

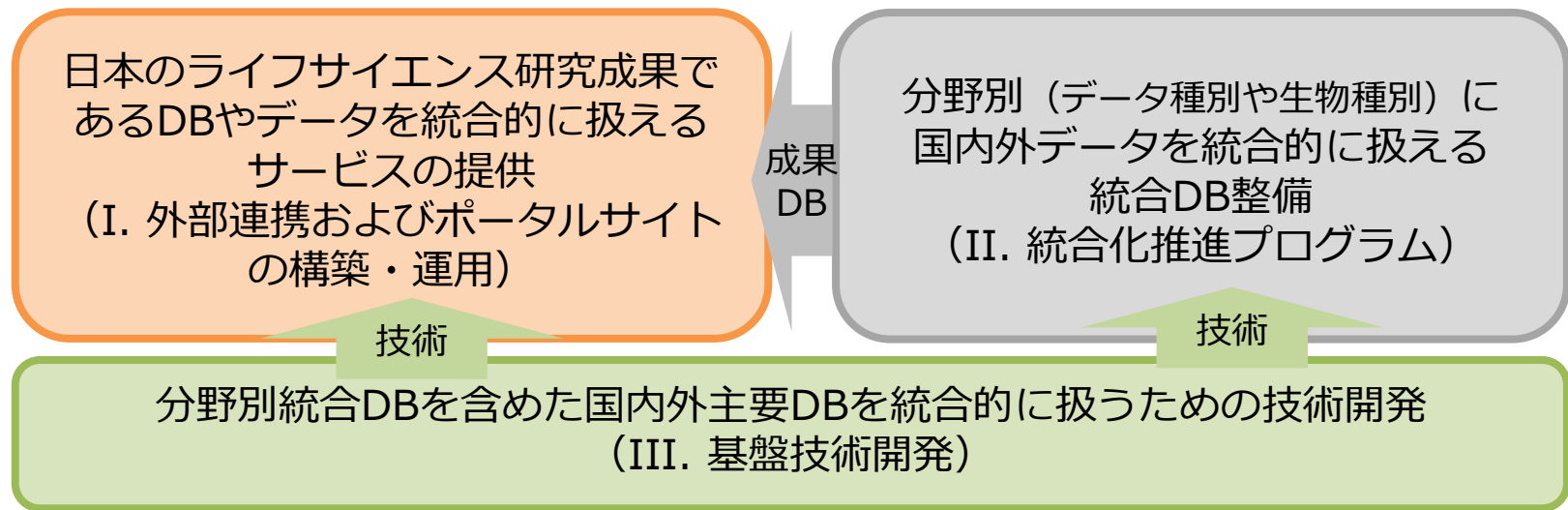
ライフサイエンスデータベース 統合推進事業 事業報告書（概要）

令和3年10月

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）

事業目的 および 事業評価経緯

国およびJSTのミッションを踏まえ、本事業では日本のライフサイエンス研究の成果であるデータ・データベース（DB）が広く研究者コミュニティに共有・活用されるよう、データベースの統合利用のための研究開発とサービス提供を実施してきた。具体的には、ライフサイエンス分野のデータ・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つの活動を組み合わせて事業を推進している。



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年度事業評価から約5年となる節目であることから、NBDCでは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。

事業成果要旨

事業の3つの活動を連携させることにより、統合データを充実させ、統合したデータをより活用しやすく提供するための新規サービス・技術開発を実施した。

3つの活動のいずれにおいても、

- **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。
- **提供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**、また**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。

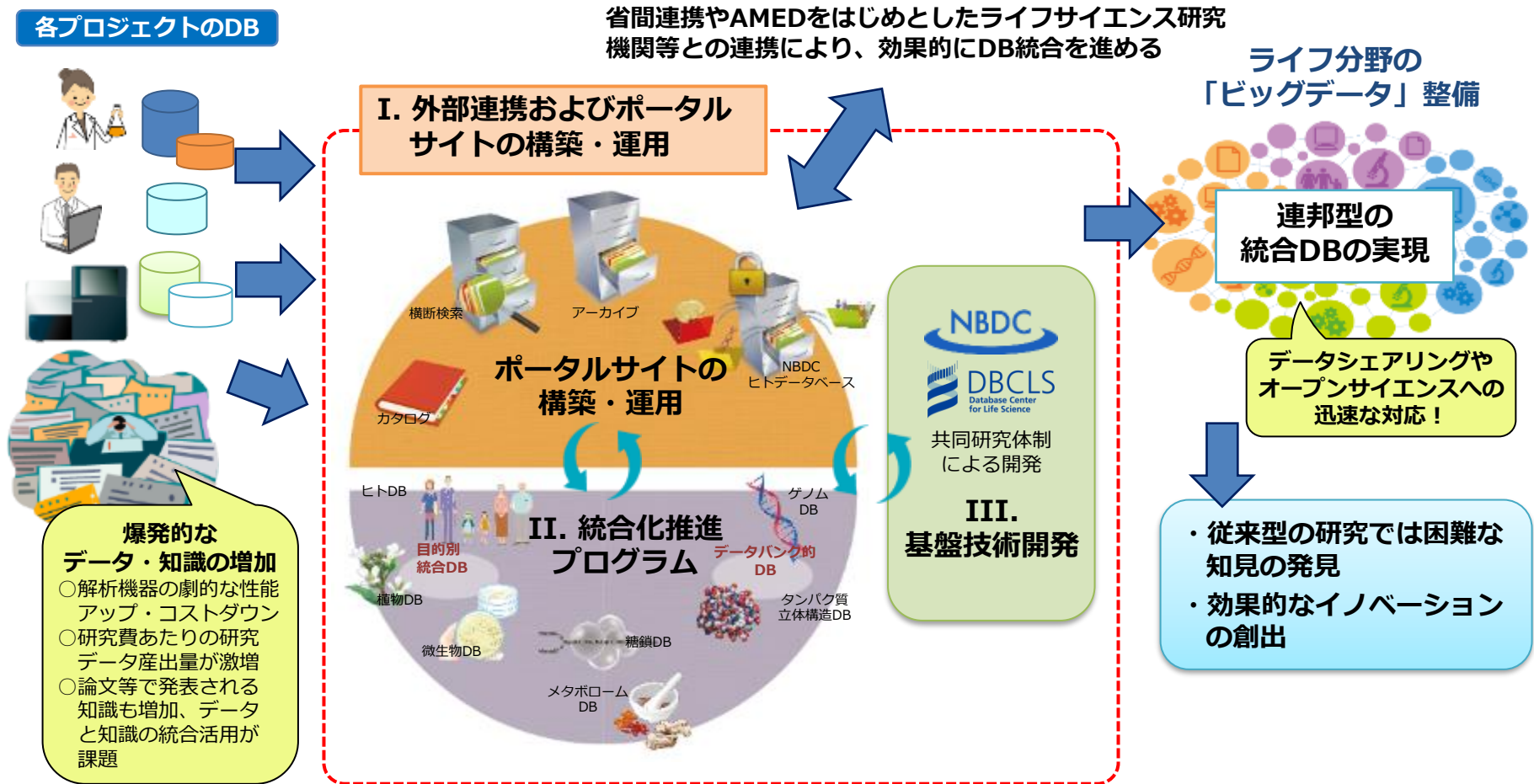
事業に共通する取組みとして、

- **RDF化による統合を進捗**させ、**統合データ基盤の拡大**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。
- 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。

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を着実に実施。今後に向けて、急速に進展する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、利用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。

事業概要

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・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する。

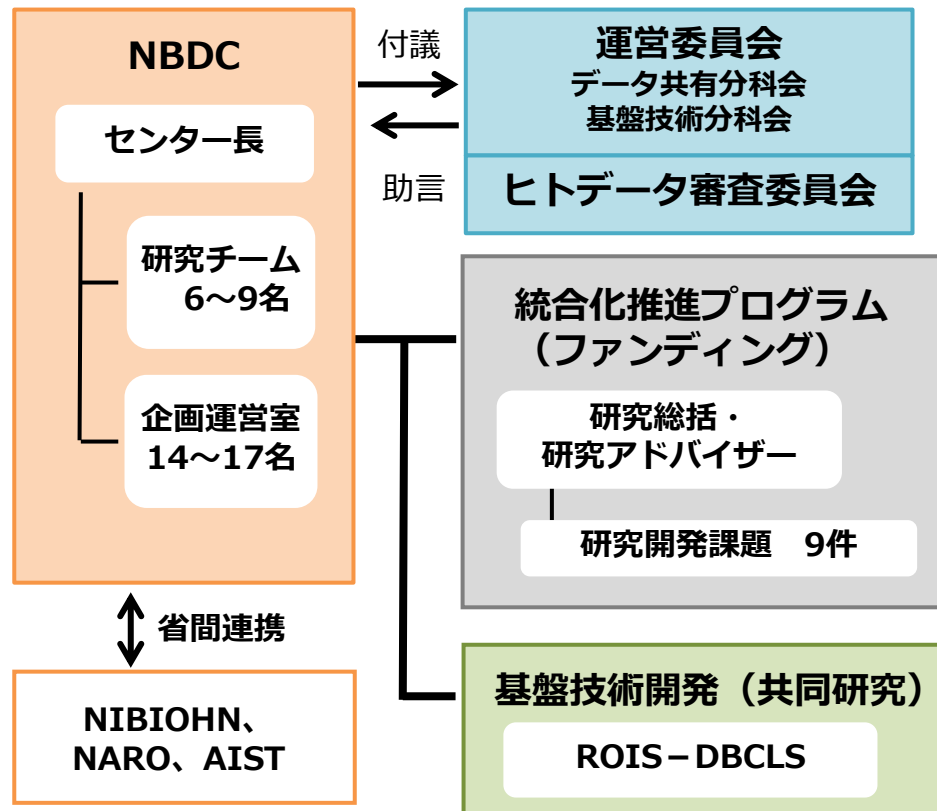


事業背景 および 体制

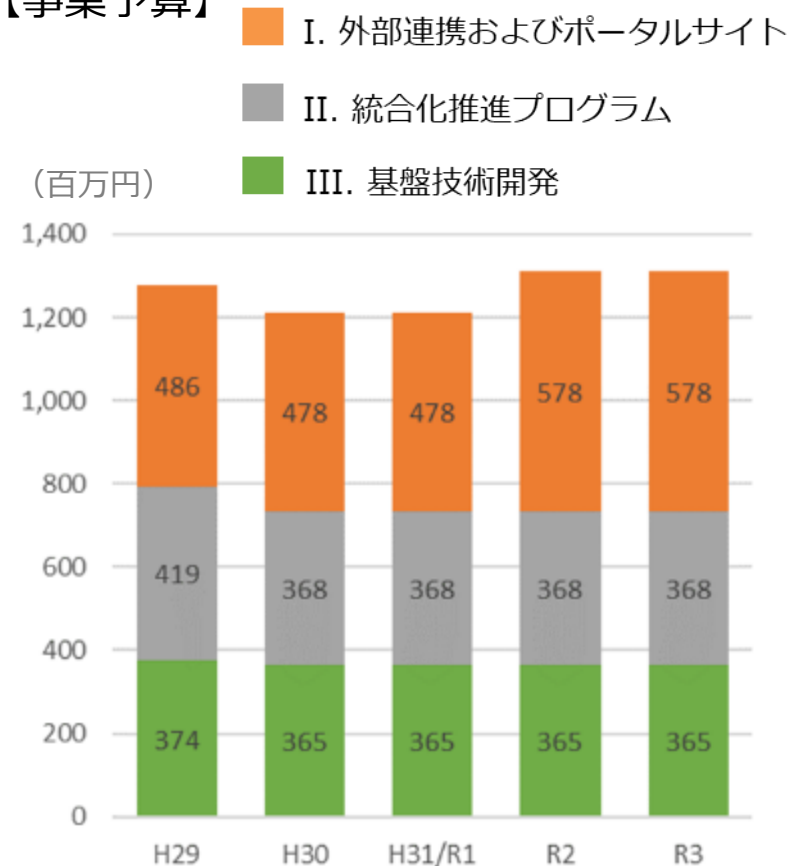
CSTPの「統合データベースタスクフォース報告書」を受け、JSTの前身事業と文部科学省の統合データベース事業※を統合する形で、我が国における統合DB整備の中核機関としてセンターが発足。

※中核機関：情報・システム研究機構（ROIS）ライフサイエンス統合データベースセンター（DBCLS）

【事業推進体制】



【事業予算】



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を利用者が統合的に扱えるようにするため、以下を推進

(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

(3) 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

- 公的資金により日々作成されるDBを広く利用可能とすることを目的として、ウェブサービスを提供。

◆ データ充実（統合の進捗）

国内DB情報を広く収集するため、厚生労働省・農林水産省・経済産業省傘下の研究機関との省間連携によるDB調査・データ充実を実施。

	サービス名	サービス概要 ※省間連携の取組み内容	収録DB数 (H29年度以降の増)	利用状況 (月平均ユニークIP) (H28年度比較)
<p>広く浅い統合 (国内公開DBをほぼ網羅)</p> <p>横断検索</p> <p>アーカイブ</p> <p>深く高度な統合 (国内外の主要DB)</p>	カタログ	DBの所在 (URL) や概要の情報を提供 ※研究成果DB調査・情報提供(NIBIOHN、AIST、NARO)	2,506 (+909)	6,514 (2.3倍)
	横断検索	カタログ収録の複数DBについて、収録情報の一括検索が可能 ※各省系検索サービスの相互連携(NIBIOHN、AIST)	727 (+115)	54,092 (5.5倍)
	アーカイブ	寄託DBのデータを保全、ダウンロードして利用可能 ※統一的なデータ加工 (表形式) (AIST)	152 (+ 23)	26,285 (2.6倍)
	RDFポータル	機械処理・高度な統合検索が可能な形式 (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)	27 (+ 10)	637 (4.0倍)

着実に統合を進捗させ、5年間の平均としてサービス利用も2倍以上に増加

- 運営委員会提言（H29）に対応する新たな取組みとして、利用者の知識発見に向けたサービスの強化・充実を実施。

◆ 国際連携によるデータ充実

利用者要望（国外DB情報充実）への対応として、英国Oxford大学との連携により、「カタログ」に国外DB情報を約620件追加（H31年3月）。

⇒ 収録増（約1.4倍）に加え、継続的に国外DB情報を入手・更新する仕組みを構築。



◆ 統合データを活用しやすく提供

「RDFポータル」等の統合データを広範なユーザに利用しやすく提供するため、統合データ活用インタフェース「TogoDX」の開発を実施（詳細はIII.基盤技術開発）。

⇒ 利用者がDBを渡り歩くことなく、且つ、DBによるデータIDの違いを意識することなくDBをまたいだデータの探索と入手が可能となるよう、利用者が複数DBを柔軟且つ簡便に組み合わせながらデータを抽出できる機能を実現。

データ・機能の両面から、利用者の知識発見に資するデータ活用基盤を向上

ヒトデータの統合と利活用推進

○ ヒトのゲノム等データについても、個人情報に配慮しつつ共有・利用を推進する「NBDCヒトDB」を、国立遺伝学研究所のDDBJと連携して提供。

◆ データ充実（統合の進捗）

利用者からのデータ提供（寄託）への対応を着実に実施し、後述の連携によるデータ充実も含め、5年間で利用可能なデータが約8倍に大幅拡大（延べ31万人分）。

◆ 利用環境整備

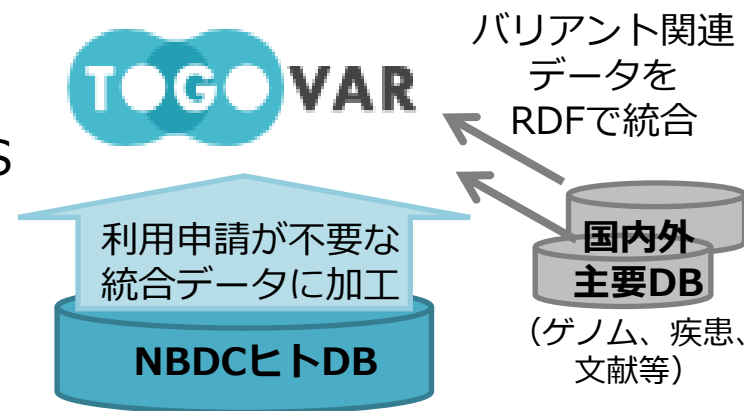
法令改正への対応、運用改善のためのガイドライン見直し、共用スパコンとの連携（機関外サーバ）の導入・拡大による解析環境の向上等も、積極的に実施。

○ 運営委員会提言に対応する新たな取組み・技術開発も実施。

◆ 日本人ゲノムを統合し利用しやすく提供

プロジェクト横断的なヒトゲノムデータ活用のための新規サービス「TogoVar」をDBCLSと共同で開発（H30年6月）。

⇒ 統合データに加工することで事前の利用申請なく利用可能とし、利用者の効率的なデータ利用に貢献。



従前のデータの受け皿に加え、利用促進のための新規の仕組み・技術開発も実施

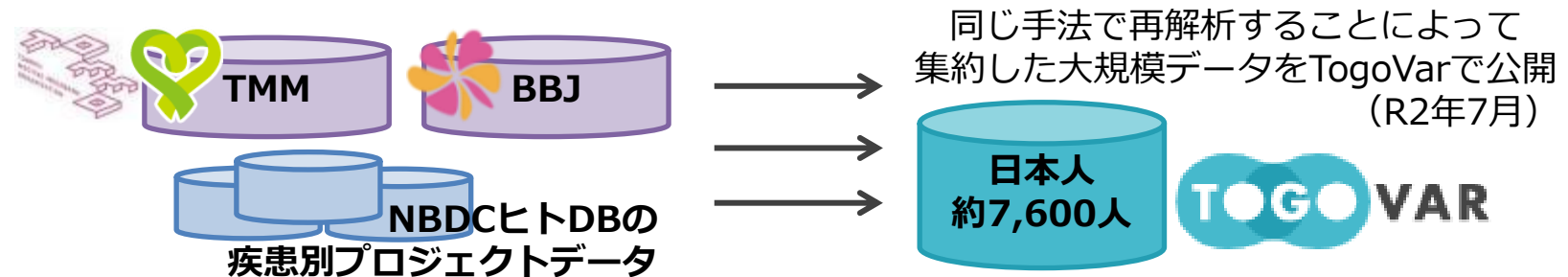
- TogoVar構築でNBDCが提供データを加工して提供する仕組み構築・技術開発を行ったことにより、外部プロジェクトとの連携が広がりデータ充実も加速。

◆ 連携によるデータ充実

AMEDの国際プロジェクト「GEM Japan」への協力として、東北メディカル・メガバンク（TMM）やバイオバンク・ジャパン（BBJ）等と連携。

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。

⇒ 日本人としてこれまでに無い規模（約7,600人）、且つ利便性の高い統合データの構築・共有に貢献。



⇒ 国内研究者による疾患研究への活用の他、ヒトゲノム等の国際的共有化を推進するGA4GH*から国際的に情報発信され、国際的にも存在感を示した。

(* Global Alliance for Genomics and Health : 54カ国、約700の産学官の機関が加盟)

統合によりデータ価値を最大化し、国際的に価値あるデータの構築・共有に貢献

- 統合の拡大を目的として、外部連携を実施。
既存連携の拡大に加え、運営委員会提言に対応する新規連携の構築にも注力した。

◆ AMEDゲノム制限共有データベース（AGD）への協力

円滑な公開（NBDCヒトDBへの移行等）に向けたプロジェクト内共有DB（AGD）について、データ提供を受け付ける等、本格稼働に協力。
AMEDが構築中のCANNDsの仕組み検討にも協力を実施している。

◆ SIP「スマートバイオ産業・農業基盤技術」への協力

外部協力機関として参画し、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（SHD）の仕組み構築・運用を支援。

◆ 連携による公開・共有の円滑化、データ価値の最大化等

上述の他、科研費「先進ゲノム支援」、AMED、NIMS、JST他事業等について、ポータルサイトの各サービスを活用しつつ、連携・支援に取り組んだ。

また、事業を超えた国レベルでのデータ共有に関して、COCNや日本学術会議の提言等への協力を実施した。



主にDB開発者との連携

外部プロジェクトも含めた連携



国レベルからプロジェクトまで、仕組みづくり・技術開発の実績を活かして貢献

次の点により、統合のための取組み・成果の双方について、ライフサイエンスの研究開発環境の整備・充実に対して寄与は十分であったと考えている。

- **用途に沿って統合したデータを提供するサービスを新規に開発。**
(ヒトゲノムバリエーション-TogoVar、RDF統合データ活用-TogoDX)
- 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。**
(TMMやBBJとの連携による日本人約7,600人のゲノムデータを統合 等)
- **ヒトデータ解析環境の向上等、利便性向上を積極的に実施。**
(機関外サーバによる共用スパコンとの連携 等)
ヒトDBやそのためのガイドライン (ノウハウ) を、
新サービス (TogoVar) 開発・外部プロジェクトとの連携につなげた。
- 省間連携による既存サービスの**着実な運用とデータ・機能充実。**
(カタログ、横断検索、アーカイブ、RDFポータル)
 - ・ 今後は、DXなどの社会状況等を踏まえ、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。
 - ・ 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公開・統合の成功事例・ノウハウを広げていくことも重要**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の、公募を通じた分野別の統合DBの開発

- ファンディングによる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、プログラム実施内容や推進方策を強化。

研究総括：長洲 毅志（元エーザイ株式会社 アドバイザー）（～R1年度）
伊藤 隆司（九州大学 教授）（R2年度～）

◆ 成果の最大化に向けた研究期間の延長と評価による計画見直し

- ・ 計画的に価値あるデータ整備・開発を実現するため、研究期間を5年に延長。中間評価により開発内容の拡充や見直しを実施。
（拡充例：データ整備の律速を解消、見直し例：利用者意見の反映強化）

◆ 利用者視点での研究開発を推進

- ・ 主要データ種ごとに利用者視点で助言できる産学の研究アドバイザーを選任。課題評価や年次の進捗確認において、開発の方向性に助言。
- ・ 各研究課題に対しアドバイザーリー委員会の設置や講習会等、利用者意見の積極的な収集と開発への反映を要請。

◆ 課題間連携による統合データの利便性向上・高付加価値化

- ・ 基盤技術開発との連携による成果DBのRDF化を引き続き推進し、各DBのデータ充実とDB間連携による統合データ基盤の拡大を企図。
- ・ 合同報告会やハッカソン等により、研究課題間のネットワーク作りを推進。

各研究課題における利用者視点の反映と課題間連携を主軸に推進

採択課題の概要

- ライフサイエンスの幅広いデータの統合化のため、オミクス等各種データや生物種を対象とする統合DB整備・開発課題を採択。

(H29年度採択7課題、H30年度採択2課題、R1～3年度は公募なし：計9課題)

期間	対象	DB名	概要・特徴
H29—R3年度	エピゲノミクス	ChIP-Atlas	<ul style="list-style-type: none"> 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、網羅的再解析データを整備・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。
	疾患パスウェイ	KEGG MEDICUS	<ul style="list-style-type: none"> 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エーションをパスウェイ上で俯瞰できるDB。 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として蓄積するDBは独自性が高い。
	糖鎖	GlyCosmos	<ul style="list-style-type: none"> 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。日本糖質学会の公式ポータルサイトとして承認されている。
	タンパク質立体構造	PDBj	<ul style="list-style-type: none"> 国際標準レポジトリ (wwPDB) を国外DBと共同で運営 (アジアオセアニア地区担当)、解析ツールや関連DBも提供。 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。

採択課題の概要

期間	対象	DB名	概要・特徴
H29-R3年度	マイクロバイオーム	Microbe DB.jp	<ul style="list-style-type: none"> 公開ゲノム・メタゲノムデータを再解析した上で統合し、菌株データ、系統分類情報とともに提供。 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も収録。
	多層オミクス	DBKERO	<ul style="list-style-type: none"> 疾患の各種オミクスを統合し、遺伝子領域、転写因子結合部位、バリエーション、エピゲノム情報等を並列表示できるDB。 多層オミクスに対応したデータビューワを開発し、日本人由来データを収録。
	植物ゲノム	Plant GARDEN	<ul style="list-style-type: none"> ゲノム情報をもとに遺伝子、転写産物、バリエーション、DNAマーカー等を閲覧・比較できるDB（R2年7月開設）。 植物の収録ゲノム数において世界最大規模。国内外公開データを解析して整備したバリエーションデータは本DB固有。
H30-R4年度	代謝物	MetaboBank	<ul style="list-style-type: none"> 国際コンソーシアム（MetabolomeXchange）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。
	プロテオーム	jPOST	<ul style="list-style-type: none"> プロテオーム情報を標準化・統合し、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DB。 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。再解析データの整備・提供も実施。

ヒト・動物以外にも植物や微生物を含め、各種オミクスを幅広く採択

- 各研究課題において、利用者意見を積極的に収集し開発に反映。国際的にも価値ある統合データを整備するため、国際連携・標準化にも取組み。

◆ 利用者視点による開発の例

- ・ 外部DBとの連携、利用者保有データの解析機能提供・手元解析用パッケージ化

◆ 国際連携・標準化による統合データ整備

- ・ 前期に引き続き、国際的なレポジトリ運用により国内外に貢献
(タンパク質立体構造DB、糖鎖DB、プロテオームDB)
- ⇒ 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
- ・ メタボロームについて、国際レポジトリ開設に向けDB開発

◆ 利活用状況の例

- ・ エピゲノミクスDB (ChIP-Atlas) : 疾患研究等多数の論文に引用
- ・ タンパク質立体構造DB (PDBj) : 国際連携で創薬を含む幅広い研究に貢献
- ・ 微生物DB (MicrobeDB.jp) : 複数の民間企業と共同研究で利活用
- ・ プロテオームDB (jPOST) : 1,000件超の国内外プロジェクトがデータ登録

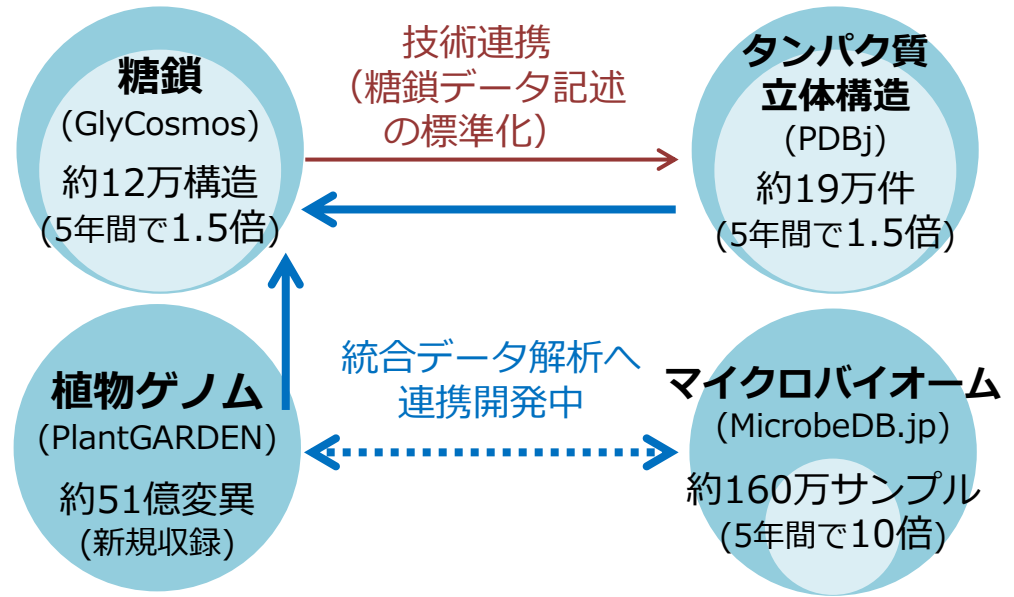
アクセス数も増加傾向にあり、産学の多様なライフサイエンス研究に貢献

課題間連携による統合化

- 課題間連携の推進により、各DBの収録データ増に加えて利用者が扱える統合データを拡大。データを効率的に扱うための技術連携も実施。

◆ 事業共通のデータ形式 (RDF) によるDB間連携の例

- ・ DBどうしの関連データの連携 (タンパク質構造—糖鎖、植物—糖鎖)
- ・ 利用者連携によるホロゲノム (植物—微生物ゲノム) 統合解析ツールを開発中



◆ 利用者のデータ解析の流れに沿ったデータ・DB機能の連携の例

- ・ DB検索結果を他DBでの検索につなげる (次ページ)



課題間のデータ・DB機能の連携でデータDBを高付加価値化

◆ 利用者のデータ解析の流れに沿ったデータ・DB機能の連携の例

Differential Expression Analysis

① jPOSTでデータセット間を比較し、増減したタンパク質のリストを表示

Enrichment Analysis

別画面が起動

② エンリッチメント解析で「ChIP-Atlas」を選択すると増減した遺伝子リストが自動的にChIP-Atlasに入力

ChIP-Atlas - Enrichment Analysis

③ ChIP-Atlasで更に解析

(jPOST→ChIP-Atlas)
機能開発に加え、受け入れ可能なデータIDを増やす等で連携を実現

次の点により、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、**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、課題間連携に主軸を置いた**研究開発推進マネジメントを実施**。
- **収録データ増、利用者視点による開発や国際連携による統合データ整備**が進捗。対象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、**産学の幅広いライフサイエンス研究に活用**され、今後の利活用の拡大が期待される。（例えば、エピゲノミクス、タンパク質立体構造、微生物、プロテオーム等）
- データ種を超えた利活用に向け、**課題間連携で統合DBを相乗的に高度化**。（微生物－植物：ホロゲノム解析、プロテオーム－エピゲノミクス：解析連携等）
 - ・ 今後は、計測技術の進展等による新たなデータ種の共有・利活用も含め、**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。
 - ・ 国の方針に沿いつつ、**国際的なオープンサイエンスへの貢献強化**の検討も必要。
 - ・ 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

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活用するための基盤的な技術開発

開発概要と重点的取組み

- オープンサイエンスに向けた統合データの利用促進のため、外部有識者意見を受けた開発重点の設定等、開発構成の見直しを実施。

具体的なデータ利用を想定した開発のため、**重点応用領域**を設定

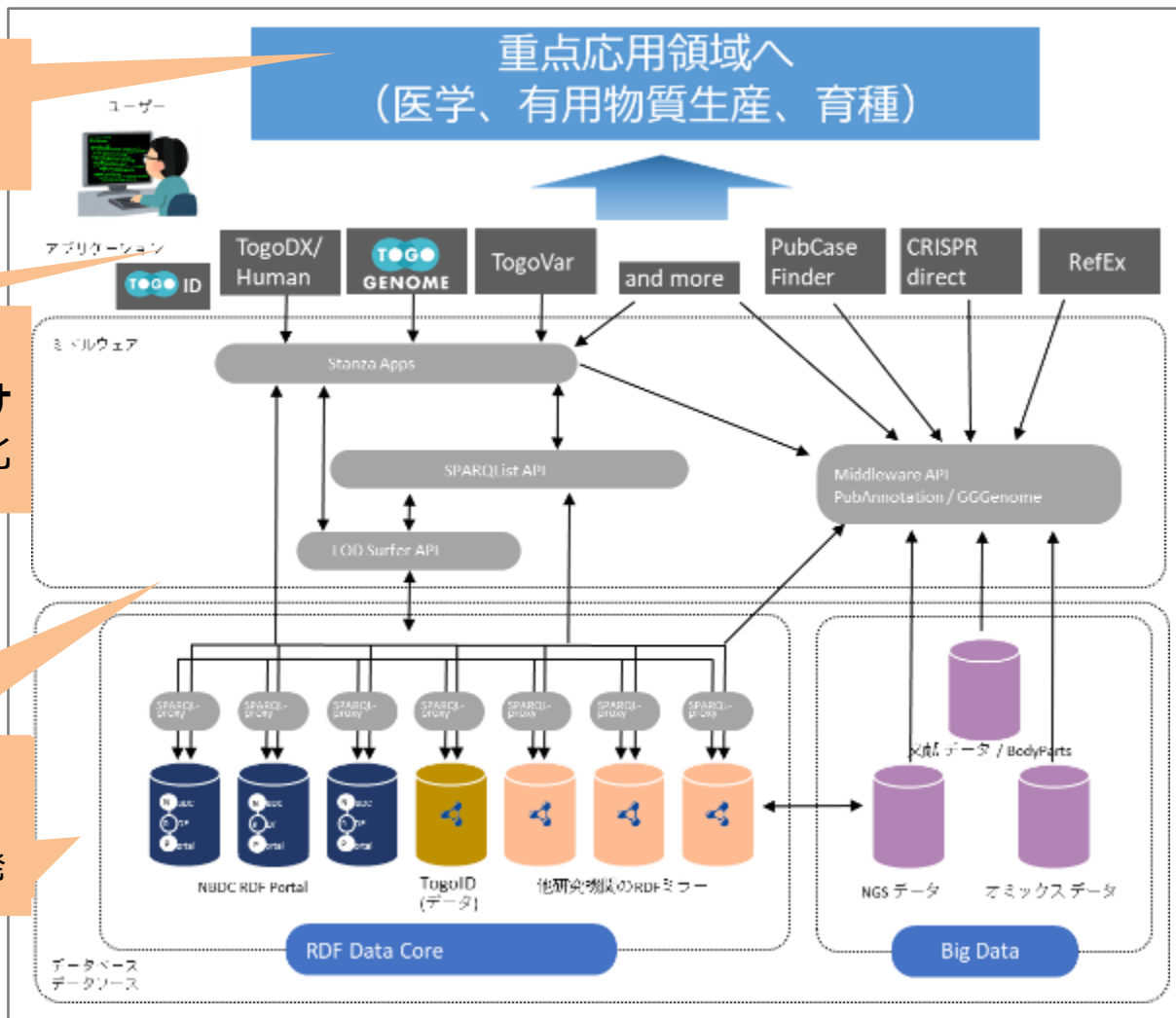
アプリケーション

統合データの利活用促進に向けて、**エンドユーザ・実験科学者に向けたアプリケーション開発**を重点化

データ基盤・ミドルウェア

アプリケーションを支える**統合データ基盤整備・関連技術（ミドルウェア）**開発

※外部の開発者も利用できるよう、データ基盤やミドルウェアもサービスとして提供



(図はDBCLS作成)

- 利用者が、データ利用目的に合わせて効率的にデータ基盤にアクセスできるように、アプリケーション開発を実施。

◆ 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

DBCLSとNBDC研究員の共同開発。RDF化によりバリエーションに関する国内外の様々なDBのデータを統合し、利用者が関連データを一括して検索・閲覧可能に。RDFの拡張性により、迅速な開発・関連データを統合して提供できる利点も。

⇒ 日本人としてこれまでに無い規模（約7,600人）、且つ利便性の高い統合データを提供する技術基盤に貢献。

◆ 希少疾患への統合データ利用に向けた「PubCaseFinder」の開発

希少疾患と症状との関係を素早く検索可能。100万件以上の症例報告から情報を抽出し、症状の類似度を計算。

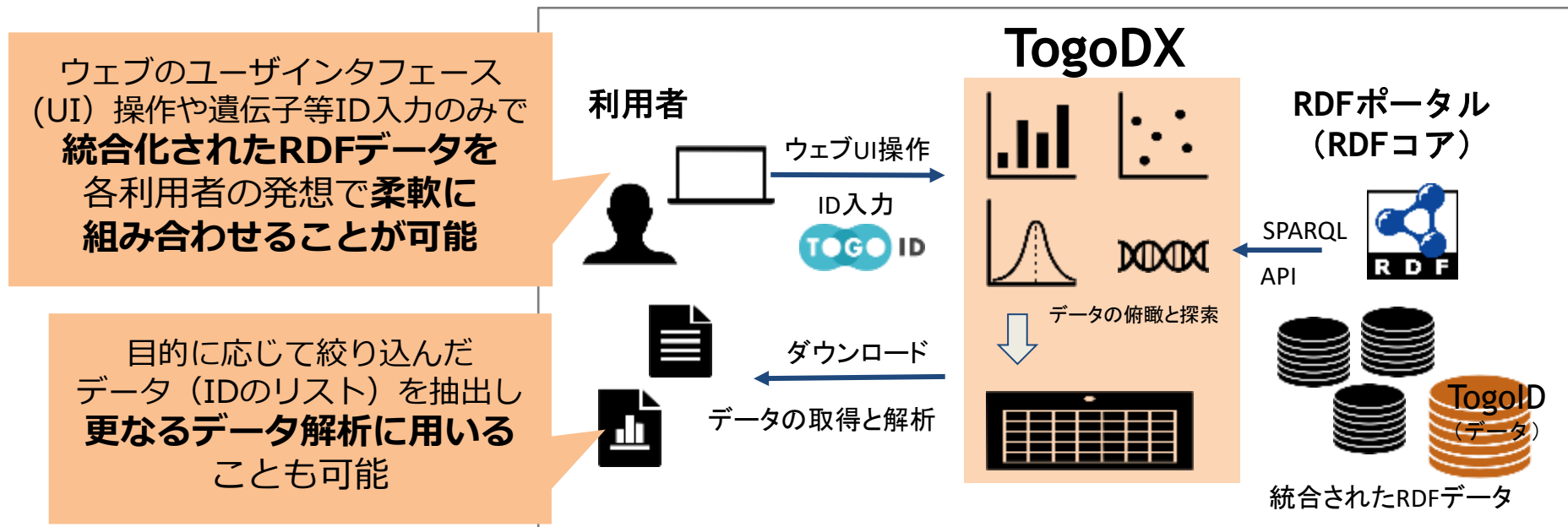
⇒ 利用者である医師・研究者と連携し、複数機関との共同研究により開発・活用。国際的な希少疾患検索システムでも活用。

上記の他、既存アプリケーションCRISPRdirect（ゲノム編集におけるガイドRNA設計システム）の核酸医薬品の安全性評価での活用、微生物資源活用のための統合データ整備・アプリケーション開発等が進捗。

利用者意見を反映しながら利用者とともに統合データの活用へ取り組み

- 本事業で大規模に統合してきた多種多様なデータを、利用者の発想で柔軟に組み合わせることが可能な「TogoDX」を新規開発（DBCLSとNBDC研究員の共同開発）。
- RDF化で統合した多様なDBを俯瞰し、目的に応じてデータの絞り込み・抽出が可能となるよう、新たなアプローチで開発
- IDつながりを利用することで、必要なデータを効率的、且つ簡単な操作で探索、取得することが可能。オープンサイエンスの価値である、複数データの組み合わせによる予期しなかった発見への貢献を目指す。
- 第1版（R3年10月公開）はヒト中心の分子・オミクス・疾患等のデータが対象だが、あらゆるデータへの応用が可能。

（図はDBCLS作成）






























アプリケーションを支えるデータ基盤

- 本事業の共通フォーマットとして、RDFによるDB間データ統合を継続。
 - 省間連携や統合化推進プログラムの成果DBを含め、国内外の主要DBを統合。
 - 後述のバイオハッカソンにより、国際連携によるRDF化も推進。
- ⇒ RDF化データを集積するRDFポータル収録DB数は、5年間で約1.6倍。収録DB数において世界有数であり、TogoDX等アプリケーションのデータ基盤。

RDFのメリット：

- 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)
- 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)
- 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)
- オントロジーの利用による語彙の共通化
- W3Cによる国際標準仕様でベンダーロックインが無い 等

- | | |
|---|--|
| <ul style="list-style-type: none"> • 塩基配列とアノテーション
INSDC (DDBJ/DBCLS)  • ゲノム情報
Ensembl (EBI) 
RefSeq (TogoGenome)  • アミノ酸配列とアノテーション
UniProt (SIB)  • タンパク質立体構造
PDB (PDBj) 
BMRB (PDBj) 
FAMSBASE (Chuo U)  • 化合物
PubChem (NCBI) 
ChEMBL (EBI) 
Nikkaji (JST)  • 遺伝子発現
RefEx, GTeX (DBCLS) 
ExpressionAtlas (EBI)  • サンプル
BioSamples (EBI/DDBJ) 
JCM (RIKEN)  | <ul style="list-style-type: none"> • 医学 (Med2RDF)
ICGC, COSMIC, CIViC 
DGIdb, OpenTG-Gates 
ClinVar, dbSNP, dbVar 
ExAC, gnomAD 
HiNT, INstruct  • 糖鎖
GlyTouCan, GlycoEpitope, WURCS, GGDonto, PAConto  • プロテオーム
jPOST 
The Human Protein Atlas  • パスウェイ
Reactome (EBI)  • その他
MeSH (NCBI) 
BioModels (EBI) 
MBGD (NIBB/DBCLS) 
Quanto (DBCLS)  |
|---|--|

既に多くの主要なDBがRDF化されている

(図はDBCLS作成)

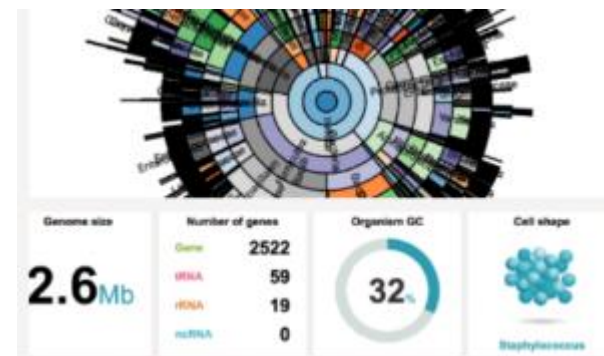
国内外ネットワークで拡大した統合データをアプリケーションへ提供

- 利用者へのアプリケーションを通じたデータ提供のため、アプリケーションとデータ基盤をつなぐ技術（ミドルウェア）開発も実施。

- **TogoStanza**

RDFデータを利用したウェブツールの開発の効率化・高速化を目的に開発。RDF化したDBの検索結果を可視化する再利用可能なモジュール（Stanza）を提供。

TogoVarや**統合化推進プログラム**の開発で利用。

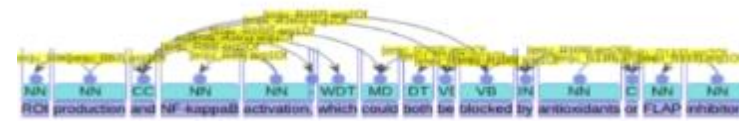


（TogoStanzaによる可視化例）

- **PubAnnotation**

文献から機械可読な情報を取り出すための注釈付（アノテーション）の世界最大規模のレポジトリ。

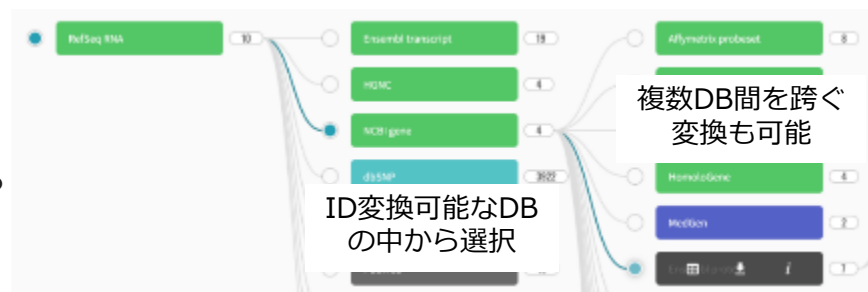
PubCaseFinderや**統合化推進プログラム**の開発において、文献からの知識抽出・DB化に貢献。



（文献アノテーションの例）

- **TogoID**

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継続し、**今期公開**。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。**TogoDX**で利用。



（TogoIDによるID変換の例）

アプリケーション・データの効率的な提供のため不可欠な技術開発も進展

- 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。

◆ バイオハッカソン（国際版・国内版）

- ・ 合宿形式でテーマ横断的に技術的課題に取り組み、開発者どうしの密度の高い情報交換と生産性の高い技術開発が可能。
- ・ DB統合化による国内外データ共有の国際連携コミュニティを形成し、技術的な課題解決のための合意形成・技術開発・実装を行う。
（前期の成果例として、国際的な研究データ共有基準「FAIR原則」への貢献）

⇒ バイオハッカソンでの開発例：

国内外DB開発者との共同によるDBのRDF化（欧州EBI、AMED事業参画者）、
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、
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

◆ 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。前期に策定した「RDF化ガイドライン」に沿ったRDF化を支援することで、他DBとの連携性が高く再利用性に優れるRDFデータ作成を支援。

データ共有のための国内外連携を形成、リーダーシップ・プレゼンスを発揮

次の点により、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**に重点を移した開発を実施。
(ヒトゲノムバリエーション-TogoVar、希少疾患-PubCaseFinder、RDF統合データ活用-TogoDX 等)
- **利用者の研究開発や国際研究ネットワークで活用**され、波及効果につながった。
(TogoVar・PubCaseFinderのGA4GHへの貢献、CRISPRdirect 等)
- 統合に向けた**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**。
(バイオハッカソン等によるDBのRDF化やオントロジー開発)
- **本事業で整備してきたデータ基盤の統合的活用インタフェースTogoDXの開発**。
 - ・ ヒトデータに関して開発を先行させたことから、今後の課題点として、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。
 - ・ 統合の国内外ネットワークが形成されてきたことを踏まえ、**統合・活用しやすいデータ・DBのための技術・ノウハウをより広く波及させていく**ことも課題。
 - ・ 利用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. 事業共通の取組み

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

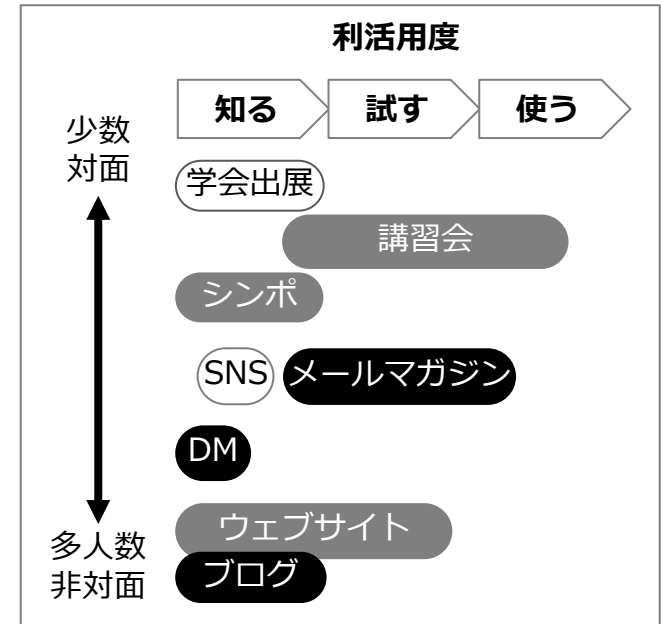
○既存取組みの改善と、既存取組みを補完する新規施策の実施により、提供サービスへの導線強化とプレゼンス向上を図った。

◆ 既存取組みの改善

- ・ **DB講習会**：5年間で23回開催し、延べ1,833名が参加（1回あたり平均で前期の約2倍）。NBDC/DBCLS企画回やオンライン開催の工夫。
- ・ **シンポジウム**：毎年1回開催し事業成果を発信、延べ1,412名（1回あたり平均で前期同様）。DB開発者と利用者の交流・議論を深める工夫等。

◆ 情報接点の追加

- ・ **ブログ**：認知獲得のため、事業取組みや解説記事を配信。ブログのうちCOVID-19関連研究データ紹介（R2年3月掲載）は、累計5万回以上閲覧。他に、シンポジウム抄録を題材にしたキーワード可視化法やDOI活用法も高閲覧。
- ・ **メールマガジン**：利用者との継続的な関係構築のため講習会情報・事業成果・ブログ掲載等を配信。



■:新規 ■:改善 □:継続実施

既存取組みの補完として、DM・メールマガジン・ブログを開始。既存取組みも改善し、全体として広報効果の増強を目指した。TogoVar公開時は新旧取組みをフル活用し、サービスの認知・利用を促した。

既存取組みの改善と新規情報接点により、前期を上回る研究者へ成果発信

事業の3つの活動のそれぞれや相互の連携により、対応を実施した。

- 未公開データを含めた大型プロジェクト支援：
2つのグループ共有DB支援（AMED、SIP「スマートバイオ産業・農業基盤技術」実施機関）、**円滑な公開・共有、連携によるデータ価値の最大化等について協力を実施。**
- 利用者視点によるデータ統合や利活用への取組み：
統合データ基盤を利用者の目的に沿って活用できるよう、開発・改善を実施。
（新規サービス開発や、統合化推進プログラムの課題間連携 等）
大規模データの利活用に向けた外部連携や技術開発も実施。
（TMM・BBJとの連携、TogoDX・TogoID・PubAnnotation 等）
- データ基盤整備の課題である継続性や発信力強化への取組み：
統合のための研究者ネットワークの拡大に向け、国内外連携、成果発信、ハッカソン等、**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。

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、本事業でも国際連携・人材支援の成果は得ているが、DX等社会状況への対応に向けて、**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。