ライフサイエンスデータベース統合推進事業

事業報告書

令和 3 年 10 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

目次

第 1　事業概要および事業評価経緯 1
第 2　取組み内容および成果 5
I. 外部連携およびポータルサイトの構築・運用 5
（1）省間連携等によるデータベース統合化およびポータルサイト構築・運用 5
（2）ヒトデータの統合と利活用推進 10
（3）その他の連携 15
（4）まとめ（I. 外部連携およびポータルサイトの構築・運用） 18
II. 統合化推進プログラム 19
（1）研究開発推進マネジメント 19
（2）各研究開発課題による統合化 22
（3）利活用に向けた取組みと利用状況 26
（4）課題間連携による統合化 29
（5）まとめ（II. 統合化推進プログラム） 30
III. 基盤技術開発 31
（1）開発概要と重点的取組み 31
（2）統合データ活用に向けたアプリケーション開発 33
（3）アプリケーションを支えるデータ基盤・ミドルウェア 38
（4）国内外との連携・情報発信 41
（5）まとめ（III. 基盤技術開発） 43
IV. 事業全体に共通する取組み 44
（1）広報およびデータベース講習会 44
（2）NBDC 運営委員会提言への対応状況 46
（3）まとめ（事業成果） 48

第１　事業概要および事業評価経緯

事業目的・趣旨

我が国におけるライフサイエンス研究の成果が、広く研究者コミュニティに共有且つ活用されることにより、基礎研究や産業応用研究につながる研究開発を含むライフサイエンス研究全体が活性化されることを目的として、科学技術振興機構（JST）では、平成 23 年 4 月にバイオサイエンスデータベースセンター（NBDC）を設置した。基礎・応用を含む研究全体の活性化に貢献するため、オープンサイエンスの動向を踏まえ、我が国のライフサイエンス分野の研究成果が広く共有・活用されるよう、様々な研究機関等によって作成されるデータベース（DB）の統合を推進している。

経緯：センター発足～前回事業評価まで（前期）について

NBDC は、「ライフサイエンス分野における我が国全体の恒久的且つ一元的な統合データベース」についての方針である「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 DB タスクフォース）（平成 21 年 5 月）を受けて発足し、同報告書を踏まえた、1）DB 統合に向けた戦略の立案、2）DB 統合システムおよび公開のためのインタフェースとしてのポータルサイトの構築・運用、3）バイオ関連データベースの統合化の推進（統合化推進プログラム）、4）データベース統合化基盤技術の研究開発（基盤技術開発）、の 4 つの事業の柱によって、DB の統合を推進してきた[1]。

センター運営 6 年目の平成 28 年度には、発足以降の事業活動内容に対する評価（事業評価）を実施し、「事業計画や運営体制、事業の推進状況や成果等、全体として優れた成果を出している」等の評価を受けた[2]。さらに、事業評価結果等を踏まえて平成 29 年度以降に活動を拡大・強化すべき活動内容について、NBDC 運営委員会から提言[3]を受けた。

事業概要

平成 29 年度以降も、センター発足当初からの事業目的に沿って、日本のライフサイエンス研究の成果であるデータや DB が広く研究者コミュニティに共有・活用されるよう、DB の統合利用のための研究開発とサービス提供を実施した。

JST の第 4 期（期間：平成 29～令和 3 年度）の中長期目標・中長期計画[4]において、本事

---

[1] 詳細は平成 23～28 年度の事業活動内容をまとめた「ライフサイエンスデータベース統合推進事業 事業報告書」（平成 28 年 11 月）を参照。URL：https://biosciencedbc.jp/gadget/unei/jigyou_houkoku.pdf

[2] 詳細は「ライフサイエンスデータベース統合推進事業 平成 23 年度～平成 28 年度 事業評価報告書」（平成 29 年 3 月）を参照。URL: https://biosciencedbc.jp/gadget/unei/jigyou_hyouka_23_28.pdf

[3] 詳細は「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）を参照。URL: https://biosciencedbc.jp/gadget/unei/unei_teigen.pdf

[4] URL: https://www.jst.go.jp/all/about/jigyou.html

業の目標等は以下のように定められている。

（ライフサイエンスデータベース統合の推進）
我が国におけるライフサイエンス研究の成果が、広く研究者コミュニティに共有され、活用されることにより、基礎研究や産業応用につながる研究開発を含むライフサイエンス研究全体の活性化に貢献するため、文部科学省が示す方針の下、様々な研究機関等によって作成されたライフサイエンス分野データベースの統合に向けて、オープンサイエンスの動向を踏まえた戦略の立案、ポータルサイトの拡充・運用及び研究開発を推進し、ライフサイエンス分野データベースの統合に資する成果を得る。

［達成すべき成果（達成水準）］
- ライフサイエンスデータベース統合化の基盤となる研究開発、分野毎のデータベース統合化及び統合システムの拡充にオープンサイエンスの観点から取り組むこと。
- ライフサイエンスデータベースに関連する府省や機関との連携等に取り組むこと。
- 連携、データ公開及びデータ共有の進展並びにデータベース利活用の観点から、ライフサイエンス分野のデータベースの統合に資する成果やライフサイエンス研究開発の活性化に資する成果を得ること。

国（センター発足経緯）および JST（中長期目標・計画）のミッションを踏まえ、ライフサイエンス分野のデータや DB を利用者が効率的・効果的に扱えるようにするため、次の 3 つの活動の取組みを組み合わせて事業を推進している。

I. 外部連携およびポータルサイトの構築・運用

日本のライフサイエンス研究成果であるデータ・DB を統合的に扱えるウェブサービスを提供。各省の公的研究資金で作成された DB を統合的に扱える省間連携等によるサービスの他、個人情報に配慮が必要な人体由来データ（ヒトデータ）を統合的に扱うためのウェブサービス、および、これらのサービスを背景とした利用者を含む外部との連携を実施。

II. 統合化推進プログラム

分野別（データ種別や生物種別）の統合 DB 整備。国内のライフサイエンス研究等によって産出された研究データを広く収集し、より多くの多様な研究者にとって価値のあるデータとして提供するため、公募による開発を実施。

III. 基盤技術開発

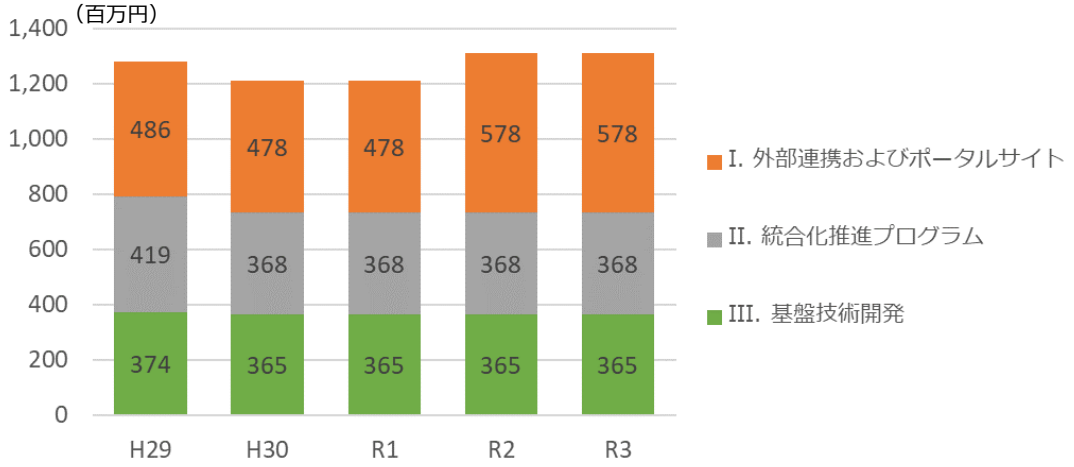
分野別統合 DB を含め、ライフサイエンスにおける国内外の多様な DB を組み合わせて統合的に利活用するための基盤的な技術開発を実施。

なお、中長期目標・計画に基づく事業の柱は、平成 28 年度以前と同様に、1）戦略の立案、2）ポータルサイトの構築・運用、3）バイオ関連データベースの統合化の推進（統合化推進プログラム）、4）データベース統合化基盤技術の研究開発（基盤技術開発）、であるが、1）の「戦略の立案」（DB 整備・統合化の企画等事業全体の運営、外部との連携構築等）と 2）の「ポータルサイトの構築・運用」は一体的に実施しているため、本報告書では両者を合わせて「外部連携およびポータルサイトの構築・運用」と記載する。

NBDC 運営委員会の提言への対応として、公開・共有データの拡大、応用を意図した統合データ基盤整備、利活用への取組みは、上記 I.～III.のそれぞれの活動の中で、重点的に実施した。

予算・体制

平成 29～令和 3 年度の事業予算は、各年度 12～13 億円であった（いずれも JST 運営費交付金中の推計額）。令和 2、3 年度にポータルサイトの構築・運用にかかるウェブサービス改良に関し予算増を受けているが、II. 統合化推進プログラム、III. 基盤技術開発については平成 30 年度に減少し、以降は同額であった。



図：事業予算（H29～R3 年度）

事業運営体制は、基本的に前期と同様とした。NBDC はセンター長の下、研究チームがサービス開発・運営を、企画運営室が事業運営にかかる企画・事務等を実施した。事業内容においても、省間連携等によるポータルサイトの構築・運用やそのための共同研究を継続するとともに、外部有識者からなる研究総括および研究アドバイザーのプログラムマネジメント体制で推進する公募型研究開発である統合化推進プログラム、共同研究による基盤技術開発を引き続き実施した。

各年度の事業運営については、外部有識者からなる NBDC 運営委員会に諮り、助言を受けて実施した。平成 29 年度には、基盤技術開発についてのより深い議論・評価のために

NBDC 運営委員会の下に「基盤技術分科会」を設置し、各年度の基盤技術開発の実施状況等の評価を実施している。同じく NBDC 運営委員会の分科会である「データ共有分科会」には、前期同様に、ヒトデータ共有にかかるガイドライン（NBDC ヒト DB の運用・利用にかかるルール）を諮り、ヒト DB への個別ヒトデータの利用・提供申請の審査はヒトデータ審査委員会で実施している。

事業評価経緯

NBDC は JST 事業の一つとして、「ライフサイエンス分野のデータベースの統合に資する成果やライフサイエンス研究開発の活性化に資する成果を得ること」を目的に掲げている。令和 3 年度は、JST の第 4 期中長期目標期間の最終年度であり、また、平成 28 年度に NBDC において実施した 6 年間の事業評価から 5 年となる節目であることから、NBDC の取組みの評価を目的として、事業評価を実施することとした。本報告書は、事業評価のために平成 29 年度から令和 3 年度までの主な成果をとりまとめたものである。

第 2　取組み内容および成果

I. 外部連携およびポータルサイトの構築・運用

日本のライフサイエンス研究の成果であるデータ・DB を利用者が統合的に扱えるようにするため、次の 3 つの活動を実施した。

（1）省間連携等による DB 統合化およびポータルサイト構築・運用
　各省の公的研究資金で作成された DB の統合化

（2）ヒトデータの統合と利活用推進
　個人情報保護に配慮が必要な人体由来データ（ヒトデータ）の統合化

（3）その他の連携
　利用者を含む外部のプロジェクト等との連携による統合の拡大とデータ共有化支援

（1）省間連携等によるデータベース統合化およびポータルサイト構築・運用

経緯：センター発足～前回事業評価まで（前期）の進捗

NBDC は、各省を跨いだ DB の統合のために、文部科学省・厚生労働省・農林水産省・経済産業省の DB 統合事業の関係者で協議・連携を行い、合同ポータルサイト「integbio.jp[5]」を開設、さらに、NBDC ポータルサイトのサービスとして「Integbio データベースカタログ[6]」「生命科学データベース横断検索[7]」「生命科学系データベースアーカイブ[8]」「NBDC RDF ポータル[9]」を開設・運用した。

NBDC における省間連携の枠組みは、関係 4 省傘下機関の協力による統合 DB ポータルサイト（上記 4 サービスによる 4 ステップの統合）の連携である。特に、厚生労働省傘下の医薬基盤研究所（当時。現：医薬基盤・健康・栄養研究所（NIBIOHN））および経済産業省傘下の産業技術総合研究所（AIST）とは共同研究を実施することとし、各サービスにおいて提供するデータ作成や、サービス間連携にかかる技術開発等を共同で実施してきた。結果として、平成 28 年度末までに、カタログ・横断検索については国内の公開 DB をほぼ網羅する状況まで、統合化を達成した。

継続的な取組みによるデータ充実

平成 29 年度以降も、各省の公的資金等により開発された DB を利用者がまとめて扱える（データ・DB がどこにあるか探せる、入手して加工できる）ようにするため、NBDC では

---

[5] URL: https://integbio.jp/
[6] URL: https://integbio.jp/dbcatalog/?lang=ja
[7] URL: https://dbsearch.biosciencedbc.jp/
[8] URL: https://dbarchive.biosciencedbc.jp/index.html
[9] URL: https://integbio.jp/rdf/

提供するサービスを充実させた。具体的には、多くの DB の統合化と多様な用途への対応のため、引き続き、広く浅い統合（カタログ：研究成果 DB の所在を網羅的に把握）から深く高度な統合（RDF ポータル：DB を再編して共通フォーマットに揃える）まで、4 つのサービスを組み合わせて提供した。

いずれのサービスも、前期同様に NIBIOHN・AIST との共同・分担によって、公的研究資金の成果報告書等の情報を利用した新規 DB 調査、サービスの相互検索連携、DB アーカイブ化、前期に策定した統一ガイドラインによる RDF 化を実施した。NBDC 独自の活動として、統合化推進プログラム・基盤技術開発の成果 DB 等の各サービスへの収録も継続して実施した。なお、農林水産省傘下の農業・食品産業技術総合研究機構（NARO）からも、カタログ収録のための DB 調査における協力を受けた。

表：省間連携 4 サービスの概要

サービス名	サービス概要 ※省間連携の取組み内容
カタログ	散在した公開 DB の中から目的の DB が探せるように、DB の所在（URL）や概要の情報を提供 ※公的資金の成果 DB 調査・情報提供（NIBIOHN、AIST、NARO）
横断検索	カタログ収録 DB のうち技術的に可能な（ログイン不要、キーワード検索可能等）DB について、DB 内の個別情報の一括検索が可能 ※各省系の検索サービスの相互検索連携（NIBIOHN、AIST）
アーカイブ	DB の寄託を受けてデータを保全し、ダウンロードして利用が可能 ※統一的なデータ加工（表形式）（AIST）
RDF ポータル	機械処理・高度な検索が可能な形式（RDF）に DB を再編して提供 ※統一的なデータ加工（グラフ形式）（NIBIOHN、AIST）

新規取組みによるデータ・機能の充実

平成 29 年度の NBDC 運営委員会提言に対応する取組みとして、NBDC では、利用者視点を踏まえつつ、データ・機能の両面でデータ活用基盤を向上させた。なお、以下の【利用者視点】は、主にアンケート調査（H28、H30 年調査、イベント時アンケート等）で得られたものであり、それぞれへの対応も付記した。

① 国際連携によるデータ充実

【利用者の視点】国内だけでなく、国外の DB もまとめて扱えるようにしてほしい。

⇒ 国際連携により、カタログの国外 DB 情報を大幅充実・継続入手の仕組みを構築

国外 DB 情報の充実のため、オープンサイエンスを推進し平成 23（2011）年から国際的

に NBDC のカタログと同様のサービスを提供してきた FAIRsharing.org[10]（運営主体：英国 Oxford 大学）との DB 情報の相互提供を、平成 29 年度に開始した。当時 FAIRsharing.org で公開していた DB 情報（約 1,000 件）のうち、カタログで未収録であった DB の情報の提供を受け、NBDC において日本語の DB 説明や、データの種類、稼働状況等のメタデータを付加したのち、617 件の海外 DB 情報を追加収録・公開した（平成 31 年 3 月）。その後も継続的に FAIRsharing.org から国外 DB 情報を入手し収録データに反映している。

初回のデータ交換による大幅な収録増（約 1.4 倍）に加えて、継続的に国外 DB 情報を入手・更新する仕組みを構築できた。

② 統合データを活用しやすく提供する「TogoDX」の開発

【利用者の視点】複数の DB を渡り歩かなければならず、目的のデータ取得に手間がかかる。DB 毎に語彙やデータ ID がそれぞれ異なり、利用者側に検索の工夫や変換の手間がかかる。

⇒ RDF 化を用いて様々な DB を統合し、データを探索するための枠組みとなる新規システムを構築

これまで省間連携を含め本事業で進めてきた RDF 化によって、国内外の主要 DB のデータ形式を統一し汎用性に優れるデータ基盤を整備しているが、より簡便なデータ検索・入手が可能となるよう、DBCLS との共同により統合データ活用インタフェース「Togo Data eXplorer（TogoDX)」の開発を進めた（詳細は『III.基盤技術開発』35 ページを参照)。

TogoDX では、多様な興味・関心を持つ研究者がそれぞれの発想で DB を柔軟且つ簡便に組み合わせて利用できるようにしている。初回公開版では、国内外 DB から収集・統合したヒトに関するデータを対象とし、50 以上の様々な観点（遺伝子発現の量や組織、タンパク質ドメイン、パスウェイ、疾患等）を自由に組み合わせながら DB をまたいで興味のあるデータを探索できる機能を実現している。

③ その他のサービス改善等

利用者意見の反映のため、平成 30 年度にユーザテスト（利用者にサービス画面を操作してもらいながら課題点を抽出）を実施し、カタログの検索機能の改良に向けた開発（類義語による検索を可能とする）や、横断検索の画面表示の改良（利用者が検索キーワードを検討しやすくなるよう、収録内容やなぜヒットしたかの情報を充実）や検索応答速度の向上につなげた。

[10] URL: https://fairsharing.org/

表：省間連携 4 サービスの収録 DB 数

各年度末の累計収録 DB 数。R3 年度は 9 月末時点。

	(参考) H28	年度					H28 比較
		H29	H30	R1	R2	R3	
カタログ	1,597	1,644	2,331	2,431	2,484	2,506	1.6 倍
横断検索	612	643	673	667	680	727	1.2 倍
アーカイブ	129	137	144	150	150	152	1.2 倍
RDF ポータル	17	20	21	25	27	27	1.6 倍

利用状況等

いずれのサービスも着実に運用し、年度やサービスにより増減の時期や程度は異なるものの、平成 29 年度以降の 5 年間の平均として、前期最終年度（平成 28 年度）の 2 倍以上の利用者に利用された（月平均ユニーク IP[11]による比較）。

表：省間連携 4 サービスの利用状況

いずれも月平均であり、R3 年度は 9 月までの月平均。

		(参考) H28	年度					H28 比較
			H29	H30	R1	R2	R3	
カタログ	ユニーク IP	2,847	4,157	5,617	9,233	7,233	6,331	2.3 倍
	アクセス数（千件）	15	24	31	35	34	26	2.0 倍
横断検索	ユニーク IP	9,889*	51,265	68,515	74,577	48,639	27,466	5.5 倍
	アクセス数（千件）	28*	132	185	203	116	70	5.1 倍
アーカイブ	ユニーク IP	10,098	13,491	16,310	41,975	35,231	24,418	2.6 倍
	アクセス数（千件）	134	318	518	535	567	745	4.0 倍
RDF ポータル	ユニーク IP	161	235	303	275	1,031	1,339	4.0 倍
	アクセス数（千件）	31	28	252	213	188	421	7.1 倍

＊改修のためアクセスが低下した期間を除いた、9 月以降の 7 ヶ月の平均。

[11] ユニーク IP とは、IP アドレスによって利用者を識別し、集計期間（月平均の場合は同じ月）中に何件の IP アドレスからのアクセスがあったかを重複無くカウントした件数である。

表：省間連携 4 サービスのアクセス数における国内・ドメイン割合

いずれも H29～R2 年度の平均。ドメイン割合は国内外全体における割合。
IP アドレスによる分類のため、利用者の所属機関の属性とは必ずしも一致しない。

	国内割合（%）	ドメイン割合（%）	
		ac/edu	co/com
カタログ	60	6	16
横断検索	91	6	8
アーカイブ	56	15	16
RDF ポータル	82	10	1

国際比較の観点では、カタログの類似サービスに、上述の FAIRsharing.org があるが、DB 収録数は約 1,800 件（日本の DB は約 80 件）であり、日本語によるサービス提供も無い。また、RDF 形式の DB は国外でも整備されているが、集積されたデータセット数として、RDF ポータルは世界有数の規模である（詳細は『III. 基盤技術開発』38～39 ページを参照）。

外部発表

研究開発やサービスにかかる情報発信のため、論文や学会等で発表した。

【論文】

Kawashima S, Katayama T, Hatanaka, H, Kushida T, Takagi T. (2018) NBDC RDF portal: a comprehensive repository for semantic data in life sciences. Database, 2018: bay123. DOI: 10.1093/database/bay123

Onami JI, Hatanaka H, Kawamoto S, Takagi T. (2019) Life science database cross search: A single window system for dispersed biological databases. Bioinformation, 15(12): 883-886. DOI: 10.6026/97320630015883

【学会発表、講演等（主なもの）】

Research Data Alliance (RDA) 4 回、
The Future of Research Communications and e-Scholarship（FORCE11）2 回、
Biocuration Conference、Japan Open Science Summit　等

（2）ヒトデータの統合と利活用推進

経緯：センター発足～前回事業評価（前期）までの進捗

NBDC では、ヒトの塩基配列や画像データ等の研究データを広く研究者間で共有するための国内で初めてのプラットフォームである「NBDC ヒトデータベース[12]」を開設し、新型シーケンサのデータ取扱いの経験のある国立遺伝学研究所の DDBJ（DNA Data Bank of Japan）と連携して運用する仕組みを構築した。NBDC はデータ共有のためのルール作りと、データ登録やデータ利用の審査を担当し、DDBJ はデータの受入・保管・配布を担当することとした。

継続的な取組みによるデータ充実

平成 29 年度以降も引き続き、DDBJ との連携により利用者からのデータ提供（寄託）申請への対応を着実に実施した。後述の外部連携によるデータ充実も含め、公開（制限公開を含む）のデータが大幅に拡大し、平成 29 年度からの 5 年間で延べ約 4 万人分から延べ約 31 万人分まで充実した。公開されたデータ数は下表の通りで、増加傾向にある。なお、データ提供申請の全体としては、平成 29 年度以降 202 件（1 年あたり 40 件程度）であり、表の公開データ数とは一致しない。これは、申請から審査の期間に加え、審査後も論文査読・公開待ち等の理由により、公開に至っていない場合が多数あるためである。

表：NBDC ヒトデータベースの公開研究データ数

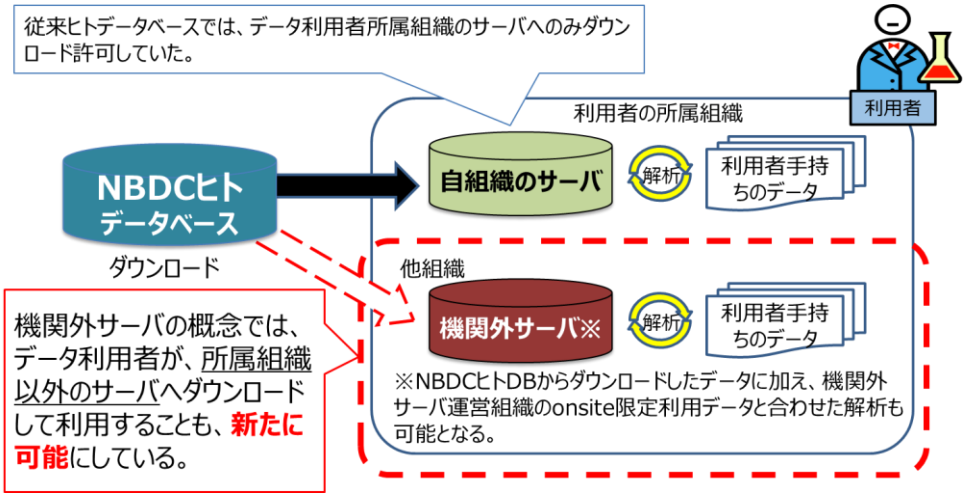
各年度末までに公開されたデータの、提供元研究課題数（累計）。制限公開を含む。
R3 年度分は 9 月末時点。

	（参考）	年度					H28
	H28	H29	H30	R1	R2	R3	比較
NBDC ヒト DB	52	73	100	130	171	188	3.6 倍

ヒトデータ利用環境の整備

NBDC ヒト DB においては、国内外の研究や法規制等を踏まえつつ、データ共有のための仕組みづくり・改善を積極的に実施した。中でも、利用者視点による運用改善・新たな仕組みづくりとして、従前はデータ利用者の所属機関のサーバに限定していた利用データのダウンロード先を、事前に NBDC がセキュリティ環境を確認した外部のスーパーコンピュータ等計算資源にも拡大する「機関外サーバ」（所属機関外利用可能サーバ）の仕組みを構築した。

[12] URL: https://humandbs.biosciencedbc.jp/



図：所属機関外利用可能サーバ（機関外サーバ）の仕組み構築

仕組みの構築にあたっては、外部有識者から構成される委員会（NBDC 運営委員会データ共有分科会）に諮りつつセキュリティ等の要件検討を実施し、平成 30 年度に NBDC ヒトデータ共有ガイドライン等に反映・施行した。現在、国立遺伝学研究所（平成 30 年 9 月から）、東北大学東北メディカル・メガバンク機構（平成 31 年 4 月から）、一般社団法人柏の葉オーミクスゲート（令和 3 年 3 月から）の計算機システムが利用可能となっている。

機関外サーバ以外にも、法令改正や運用改善のため、NBDC ヒト DB に係るガイドライン等を改正し、データ利用環境整備も実施した。具体的には、個人情報保護法改正に伴う研究指針の改正（平成 29 年 5 月施行）や利用者からの改正内容・経過措置等への問い合わせに対応した。また、寄託データに対する定型解析結果を提供する取組みや、制限公開データの概要を閲覧するための新たな公開レベルとして「登録者公開データ」の導入を行う等、利便性向上のための変更を実施した。

新規取組みによるデータ・機能の充実

平成 29 年度の NBDC 運営委員会提言に対応する取組みとして、NBDC として、利用者レビューや利用者との連携も取り入れつつ、利用者視点に立つことで、データ・機能の両面でヒトデータ共有・活用の基盤向上のための取組みを実施した。

① 日本人ゲノム情報を統合し利用しやすく提供する「TogoVar」の開発

【利用者の視点】ゲノム配列データの利用には倫理審査の手間と時間がかかる。DB や文献でのゲノム変異（バリアント）の記載方法が様々あり、相互関連付けに手間がかかる。

⇒ 利用申請不要且つ相互関連付け済みデータを作成して提供する、新規サービスを構築

NBDC ヒト DB のデータを含めて国内外の主要なヒトゲノム関連データを一括で検索・比較できる「日本人ゲノム多様性統合データベース （TogoVar)」を DBCLS と共同で開発

した（平成 30 年 6 月）[13]。NBDC ヒト DB のゲノムデータを研究プロジェクト横断的に集約してゲノムの個々人による違い（バリアント）の頻度情報とすることによって、次に挙げるヒトデータの利活用にかかる課題点に対応し、国外のデータと比較しやすく、また、事前の利用申請なしにデータの概要を閲覧可能とした。

・配列解析技術の進展により rs 番号等の識別子[14]が付かないデータが大量に産出されるようになり、様々な DB や論文に掲載されたバリアント情報の相互関連付けに手間がかかる。

⇒TogoVar では、新たな識別子により同じバリアントには同じ ID を付与。

・バリアントは解析毎に多数検出されるため、研究対象集団と他の集団の頻度情報を比較する際、バリアント毎に様々な DB に散在する情報を収集すると手間がかかる。

⇒TogoVar では、日本人集団と世界の様々な集団のデータを一括で参照可能。

開発に当たっては、サービス開設前から学会等での聞き取りやユーザとなり得る産学の研究者にレビューを依頼する等、利用者意見を反映したサービスとなるよう取り組んだ。

② 外部連携によるデータ充実

【利用者の視点】入手可能な大規模ゲノムデータが欧米人種に偏り、日本人データが少ない。

⇒ 大規模ゲノムコホートとの連携により、これまでに無い規模のデータ統合を実現

TogoVar 構築で NBDC が提供データを加工して提供する仕組み構築・技術開発を行ったことにより、外部プロジェクトとの連携が広がりデータ充実も加速した。

具体的には、TogoVar が参画する「GEM Japan」プロジェクトを日本医療研究開発機構（AMED）がヒトゲノム・医療データの国際的な共有に取り組む Global Alliance for Genomics and Health（GA4GH）[15]に提案し、日本発のプロジェクトとしてアジアで初めて採択された。NBDC は、データ共有促進の一環として GEM Japan プロジェクトへ協力し、東北メディカル・メガバンク（TMM）、バイオバンク・ジャパン（BBJ）等との連携により、これらのプロジェクトが産出するゲノム情報を統合して TogoVar から公開した（令和 2 年 7 月）。NBDC 研究員が核となって調整することで、プロジェクト間のデータ統合に加え、データ利用ライセンスの整理・一本化を実施した結果、日本人に関してこれまでに無い規模（約 7,600 人）で、且つ利便性の高い形でのデータ公開を実現した。GA4GH での紹介や、欧州のゲノム DB（Ensembl）での活用[16]により、TogoVar が AMED の国際連携におけるデ

---

[13] URL: https://togovar.biosciencedbc.jp/

[14] バリアントの識別子（他のものと区別するために付けられる文字列）として、一塩基多型等は米国 NIH の dbSNP への登録時に発行される rs 番号が、構造多型等は、dbVar への登録時に発行される nsv/esv 番号が、よく利用されている。

[15] ヒトゲノム等の国際的共有化を推進する国際協力組織で、54 カ国、約 700 の産学官の機関が加盟（R3 年 10 月時点）。

[16] URL: https://asia.ensembl.org/info/genome/variation/species/populations.html

ータ公開プラットフォームとして国際的に存在感を示した。

利用状況等

ヒト DB および TogoVar の利用状況は以下のとおり。ヒト DB については利用申請数が大幅に増加した他、TogoVar へのアクセス数も増加傾向にある。特に TogoVar は、国内のゲノム医科学分野の研究者から「難病患者さんの診断に直接的に役立てられる」との評価を受けている他、複数の研究論文においてリファレンスデータとして利用されている。

一層の利便性向上のため、ヒト DB では大規模データを手軽に利用できる解析環境・データ整備（機関外サーバ上でデータ解析が完結できる環境整備、寄託データを予め定型解析したデータも提供）を進めている。TogoVar においても関連する国内 DB（MGeND 等）とのデータ連携の計画があり、今後更なる利用の拡大が見込まれる。

表：NBDC ヒトデータベースへの利用申請数

制限公開データについて、各年度末の累計の利用申請件数。R3 年度分は 9 月末時点。

	（参考）	年度					H28
	H28	H29	H30	R1	R2	R3	比較
NBDC ヒト DB	26	48	91	128	164	187	7.2 倍

表：TogoVar の利用状況（アクセス数）

いずれも月平均であり、R3 年度は 9 月までの月平均。

		年度				
		H29	H30	R1	R2	R3
TogoVar	ユニーク IP	−	595	610	795	724
	アクセス数（千件）	−	30	94	89	75

表：TogoVar のアクセス数における国内・ドメイン割合

H30～R2 年度の平均。ドメイン割合は国内外全体における割合。
IP アドレスによる分類のため、利用者の所属機関の属性とは必ずしも一致しない。

	国内割合（%）	ドメイン割合（%）	
		ac/edu	co/com
TogoVar	95	18	2

国際比較の観点では、ヒト DB に類似のサービスとしては、米国 National Institutes of Health（NIH）が運用する dbGaP[17]、欧州 European Bioinformatics Institute（EBI）等が運用する EGA[18]がある。データ件数（Study 数での比較）は NBDC ヒト DB の方が少ないが、

[17] URL: https://www.ncbi.nlm.nih.gov/gap/

[18] URL: https://ega-archive.org/

dbGaP・EGA とともに論文投稿時の推奨データ登録先となっている。

TogoVar の類似サービスとして gnomAD や dbSNP があるが、対象としている日本人の人数や集計前データの取得において、TogoVar が優れる（詳細は『III. 基盤技術開発』33〜34 ページを参照)。

外部発表

研究開発やサービスにかかる情報発信のため、学会発表や書籍への寄稿等を実施した。

【論文】

箕輪真理（2019）「NBDC ヒトデータベースと日本人ゲノム多様性統合データベース "TogoVar”」家族性腫瘍, 19(1): 45-49. DOI: 10.18976/jsft.19.1_45

【学会発表、講演等（主なもの)】

日本人類遺伝学会　4 回、アメリカ人類遺伝学会　4 回、
日本家族性腫瘍学会学術集会、日本生化学会、Japan Open Science Summit、
バイオバンク・ネットワークウェビナー　等

【書籍（主なもの)】

川嶋実苗、児玉悠一、高木利久（2017）「国際的なデータシェアリングの加速と国内の取り組み」実験医学増刊, 35(17): 23-29.

川嶋実苗、児玉悠一、高木利久（2017）「NBDC ヒトデータベースとグループ共有への展開」実験医学増刊, 35(17): 90-97.

豊岡理人、高木利久（2018）「オープンサイエンスと NBDC ヒトデータベース」医学のあゆみ, 266(5): 377-382.

山本奈津子、川嶋実苗、清水佳奈、片山俊明、荻島創一（2018）「<続>改正個人情報保護法でゲノム研究はどう変わるか？」実験医学, 36(13): 2260-2268.

山本奈津子、川嶋実苗（2019）「関係ないと思っていませんか？臨床研究法について知るべきこと」実験医学, 37(1): 81-85.

川嶋実苗、豊岡理人、三橋信孝（執筆分担）坊農秀雅／編（2020）実験医学別冊「バリアントデータ検索＆活用　変異・多型情報を使いこなす達人レシピ」

豊岡理人（2021）「NBDC ヒトデータベースと TogoVar」実験医学増刊, 39(7): 85-89.

（3）その他の連携

経緯：センター発足～前回事業評価（前期）までの進捗

省間連携 4 サービスについて、NBDC では、JST 情報事業が提供する J-GLOBAL 等との相互連携を構築し、JST の研究資金事業との連携として、各事業の公募要領に NBDC へのデータ提供協力についての記載が盛り込まれるよう働きかけを実施するとともに、各事業からのライフサイエンス分野データ共有動向に関する助言や問い合わせへの対応を実施した。

また、AMED への協力として、ヒトデータの新たなデータ共有のカテゴリとして、公的 DB への登録に先立つ「制限共有」（研究グループ内での共有）を実現するための DB 運用ルール策定等、DB 開設への支援を実施した。

外部プロジェクトのデータ共有化支援

広くデータを統合するため、また、NBDC 運営委員会提言を受けた NBDC による外部プロジェクトのデータ共有化支援として、大規模にデータを産出する事業やその関係機関に対しては、データ公開前から連携して研究の進捗と共に生じる課題に対応できるよう取り組んだ。次の①、②のようなグループ共有 DB の仕組み構築により、公開に先んじた研究グループ内限定でのデータ共有・利活用・付加価値付けの検討や、論文公開に合わせた早期データ公開のため、NBDC ヒト DB と連携したグループ共有段階でのデータ整備・データ提供申請・審査の実施、といった仕組み上の利点を提供した。

① 日本医療研究開発機構（AMED）との連携

平成 28 年度に締結した JST（NBDC）と AMED との基本連携協定に基づく具体事案として、公開に先立つ研究プロジェクト内や研究グループ内でのデータ共有の仕組みである AMED ゲノム制限共有データベース（AGD)の運営への協力を継続した。平成 29 年 2 月に運営を開始した AGD には平成 29 年度以降に 11 件の研究課題・プロジェクト等からデータ提供申請があり、論文公開等に伴って順次 NBDC ヒト DB へデータ移行が行われている。また、AMED が構築中の、公開が困難なデータを共有・利活用する仕組みである CANNDs について、データ共有の仕組み検討への協力を実施している。

② 戦略的イノベーション創造プログラム（SIP）スマートバイオ産業・農業基盤技術との連携

NBDC は、研究開始当初の平成 30 年度から外部協力機関として参画している。NBDC ヒト DB へのデータ提供につながる取組みとして、民間企業を含む 30 件以上の研究機関からなる研究コンソーシアムにおいてデータ共有を行う DB（SIP Healthcare Group Sharing Database、SHD）の仕組み構築を支援し、令和 3 年 9 月に運営を開始した。

この他、円滑なデータ公開・共有、連携によるデータ価値の最大化等について、以下のプロジェクト等との連携・協力を実施した。

・文部科学省科学研究費助成事業「先進ゲノム解析研究推進プラットフォーム」

「説明文書及び同意文書のモデル書式」等を作成するにあたり、NBDC ヒト DB へのデータ登録と共有が円滑に行えるように協力を実施。

・バイオバンク・ジャパン（BBJ）

ゲノムデータと臨床情報の一体的な活用に向けた取組みとして、ゲノムデータは NBDC ヒト DB、臨床情報は BBJ で保管し、利用審査を連携する体制を構築。

・東北大学東北メディカル・メガバンク機構（ToMMo）

NBDC ヒト DB のサテライト拠点を ToMMo に設置し、ToMMo のスーパーコンピュータ上でのみ解析可能なデータと NBDC ヒト DB 上のデータを利用者がまとめて解析するための仕組みを構築。

・NIMS マテリアルデータプラットフォームセンター

高分子化学における材料設計のための DB 整備とデータ連携を NIMS が実施するための協力として、DBCLS と共同で RDF 化を支援。

・AMED バイオバンク横断検索システム

国内バイオバンクの試料・情報と NBDC ヒト DB のゲノム情報（TogoVar）とを利用者が一括して検索できるよう、連携に向けた協議・検討を実施している。

また、センター長を中心に、事業を超えた国レベルでのデータ共有にかかる提言や調査への協力も実施した。産業競争力懇談会（COCN）の 2018 年度推進テーマにかかる報告書「デジタル・バイオエコノミーの実現に向けて」（平成 31 年 2 月）[19]、日本学術会議バイオインフォマティクス分科会による提言「持続可能な生命科学のデータ基盤の整備に向けて」（令和元年 11 月）[20]である。

継続的な取組みによる JST 内連携

JST 組織内連携として、引き続き、NBDC が提供するカタログ・横断検索・アーカイブの各サービスについて、JST 情報事業が提供する J-GLOBAL、J-STAGE、researchmap との相互連携を実施した他、形態素解析用に活用しやすく整備した辞書の作成・公開（科学技術用語形態素解析辞書）、日本化学物質辞書（日化辞）RDF の充実を実施した。

また、JST 科学技術用語辞書のうちライフサイエンス分野に関連する用語について、用語同士の関係性を高度に分類・整理するオントロジー化を実施し、「生物学概念相互関係オン

[19] URL: http://www.cocn.jp/report/thema108-L.pdf

[20] URL: https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t279-1.pdf

トロジー（IOBC)」として平成 30 年 6 月に公開した。従前は 1 種類であった関係性を 31 種類に細分類し RDF 形式にしたことで、疾患関連遺伝子の中でも異なる疾患に共通する因子を探索する等、高度な検索が可能となった。

JST の研究資金事業との連携として、前期からの継続的な取組みにより各事業の公募要領に NBDC へのデータ提供協力についての記載が盛り込まれるようになった他、各事業からのライフサイエンス分野データ共有動向に関する助言や問い合わせへの対応を実施した。

（4）まとめ（I. 外部連携およびポータルサイトの構築・運用）

日本のライフサイエンス研究の成果であるデータ・DB を利用者が統合的に扱えるようにするため、継続的な取組みによる統合データの充実に加え、新規の取組みとして、規模や独自性において国際的に価値ある統合データを構築するとともに、データ利用障壁を低減するための取組み・開発を実施し、データ・機能の両面でデータ活用基盤を向上させた。次の 4 点により、統合のための取組み・成果の双方について、ライフサイエンス研究を効率的・効果的に進めるための研究開発環境の整備・充実への寄与は十分であったと考えている。

○ より活用しやすいデータ基盤を整備するため、従前の利用者が手元で統合しやすいデータを提供するサービスに加え、用途に沿って統合済みのデータを提供するサービスを新規に開発した。

具体的には、DBCLS との共同により、ヒトの機微データをアクセス制限不要な集計データに加工して提供する「TogoVar」や、RDF 化により DB を超えて利用者が柔軟にデータを組み合わせ・抽出できる「TogoDx」の開発を実施した。

○ 統合の拡大のため、外部の研究プロジェクトとの連携を拡大し、これにより、NBDC の各サービスのデータ充実や外部プロジェクトにおけるデータ共有化・公開データの高付加価値化につなげた。

特に、TogoVar に関し TMM や BBJ 等との連携により、日本人として前例がない規模（約 7,600 人）のデータを統合し、国内外での活用につながった。

○ データ利用環境整備として、NBDC ヒト DB に関し、法令改正への対応、運用改善のためのガイドライン見直し、共用スパコンとの連携の導入（機関外サーバの仕組み構築）・拡大による解析環境の向上等、利便性向上を実施した。

ガイドラインを含む前期からの NBDC ヒト DB 運営の実績・ノウハウを、上記の TogoVar 開発や外部プロジェクト（AMED、SIP、GA4GH 等）との連携につなげた。

○ 省間連携による既存サービスについては、着実に運用を行うとともに利用者要望を踏まえつつデータや機能を充実させ、上記 TogoDX や外部連携の基盤とした。

なお、前期提言を踏まえつつも、ヒトデータを中心に、前述の通りの成果を得た。一方で、今後は、急速に進展しつつあるデジタル・トランスフォーメーション（DX）など社会状況を踏まえ、ヒトデータ以外で新たに検討すべき領域があれば、対応していく必要がある。また、より多くの外部プロジェクトを支援することを通して、本事業が蓄積してきた研究成果の公開や統合の成功事例、それらのためのノウハウを広げていくことも、事業の成果を最大化するためには重要と考えている。

II. 統合化推進プログラム

国内のライフサイエンス研究等によって産出された研究データを広く収集し、より多くの多様な研究者にとって価値のあるデータとして提供するため、公募を通した分野別の統合 DB の開発を実施している。次の 4 構成により取組みと成果の概要を記載する。

（1）研究開発推進マネジメント
　研究総括による公募企画・選考、進捗管理や課題間連携の推進、課題評価
（2）各研究開発課題による統合化
　分野別（データ種別や生物種別）に国内外データを統合する DB の開発
（3）利活用に向けた取組みと利用状況
　利用者との連携関係の構築による利活用の推進
（4）課題間連携による統合化
　研究開発課題を連携させることによる更なる統合化の推進

（1）研究開発推進マネジメント

経緯：センター発足～前回事業評価まで（前期）の進捗

統合 DB のより使われる形での整備に向けて、プログラム全体の運営やそのための適切な評価が実施できるよう、プログラムオフィサー（研究総括）・研究アドバイザーを選任してプログラムを運営した。前期期間中は、4 回の公募および選考、採択課題の年次の進捗管理、事後（当時は研究期間が 3 年のため事後評価のみ）を実施した。進捗に応じた助言や課題間の連携の推進のため、採択時のキックオフ会議、サイトビジット等も開催した。

推進体制・基本方針

今 5 ヶ年においても前期同様に、研究総括が研究アドバイザーの協力の下に研究開発を推進するためのマネジメントを実施する体制とした。研究総括は、選考・研究マネジメント経験やデータ公開・利用への積極的な取組み経験を踏まえ、令和元年度までは前期に引き続き長洲毅志氏（元 エーザイ株式会社アドバイザー）が担当したが、基本方針を引き継ぐ形で令和 2 年度以降は伊藤隆司氏（九州大学 教授）に交代した。研究アドバイザー[21]は、研究総括が幅広い研究者の視点をプログラム運営に反映できるよう、産学やダイバーシティ（男女、年齢層）の観点も踏まえつつ、バイオインフォマティクスの専門家に加えて、後述のとおり各種データ利用の専門的見地から助言が可能な外部有識者も選任した。

今期における研究総括の基本方針として、次の 3 点によるマネジメントを実施した。

・データ利活用の推進に向けた、日本として独自に整備すべきデータ基盤の構築
・国際的なオープンサイエンスの潮流を踏まえた、国際連携によるデータ共有の標準化や国内のみならず国外からの利用者にも活用される DB の開発

[21] 研究アドバイザーの名簿は、https://biosciencedbc.jp/funding/program/dicp/ に掲載。

・企業研究者を含む DB 利用者との連携に係る活動の重視

公募企画・選考におけるマネジメント

平成 29 年度[22]、平成 30 年度[23]の公募にあたり、NBDC 運営委員会の提言も踏まえ、より活用される DB 開発に向けて次の取組みを実施した。

・利用者の研究開発へのより一層の貢献を目指し、募集対象となる DB の要件として、我が国において一定以上の規模の利用が見込めることに加え、我が国のデータ基盤として整備すべき独自性、または、国際的なデータ共有における中核としての実績、のいずれかを必須とした。
・DB 利用者（学協会・産業コミュニティを含む、産学のデータ提供者・利用者）との連携・協業を必須とし、開発に利用者意見を反映するための計画立案を求めた。
・選考にあたり対象分野と扱うデータ種のポートフォリオを実施し、ヒト・動物以外にも植物や微生物を含め、各種オミクスの主要なデータを幅広くカバーした。
・計画的なデータの統合化・データ提供のための機能開発を実施するため、前期は 3 年であった研究期間を 5 年に延長した。これに伴い 3 年次に中間評価を実施することとし、研究提案書において研究終了時までの達成目標に加え、3 年次までの達成目標の記載を求めた。

進捗管理や課題評価におけるマネジメント

前期と同様に、毎年の研究計画書と年次報告書の確認、サイトビジットや研究総括と研究代表者との面談・意見交換を適宜実施し（5 年間で延べ 100 回以上)、進捗状況の把握とともに研究開発の方向性を議論し、研究遂行上の課題解決を支援した。

3 年次の中間評価においては研究課題毎に研究開発の進捗状況や成果見込みの評価を行い、必要に応じて、以降の計画の見直しを実施した。順調な進捗が認められるものの、データ整備・提供体制に改善の余地がある研究課題への対応の他、利用が進んでいない、開発が遅れている等の問題点に対して、一部目標変更による特定テーマ・機能改善への集中的取組み検討、想定ユーザとの連携の橋渡しを、NBDC 担当者、研究アドバイザーも関与して実施した。

また、外部状況変化への対応のため、研究総括・研究アドバイザーの合意の下 NBDC 担当者が研究代表者を支援し、計画や研究開発費の柔軟な変更を行った。例としては、新型コロナウイルス感染症等感染症関連データを扱う特設サイトやデータセットの整備、想定を上回る急速なデータ増加への対応のための計画変更が挙げられる。

---

[22] 公募要領は、https://biosciencedbc.jp/gadget/fund/h29guide-togo.pdf に掲載。

[23] 公募要領は、https://biosciencedbc.jp/gadget/fund/h30guide-togo.pdf に掲載。

産学のデータ利用者との連携強化

民間企業の研究者を含む DB 利用者との連携を強化するため、各研究課題に共通的な取組みとして、以下を実施した（各研究課題における開発・取組みの事例は後述『(3) 利活用に向けた取組みと利用状況』を参照)。

・研究アドバイザーに、バイオインフォマティクスの専門家、さらには支援 DB が対象とする分野のデータやその利用に専門的な知識をもつ産学の研究者を加え、より DB 利用者の観点から開発の方向性に助言できるマネジメント体制とした。
・各研究課題に対して、利用者意見の積極的な収集と研究開発への反映を要請した。具体的には、DB が対象とする研究領域の専門家や関連学会からなるアドバイザリー委員会の設置や講習会の開催等、利用者と活発なコミュニケーションを取りながら開発を進められる体制・計画の整備を求めた。
・開発への利用者意見の反映を目的として、各 DB のユーザテスト（利用者にサービス画面を操作してもらいながら課題点を抽出）の実施を支援した。この実施に当たっては、テスト参加者のリクルートや、実際の利用を想定したタスク設定等を NBDC 側が支援した。

課題間連携におけるマネジメント

前期と同様に、DB を超えてデータ連携を形成し利用者が円滑に多様なデータを扱えるよう、各研究課題で開発する DB どうしの連携を推進した（連携の事例は後述『(4) 課題間連携による統合化』を参照)。

・基盤技術開発との連携による開発 DB の RDF 化を引き続き推進した。これにより、他 DB との連携による各 DB のデータ充実・高付加価値化、および、基盤技術開発や省間連携等との共通フォーマットによる統合データ整備・拡大を図った。
・採択時のキックオフ会議、合同成果報告会の開催、基盤技術開発で開催するバイオハッカソン・スパークルソンへの参加を促すことにより、研究課題に参画する研究者の交流・ネットワーク作りを推進した。

（2）各研究開発課題による統合化

上述のとおり、平成 29、30 年度にそれぞれ公募を実施し、平成 29～令和 3 年度の研究課題として 7 件[24]、平成 30～令和 4 年度の研究課題として 2 件[25]を採択した。前期に引き続き、ヒト・動物、植物、微生物の生物別、ゲノム、エピゲノム、プロテオーム、糖鎖、メタボロームやパスウェイ情報等の幅広いオミクス単位でのデータ統合を目指している。課題中間評価時点において、9 課題中 7 課題が概ね順調に研究成果を上げている[26]。

各研究開発課題の概要

【平成 29 年度採択】

① 課 題 名 ： エピゲノミクス統合データベースの開発と機能拡充
研究代表者： 沖 真弥（京都大学　大学院医学研究科　特定准教授）
主な開発 DB： ChIP-Atlas （https://chip-atlas.org/）

DB 概要	個別に公開された ChIP-seq 等データを抗原別、生物種別に整理して収載。 再解析データの整備により、比較解析や転写因子の予測が可能。
データ収録数	ChIP-seq、DNase-seq、ATAC-seq、BS-seq データ： 約 22 万件（5 年間で約 3.7 倍）
国際比較・ 独自性	公共 DB（NCBI、EBI、DDBJ）の ChIP-seq 等の網羅的再解析データを提供。 ユーザの遺伝子リストから転写因子を予測する独自性の高い機能を提供。
利用状況	平成 29 年度以降、国内外の 240 件以上の論文に引用されている。 産学の複数研究機関と DB 収録データを使った共同研究を実施中。
主な論文報告	Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, Kawaji H, Nakaki R, Sese J, Meno, C (2018) ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO Rep. 19(12): e46255. DOI:10.15252/embr.201846255　他

② 課 題 名 ：ゲノム・疾患・医薬品のネットワークデータベース
研究代表者： 金久 實（京都大学　化学研究所　特任教授）
主な開発 DB： KEGG MEDICUS （https://www.kegg.jp/kegg/medicus/）

DB 概要	医薬品が作用するパスウェイや標的分子の検索、疾患や薬剤の作用に関わるバリアントをパスウェイ上で俯瞰し作用機作の解釈ができる DB。
データ収録数	疾患ネットワークバリアントの要素数：約 1,300 件（今期新規収録） その他：バリアント約 400 件、疾患約 2,500 件、医薬品約 1 万件 等
国際比較・ 独自性	国際的に認知度が高い DB である KEGG をベースとして開発。 ヒトゲノムの多様性を、生体システムを構成するネットワーク要素の多様性

[24] 採択課題・公募の概要は https://www.jst.go.jp/pr/info/info1248/index.html に掲載。
[25] 採択課題・公募の概要は https://www.jst.go.jp/pr/info/info1309/index.html に掲載。
[26] 平成 29 年度採択 7 課題の中間評価結果は https://biosciencedbc.jp/funding/evaluation/dicp2017midterm.html に掲載。平成 30 年度採択 2 課題の中間評価結果は https://biosciencedbc.jp/funding/evaluation/dicp2018midterm.html に掲載。

	として蓄積する DB は独自性が高い。
利用状況	後述のとおり多くのアクセスがある（月平均ユニーク IP：約 200 万件）
主な論文報告	Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe, M (2019) New approach for understanding genome variations in KEGG. Nucleic Acids Res., 47(D1): D590-D595. DOI:10.1093/nar/gky962　他

③ 課 題 名 ： 糖鎖科学ポータルの構築
研究代表者： 木下 聖子（創価大学　糖鎖生命システム融合研究所　副所長・教授）
主な開発 DB： GlyCosmos Portal （https://glycosmos.org/）

DB 概要	国内の糖鎖関連 DB をまとめ、H31 年 4 月に開設。糖鎖分子構造・糖鎖遺伝子関連パスウェイ情報を整備。国外 DB と標準化・品質管理で連携。
データ収録数	糖鎖構造：約 12 万件（5 年間で約 1.5 倍） その他：糖鎖関連遺伝子約 3 万件、糖タンパク質約 5 万件 等
国際比較・ 独自性	日本発の糖鎖構造記述法で国際糖鎖構造レポジトリを運用（前期成果）。 遺伝子発現情報から糖鎖構造を予測する機能を国際共同開発。
利用状況	国外 DB からのリンクや複数の文献成果への貢献。 日本糖質学会（JSCR）の公式ポータルサイトとして承認されている。
主な論文報告	Yamada I, Shiota M, Shinmachi D, Ono T, Tsuchiya S, Hosoda M, Fujita A, Aoki NP, Watanabe Y, Fujita N, Angata K, Kaji H, Narimatsu H, Okuda S, Aoki-Kinoshita KF. (2020) The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences. Nat Methods, 17(7): 649-650. DOI:10.1038/s41592-020-0879-8　他

④ 課 題 名 ： 蛋白質構造データバンクのデータ検証高度化と統合化
研究代表者： 栗栖 源嗣（大阪大学　蛋白質研究所　教授）
主な開発 DB： PDBj (Protein Data Bank Japan) （https://pdbj.org/）

DB 概要	タンパク質等の生体高分子の国際標準レポジトリ（wwPDB[27]）を国外 DB と共同で運営、解析ツールや関連 DB も提供。
データ収録数	タンパク質等構造データ（wwPDB 全体）：約 19 万件（5 年間で約 1.5 倍） ※PDBj によるデータ登録処理は、wwPDB 全体の約 24%
国際比較・ 独自性	日米欧 3 極連携による国際標準レポジトリ（アジアオセアニア地区担当）。 RDF 化や糖鎖構造表記の標準化等、3 極全体のデータ高度化に貢献。
利用状況	多数の論文への貢献の他、製薬企業もデータを多用。wwPDB 全体におけるデータダウンロードは、1 日あたり 200 万回以上。
主な論文報告	Kinjo AR, Bekker GJ, Wako H, Endo S, Tsuchiya Y, Sato H, Nishi H, Kinoshita K, Suzuki H, Kawabata T, Yokochi M, Iwata T, Kobayashi N, Fujiwara T, Kurisu G, Nakamura H. (2018) New tools and functions in data-out activities at Protein Data Bank Japan (PDBj). Protein Science, 27(1): 95-102. DOI:10.1002/pro.3273　他

[27] URL: https://www.wwpdb.org/

⑤課 題 名 ： データサイエンスを加速させる微生物統合データベースの高度実用化開発
研究代表者： 黒川 顕（国立遺伝学研究所 生命情報研究センター　教授）
主な開発 DB： MicrobeDB.jp （https://microbedb.jp/）

DB 概要	公開されているゲノム・メタゲノムデータを再解析した上で統合し、菌株保存機関の菌株データ、サンプルの分離源、系統分類情報とともに提供。
データ収録数	マイクロバイオームデータ：約 160 万サンプル（5 年間で約 10 倍）等
国際比較・独自性	ゲノム・菌株・系統情報を統合した DB として世界的に独自性が高い。日本の微生物資源情報（JCM）も収録。
利用状況	学術研究機関の他、民間企業（7 社）のデータ解析でも活用。
主な論文報告	Higashi K, Suzuki S, Kurosawa S, Mori H, Kurokawa K (2018) Latent environment allocation of microbial community data. PLoS Comput. Biol., 14(6): e1006143. DOI:10.1371/journal.pcbi.1006143 他

⑥ 課 題 名 ： 疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オーミクスデータの統合
研究代表者： 菅野 純夫（東京医科歯科大学　難治疾患研究所　非常勤講師）
主な開発 DB： DBKERO （https://kero.hgc.jp/）

DB 概要	疾患の各種オミクスを統合し、遺伝子領域、発現量、転写因子結合部位、バリアント、エピゲノム情報等のデータトラックを並列表示できる DB。
データ収録数	ゲノム・エピゲノム・トランスクリプトーム：約 5,500 データセット
国際比較・独自性	転写開始点やパスウェイ等、多層オミクスに対応したデータビューワ。日本人由来データ（iJGVD、HGVD、JPDSC、CREST-IHEC 等）を収録。
利用状況	中間評価結果を受けて、利用者を呼び込むための開発・活動を実施中。
主な論文報告	Suzuki A, Kawano S, Mitsuyama T, Suyama M, Kanai Y, Shirahige K, Sasaki H, Tokunaga K, Tsuchihara K, Sugano S, Nakai K, Suzuki Y. (2018) DBTSS/DBKERO for integrated analysis of transcriptional regulation. Nucleic Acids Res., 46(D1): D229-D238. DOI:10.1093/nar/gkx1001　他

⑦ 課 題 名 ： 個体ゲノム時代に向けた植物ゲノム情報解析基盤の構築
研究代表者： 田畑 哲之（かずさ DNA 研究所　所長）
主な開発 DB： Plant GARDEN （https://plantgarden.jp/）

DB 概要	前身 DB を全面改修し、ゲノム情報をもとに遺伝子、転写産物、バリアント、DNA マーカー等を閲覧・比較できる DB を開設（R2 年 7 月）。
データ収録数	植物ゲノム・バリアント：約 120 植物種・約 51 億件（今期新規収録） その他：遺伝子、DNA マーカー、QTL 等
国際比較・独自性	植物の収録ゲノム数において世界最大規模。公共 DB の国内外公開データを解析して整備したバリアントデータは本 DB 固有。
利用状況	開発と併行して運用した前身 DB（PGDBj）が育種研究論文に貢献。利用者意見を反映した開発を進めていることから、今後の利用成果が期待される。
主な論文報告	Ghelfi A, Shirasawa K, Hirakawa H, Isobe S (2019) Hayai-Annotation Plants: an ultra-fast and comprehensive gene annotation system in plants. Bioinformatics, 35(21): 4427-4429. DOI:10.1093/bioinformatics/btz380

【平成 30 年度採択】

⑧ 課 題 名 ： 物質循環を考慮したメタボロミクス情報基盤

研究代表者： 有田 正規（情報・システム研究機構 国立遺伝学研究所　教授）

主な開発 DB： MetaboBank （https://mb.ddbj.nig.ac.jp/search）

DB 概要	国際コンソーシアム MetabolomeXchange[28]と連携し、アジア初となるメタボロームの世界標準のデータレポジトリ構築を目指して開発を実施中。
データ収録数	ー（開発中のため該当なし）
国際比較・独自性	国際コンソーシアムと連携し、アジア初のメタボロームデータの国際標準レポジトリを開発中。
利用状況	ー（開発中のため該当なし）
主な論文報告	ー（開発中のため該当なし）

⑨ 課 題 名 ： プロテオームデータベースの機能深化と連携基盤強化

研究代表者： 石濱 泰（京都大学　大学院薬学研究科　教授）

主な開発 DB： jPOST （https://jpostdb.org/）

DB 概要	プロテオーム情報を標準化・統合・一元管理し、多彩な生物種・翻訳後修飾・絶対発現量も含めた横断的解析ができる統合プロテオーム DB。
データ収録数	プロテオームデータ：約 600 プロジェクトが公開（4 年間で約 6 倍） ※H30 年度採択のため、H30 年度以降の 4 年間の増を記載
国際比較・独自性	アジアオセアニア地区初のプロテオームの国際標準レポジトリ[29]（前期成果）。再解析データの整備・提供により実験を跨いだ比較解析も可能。
利用状況	国内外の 1,000 を超えるプロジェクトがデータ登録（未公開を含む）
主な論文報告等	Moriya Y, Kawano S, Okuda S, Watanabe Y, Matsumoto M, Takami T, Kobayashi D, Yamanouchi Y, Araki N, Yoshizawa AC, Tabata T, Iwasaki M, Sugiyama N, Tanaka S, Goto S, Ishihama Y. (2019) The jPOST environment: an integrated proteomics data repository and database. Nucleic Acids Res. 47(D1): D1218-D1224. DOI: 10.1093/nar/gky899　他

[28] URL: http://www.metabolomexchange.org/site/

[29] 国際コンソーシアム ProteomeXchange（URL: http://www.proteomexchange.org/）に参画。

（3）利活用に向けた取組みと利用状況

経緯：センター発足～前回事業評価まで（前期）の進捗

研究提案の選考において、データ産出プロジェクトとの連携状況、関連の研究コミュニティ・学会からの支援が得られること、利用者確保の見通しがあることを基準に含めた。

利活用に向けた取組み

多くの研究者が多様なライフサイエンス関連の研究データを幅広く利活用できるよう、上述『(1) 研究開発推進マネジメント』のとおり、産学のデータ利用者との連携・協業を推進し、各研究課題において次のような取組みを実施した。

・各研究課題において、関係学会の有識者や利用者をメンバーに含むアドバイザリー委員会の設置、講習会の開催等により、DB への意見を収集（9 課題全体で年 6 回程度の委員会開催）。例として、糖鎖 DB（GlyCosmos）は、令和元年度に日本糖質学会の公式ウェブポータルに採用され、定期的にデータの正確さや利便性について学会の確認を受けることとなった。植物 DB（Plant GARDEN）でも、アドバイザリー委員会において定期的に意見聴取を行っている。
・NBDC が各課題へ広報促進を要請し、多数の学会展示、チュートリアル動画の配信等の広報活動を実施。特に、日本分子生物学会では毎年の年会において、NBDC・DBCLS の働きかけにより各課題が合同でフォーラム（DB 紹介）や展示を実施した。

利用者視点による研究開発として、次のような事例が挙げられる。

・タンパク質立体構造 DB（PDBj）では、利用者視点で使いやすいデータの整備・提供に向けて、低分子化合物構造 DB（Cambridge Structural Database）との連携を実施。
・植物 DB（Plant GARDEN）では、利用者が保有データを Plant GARDEN の情報と比較して配列変異等を表示する機能を開発。
・DB 側へデータをアップロードせずに利用者が保有するデータを解析できるよう、植物 DB（Plant GARDEN）と微生物 DB（MicrobeDB.jp）では、利用者の計算機環境で利用できる解析パイプラインのコンテナ化に対応。
・微生物 DB（MicrobeDB.jp）では、ホロゲノム解析支援ツールの解析に向けて、植物共生細菌の研究者との連携も行いつつ、宿主である動植物と微生物のゲノムを統合的に解析・閲覧可能となるよう開発を実施中。
・ユーザテストの結果を受けて、複数の DB でユーザインタフェースのわかりづらさを改善するための変更を実施。

国際連携・標準化による統合データ整備

国際的なオープンサイエンスの潮流を踏まえ、国際的にも価値あるデータを整備することによりデータ利用の拡大に取り組んだ。各研究課題において、次のような取組み・開発を

実施した。

・国際協調・標準化によるデータ共有について、前期に引き続いて国際レポジトリを運用し（タンパク質立体構造 PDBj、糖鎖 GlyCosmos、プロテオーム jPOST）、メタボロームについても新規の国際レポジトリを開発している（MetaboBank）。特に、タンパク質立体構造やプロテオームは、多くのジャーナルが論文投稿時に PDBj や jPOST が参画する国際コンソーシアム（wwPDB、ProteomeXchange）へのデータ登録を求めていることから、国内外の幅広い研究者へ貢献している。

・前期から継続して運用している国際レポジトリでは、国内外からのデータ受入・公開を着実に実施しつつ、更なるデータフォーマットの国際標準化に向けた開発・活動も実施した。タンパク質立体構造については、wwPDB が公式データフォーマットを定めるにあたり、PDBj が各登録データの品質レポートで整備を進めてきた RDF 形式も二次フォーマットの一つとして認められ、統合利用に優れたデータ整備が今後も見込まれる。また、近年急増する中国からのデータ登録に対応するため、PDBj が PDBchina の立ち上げ支援も行った。プロテオームについても、Human Proteome Project と連携し、ProteomeXchange の各 DB の登録データを一意に識別でき各論文との対応づけができる仕組み構築に、jPOST も貢献した。

利用状況

各 DB の利用状況のうち、主要なものとして次のような事例が挙げられる（DB 毎の利用状況の概略は『(2) 各研究開発課題による統合化』に記載）。

・エピゲノミクス DB（ChIP-Atlas）は、多数（5 年間で 240 報以上）の論文に引用され、疾患研究や動物の進化に関し遺伝子の発現を制御する転写因子やゲノム領域の解析等に使用されている。また、複数の民間企業と DB 収録データを利用した共同研究を実施している。

・タンパク質立体構造 DB（PDBj）は、国際連携の全体（wwPDB）のうち約 24%のデータ登録を担当（年約 3,000 件）。wwPDB 全体として、1 日あたりのデータダウンロードが 200 万回以上あり、創薬を含む幅広い研究に貢献している。

・微生物 DB（MicrobeDB.jp）は、学術機関で菌株の分離環境のアノテーションやマイクロバイオーム研究に用いられている他、食品会社等との共同研究による活用が進んでいる。

・プロテオーム DB（jPOST）は、国際コンソーシアム ProteomeXchange のメンバーとして国内外のデータ登録を受け付け、1,000 を超えるプロジェクトがデータ公開に向けて登録済みである。

各 DB のアクセス数の推移は次の表のとおり。一部の DB において減少も認められるが、概ね増加傾向にあり、アクセス数の面からも、利用は拡大傾向にある。

表：統合化推進プログラム各データベースの利用状況

いずれも月平均であり、各研究開発課題から提出された
終了報告書または中間報告書に基づく。

		年度				
		H29	H30	R1	R2	R3
ChIP-Atlas	ユニーク IP	2,589	694	1,452	4,360	4,613
	ページ数（千件）	79	97	126	144	211
KEGG MEDICUS	ユニーク IP	695,248	1,341,147	2,425,518	1,956,422	1,583,110
	ページ数（千件）	2,117	3,734	7,658	6,520	5,063
GlyCosmos	ユニーク IP	未公開	未公開	665	1,198	797
	ページ数（千件）	未公開	未公開	27	230	367
PDBj	ユニーク IP	42,050	51,359	65,863	91,427	75,870
	ページ数（千件）	870	2,471	2,335	3,088	2,919
MicrobeDB.jp	ユニーク IP	6,764	5,927	5,497	2,732	1,976
	ページ数（千件）	28	29	31	104	176
DBKERO	ユニーク IP	171	243	223	1,230	1,423
	ページ数（千件）	2	7	11	22	13
Plant GARDEN	ユニーク IP	未公開	710	8,573	9,751	6,016
	ページ数（千件）	未公開	4	193	80	64
jPOST	ユニーク IP	320	456	472	545	−
	ページ数（千件）	633	673	413	1,593	−

MetaboBank は開発中のためデータなし。

（4）課題間連携による統合化

経緯：センター発足～前回事業評価まで（前期）の進捗

データ種や生物種を超えたデータ利用に資するため、研究開発課題間の連携を推進した。取組みとしては、各研究課題参画者の交流・ネットワーク形成のため、キックオフ会議や合同成果報告会を実施した。また、技術面では、事業としての統合データ基盤構築のため、基盤技術開発との連携により統合化推進プログラムで開発した DB の RDF 化に取り組んだ。

RDF 化による汎用統合化データの整備・充実

公募要領で明記するとともに、中間評価会・サイトビジット等において、開発 DB の RDF 化を各研究課題へ要請した。また、相互連携性が高い RDF データを作成するため、前期に基盤技術開発で策定した「DBCLS データベース RDF 化ガイドライン」に沿った RDF 化の働きかけを継続した。また、DBCLS が実施する基盤技術開発と連携した RDF 化を効率的に実施するため、前期同様に、バイオハッカソンやスパークルソン（詳細は、『III. 基盤技術開発』41 ページを参照）をリアルタイムな開発の場として活用した。

RDF 化による DB 間データ連携の事例として、以下が挙げられる。植物と根圏微生物、疾患と腸内細菌叢といった生物種をまたぐ複雑な相互作用を、遺伝子、タンパク質、糖鎖等の様々な情報を包括して統合解析するための情報基盤となっていくことが期待される。

・植物 DB（Plant GARDEN）の糖鎖関連遺伝子を、糖鎖 DB（GlyCosmos）に収録
・タンパク質立体構造 DB（PDBj）の複合糖鎖立体構造を、糖鎖 DB（GlyCosmos）に収録
・微生物 DB（MicrobeDB.jp）と植物 DB（Plant GARDEN）が、植物共生細菌の研究者の意見を取り入れつつ、ホロゲノム解析機能を開発中

その他のデータ利用の円滑化に資する連携

各 DB の利便性向上の観点から、データ連携に加えて、課題間の技術連携や DB の機能面での連携も実施した。事例として、以下が挙げられる。

技術連携：

・糖鎖 DB（GlyCosmos）の協力による、タンパク質立体構造 DB（PDBj を含む wwPDB 全体）における糖鎖構造記述法の国際標準化

データ解析に即したデータ・DB 機能の連携：

・プロテオーム DB（jPOST）の比較解析結果を基に、エピゲノミクス DB（ChIP-Atlas）またはパスウェイ DB（KEGG）でエンリッチメント解析を行える機能の開発
・エピゲノミクス DB（ChIP-Atlas）の転写因子結合サイトの予測結果を、多層オミクス DB（DBKERO）上で閲覧可能にする機能の開発

（5）まとめ（II. 統合化推進プログラム）

国内のライフサイエンス研究等によって産出された研究データを広く収集し、より多くの研究者にとって価値あるデータとして提供するため、公募により 9 件の統合 DB の開発を実施した。より使われる統合 DB の開発のため、幅広く分野別（データ種別や生物種別）に統合 DB を整備・拡大するとともに、利用者との連携構築、課題間連携の推進によって、統合データ基盤を更に強化・充実させた。次の 3 点により、統合のための取組み・成果の双方について、ライフサイエンス研究を効率的・効果的に進めるための研究開発環境の整備・充実への寄与は十分であったと考えている。

○ DB 統合の一層の進捗と、より活用されるデータ基盤の形成に向けて、研究分野とデータ種類のポートフォリオを実施してライフサイエンスの幅広いデータを対象とした統合 DB を公募・採択し、利用者との連携、および DB を超えた課題間連携に主軸を置きつつ、進捗管理・中間評価等の研究開発推進マネジメントを実施した。

○ 各研究課題の DB において収録データを増加させるとともに、利用者視点による開発や国際連携・標準化による統合データ整備も進捗させた。各研究課題で見ると統合対象のデータによる違いや今後の取組み・開発が必要な部分はあるものの、全体として産学の幅広いライフサイエンス研究に活用されていることから、ニーズに沿ったデータ整備・開発により、今後の利活用の拡大が期待される。

○ データ種別を超えたデータ利活用に向けて、RDF 化による DB を超えた統合解析ツールの開発や、利用者のデータ解析に即した DB 間のデータ・機能連携を実施し、課題間連携によって統合 DB を相乗的に高度化した。

前期に引き続いて支援したテーマも多く、結果的に統合 DB の高度化やテーマ間の連携の進展につながった。また国際的な基盤整備にも貢献する DB 支援も実施できた。一方で、研究や計測技術の進展による新たなデータ種の共有・利活用が必要との認識があり、より多様なデータ種について整備を進められる仕組みを検討していく。加えて、国の方針に沿いながら、オープンサイエンスの観点で国際的により貢献する統合 DB 開発を進められるよう、検討することも必要である。さらに、今期に推進してきた利用者との連携をより一層推し進め、利用者の知識発見に貢献し、研究基盤としての価値を高めていくことを目指す。

III. 基盤技術開発

上述 II. 統合化推進プログラムで開発する DB を含め、ライフサイエンスにおける国内外の多様な DB を組み合わせて統合的に利活用するための基盤的な技術開発を実施した。

（1）開発概要と重点的取組み

開発の効率的な推進のため、開発の重点領域の設定、および、年度評価におけるマネジメントの強化を実施

（2）統合データ利活用に向けたアプリケーション開発

エンドユーザのデータ基盤利用の入り口となるアプリケーションの開発

（3）アプリケーションを支えるデータ基盤・ミドルウェア

共通フレームワークによる統合データ整備とアプリケーション開発者向けの技術開発

（4）国内外との連携・情報発信

大規模データ整備のための国内外連携および利活用に向けた情報発信

（1）開発概要と重点的取組み

経緯：センター発足～前回事業評価まで（前期）の進捗

国内外で取得される多種多様なデータを入手・利用するために必要なデータインフラ整備として、DB 収録データの記述に RDF（Resource Description Framework）を採用し、国際的な開発者ネットワークも形成しながら、統合データ整備およびデータ利用のための技術開発を実施した。また、統合化推進プログラムの成果 DB についても、複数 DB の統合利用のための RDF 化支援を実施した。主な成果として、DB 統合のための RDF 基盤の整備とその標準化のための国際連携が挙げられる。

センター発足時に 3 年間（平成 23～25 年度）の公募プログラムとして情報・システム研究機構ライフサイエンス統合データベースセンター（DBCLS）を採択し、本事業の第二段階への移行[30]に伴い、平成 26 年度以降は NBDC と DBCLS との共同研究により実施した。

外部有識者意見を受けた開発重点の設定

平成 29 年度以降も引き続き NBDC と DBCLS との共同研究により技術開発を実施した。実施にあたっては、平成 28 年度の事業評価・NBDC 運営委員会提言や後述の基盤技術分科会の評価を受けて、オープンサイエンスに向けた統合データの利活用促進のため、開発重点の設定等、開発構成の見直しを行った。具体的には、次の 3 点である。

① 利用者の属性によってニーズが異なることから、ニーズに沿った開発のため、主要な利用者層に対応する 4 つの開発レイヤーを設定

[30] 第 70 回ライフサイエンス委員会（文部科学省、平成 25 年 8 月 12 日開催）に報告を行い策定した平成 26 年度以降の事業の進め方については、『ライフサイエンス分野の統合データベース整備の第二段階の推進戦略』（URL: https://biosciencedbc.jp/gadget/unei/suishin_senryaku.pdf）を参照。

アプリケーション層：主な利用者層は、エンドユーザ・実験科学者
ミドルウェア層：主な利用者層は、アプリケーション開発者
データベース層：主な利用者層は、大規模データを活用する科学者（データ科学者）
データソース層：主な利用者層は、DB 構築者

② データ利用の具体的な出口設定に基づく開発のため、重点応用領域（医学、有用物質生産、育種）を設定

③ 重点応用領域での統合データの利活用促進に向けて、利用者との連携によるアプリケーション開発、および、関連する技術開発・データ整備を重点的に実施

基盤技術分科会の設置による年度評価の導入

平成 26 年度以降の共同研究については、各年度の研究開発計画を NBDC 運営委員会に諮り承認を得て実施しており、今期も同様であるが、多岐にわたる開発内容についてのより専門的な視点での議論・評価のため、平成 29 年度より、NBDC 運営委員会の下に「基盤技術分科会」を設置している。

基盤技術分科会は、バイオインフォマティクスや情報科学の技術開発の専門家だけでなく、多様なライフサイエンスデータをまとめて扱う利用者の視点での議論のため、システム生物学や大規模データ解析の知見を有する外部有識者で構成した[31]。研究開発全体の方針・テーマバランス、個々のテーマ見通しに対する評価・助言を受けており、総合的にはいずれの年度も「優れている」との評価であった。今後の研究実施に関して指摘された課題点については、次年度以降の開発計画に反映した。

[31] 委員名簿は、https://biosciencedbc.jp/about-us/organization/ に掲載。

（2）統合データ活用に向けたアプリケーション開発

経緯：センター発足～前回事業評価まで（前期）の進捗

技術開発によって統合したデータを使いやすく提供するため、TogoGenome（生物種とゲノムの RDF 検索システム）、RefEx（遺伝子発現のリファレンスデータセット検索システム）、CRISPRdirect（ゲノム編集のためのガイド RNA 設計システム）等の開発・サービス提供を実施した。

重点応用領域での統合データ利活用促進に向けたアプリケーション開発

統合データ基盤をライフサイエンスの研究現場へ展開するため、利用者の研究目的に合わせた効率的なデータ利用のためのアプリケーションの開発を進めた。具体的には、アプリケーション経由でデータを利用し自身ではプログラミングを行わない実験科学者も想定利用者に含め、データ基盤にアクセスするための使いやすいウェブインタフェースを提供するとともに、利用者意見を開発につなげることを目的とした。

新規を含め、5 年間を通して約 10 のアプリケーションを開発したが、3 つの重点応用領域のうち、ゲノム医療・個別化医療に向けた国際的な研究動向を踏まえた高い需要見込みから、先行して開発を推進した「医学」領域を中心に、概略を記載する。なお、「有用物質生産応用」、「育種」については、『④ その他のアプリケーション開発』で述べる。

① 日本人ゲノム解析への統合データ利用に向けた「TogoVar」の開発

医学応用のうち、日本人ゲノム解析への統合データの利用に向けて、NBDC 研究員と共同で、「日本人ゲノム多様性統合データベース （TogoVar)」を開発した（ヒトゲノムデータ利用における課題への対応については、『I. 外部連携およびポータルサイトの構築・運用』11～13 ページを参照)。

TogoVar では、利用者がバリアントデータを利用する際に関連する情報も参照できるよう、様々な DB 由来のデータを RDF 化により統合して提供している。RDF 化により統合して提供しているデータには、国内外のバリアント頻度情報（jMorp、HGVD、gnomAD）だけなく、バリアントの臨床的意義（ClinVar）や文献で言及されているバリアント（PubMed、PubTatorCentral)、被引用文献（Colil)等がある。これらの RDF 化したデータを後述の TogoStanza 等のミドルウェアにより画面表示し、アプリケーションを通じた統合データ提供を実現している。これらの国内外の関連データを一括で検索・比較を可能とする RDF 化によるデータ整備・ミドルウェアの応用が、『I. 外部連携およびポータルサイトの構築・運用』（11～13 ページ）に記載のとおり、国内外で活用される統合データの整備・提供の基盤となっている。

国際比較の観点では、サンプル数、日本人サンプル数、頻度集計前の生データ（個人ゲノム）の取得、検索条件の種類に関して優位性を持っている。関連データとの連携についても、後述の RDF の柔軟性・拡張性を活かしつつ、MGeND 等の DB との連携に今後取り組む計

画である。

表：TogoVar と類似するサービスとの比較

	TogoVar	gnomAD（米国）	dbSNP (ALFA)（米国）
全サンプル数	約 18 万人	約 14 万人	約 10 万人
日本人サンプル数	◎（約 18 万人）	△（約 80 人）	不明（5 千人以下）
集計前の生データ取得	◎（NBDC ヒト DB を通じて取得可能）	△	△
関連データとの連携	△(連携途上)	△	◎（NCBI の多様なデータと連携）
検索条件	◎（6 種類）（遺伝子名、変異 ID・位置・頻度、病原性、疾患名）	○（3 種類）	◎（6 種類）

（DBCLS による調査に基づき NBDC で作成）

② 希少疾患の診断率向上への統合データ利用に向けた「PubCaseFinder[32]」の開発

医学応用として、希少疾患の疾患解明や診断率向上（遺伝学的検査結果の解釈）に利活用できる統合データの整備・提供に向けて、希少疾患検索システム「PubCaseFinder」を開発した（平成 29 年 9 月）。

遺伝学的検査の解釈の際には、疾患候補を数十から数百に絞り込んだ後に、それら疾患において患者と同様の症状が報告されているかを、疾患 DB や文献を元に調べる必要があり、多くの手間が必要となっている。この課題への対応として希少疾患と症状との関係を素早く検索できるアプリケーションを実現するために、過去の文献（PubMed に収録されている 100 万件以上の症例報告）から希少疾患に関連する「兆候・症状」を機械的に（自然言語処理で）抽出して類似度を計算する技術開発を行った。

後述の比較表にある優位性等から、国内外での利用が進んでいる。特に、複数の研究機関（慶應義塾大学、東京医科歯科大学、東京大学、東北大学）で、医師や研究者に利用してもらいながら、利用者意見を開発に反映している。また、国際的には、GA4GH で世界中の研究機関から遺伝子型と表現型の一致する希少疾患の症例を探索する MatchmakerExchange におけるひとつのノードとして採用されている。

国際比較の観点では、対象疾患として希少疾患を含む、診療録からの入力が可能、日本語対応等において優位性がある。希少疾患は国際的なキュレーションが進んでいないが PubCaseFinder はそれ自身で文献から「兆候・症状」を抽出するため対応可能となっている。

[32] URL: https://pubcasefinder.dbcls.jp/

表：PubCaseFinder と類似するサービスとの比較

	PubCaseFinder	Phenomizer（ドイツ）	Face2Gene（米国）
提供環境	◎（ウェブ、API 有）	○（ウェブ、API 無）	◎（アプリ）
入力の容易さ	◎（症状リスト、診療録）	○（症状リスト）	◎（顔写真）
対象疾患の多さ	◎（遺伝性疾患、希少疾患）	○（遺伝性疾患）	△（顔貌に関連のある疾患）
検索精度	○（中）	△（低）	◎（高）
多言語化	○（日、英）	△（英）	△（英）

（DBCLS による調査に基づき NBDC で作成）

③ Togo Data Explorer（TogoDX）の開発

上述のような特定の用途に合わせたアプリケーションでは、多くの利用者が同じように繰り返し行っているデータ利用（主に既存データとの比較・参照用途）に対応しているが、アプリケーションを通じて提供できるのは本事業で大規模に統合してきた多種多様な統合データの一部にすぎない。オープンサイエンスの価値のもう一つの側面である、公開 DB を利用者の発想で複数組み合わせることによって予期していなかった発見を得るためには、従前とは異なる新たな開発アプローチが必要であった。

この課題の解決のため、DBCLS と NBDC の共同開発として、国内外の主要 DB を RDF 化や ID つながり（データ間のつながりや、ID の互換性）を利用し、大規模に統合した多様なデータを俯瞰しながら、利用者の研究目的に応じて絞り込んでいくことができる「Togo Data eXplorer（TogoDX）」の開発を進めた。TogoDX は、これまで基盤技術開発によって整備してきたデータ基盤（RDF 化によるデータ形式の統合）を活用しつつ、利用者自身の発想で柔軟に複数 DB 由来のデータを組み合わせられるよう、新規（後述の TogoID）や既存のミドルウェアを活用することで実現した。

令和 2 年度から開発を開始し、令和 3 年 10 月にヒト中心の分子・オミクス・疾患を対象とする第 1 版「TogoDX/human[33]」を公開した。TogoDX/human では、遺伝子、タンパク質と構造、化合物、糖鎖、疾患、バリアント等に関わるデータセットをつなぐことで、例えば、疾患関連遺伝子や疾患関連遺伝子でのバリアントの有無の探索、疾患に適応可能な化合物（薬剤）の検索、薬剤の開発状況の調査等、様々な応用が可能であるとともに、それらに必要なデータを簡単な操作で探索、取得することが可能である。

また、TogoDX は RDF により統合したあらゆるデータへの応用が可能であり、ヒト中心に限らず広範なデータを対象とすることができる。

[33] URL: https://togodx.dbcls.jp/human/

④ その他のアプリケーション開発

上記の他、医学以外の有用物生産・育種への重点応用領域での統合データ利用促進も念頭に、以下の開発を実施した。

・ゲノム編集におけるガイド RNA 設計システム「CRISPRdirect/GGGenome[34]」の改良

ゲノム編集技術の進展に対応し、設計対象生物種を拡大した他、Cas9 以外のヌクレアーゼを利用したゲノム編集への対応、オフターゲットサイトの検索機能の改良等を実施。また、秘匿情報を含む配列入力が難しいという民間企業利用者意見への対応として、株式会社レトリバによるパッケージ版商品の開発（令和 2 年 8 月）に合わせた従来のオープンソース部分の提供形態の策定を行った。

基礎研究での利用の他、少なくとも 3 つの核酸医薬品において医薬品医療機器総合機構（PMDA）での承認審査で安全性評価（オフターゲット検索）のため利用されている。

・微生物資源活用のための培地データの統合・提供（TogoMedium[35]）

有用物質生産応用への統合データ利活用促進に向けて、実験研究者と共同でデータ整備・アプリケーション開発を行い、令和 2 年度に公開した。国内の微生物資源（JCM、NBRC）を培養するための培地成分を RDF 化により統合することによって、微生物毎に異なる培地成分の比較を可能としており、有用物質生産のための培養条件検討への活用が見込まれる。

・育種関連データの活用に向けた技術検討

育種応用については、アプリケーション開発に向けて他機関の育種関連データの RDF 化の支援等を進めたが、平成 30 年度から DBCLS が実施機関として参画する「戦略的イノベーション創造プログラム（SIP）スマートバイオ産業・農業基盤技術」の中で、産業との連携に基づくデータ利活用の取組みを進めることとなったため、本共同研究においては技術検討（国内外 DB の RDF 化のための語彙等整備）のみとした。

[34] CRISPRdirect のサービス提供のために、別途のサービスとしても提供している GGGenome の配列検索技術を利用している。CRISPRdirect URL: https://crispr.dbcls.jp/

[35] URL: http://growthmedium.org/

表：主なアプリケーションの利用状況（アクセス数）

いずれも月平均であり、R3 年度は 9 月までの月平均。

		年度				
		H29	H30	R1	R2	R3
TogoVar	ユニーク IP	–	595	610	795	724
	アクセス数（千件）	–	30	94	89	75
PubCaseFinder	ユニーク IP	–	399	486	470	2,160
	アクセス数（千件）	–	32	42	27	48
CRISPRdirect	ユニーク IP	2,479	2,879	2,768	2,942	2,986
	アクセス数（千件）	21	24	23	24	23

TogoDX は R3 年 10 月開設のため、データなし。

（3）アプリケーションを支えるデータ基盤・ミドルウェア

経緯：センター発足～前回事業評価まで（前期）の進捗

ライフサイエンス関連データの統合技術として RDF を採用し、そのフレームワークに基づき、多様なデータの統合（後述のスパークルソンによる統合化推進プログラム DB の RDF 化支援を含む）およびデータ整備・利用を効率的に進めるための技術・ツールの開発を着実に進めた。RDF 採用のメリットとして、次のようなものが挙げられる。①どのような情報でも表現できる柔軟さ（ライフサイエンスの多様な DB どうしの統合に適する）、②複雑なデータへのスキーマレスな拡張も容易（DB・アプリケーション開発が迅速に行える）、③同じものには世界中で同じ ID が用いられるため、分散データの統合が容易（公開されれば理論上は統合利用可能）、④オントロジーの利用による語彙の共通化、⑤W3C による国際標準仕様でベンダーロックインが無い、等。

統合データ利活用促進に向けたデータ基盤の充実とミドルウェア開発

アプリケーションの開発を通じてライフサイエンスの研究現場に必要な情報を展開するためには、アプリケーション開発の土台となる RDF 化データ基盤のさらなる充実と、より効果的なデータ統合のための DB の整備、それらにアクセスするために必要なデータとアプリケーションをつなぐミドルウェアの開発が不可欠であると考えられた。

本項では、この 5 年間に拡充した RDF データと、10 以上あるミドルウェアの中で外部のアプリケーション開発者にも有効利用可能なツールである「TogoStanza」、テキストデータと RDF データをつなぐ仕組みとして文献等に含まれるライフサイエンス知識の RDF 化を支援する「PubAnnotation」、さらには上述の TogoDX の開発に不可欠な「TogoID」について概要を記載する。

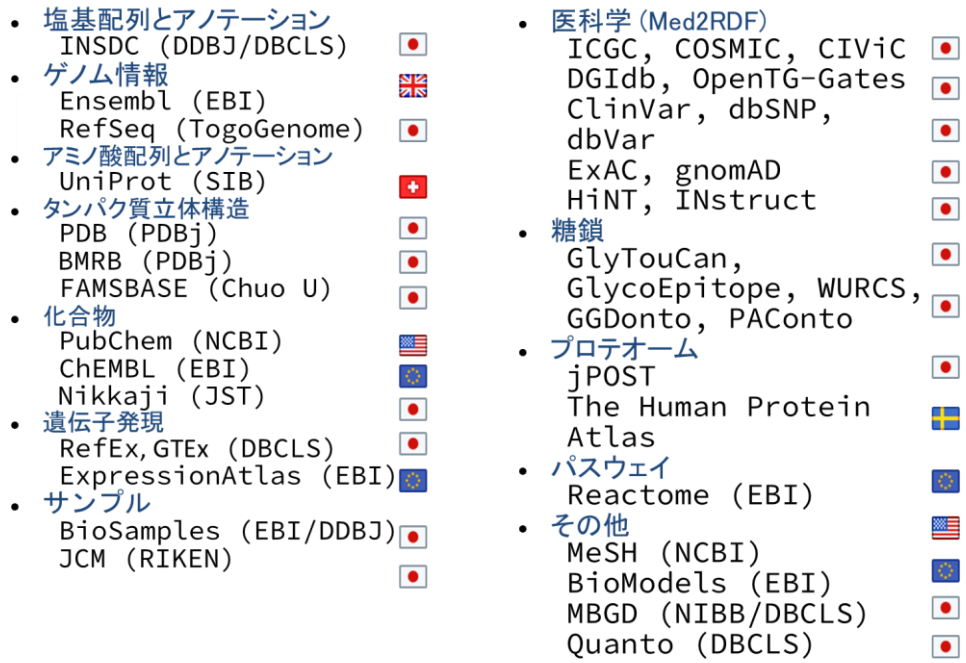
① ライフサイエンス知識の RDF 化の拡充

重点応用領域に寄与する DB の RDF 化を推進するため、前期と同様に後述のスパークルソン等を通じて統合化推進プログラムの成果 DB の RDF 化を支援しつつ、アプリケーションを通じたデータ提供のため、国内外連携により外部 DB の RDF 化を継続した。データを標準化し再利用性を高めつつ持続的に統合していくためには、EBI、SIB、NCBI 等の国際的なバイオインフォマティクスの DB 機関等との連携が不可欠である。そのため、これまで国際標準化に多くの成果をあげてきた後述の国際版バイオハッカソンを継続し、RDF 基盤の整備とその標準化のための国際連携を図った。

また、RDF 化データを効率よく利用できる DB インフラの整備として、NBDC と DBCLS の連携により平成 27 年度に開設した「NBDC RDF ポータル[36]」へのデータ収録を継続した。平成 28 年度末の 17 データセットに対して、令和 2 年度末までに 10 件のデータセッ

[36] URL：https://integbio.jp/rdf/dataset

トを新規に収録し、データセット数は 5 年間で 1.6 倍に拡充した。また、年度あたり 3〜7 件の既存データセットの更新や、大容量の RDF データを効率的に維持・管理・提供するための研究開発も併せて実施した。データセット数において、世界有数のライフサイエンスの RDF データリソースとなっている。



（DBCLS 作成）

図：国内外で RDF 化されたデータベース

② RDF データの検索結果を可視化するミドルウェア「TogoStanza[37]」

RDF データを利用したウェブツールの開発の効率化、高速化に貢献することを目的に、平成 27 年に公開したミドルウェア。RDF DB を SPARQL クエリで検索した結果を、グラフやマップ等で可視化する再利用可能なモジュール（Stanza）を提供する。

平成 29 年度以降の開発として、アプリケーション間の連携やウェブツール開発の一層の効率化を目指し、新機能の開発等を実施した。上述の TogoVar 等、他の DBCLS のサービスの開発や、統合化推進プログラムの DB（MicrobeDB.jp や関連 DB）の開発でも利用されている。

[37] URL: http://togostanza.org/

③ テキストデータと RDF データをつなぐ「PubAnnotation[38]」

文献のような自然文から機械可読な情報を取り出すためには、データとして抽出できるような注釈付（アノテーション）が必要である。アノテーションデータは SPARQL 検索が可能であることから、アノテーションデータの整備・集積により、文献に記載された知識とオミックス等計測データとの統合検索が可能となる。

PubAnnotation は、様々な文献アノテーションデータを統合する共通レポジトリとして開発し（平成 24 年)、すでに世界最大規模の文献アノテーションデータの集積拠点となっている。今期においては、多種多様な文献アノテーションデータの蓄積に伴うデータの比較や評価への対応、エンリッチメント解析に応用するための仕組みの開発、海外研究者との共同による機械学習への活用に向けたデータ・ツールの検討、等を実施した。また、PubCaseFinder や統合化推進プログラムの糖鎖 DB「GlyCosmos」と連携し、文献からの知識抽出・DB 化を支援した。

④ 様々な DB で用いられるデータ ID の相互変換を行う「TogoID[39]」

複数の RDF 化 DB をまとめて利用するためのデータ基盤技術として開発を継続し、公開版として今期に新規開発。国内外の 60 以上の DB で用いられているデータ ID 間の対応関係を検索および変換することができる。1 対 1 の ID 変換だけでなく、疾患の ID から関連する遺伝子の ID への変換等、数珠つなぎのように複数の DB 間をまたぐ変換も可能としている。

上述の TogoDX では、RDF 化による DB 形式の統一に加えて TogoID による ID の相互関係を利用することによって、アプリケーションの機能を実現している。

[38] URL: http://pubannotation.org/

[39] URL: https://togoid.dbcls.jp/

（4）国内外との連携・情報発信

経緯：センター発足～前回事業評価まで（前期）の進捗

GA4GH への加入によるヒトデータの国際的な共有の動向把握を行うと共に、開発者が合宿形式で問題解決にあたるバイオハッカソン等により国内外開発者のネットワークを形成し、「FAIR 原則」への貢献や RDF 化による DB 統合の標準化に向けた共同開発・連携につなげた。また、統合化推進プログラムにより開発した DB の RDF 化を進めるためにスパークルソン（SPARQLthon）を開催して支援し、効率的で相互連携性の高い RDF 化のための「DBCLS RDF 化ガイドライン」を策定した。

DB 統合化に向けた国内外連携

前期に引き続き、バイオハッカソン等を通じてデータモデルやオントロジー等の標準化に向けた国内外の開発者コミュニティの形成と共同開発を実施した。

具体的には、技術的な課題を克服すると同時に、標準化のための合意形成とその実装にあたる国内外の開発者コミュニティ作りのために、オントロジー開発・辞書開発・ガイドライン作成・RDF 化推進・システム開発に関する国際版および国内版のバイオハッカソンを継続的に開催し、当該分野のハブとして世界的なプレゼンスを目指した。文献からの情報抽出とその利活用のための技術開発を行うハッカソンである BLAH も継続して開催した。加えて、主に統合化推進プログラムを対象とした RDF 化支援を行うスパークルソンも継続して毎月開催した。これらの活動は、開発者どうしのネットワーク・交流の場を提供することによって、DB を支える人材の支援にも貢献していると考えている。

・国際版バイオハッカソン：

国内外の DB 開発者が一堂に会し、合宿形式で共同開発を実施。

H29～R1 年度は毎年開催、3 回の延べ参加者数は 388 名

（1 回あたりの参加者数として、前期の約 1.5 倍）。

R2、3 年度は、新型コロナウイルス感染症拡大防止のため開催見送り。

・国内版バイオハッカソン：

国内の DB 開発者とのネットワーク構築と共同開発のため実施。

毎年開催し、5 回の延べ参加者数は 283 名（※）。

・Biomedical Linked Annotation Hackathon（BLAH）：

国内外の研究者が共同で文献からの情報抽出とその利活用のための技術開発を行う。

H29～R2 年度は毎年開催、4 回の延べ参加者数は 158 名（※）。

・スパークルソン：

主に統合化推進プログラムに参画している研究者を対象に、DB の RDF 化を支援。

前期同様に毎月開催し、H29～R2 年度の延べ参加者数は 1,329 名（※）。

※1 回あたりの参加者数として、前期と同水準

国際版バイオハッカソンを通した開発連携として、次のような例が挙げられる。

・国内外 DB 開発者との共同による DB の RDF 化
（INSDC 関連 DB について欧州 EBI と、医学関連 DB について AMED 事業参画者との共同）
・国内外 DB 開発者との共同による RDF 化に向けたデータモデルの検討
（育種関連 DB についてフランスの AgroLD との共同）
・国際コンソーシアム（GA4GH）との連携に向けたアプリケーション機能開発
（GA4GH プロジェクト参画の MatchMakerExchange とのシステム連携）

利活用に向けた情報発信

上述のハッカソン等開催の他、前期に引き続き、ライフサイエンスの DB やツールの動画マニュアル等を掲載する「統合 TV」の運営を通じて、本事業で開発した DB・サービスの使い方や講習会資料を配信した。また、NBDC との共同により、DB 講習会（AJACS）・学会展示によるサービス紹介も実施した（詳細は『IV.（1）広報およびデータベース講習会』44〜45 ページを参照）。

論文発表

Kawashima S, Katayama T, Hatanaka, H, Kushida T, Takagi T. (2018) NBDC RDF portal: a comprehensive repository for semantic data in life sciences. Database, 2018: bay123. DOI: 10.1093/database/bay123　※NBDC 研究員との共著のため、再掲

Katayama T, Kawashima S, Okamoto S, Moriya Y, Chiba H, Naito Y, Fujisawa T, Mori H, Takagi T. (2019) TogoGenome/TogoStanza: modularized Semantic Web genome database. Database, 2019: bay132. DOI: 10.1093/database/bay132

Kim JD, Wang Y, Fujiwara T, Okuda S, Callahan TJ, Cohen KB. (2019) Open Agile Text Mining for Bioinformatics: The PubAnnotation Ecosystem. Bioinformatics, 35(21): 4372–4380. DOI: 10.1093/bioinformatics/btz227

Garcia L, Antezana E, Garcia A, Bolton E, Jimenez R, Prins P, Banda JM, Katayama T. (2020) Ten simple rules to run a successful BioHackathon. PLoS Comput. Biol., 16(5): e1007808. DOI: 10.1371/journal.pcbi.1007808

他 58 報。

（5）まとめ（III. 基盤技術開発）

ライフサイエンスにおける国内外の多様な DB を組み合わせて統合的に利活用するための基盤的な技術開発を実施した。開発の効率的な推進のため、利用者意見を取り入れつつデータ基盤の利活用のためのアプリケーション開発に重点的に取り組んだ。アプリケーションを支えるデータ整備・技術開発、および、国内外連携によるデータ統合に向けた研究者ネットワークの形成も、引き続き実施した。次の 4 点により、統合のための取組み・成果の双方について、ライフサイエンス研究を効率的・効果的に進めるための研究開発環境の整備・充実への寄与は十分であったと考えている。

○ 統合データ基盤をライフサイエンスの研究現場へ展開するため、DB 開発者向けの研究開発からより広範な利用者に向けたアプリケーション開発に重点を移し、重点応用領域（医学、有用物質生産、育種）でのデータ利用促進に向けた開発を実施した。

具体的には、日本人ゲノム解析への統合データ利用に向けた「TogoVar」、希少疾患の研究・診断への統合データ利用に向けた「PubCaseFinder」を新たに開発する等、約 10 のアプリケーション開発を実施した。

○ 本事業で開発した技術が、利用者の研究開発や GA4GH 等国際研究ネットワークで活用されており、科学研究のみならず創薬等の産業分野においても波及効果につながった。

○ バイオハッカソン等により国内外の研究者とのネットワークを拡大し、これにより、DB の RDF 化による統合やオントロジー開発等の国際的な協力体制が構築され、DB を支える人材の支援にも寄与した。統合化推進プログラム等国内研究者への支援により、RDF ポータルに収録したデータセット数は約 1.6 倍に拡大した。

○ 従前とは異なる開発アプローチによって、多様なデータを俯瞰しながら利用者が柔軟に複数 DB のデータを組み合わせて抽出できる「TogoDX」の開発も実施した。

TogoDX は、本事業で整備してきたデータ基盤の活用インタフェースとして、ヒト中心の統合データに限らず、多様なライフサイエンスのデータへ応用が可能である。

ヒトデータに関して開発を先行させたことから、今後に向けた課題点として、ライフサイエンスの他の領域（育種、有用物質生産等）に向けた技術開発が挙げられる。加えて、多様な DB の統合化を進捗させ、統合化に向けた国内外ネットワークが形成されてきたことを踏まえ、今後、統合・活用しやすいデータ・DB を整備するためのツール・技術・ノウハウをより広く波及させていくことも課題となる。データ利活用により知識発見を行う研究者との連携を引き続き実施しながら、統合 DB の価値を高めるために、データ駆動型科学の成功事例を積極的に発信することの必要性も高いと考えている。

IV. 事業全体に共通する取組み

上述 I.〜III.の各項目全体に共通する取組みとしての広報・講習会等、および、事業全体における取組み・成果のまとめを次のとおり記載する。

（1）広報およびデータベース講習会

経緯：センター発足〜前回事業評価まで（前期）について

毎年 10 月 5 日を「トーゴーの日」として DB 統合の成果を報告する公開シンポジウムを開催するとともに、DBCLS との共同により DB 統合の活動を紹介する初心者向けの講習会「AJACS」や学会出展を実施した。AJACS は、平成 24 年度以降は講習会の受入機関を募集し、受入機関が希望する講習内容で講師派遣を行う形式で実施した。

既存取組み（講習会、シンポジウム等）の改良、新規取組み（ダイレクトメール、ブログ、メールマガジン）の導入により、各活動の発信力向上、発信媒体の多層化、活動間の連携強化を図った。多様な取組みを通じて、本事業の認知度向上、DB の利活用方法・統合の成果発信等を広く進めた。

既存取組みの改善

① 統合データベース講習会（AJACS）

前期同様に DBCLS との共催により、本事業成果 DB の講習会を開催した。平成 29 年度以降、令和 3 年 8 月までに 23 回開催し、延べ 1,833 名が参加した（前期 6 年間 33 回の延べ参加者数の約 1.4 倍、1 回あたりの平均で前期の約 2 倍）。今期の開催において、以下の取組みの工夫を実施した。講義資料および講義動画は後日公開し、多くの研究者・学生が DB の使い方を自習できるようにしている。

・NBDC/DBCLS 企画形式による開催（平成 30 年度〜）

従来の開催受入機関の希望に沿った講習内容で開催に加えて、本事業として重点的に広報したいサービスを取り上げる企画回を年 1 回以上開催した。従来形式の 1.3 倍の参加者を得たことから、受講者の要望に沿った企画であったと考えている。

・オンライン開催（令和 2 年度〜）

新型コロナウイルス感染症拡大防止の観点から開催受入機関の募集は中止し、全回オンライン開催とした。集中力が途切れやすいオンライン受講の特性を踏まえ、開催時間の短縮や、回毎にテーマ性を持たせてより深い理解につながるように、工夫を行った。オンライン化で参加者数が約 2 倍となり、地域的な制約なく受講者を取り込めた。

② トーゴーの日シンポジウム

引き続き、事業成果を発信する一般公開シンポジウムを毎年 1 回開催した。今期の開催においては、DB 開発者と利用者との交流・議論を深めるため、DB 開発者に加えて事業成果 DB の利用者にも口頭発表を依頼する等の工夫を行った。なお、上述の DB 講習会

と同様に、令和 2、3 年はオンライン開催とした。平成 29 年度以降の 5 回で延べ 1,412 名が参加し、1 回あたりの平均参加者数は前期と同水準であった。

③ その他

前期と同様に、上述の講習会や学会出展（年 2〜5 回）の機会を利用する等して、サービスへの意見収集を実施した（年平均 50 件程度)。

また、前期から運用してきた NBDC ウェブサイト（コーポレイトサイト）について、事業内容や成果を明瞭に伝えられる基盤とすることを目指し、平成 31 年 4 月に大幅改修を実施した。情報を構造化して機械可読性を高めることで人にも機械にも見やすいものとし、改修前後半年の比較でアクセス数は約 1.5 倍に増加した。

新規取組みによる情報接点の追加

発信媒体の多層化により、認知度の向上や利用者との継続的な関係性構築を目指した。既存取組みとの連携により、イベント開催情報をメールマガジンやダイレクトメールで紹介したり、シンポジウムやウェブサービスを題材にしたブログ記事を公開したりする等、発信媒体を連携させ、相乗的な広報効果が得られるように取り組んだ。

① ブログ（平成 30 年度〜）

DB に詳しい研究者以外にも活動をわかりやすく伝え、また親近感を持ってもらうための取組みとして導入。NBDC、DBCLS 両所属員が、様々な活動の紹介、解説等を掲載した。4 年間で 40 件以上の記事を公開し、累計 9 万回以上閲覧されている。特に閲覧が多いのは、令和 2 年 3 月に公開した新型コロナウイルス感染症（COVID-19）に関する国内外の研究データの紹介記事であり、累計 5 万回以上閲覧されている。他に、トーゴーの日シンポジウム抄録を題材にしたキーワードの可視化法、DOI (Digital Object Identifier)の活用法が、それぞれ 5,000 件以上閲覧されている。ブログ記事へのアクセスの約 75%がウェブ検索サービス由来であり、NBDC に接点を持っていなかった潜在的利用者を含む幅広い層へのアプローチに貢献していると言える。

② メールマガジン（平成 30 年度〜）

講習会・シンポジウムの参加者やコーポレイトサイトへの来訪者等、NBDC を認知している層に対し、継続的に情報を配信する目的で実施した。登録者は約 3,300 人で、4 年間で 60 回以上を配信した。配信内容は、講習会・シンポジウムの開催情報、新規サービス等事業成果、ブログ新規掲載等とした。メールマガジン経由のコーポレイトサイトへのアクセスの約 10%が、講習会・シンポジウム参加登録フォームを開いていることから、イベント参加に意欲的な層への情報提供ツールとなったと言える。

③ その他

NBDC を知らなかった層の掘り起こしを目的として、平成 30 年度から講習会やシンポジウムの開催情報のダイレクトメール（DM）送付を実施した。民間企業を含む全国の研究機関を NBDC において調査し、1 回あたり約 500 カ所、計 14 回の送付を行った。

（2）NBDC 運営委員会提言への対応状況

平成 29 年 3 月の NBDC 運営委員会提言については、上述の I.～III.のそれぞれの活動や相互の連携により対応を実施した。

提言 1 への対応（未公開データを含めた大型プロジェクト支援）

AMED、SIP「スマートバイオ産業・農業基盤技術」実施機関との連携による 2 つのグループ共有 DB の構築支援と運営により、延べ 40 件以上の研究機関が関わるデータ共有体制の構築に貢献した。AMED ゲノム制限共有データベース（AGD)では、「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」や BBJ などからのゲノム医療研究におけるデータを受け入れ、SIP Healthcare Group Sharing Database（SHD）では、「食によるヘルスケア産業創出」に向けて集積されたデータの共有のための支援を実施した。この他、円滑なデータ公開・共有、連携によるデータ価値の最大化等について、主な事例のみで 5 件以上のプロジェクト等との連携・協力を実施した。（詳細は、15～16 ページ）

提言 2 への対応（利用者視点によるデータ統合）

事業の 3 つの活動（I.～III.）のそれぞれにおいて利用者意見を踏まえた開発・改善を行うと共に、RDF 化により 3 つの活動の連携によるデータ基盤形成を推進した。具体的には、

・NBDC ポータルサイトのサービス充実（TogoVar の新規開発、カタログ等既存サービスの改善）に利用者意見・レビュー結果を反映（詳細は、6～7、11～13 ページ)。
・統合化推進プログラムにおいて、糖鎖 DB や植物 DB など各課題でのユーザ意見聴取、研究アドバイザーに利用者側からの助言ができる有識者を選任（19～21、26 ページ)。
・基盤技術開発において、TogoVar や PubCaseFinder など具体的なデータ利用目的に合わせたアプリケーション開発に重点を置くとともに、開発において利用者との連携を実施（33～36 ページ)。
・DB 間データ連携を進捗させ、様々なオミクス、生物種、知識（テキストデータ）を超えたデータ活用基盤の整備を前進させた（38～39 ページ)。データ活用基盤に加え、統合データ活用インタフェース TogoDX の開発（35 ページ）や、統合化推進プログラムの微生物－植物やプロテオーム－エピゲノミクスの DB 間連携（29 ページ）など、データ活用のための技術開発も実施した。

提言 3 への対応（大規模データを扱う研究者とのデータ統合・利活用への取組み）

上述の提言 1、2 への対応と重複する部分もあるが、事例として、国内バイオバンクと共同してこれまでにない規模の日本人ゲノムデータを統合し、国際的に価値あるデータの整備・提供を実施したことが挙げられる（12～13 ページ)。また、個別のデータ利用目的に合わせたアプリケーションの提供に加え、広範な利用者が大規模に統合した多様なデータセ

ットを利用者の発想で複数組み合わせることによって予期していなかった発見を得られるよう、TogoDX の開発を実施した（35 ページ）。TogoDX から抽出したデータの機械学習等データサイエンスでの活用、TogoID や PubAnnotation などミドルウェア（39～40 ページ）も、大規模データを扱う研究者に貢献するものと考えている。

これら成果が利活用により強く結びつくためには、データ駆動型科学の成功事例等、DB 統合がもたらす価値の発信が必要である。外部との密接な共同体制を構築することはもとより、統合データ基盤を構築・活用できる人材の育成も大きな課題であると考えている。本事業を通じて国内外連携や人材支援の成果を得ているところだが、研究や社会における DX への対応にはまだ十分とは言えない。なお人材に関しては、国としても課題となっており議論が進められている。

提言 4、5 への対応（継続的なデータ整備・統合、国内外プレゼンス強化）

提言に記載のとおり NBDC 単独では解決が困難であるものの、関係する機関と連携し、次のような取組みを実施した。

・統合の拡大に向けた研究者ネットワークの拡大：
従前の国内 DB センター（DBCLS、DDBJ）や国内 DB 開発者との連携に加え、国内のバイオバンク等、大規模にデータを生産する外部プロジェクトとの連携を構築した。バイオハッカソン等による開発者のネットワーク形成や講習会も引き続き実施し、データの共有・統合に関わる意識醸成や人材支援に取り組んだ。

・国内外プレゼンス強化：
事業の 3 つの活動（I.～III.）のそれぞれにおいて、国内外 DB 連携や日本人データ統合等、国際的なデータ共有の課題解決に向けたデータ基盤を形成し、国内外に向けて成果発信を実施した。上述の研究者ネットワークの拡大と合わせ、国内外のデータ共有プロジェクトに貢献することにより、プレゼンス強化を図った。

なお、国内 DB センターとの関係性について、文部科学省ライフサイエンス委員会基礎・横断研究戦略作業部会の本年 6 月の報告書「今後のライフサイエンス研究支援基盤の在り方について」[40]において、今後の対応の方向性として次のとおり記載されている。

（対応の方向性）
○ NBDC は、ファンディングエージェンシーとして、国内外の統合データベースの連携や、データ共有のネットワーク構築及び国際的にも認知され高い価値を有するデータベースを対象にファンディング等に注力するなど、JST と ROIS のそれぞれの役割分担を明確にし、運用していくべきである。

[40] URL: https://www.lifescience.mext.go.jp/files/pdf/n2272_14.pdf

（3）まとめ（事業成果）

国内のライフサイエンス研究等の成果であるデータ・DB を統合的に扱うためのデータ整備・技術開発に対して、3 つの活動（I. 外部連携およびポータルサイトの構築・運用、II. 統合化推進プログラム、III. 基盤技術開発）を組み合わせて取り組んだ。統合データを充実させ利用者が活用できるデータを拡大するとともに、より活用しやすく統合したデータを提供するための新規サービス・技術開発により、データ活用基盤としての機能を向上させた。

3 つの活動のいずれにおいても、以下 2 つのことが言える。

○ より活用される統合データ基盤の整備のため、利用者や外部プロジェクトとの関係強化に取組み、統合の拡大と応用を意図したデータ整備・技術開発を実施した。

具体的には、利用者のデータ用途に合わせた新規サービス開発・データ統合（TogoVar、PubCaseFinder 等）、統合化推進プログラムの各研究課題における利用者との連携強化、が挙げられる。

○ 利用者に提供するデータを着実に充実させ、多種多様なデータについて国際的な統合化・共有に貢献するとともに、産学の幅広いライフサイエンス研究でのデータ利用に貢献した。

特に、前例がない規模の日本人ゲノムデータの統合に寄与し国内外での活用につながった他、統合化推進プログラムにおいて国際連携・標準化による統合データ整備が進捗した。

その他、事業全体に共通する取組みとして、以下 2 つのことが言える。

○ RDF 化した DB を着実に増やし、統合データ基盤を拡大させた。基盤の拡大のためにバイオハッカソン等による国内外研究者のネットワーク・協力体制を形成、活用した。加えて、統合したデータ基盤を広範な利用者に活用しやすく提供するため、TogoDX や関連技術の開発にも取り組んだ。

○ 統合データ基盤整備のすべての取組みを通じて、データ共有・統合に向けた意識醸成やデータ基盤を構築・活用する人材への支援に取り組んだ。

結論として、多様な DB の統合化が進捗し統合化に向けた国内外ネットワーク形成が着実になされている。一方で、今後に向けては、3 つの活動を通して、事業の一つの目標であるオープンサイエンスへより一層貢献するべく、DX 等の社会状況を踏まえつつ、より多様なデータ種、国際的なニーズや連携を視野に入れた統合 DB 整備を進めること、データ利活用により知識発見を行う研究者との連携をより積極的に実施し本事業の技術・ノウハウを広め、データ駆動型科学の成功事例の創出と発信に力を入れることが必要と考えている。