

令和4（2022）年2月

ライフサイエンスデータベース統合推進事業
平成29年度～令和3年度 事業評価報告書

国立研究開発法人科学技術振興機構（JST）
バイオサイエンスデータベースセンター（NBDC）運営委員会

本報告書は、JSTがNBDCにおいて実施するライフサイエンスデータベース統合推進事業の平成29年度から令和3年度までの取組みについて、NBDC運営委員会が行った事業評価の結果をまとめたものである。

1. 事業概要

本事業は、総合科学技術会議ライフサイエンスプロジェクトチーム統合データベースタスクフォースにより取りまとめられた「統合データベースタスクフォース報告書」の内容を受けて、平成23年度に開始した。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の統合を推進している。

具体的には、ライフサイエンス分野のデータやデータベースを利用者が効率的・効果的に扱えるようにするため、次の3つの活動を組み合わせて事業を推進している。

1) 外部連携およびポータルサイトの構築・運用

ライフサイエンス研究の成果であるデータやデータベースを利用者が統合的に扱えるようにするためのサービス提供及び外部プロジェクト等との連携

2) 統合化推進プログラム

公募を通じた分野別（データ種別や生物種別）の統合データベース開発

3) 基盤技術開発

国内外の多様なデータベースを組み合わせて活用するための基盤的な技術開発

2. 事業評価の目的

平成 28 年度に NBDC において実施した前回事業評価以降の 5 年間の事業運営状況を検証し、今後の改善に資することを目的とする。

3. 評価者

NBDC 運営委員会が評価を実施した。令和 3 年度の委員会構成は以下のとおり。

	氏名	所属
○	城石 俊彦	国立研究開発法人理化学研究所 バイオリソース研究センター センター長
	相澤 彰子	大学共同利用機関法人情報・システム研究機構 国立情報学研究所 コンテンツ科学研究系 教授
	大江 和彦	国立大学法人東京大学大学院医学系研究科 教授
	小田 吉哉	国立大学法人東京大学大学院医学系研究科 特任教授
	加藤 規弘	国立研究開発法人国立国際医療研究センター 研究所 遺伝子診断治療開発研究部 部長
	後藤 由季子	国立大学法人東京大学大学院薬学系研究科 教授
	末吉 亙	KTS 法律事務所 弁護士
	諏訪 牧子	学校法人青山学院 青山学院大学大学院 理工学研究科 教授
	原田 秀幸	日本製薬工業協会 研究開発委員会 副委員長
	松田 秀雄	国立大学法人大阪大学大学院情報科学研究科 教授

○：委員長

(委員長以外は 50 音順)

4. 評価方法

4. 1 評価手順

1) 評価項目の設定

JST 第4期中長期目標における本事業の評価軸を踏まえた上で、委員会の検討により NBDC 発足当初想定された国の課題対応や先進性の視点も加え、設定した。

2) 書面評価

NBDC が作成した事業成果報告書による書面評価を、令和3年10月下旬から11月中旬に実施した。

3) ヒアリング

同年11月16日に NBDC センター長へのヒアリングを実施した。

- ・NBDC センター長からの事業報告と質疑応答
- ・委員による議論

4) 評価報告書の作成

書面評価ならびにヒアリングに基づき、委員全員の確認・合意を得て委員長が評価結果を取りまとめた。

4. 2 評価項目

1) 事業の各取組みそれぞれについての評価項目

- ・事業の趣旨に照らして適切な取組み（内容・活動レベル）であるか。
- ・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が得られているか。
- ・活動の成果は、利用につながっているか（先進的な取組みの場合、期待される効果は十分か）。

2) 事業全体についての評価項目

- ・事業の趣旨に照らして適切な取組み（内容・活動レベル）であるか。
- ・国として対応が必要な課題（恒久化・一元化、人材育成、国内外プレゼンス強化）につながる取組みなど、取組みの工夫において評価すべき内容はあるか。
- ・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・充実に寄与しているか。

上記の項目による評価に加え、「今後、強化・改善すべき点」についても委員意見を取りまとめた。

5. 評価結果

5. 1 総合評価

平成 28 年度の前回事業評価や NBDC 運営委員会提言も踏まえつつ事業趣旨に沿った取組みが適切に実施され、国内外のデータベース統合化、およびそれを利用者が活用するための技術開発により統合データ基盤が着実に整備・運用されている。事業共通のデータ形式（Resource Description Framework、RDF）を採用するなど事業の各取組みを連携させて多様なデータベースを利用者が横断的に活用できるよう提供し、事業成果の利用も全体として拡大している。加えて、データベース統合に向けた国際的な協力体制の構築や利用者視点の取組み・情報発信も適切であり、今後のさらに広範な統合への寄与や一層の利用拡大も期待できる。

特に、国内外研究者とのネットワークの形成・拡大に向け、外部プロジェクト連携やハッカソンなど優れた取組みを実施してきた。これらは、TogoVar に代表される利便性の高い統合データの構築・共有への貢献や国内外連携による統合データ基盤の拡大の他、データ共有に関わる人材の交流・支援、ならびに、本事業の国内外プレゼンスの強化にもつながるものである。本事業のみでは対応が難しい恒久化・一元化に向けても、国内事情を踏まえ、各研究機関の公開データベースを分散的に統合するという方向性に沿って技術基盤を整えていると評価できる。また、新規サービスの開発やデータベース統合のための各種取組みに利用者視点の反映を目指し、事業成果の認知・利用拡大に向けた講習会をはじめとする各種情報発信にも工夫を凝らしている。

以上より、事業全体として、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環境の整備・充実への寄与に向けた積極的な取組みにより、優れた成果が得られたと評価する。

5. 2 事業の各取組みについての評価

1) 「外部連携およびポータルサイトの構築・運用」について

各省の公的資金等により産出されたデータや開発されたデータベースを利用者がまとめて扱えるようにするという趣旨に沿った取組みが適切に行われている。前期からの継続的な取組みに加え、利用者視点を踏まえた新たな取組みによりデータ・機能を共に充実させ、データ活用基盤の向上を進めた。特に、ヒトデ

ータに関しては、寄託された機微データを誰でも閲覧できるよう加工して提供する新規サービス TogoVar を構築したことにより、東北メディカル・メガバンク、バイオバンク・ジャパンなどとの外部連携によるデータ充実の加速につながった。

成果の面では、各サービスにおいて利用者が活用できるデータベースやデータを増加させており、また4段階の統合（Integbio データベースカタログ、生命科学データベース横断検索、生命科学系データベースアーカイブ、NBDC RDF ポータル）のいずれにおいても利用者数の指標（月平均ユニーク IP 数）が前期（平成28年度）の2倍以上となるなど、着実に拡大した。新規取組みの中で、外部連携により日本人としてかつてない規模（約7,600人）の全ゲノムデータをプロジェクト横断的に統合したことは、国際的に価値ある統合データの構築・共有への貢献として優れた成果と言える。

更なる認知度向上・利用促進が課題ではあるものの、全体として、限られた予算の中での効果的な取組みが成果につながっていると評価する。

2) 「統合化推進プログラム」について

公募を通じて多様なデータ種を対象に研究開発課題を採択しデータベース統合を推進することは、取組みとして特に重要度が高い。また、研究アドバイザーに各分野のデータやその利用に専門的な知識を持つ産学の研究者を加えて利用者視点を反映するための研究開発マネジメントを実施したこと、および事業共通のデータ形式である RDF などにより課題間連携を推進したことは、優れた取組みである。データ利用者との連携・協業を推進すると共に、国際的なレポジトリの運用など国際的にも価値あるデータ基盤の整備により、データ利用の拡大にも取り組んだ。

成果の面では、データベースの統合化に向けて、各採択課題におけるデータベース開発・収録データの充実に加え、RDF 化や利用者のデータ解析に即した課題間連携による各データベースの高付加価値化や相乗的な高度化も優れている。月数万件（ユニーク IP）のアクセスがある KEGG MEDICUS や PDBj に代表されるように、採択課題の多くについて成果データベースはよく利用され、多数の論文成果や産学共同研究につながっている。国際連携によるレポジトリの運用は、国内外のオープンサイエンスに貢献していると評価する。

しかしながら、採択課題の一部については成果・利用がまだ十分とは言い難い。

対象となる分野での研究進展が速いことを考えれば、適時に質の高いデータベースを公募・支援できるような取組みの検討も必要だろう。このように課題はあるが、プログラム全体としての取組みならびに成果・利用の状況は優れ、課題間連携が国内のデータベース統合の推進に大きく貢献していることを考えれば、今後も長く継続する必要がある。

3) 「基盤技術開発」について

データ基盤開発から普及活動にいたる広範かつ多様な課題に対応する中で、優先度を付けて利用者視点のアプリケーション開発に重点的に取り組んだこと、および、ハッカソンなどにより国内外研究者とのネットワークを形成・拡大したことは、取組みとして特に高く評価できる。データベースの RDF 化による統合やオントロジー開発等を通じて国際的な協力体制が構築されたことは、データの共有・利用に携わる人材育成への寄与につながっている。また、アプリケーション開発の他、アプリケーションとデータ基盤をつなぐミドルウェア開発の取組みも、データの統合利用の裾野と恩恵を拡大するものと言える。

成果の面では、複数のアプリケーション（TogoVar、PubCaseFinder）が国際研究ネットワーク（Global Alliance for Genomics and Health、GA4GH）でも活用されている他、国際的にも多様なデータベースをつなぐ手法となっている RDF に関する技術の開発・公開によって統合データ基盤拡大における先導的な役割を果たしている。

TogoDX など提供開始から間もないサービスもあるため、継続的に利用者の声に耳を傾け使いやすさを向上させていく必要があるが、上述のとおり国内外の研究者とのネットワークが形成され、利用に向けた情報発信も積極的に実施していることから、今後の利用拡大が期待できる。

4) 上記以外の事業共通の取組みについて

認知度向上や利用促進に向けて、データベースの利用方法や統合の成果など事業取組みの発信を前期より広く積極的に進めている。特に、データベース講習会は、データの共有・利用に携わる人材への支援の観点からも有効である。

また、前回事業評価等を踏まえた平成 29 年 3 月の NBDC 運営委員会提言にも事業の各取組みの中で適切に対応している。

6. 今後、強化・改善すべき点についての意見

- 1) 既に改善に向けて取組みが実施されているが、認知度や使いやすさの向上に引き続き取り組む必要がある。

特に、本事業で開発した統合データベースを活用したからこそ得られるライフサイエンス分野の学術的な研究成果を具体的に示すことが、認知度向上に効果的、かつ重要である。その過程において、利用者との協業の強化や研究者ネットワークの更なる拡大が期待される。加えて、画面デザインや機能による使いやすさととどまらないユーザー体験(ユーザー・エクスペリエンス)の向上、研究者のデータ解析リテラシー向上の支援、専門領域の研究者以外のメディアや社会の認知度向上のための働きかけも、考慮してほしい。

また、データベース利用による研究成果を漏れなく把握することは困難であるが、データベース利用成果を効率的に把握する方法について、引き続き検討を進めてほしい。

- 2) 今後の事業推進においても国内外の技術動向や研究トレンドを積極的に収集し、急速な進展や変化に対応できるデータ基盤の整備が必要である。

多様なデータ種への対応を可能とする仕組みの他、大規模化するデータベースシステムの維持・更新への対応についても、検討してほしい。さらに、ビッグデータ解析が今後の科学の重要なポイントとなることを踏まえ、技術開発やデータ整備を充実させるために予算増等も求めていくべきではないか。

統合化推進プログラムについては、課題間連携が融合領域の理解につながる取組みでもあることから、長期的な支援と新規課題公募の双方をバランス良く進められるよう強化を期待する。

- 3) 国際的なオープンサイエンスへの貢献強化と共に、日本としての強みにつながるデータ基盤整備や利用への貢献の強化も必要である。

データベースには国際的な連携も必要であるとともに、日本固有のサービス提供のための取組みや、日本語によるサービス提供等の日本人による利用への貢献も、引き続き実施、強化していくことを期待する。

なお、評価に際し、本事業の成果を発展させデータの共有・活用を一層促進す

るための課題点として、データベース事業の重要性に即した事業規模の拡大と人材を含めた恒久性の確保、および、国として一元的にデータを集積し活用を推進する制度設計、の 2 点が挙げられた。事業の枠組みを超えることから本事業に対処を求めるものではないが、委員意見として付記する。

以上