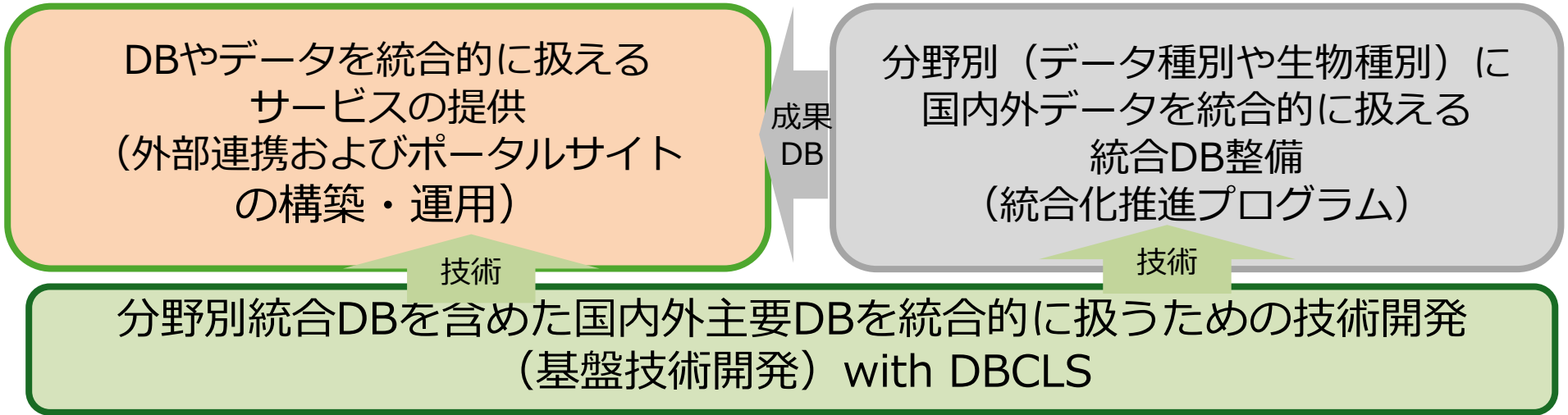
2026年10月5日
トーゴーの日
シンポジウム2026

NBDCの15年の活動を振り返って

NBDCライフサイエンスデータベース特別主監
高木利久

NBDC設立の経緯と目的

- JST BIRD(2001～2011)　バイオインフォマティクス推進センター事業
- ROIS DBCLS(2007～)　統合データベース事業
- JST NBDC(2011～)　ライフサイエンスデータベース統合推進事業
- 文科省NLDP事業(2025～)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
- NBDCの目的
  - データやDBの共有・統合による研究者の利便性向上
  - データの価値の最大化、データ駆動型の生命科学への貢献
  - 我が国の中核センターとして機能する(外部との連携)
- NBDCの3つの柱

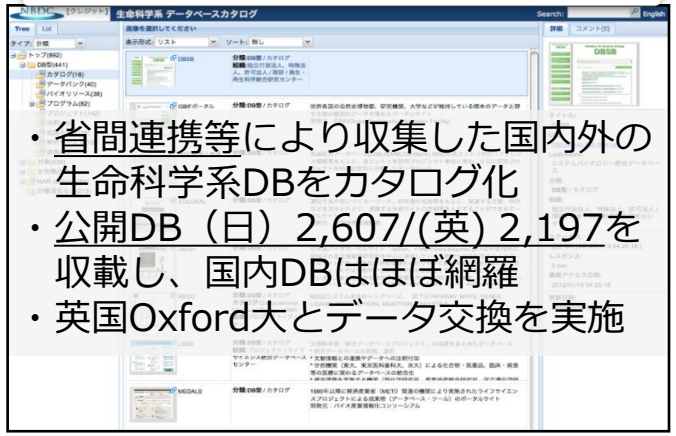


ライフサイエンスデータベース統合推進事業の成果（一部）

- DBサービス開発：
  - ヒトDB(提供申請約1600, 利用申請約700 – 半数以上が海外から)、
  - TogoVar(22DB, 11億バリアント, 25.8万人)、PubCaseFinder(17DB, 33万症例, 疾患検索38万回/年)、
  - DBカタログ(2,607DB)、DB横断検索(837DB)、アーカイブ(157DB)、
  - TogoTV(2347動画, 367.7万回視聴, 11,697人登録)、など
- データセット整備：
  - TogoID(119DB, 86億IDペア)、RDFポータル(70DB, 2299億トリプル)、
  - PubAnnotation(1,700万文献, 744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
- 国際連携
  - BioHackathon16回100人, BLAH10回50人, グラフサミット6回30人, 20ヶ国、など
- 国内連携
  - バイオハッカソン17回80人, Togothon168回60人、など
- トーゴーの日シンポ(約500名)、AJCAS(5回, 1,700名)、など
- 統合化推進プログラムにより構築されたDB群
  - （本格型）PDBj, KEGG MEDICUS, jPOST, GlyCosmos, Shin-MassBank, SSBD, INTRARED, MicrobiomeDatahub
  - （育成型）ATTED-II, DeepspaceDB, MIIB-AI, Cell IO, integMet, SSCV DB, PHi-C DB, Cura Toxii, JoGo-Next, RNATDB, MDDB-AI Hub

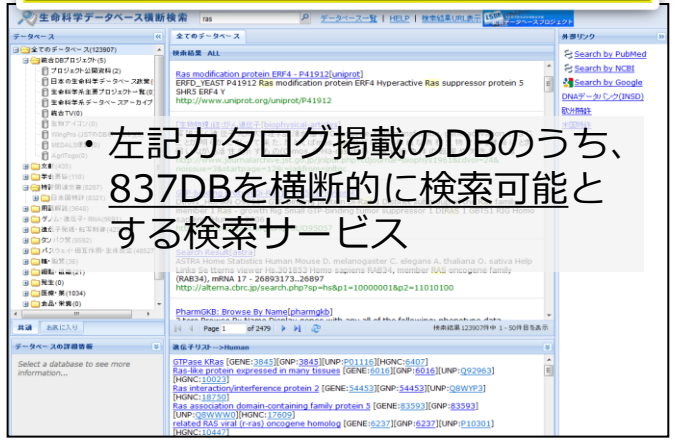
開発されたDBサービスの一例

生命科学系データベースカタログ



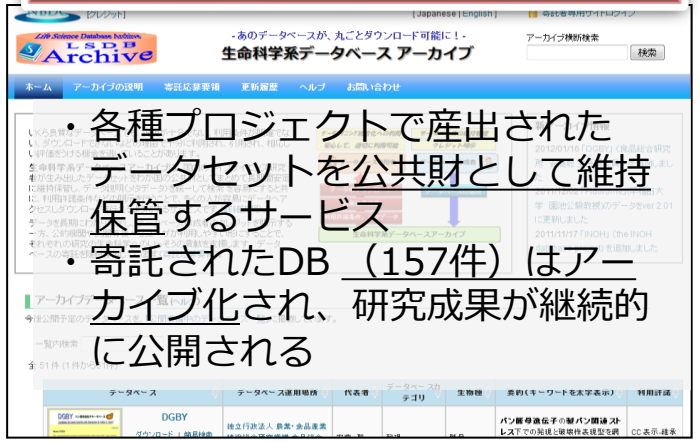
・省間連携等により収集した国内外の生命科学系DBをカタログ化
・公開DB（日）2,607/(英) 2,197を収載し、国内DBはほぼ網羅
・英国Oxford大とデータ交換を実施

生命科学データベース横断検索



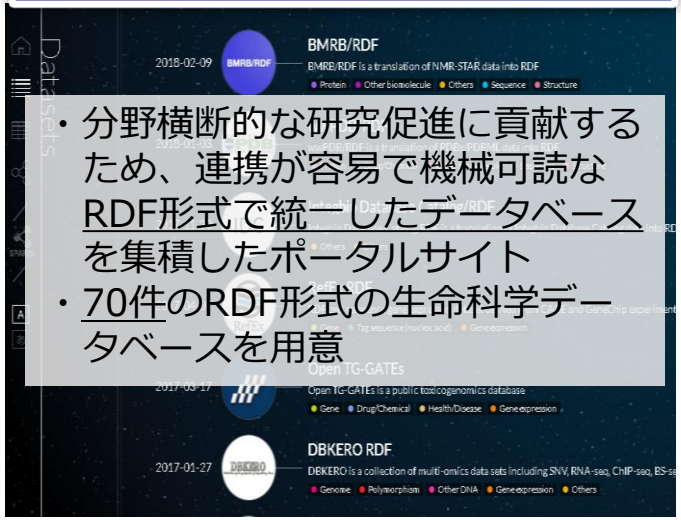
・左記カタログ掲載のDBのうち、837DBを横断的に検索可能とする検索サービス

生命科学系データベースアーカイブ



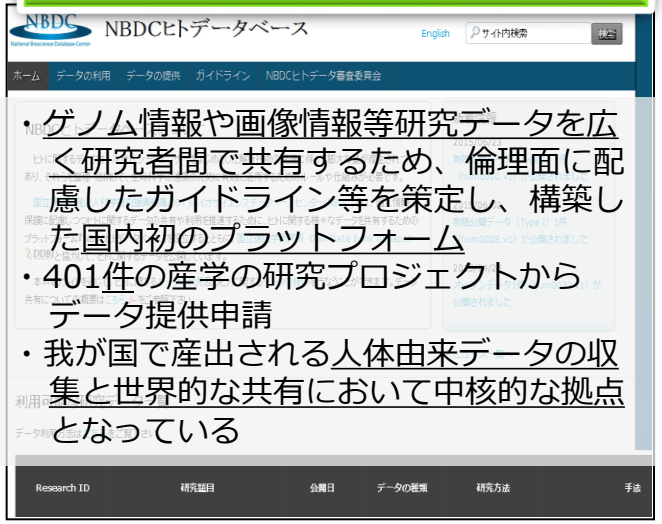
・各種プロジェクトで産出されたデータセットを公共財として維持保管するサービス
・寄託されたDB（157件）はアーカイブ化され、研究成果が継続的に公開される

RDFポータル



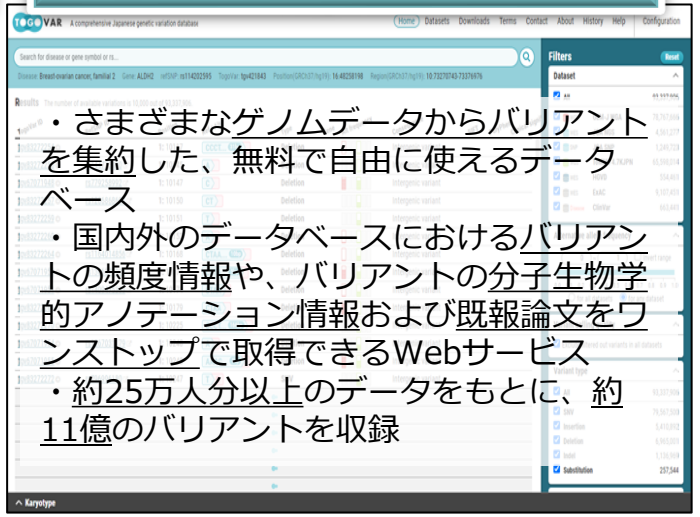
・分野横断的な研究促進に貢献するため、連携が容易で機械可読なRDF形式で統一したデータベースを集積したポータルサイト
・70件のRDF形式の生命科学データベースを用意

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
・401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11億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



国内の頻度DB統合
- ToMMo
- BBJ
- MGeND
- JGA
- :



ハプロタイプ
- 変異の組合せ
- TogoVarで詳細

JoGo: ハプロタイプDB
(九州大学)

TogoID: DB間のID対応サービス

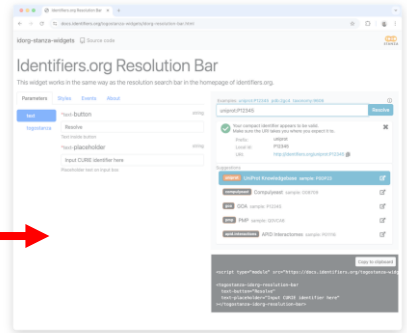


ID間の意味関係
- 119DB
- 86億ID変換ペア



海外サービス機能拡張
- TogoIDのAPI利用
- 関連DBも検索可

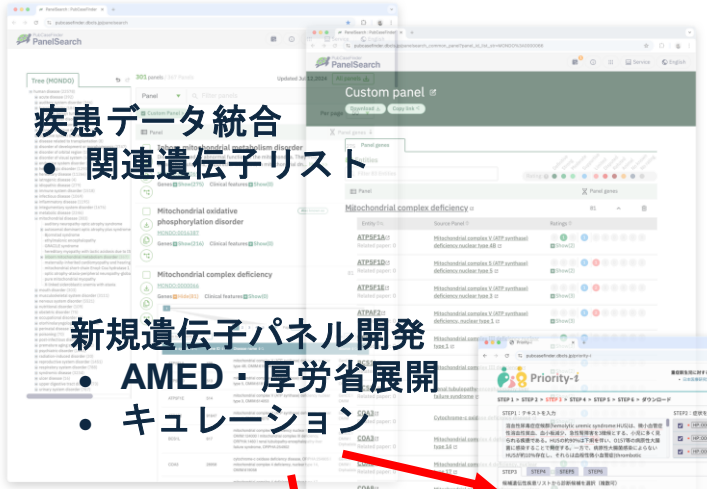
Id.org に変換機能提供
(欧州EBI)



汎用モジュール化
- TogoStanzaを利用

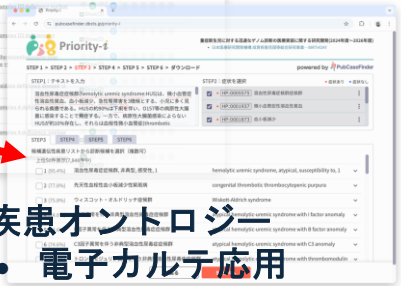
DBCLS技術提供
(欧州EBI)

PubCaseFinder: 疾患検索統合DB



疾患データ統合
- 関連遺伝子リスト

新規遺伝子パネル開発
- AMED・厚労省展開
- キュレーション

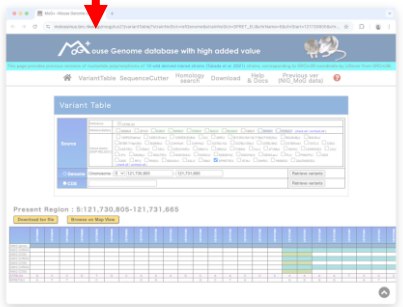


疾患オントロジー
- 電子カルテ応用
- 診断支援



難病オントロジー
- 疾患関連DB統合

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



実験用モデルマウス検索
- ヒトと同じ変異を持つ

MoG+: モデルマウスDB
(RIKEN)

統合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。

NBDC統合化推進プログラム（本格型）によるDB開発

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 チームディレクター	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 チームディレクター	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	

NBDC統合化推進プログラム（育成型）によるDB開発

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
長﨑 正朗	九州大学　教授	JoGo-Next：国際標準に向けたヒトハプロタイプデータベース構築とAI活用	JoGo-Next（Joint Open Genome and Omics Platform Next）	令和8年4月 ～ 令和11年3月
浜田 道昭	早稲田大学　教授	AI駆動型RNA創薬を支援するマルチモーダルデータベースの構築	RNATDB	
松永 康佑	埼玉大学　准教授	MDデータの共有と再利用を支えるデータ・モデル基盤の構築	MDDB-AI Hub	
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	
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	CuraToxii	
池田　和由	理化学研究所 ユニットリーダー	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	MIIDB-AI	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	
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	

ライフサイエンス分野のデータの課題

- データが多種多様で、文脈依存性、不均質性、複雑性、などの面倒な性質
- 世界には1万から2万ものデータベース
- 測定装置の革新により新たなデータが続々と登場

- データ提供が義務化されていない（インセンティブの不足も）
- どこにどういうDBがあるかわからない
- しかもそれらはフォーマットもオントロジーも権利関係もバラバラ
- 複数のデータベースの統合的利用が困難
- データ収集や標準化のコストが大きい

- これらの問題を解消するために統合データベース事業が2006年にスタート

4省連携によるデータベース統合化の4つのステップ

第一期中（平成25年度まで）に実施予定

（１）ポータル構築連携
「生命科学系データベースカタログ」への記載、リンクの実施

新センターのポータルサイトに、各省のデータベースに関する情報を記載・登録し、
各省のデータベースに対して、**リンクが行われる状態**を実現する。

（２）横断検索連携
「横断検索サーバー」の設置、一括横断検索の実現

各省における適切なサイトに、横断検索サーバーを設置するとともに検索インデックスを統一化し、
各省のデータベースに対して、**相互に一括横断検索が可能となる状態**を実現する。

（３）アーカイブ構築連携
「統一フォーマット」でのダウンロードの実現

新センターのポータルサイトに、各省のデータベースをガイドラインに沿ってアーカイブ化して保存し、
各省のデータベースに対して、**統一形式でのデータダウンロードが可能となる状態**を実現する。

第二期以降を想定

（４）データベース構築連携
「データベースの再構築」による統合の実現

新センターのポータルサイトに、各省から譲り受けたデータベースを再編して収容し、
各省のデータベースが、**統一化したフォーマットで再構築された状態**を実現する。

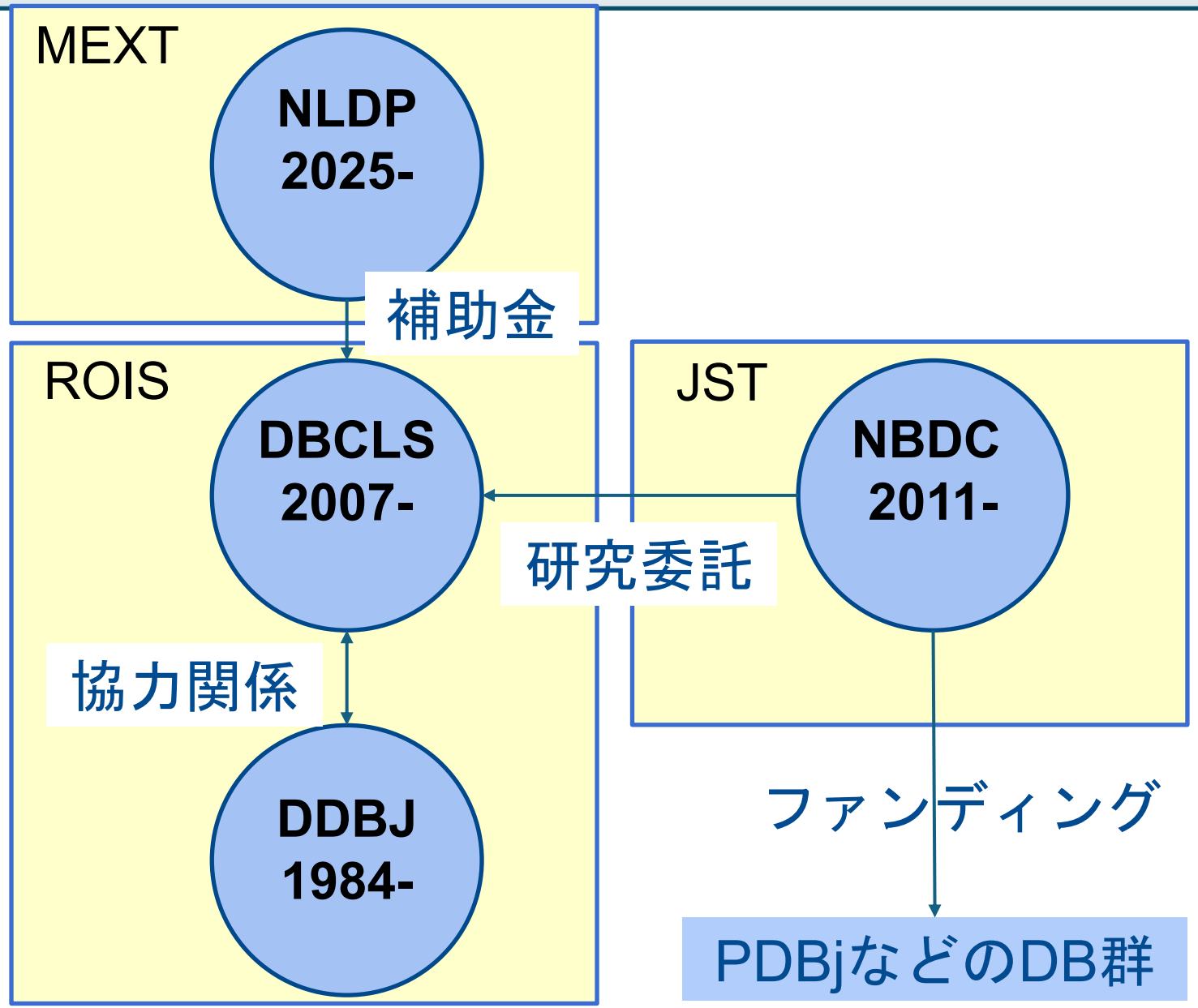
FAIR原則(Findable, Accessible, Interoperable, Reusable)に対応

AI-readyなデータ

- FAIR原則
  - グローバルで永続的なID
  - 適切な手続きや標準的なプロトコル
  - 標準的なフォーマットやオントロジー
  - 適切なメタデータ（実験条件など）やデータ使用条件

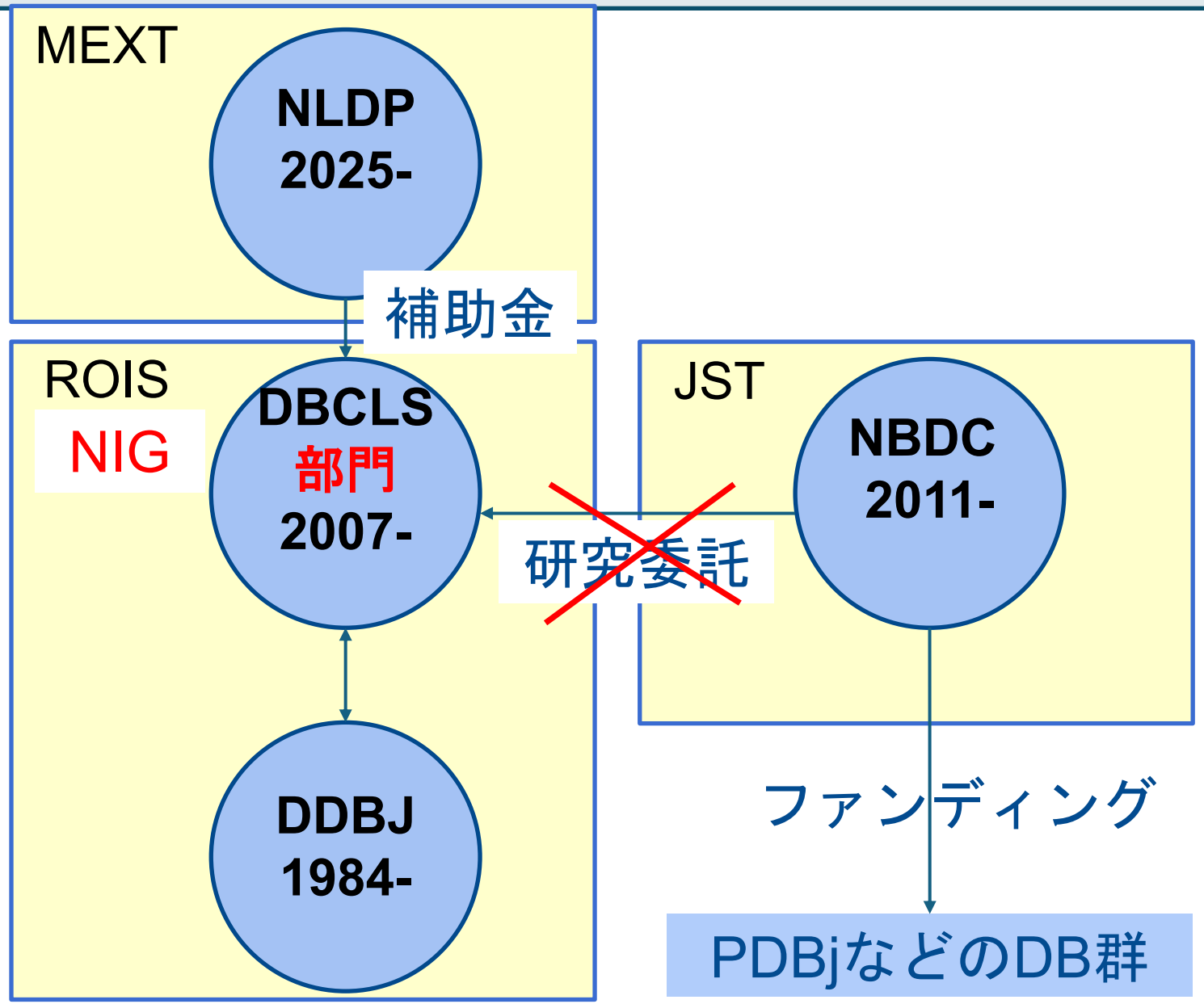
- これにプラスして
  - 国際標準なIDやオントロジー
  - 実験条件、メタデータなどの構造化とコンテキスト保持
  - AI解析用データへの変換
  - 高精度で信頼度の高いアノテーション
  - データの品質管理
  - ネガティブデータの整備
  - 統計的バイアスの除去
  - AI倫理や規制のクリア

NBDCからNLDP(DBCLS)へ



- 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

NBDCからNLDP(DBCLS)へ



- 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
  - DBCLSは遺伝研の一部門に
- 2027年度以降

これからの課題

- AI-readyなデータの拡大に向けて
  - ナショナルセンターの整備
  - データシェアリングポリシーの整備、データ提供インセンティブの設計・実装
  - 一次DBと二次DBとの一括管理
  - 高精度アノテーションのためのAI技術開発
  - DICP（統合化推進P）、DB整備拠点、大規模データ生産拠点との連携強化
  - FIAR原則準拠データの拡大
  - 知識グラフ基盤とAIとの連携強化
  - 計算機資源（処理能力、記憶容量）の大幅な増強
  - 人材育成
- 研究者のためのDBの整備
  - この後のパネルディスカッションで

トーゴーの日シンポジウム2026 パネルディスカッション　論点(1)

前半「AIが研究する時代の研究や研究者のあり方について」

(1)「AIが研究する時代」という言葉の意味をどう捉えるか？
「テーマ選び、仮説生成、実験デザイン、実験、検証、論文作成」のどこをAIが？
これらのステージすべてを担えるのか？そのような時代は来るか？来ないか？
来るとしたらいつ頃？

(2)そのような時代には研究スタイルや研究者はどうなっているか？
現在の研究のやり方と本質的にどこがどう違ってくるのか？
そういう時代の研究者は何をメインにする人？
若手研究者はそう言う時代に向けて、これから何を身につけるべきか？

トーゴーの日シンポジウム2026 パネルディスカッション　論点(2)

後半「AIが研究する時代のDBのあり方について」

(3) AIが研究する時代に必要なDBは？

・AIのためのDBはどうあるべきか？、AI-readyなデータとは？

・研究者も依然としてDBを使うとすれば、それは、どのようなDBであるか？

• 現在は1万から2万ものDBがあるが、その数や内容は大きく変わるのか？

(4) 上記のDBの開発や利用のために必要なことは？

データシェアリング、データ提供のインセンティブなど制度面

DB開発や利用の技術面

DBセンターのあり方などの体制面

(5)適宜フロアーからの質問受付