目指す

https://ngdc.cncb.ac.cn/

日本

- DDBJ/DBCLS
  - 人員 70名 (2025年)
  - 予算 15億円 (2025年)
  - ストレージ 50PB (2025年)
  - 知識グラフによる統合技術開発に長年取り組んできた
  - 国内の各DBは様々な組織で分散開発、公開されないデータも
  - 生命科学研究の促進に戦略的なAI開発と国際的なデータ連携

統合化事業により統合的に扱えるようにしたデータベースの数とデータ量



１次データベース（ゲノム配列や画像等の生データ）　**２次データベース**（研究者/AI用の知識ベース）

2025年11月現在

我が国の課題 –欧米との比較、AI連携など–

- 運営、体制、人員、計算機資源、その他：
  データシェアリングポリシー（義務化は欧米に比べて弱い）
  DMP（実践が不十分、評価の仕組みの問題）
  スパコン計算資源（現状でも計算資源が大きく不足、今後のAI解析には耐えない）
  ディスク容量（大幅に不足、例えば欧米のDBセンターでは100PB超、EBIは390PB）
  DB・BI人材（大幅に不足、欧米のDBセンターは数百人規模）
  DBの多くは一元管理、（ファンディングではなく）DBセンターでの内製
  AIプロジェクトとDB構築の一体的戦略
  知識グラフの活用は日本がリード
- 技術的課題(一部)：
  AIとの連携強化、一次DB（レポジトリ）と二次DB（知識ベース）との連携
  高度なアノテーションおよびAI導入によるDBの高度化
  日本人の疾患研究や国内の創薬に繋がるデータ統合と解析環境の整備

今後の検討課題と事業展開の方向性（提案）-統合からAI基盤へ-

- AI for Science時代におけるライフサイエンス研究データの集積・活用のあり方？
  - **知識グラフを基盤とした、AIとの連携強化**：AIを活用したDBの効率的な構築、AIへのデータや知識の提供のためのシステム開発強化
  - **一次DBと二次DBとの一体的運用**：DDBJ事業などの一次DBとの統合・連携強化
  - **新たな連携拡大**：AI戦略DB戦略に基づく、AI活用プロジェクトや多様・網羅的なデータ生産プロジェクトとの緊密な連携

- 欧米に伍した研究基盤としてのDB、（AI・IT非専門家のための）解析環境、ファンディングのあり方
  - **計算機資源および人材の大幅拡充**：大規模なDB群の一括運用、統合解析環境構築、DBの高度アノテーションと内製化促進、などのため
  - **DBとツールの統合解析環境の整備**：AI・IT非専門家のための統合解析環境整備
  - **DBの内製化の強化**：ファンディングから内製化へ
  - **基盤的・多様なDBのテナント型一元管理・運用**：AI基盤強化の一環として基盤的・多様なDBの集約と一元管理

- 持続可能な体制の構築に向けて
  - **ナショナルDBセンターとしての事業展開**：分散したDBセンターの集約・大規模化によるAI基盤の整備、DB戦略（オープン・クローズ、国際協調、データ提供インセンティブの付与の仕組み、など）の立案と実施

今後の事業展開の方向性（提案）-統合からAI基盤へ-

1次2次データの運用とAI利用に向けたデータ基盤の開発

1次データのレポジトリ事業を高度化

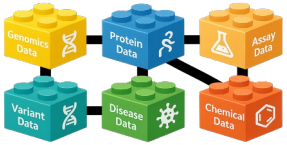
- NBDCヒトデータベースやDDBJへの**データ登録・利用申請業務をAIで支援**
- **ペタバイト規模のマルチオミックスデータを運用**するレポジトリシステムを開発
- テナント型の開発で統合化推進プログラム等のデータベース構築に必要な技術を提供

2次データの生成・提供とLLM・LMMへの対応

- **AI学習用・検証用の高品質な正解セット**を数千億項目の知識グラフから生成
- AIの推論結果に対しデータベースにある科学的根拠（エビデンス）を付与
- データを定期的に更新し、最新の知見をAIエージェントと研究者に提供
- メタデータ不足をAIで補正する技術開発とデータ再解析で付加価値の創出
- 実験由来データとAI由来データを区別し、安全に利用できるシステムを開発

大規模1次DBの開発・運用

AIの基盤データ



生命科学の基盤となるデータベースと大規模AIデータ解析環境の提供

ナショナルデータベースセンター

- **日本の全生命科学研究データを集積する共通データベース基盤**を運用
- 国の研究費によるデータを保全するためのDMPを策定・実施
- 国際連携によるAI時代における生命科学データの標準化

大規模データサイエンスのインフラ構築

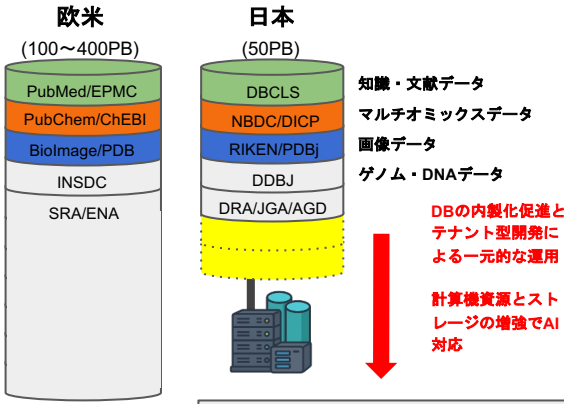
- **大規模データを大規模計算機に配置**し共用施設として利用者に提供
- GPU・CPU増強によるゲノム解析・マルチオミックスAI解析の人材育成
- ゲノム医療・創薬・バイオ産業などに向けた大規模解析のための環境を整備

次世代の大規模ゲノムデータ解析基盤整備

- ゲノムグラフを用いた完全長ゲノム解析と構造多型解析技術の提供
- スパコンを活用した高速解析ワークフローの開発と提供

大学共同利用機関法人としての大規模データサイエンス基盤

ゲノムグラフ構築・解析支援



マウスの肝臓のRNA-seqのデータを探したい。マウスの系統、オスなのかメスなのか、肝臓のどの部位から得られたデータかで分類したい。

既存の手法では困難

1. DDBJの検索ページを開く
2. 学名で生物種を指定
3. 種別でRNA-seqを指定
4. 肝臓 (liver) で検索
5. 数千件のリスト
↓
6. それぞれ文献を参照
7. 性別や臓器の部位を確認
8. 手作業で絞り込み

AI統合検索

- 6408件あります。
- 論文を参照して性別と部位で整理して。
- 論文に記載のあったものが3802件でした。
- CSV形式の表にして。