

区 分	☒ 中核機関（ ☒ 代表機関／ ☐ 参画機関） ☐ 分担機関（ ☐ 代表機関／ ☐ 参画機関） ☐ 補完課題実施機関
課 題 名	「ライフサイエンス統合データベース開発運用」 （戦略立案・実行評価／統合データベース開発／統合データベース支援）
実 施 機 関 名	大学共同利用機関法人 情報・システム研究機構
代表研究者名	高木利久

1. 課題開始時における達成目標
我が国におけるライフサイエンス分野のデータベースの抱える問題点として、データベースが散在していて所在情報や利用法が分からない、データベースやデータベースのエントリーに信頼性の高い注釈がついていない、大型プロジェクトの成果公開が不十分で、またプロジェクトが終了すると更新、メンテナンスが終了してしまう、ばらばらに構築・管理されていて検索・解析・応用が困難といった事項が挙げられるが、根本的な問題としてライフサイエンスの発展やバイオ産業の成長にとって不可欠な基盤であるにもかかわらず、データベースに関わる戦略がない、との文部科学省ライフサイエンス委員会データベース整備戦略作業部会での検討結果に基づいて、本プロジェクトはスタートした。そうした経緯を踏まえ、中核機関としてのプロジェクト当初の課題としては、（１）我が国におけるライフサイエンスデータベースの整備戦略の立案と実行評価のための体制（ライフサイエンス、情報技術、ライフサイエンス DB の 3 分野の専門家からなる研究運営委員会と常勤の研究者チーム）を構築する。（２）上記戦略に基づき統合データベースとそのための情報技術を開発し公開する。（３）統合データベースの構築を支援し、かつ、有効活用するために、ポータルサイトの構築と運用、プロジェクトデータベースの受入れと運用、人材の育成を行う、の三点を掲げた。この中で、情報・システム研究機構としては、上記（１）に関わる戦略立案・実行評価、（２）に関わる共通基盤技術開発、ヒト統合 DB の開発・運用、モデル生物・産業応用微生物統合 DB 開発・運用、および（３）に関わるポータル整備・運用、広報、普及啓発、データベースの受入れと運用を担当した。具体的な目標設定は、平成 18 年度の基本検討を経て、平成 19 年度から本格的にスタートしたプロジェクトの初年度に行った。特に DB 統合化のスキームの検討とこれに基づくサービスの実現方法の検討を行い、代表的なサービスについては、数値的な達成目標を示した年次計画を策定した。 統合化のスキームについては、主に検索のタイプから、DB カタログ、横断検索及び統合検索の三つのケースについて検討した。DB カタログはライフサイエンス分野の DB のメタデータを集め、これに注釈づけを施した DB 情報に網羅的にアクセスできるものである。横断検索は、クローリングなどにより対象となる DB のインデックス情報を集め、これによりライフサイエンス分野の DB や文献の横断的なキーワード検索を実現するものである。また、統合検索は、外部にある DB から API 等により入手したデータ、あるいは統合データベースセンター内に受け入れた DB を処理して得たデータに対して検索を行うもので、横断検索より、より高度かつ詳細な情報を獲得できるものである。統合化の年次計画としては、DB カタログについては、平成 19 年度中に、内閣府連携施策群「ライフサイエンス分野のデータベースの統合化に関する調査研究」の調査結果に基づき、我が国に存在するとされるゲノム、ポストゲノム関係の 250 の DB のほとんどを網羅することにした。また、横断検索については、平成 19 年度中に 20 件程度の代表的な DB を掲載し、引き続き平成 20 年度中に我が国の関連 DB の大部分と海外の主要 DB を掲載することとした。さらに、統合検索については、国内 DB の受入れと併せて、順次対象

となる DB 数を増やしていくこととした。策定した年次計画を図 1 に示す。また、当面の想定ユーザは、ライフサイエンスの研究者および医療、創薬などバイオ産業従事者、ライフサイエンスプロジェクトの企画立案や評価にかかわる人々、およびライフサイエンスデータベースの構築者とした。

さらに、平成20年度に上記検討結果をベースに統合化の全体像とそれに至るステップを整理し、最終目標であるイノベーション、新たな知識発見、データベース生物学の構築に向け、1. データベース、ツール、プロジェクトのカatalog化、ポータルサイト構築、2. データベースやツールの使い方、使い分け情報の提供、3. データベースやツールの統一的、シームレスな検索、利用、4. 知識発見支援のためのデータベース統合化、解析ワークフローの基盤整備、5. 目的、用途ごとのデータベース統合化、解析ワークフローの構築の5段階を統合化へのステップとした。この中の4段階までを本プロジェクトで実現することとし、5番目のステップは、ライフサイエンス全体というよりはその中の特定の分野ごとに必要となるサービスであり、次のステップと考えた。

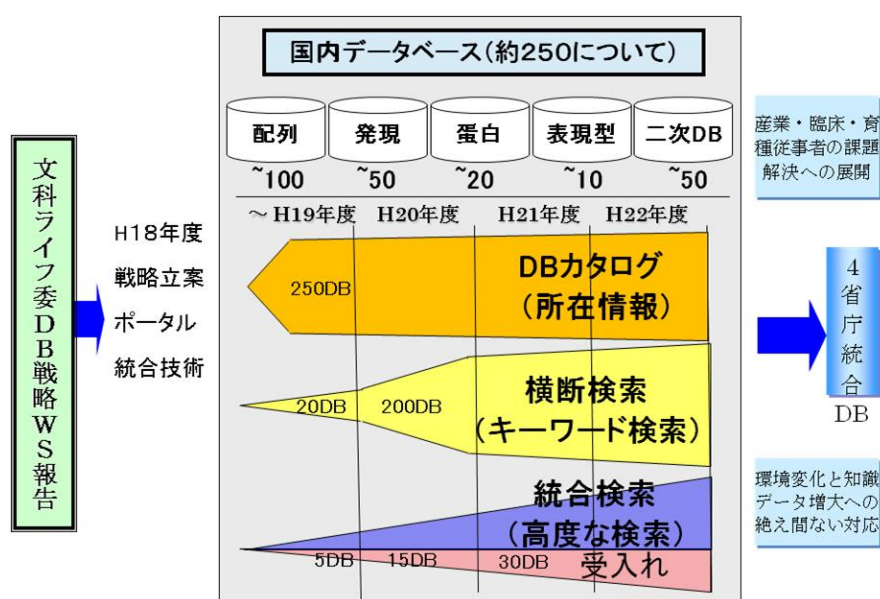


図 1 データベース統合化の年次計画

2. 平成22年10月末時点における事業計画に対する成果

(1) 成果概要

前項で述べたとおり、年次計画を設定し、数値的目標設定を行った本プロジェクトにおける主要開発ターゲットは、生命科学系データベースカタログ、生命科学系データベース横断検索、データベース受入れ、および統合検索である。これらに関する平成22年10月末時点の達成状況は以下のとおりである。なお、本プロジェクトの公募要領に記載された内容は多岐にわたったため、その他の成果の状況を含めた一覧は、(2) 進捗及び成果に示す。

1. 生命科学データベースカタログ

「データベースが散在していて所在情報や利用法が分からない」という問題点に応えるための、注釈やメタデータを付けたDBのカタログで、DBサイトのトップページをサムネイルで表示し、DBの稼働状況をモニターして表示する。継続的調査の結果、当初250余りと言われた国内ライフサイエンス関連DB数が増えたこと、また主要な国外DBも対象にしたことで、登録DB数としては目標を大幅に上回る912件を実現した。内訳は、国内DBが620件で、国外DBが292件である。

2. 生命科学系データベース横断検索

「ばらばらに構築・管理されていて検索・解析・応用が困難」という問題点に応えるもので、生命科

学分野の国内外のデータベースと文献を網羅的にキーワード検索する。国内外の主要 DB に加え、総説誌、プロジェクト報告書、学会要旨、特許などの日本語文献も同時に検索可能。検索対象 DB 数は、目標の 250 件を上回る 270 件を実現した。内訳は、国内 150 件、国外 120 件である。国内 DB については、クローリング可能でキーワード検索に適したものを選択した。

3. データベース受入れ

「プロジェクトが終了すると更新、メンテナンスが終了してしまう」という問題点に応えるもので、これまで、終了予定プロジェクトへのアンケートや個別交渉により、目標の 38 件の DB を受入れている。プロジェクト終了までに最終目標の 45 件の受入れを実現する。なお、受入れた DB は順次、各 DB に対して作成した DB そのものと格納データのメタデータ及び利用許諾と共に、データのダウンロードが可能なアーカイブサイトに登録している。

4. 統合検索

統合検索は、イノベーションや新たな知識発見を実現するための次のステップにつながる技術と位置付け、ここではプロトタイプングを行うこととした。一つは、横断検索の高度化に関するものであり、キーワード検索に加え、検索キーワードの翻訳結果や関連語表示機能、配列検索の実行機能を付加した。もう一つのアプローチは分野を絞って高度な検索を実現するものであり、40 余りの蛋白質関連 DB を対象に、蛋白質あるいは蛋白質ファミリーをキーワードに、各種データを DB 横断的に取得することを可能とした。

図 2 に各課題の達成状況とサービス画面を示した。画面からも分かるように、ユーザインターフェイスもカタログ系、検索系で統一感がもてるようにした。



図 2 主要サービスの達成状況

(2) 進捗及び成果

以下、提案書の項目に従い、得られた成果を示す。なお各成果の詳細については、別途添付した参考

資料ならびに別紙 1 の整備実績一覧を参照いただきたい。なお、本項ではサービスとして公開しているものを対象としたので、その他の成果については、「4. 他機関との連携」あるいは「8. 特記事項」の記載内容も参照いただきたい。なお、右端の利用状況（月間）の上段は一意な訪問者、下段はページビューで、数字は平成 22 年 1 月から 10 月の間のピーク値を示した。すべてのサービスが統合ホームページ <http://lifesciencedb.jp/> からアクセス可能である。

I. 戦略立案・実行評価

項 目	概 要	利用状況 (月間)
生命科学系 主要 PJ 一覧	ゲノム、ポストゲノム関連を中心に 90 年代以降の関連 4 省の生命科学系の主要なプロジェクトの詳細情報を収集；登録 PJ 数 69、H20 年 4 月開始	19 30
外部評価	登録評価者によるサービスへの評価内容公開するサイト。H19 年から前年度の公開サービスへの評価を実施。最終年度は全体評価予定；評価者数 500 人程度	225 8,675
生命科学 DB 政策	我が国の生命科学に関わる総合科学会議、文部科学省との審議会、委員会等で配布、公開された報告書、議事録等のアーカイブサイト；H21 年 2 月開始	101 373
運営委員会 等の資料	本プロジェクトの研究運営委員会、同作業部会で配布された資料ならびに議事録の公開サイト；H20 年 5 月開始	270 1,037

II. 共通基盤技術開発

項 目	概 要	利用状況 (月間)
OReFiL	オンライン上に存在する多数の生命科学系の資源(データベースやソフトウェアなど)を効率的に見つけるための検索システム；H19 年 8 月開始	449 10,239
Allie	MEDLINE を対象とし、出現する略字とその正規系のペアを検索するシステム。略字を入力することで、その使われ方を一覧表示する；H20 年 4 月開始	1,235 6,659
inMeXes	MEDLINE を対象とし、利用者が入力した文字列とマッチする表現を逐次的(1 文字入力毎に)に検索するシステム。論文中の英語表現を容易に検索；H21 年 7 月開始	751 17,722
DBCLS Galaxy	生命科学データに特化したウェブベースの対話的ツール組み合わせインタフェース。DBCLS で開発されたツール群も組み込んだ；H21 年 10 月開始	202 7,013
TogoDB	様々な形式の様々な種類のデータを簡単に読み込み、DB 化し、自動的に共通のウェブ検索インタフェースを生成するシステム；H20 年 4 月開始	1,348 114,779
TogoWS	国内各サービスの稼働状況を継続的に監視し、サービスの連携に必要なデータ形式変換機能等を提供するウェブサービスを統合するシステム；H20 年 4 月開始	516 399,873
OpenID 認証 システム	一つの ID で複数のサイトを認証できるシステム。各サイトで認証サービスを用意する必要がなく、サイト間のユーザ情報の集約 容易になる；H20 年 4 月開始	387 24,417
BioHackathon	最先端の研究開発者を招へいた国際的なソフトウェア開発会議。H19 年度からウェブサービス標準化、統合利用環境構築、セマンティックウェブで 3 度開催。DB 統合利用に関する内外研究者の国際的ハブとしての地位も確立しつつある。	1,969 22,623
TogoDoc	文献情報及び論文 PDF を管理し、また、特定の文献情報群に関連する論文情報を提示するシステム；H21 年 12 月開始	inMeXes の内数
Gendoo	遺伝子や疾患について、関連する疾患や薬剤、生命現象、臓器などの情報をキーワードで表示するシステム；	1,888 8,640

III. ヒト統合 DB の開発・運用、モデル生物・産業応用微生物統合 DB 開発・運用

項 目	概 要	利用状況 (月間)
DNA DB 総覧	国際 DNA バンクを高速に検索し生物種、分子種、プロジェクトに分類して表示。研究単位でのデータ取得を実現；検索対象 1 億 300 万レコード、H19 年 3 月開始	7,015 244,770
遺伝子発現 データ統	GEO に登録されている遺伝子発現データを測定技術とサンプルの属性に基づいて整理して表示；H19 年 3 月開始	1,564 12,629
togoexp (RefEx)	4 種類の異なる手法 (EST, GeneChip, iAFLP, CAGE) によるヒトおよびマウスの遺伝子発現データのリファレンスデータセットを構築；H20 年 3 月試験公開	159 1,717
BodyParts3D /アナトモ	解剖学用語をモデル人間中の 3 次元座標で定義した辞書とこれを部品として自由に人体の部分のモデル図を描画できるシステム；パーツ数 1315、H21 年 2 月開始	3,879 248,853
Wired-Marker	情報共有ツール、アノテーション支援ツール。オンラインで閲覧したページ、論文のテキストや図をブックマークすることが可能；H19 年 12 月開始	366,163 ダウンロード数
辞書の構築 と公開	日英の翻訳情報の各々の出典での記載状況を示したメタ用語集や、基本概念での整理を可能にする用語シソーラの作成と構造化の推進；H19 年 3 月開始	893 10,549

生物 アイコン	HPなどで自由に利用可能な生物のアイコン集。目的とするイラストをキーワード検索でき、生物に関する外部情報へのリンクも付加；アイコン数217、H20年6月開始	871 2,149
MiGAP	配列の大量生産を睨み微生物ゲノムの利用価値を高めるためにアノテーションを自動的に行うパイプラインを構築；解析総塩基数1G bases 余、H21年3月開始	196 46,700

IV. ポータル整備・運用、広報、普及啓発

項 目	概 要	利用状況 (月間)
統合 PJ ポータルサイト	統合 DB プロジェクト成果公開サイト。DBCLS 含めプロジェクト参加機関の公開サービスを網羅的にリンク；提供サービス 57、H20 年 6 月開始；右記は DBCLS 分	49,639 963,582
DB カタログ	DB サイトのトップページをサムネイルで表示。DB 稼働状況をモニターして表示。利用者からのコメントを受け付ける；登録 DB 数 912、H20 年 2 月開始	300 626
横断検索	生命科学分野の国内外のデータベースと文献を網羅的に検索。経産省統合 DB プロジェクトとの相互検索を実現；対象 DB 数 270、H20 年 3 月開始	13,632 39,787
日本語コンテンツ公開	総説誌である蛋白質核酸酵素のバックナンバーや文科省科研費ゲノム特定領域研究の年次報告書など良質な日本語文献の全文検索サービス；H20 年 6 月開始	4,904/4,609 計 32,373
新着論文レビュー	トップジャーナルに掲載された日本人を著者とする生命科学分野の論文について、著者自身による日本語レビューを公開；掲載論文数 39、H22 年 9 月開始	3,645 44,484
学協会 カタログ	医学、薬学、農学、理学など生命科学に関わる計 600 の学協会情報をカタログ化。学会ホームページのアドレスを始め、会員数、会費等を記載；H20 年 4 月開始	74 3,212
学会要旨 統合検索	学会要旨集の検索サービス。著者名や施設名の表記揺れを吸収する辞書を搭載し、研究者の研究履歴をたどることが可能；H20 年 4 月開始	88 132
統合 TV	データベースやウェブツールの使い方、講演、講習内容などを動画で発信するウェブサイト；作成コンテンツ 366、H19 年 7 月開始	4,326 77,700
講習会 (MotDB)	プロジェクト成果普及のための、端末を使った講習を全国で実施。H19 年度 2 回、H20 年度 6 回、H21 年度 9 回、H22 年度 5 回（9 月現在）開催；H20 年 1 月開始	1,380 20,669

V. データベースの受入れと運用

項 目	概 要	利用状況 (月間)
DB アーカイブ	各登録 DB に対してデータベースとデータのメタデータ、利用許諾を確認した上でアーカイブデータのダウンロードが可能；登録 DB37、H21 年 3 月開始	1,063 6,643
TogoProt	蛋白質関連 DB 統合検索システム。蛋白質に ID 経由で関連づけられる各種のデータを DB 横断的にまとめて表示する；対象 DB 数 41、H 2 年 6 月開始	134 856

3. 当初目標に対する達成度

（1）項で記述したように、DB カatalog、横断検索に関しては当初の数値目標を達成した。また、DB 受入れについても達成見込みである。また、統合検索については、横断検索の高機能化、蛋白質関連 DB に絞った統合検索システム構築など、そのプロトタイプ化が実現できたものとする。

また、本プロジェクトで開発したサービスの利用状況については、上記（2）項に示した表形式の右端に具体的数値を示した通りである。本プロジェクトの他の参画機関のサービスへの入口の役割も果たしている統合 PJ ポータルサイトへのアクセス数は、本機関の提供しているサービスに限っても、一月当たりの一意の訪問者、すなわちユニークユーザ数は 50,000 人近くに達しており、依然増加傾向である。また、ページビューも最大で 1,000,000 ページに達しようとしている。

個別にサービスを見ると、横断検索が一月当たりの一意の訪問者数がコンスタントに 10,000 人を超えており、いちばんよく利用されており、当初の課題設定が間違っていなかったことが分かる。また、蛋白質核酸酵素のバックナンバーやゲノム特定領域研究の年次報告書など、良質な日本語文献の全文検索サービスへのニーズが高いことが分かる。また、特筆すべきは始まって 2 カ月しかたっていない、トップジャーナルに掲載された論文の、日本人著者自らによる日本語レビューである新着論文レビューが早くも 3,000 人を超えるユーザを獲得していることである。これも当初よりとっていた日本語による情報

源を重視するという方向性が間違っていなかったこと示している。その他ユーザ数の多いサービスは、DNA データベース総覧や、BodyParts 3D/アナトモグラフィ、TogoTV というところで、これらもプロジェクトの初期からサービスを開始したものであり、蓄積があるということもあるが、基本的なユーザニーズの把握の方向性が正しかったものとする。その他、基盤技術に関わる検索系、ツール系のサービスも、絶対的ユーザ数は必ずしも多くないもの、ページビューの多さからみて有効に活用されているものとする。また、微生物ゲノム配列のアノテーションパイプラインである MiGAP も解析総塩基数が 10 億を超えるまでになっており、よく利用されるサービスとなっている。

4. 中間評価に対する対応

「より中核機関が主導的にプロジェクト全体を管理・運営できる体制となるよう、後半期は見直すべきである。」および「研究推進に当たっては、研究運営委員会をより効果的に活用することで、中核機関、分担機関、補完課題実施機関といった現行の組織構造のあり方についても十分に議論し、抜本的な体制の見直しに向けた検討を早急に進めるのが望ましい。」との提言を受け、研究運営委員会と作業部会の在り方を見直し、研究運営委員会ではより広い視点からの議論を行い、具体的なプロジェクトの進捗に関わる議論は作業部会で行うこととした。具体的には、プロジェクト参加機関の関係者は、研究運営委員会から作業部会に移っていただき、研究運営委員会には、産業界の代表者を充実させると同時に、産総研、理研など DB に関わる大きな組織の代表者に新たに加わっていただいた。これにより、プロジェクト終了後の統合 DB のあり方等に関する議論が深まったものとする。一方、作業部会においては、中核機関が主導的に管理・運営できる体制構築も含め、プロジェクト進捗に関わる実効的な議論をする場として、年 2 回程度作業部会分科会を開催することとした。これは、中核機関、分担機関、補完課題実施機関といった区分けではなく、関連するテーマごとに 3～4 のグループに分けて集中的に議論する場を設けたもので、関連する参加機関ごとの相互理解も含め、中核機関の意向を反映する場として、また「個別データベースの必要性について精査せよ」との提言に対する検討の場としても機能したものとする。

「プロジェクト終了後のあり方」については、文科省ライフサイエンス委員会 ライフサイエンス情報基盤整備作業部会作業部会（平成 20 年 10 月～平成 21 年 1 月）の報告書「ライフサイエンスデータベースの統合・維持・運用の在り方」や内閣府ライフサイエンス PT 統合 DB タスクフォース（平成 20 年 12 月～平成 21 年 4 月）の「統合データベースタスクフォース報告書」の提言も踏まえ、研究運営委員会では、内閣府に設置される予定であった統合データベース推進本部（仮称）への提言も念頭に、誰でも理解できるライフサイエンス統合データベースのアウトラインの議論を行い、9 項目の提言内容をまとめた。一方、本プロジェクトと科学技術振興機構（JST）のバイオインフォマティクス事業を一本化し、新たに JST 内にセンターを設置せよ、との提言を受け、中核機関でも新センターへの移管に伴う問題点の整理や費用の見積り等を行った。併せて、「省庁連携統合データベースを促進できる体制を担保すべきである。」との提言に対しては、JST と連携して平成 23 年度以降の 4 省連携に関わる方向性と具体的進め方に関する検討と提言を行った。

「本データベースの利用者を増やし、利便性を高めるために、広くモニターを募って定期的に報告を求める制度を始めてはどうか」については、取り組み内容の説明に加えてテスト検索が実行できるページや実施内容を表す模式図を見せるなどの工夫をした、サービス内容の紹介も兼ねたウェブサイト記入方式のユーザ評価のシステムを構築し、平成 20 年度以降、アカデミア、産業界の有識者を対象に評価者を募り、前年度開発したサービス内容を対象にユーザ評価を行うこととした。平成 20 年度は中核機

関の提供サービスを対象とし、平成 21 年度以降はプロジェクト全体の成果についてのユーザ評価を実施している。平成 22 年度分は 470 名余の評価者に依頼し、現在評価作業中である。

「計算機資源の不足問題」については、東京大学グループの分を中核機関で融通するとともに、中核機関分は、一部遺伝研の資源を融通してもらうこととした。

「市民向けコンテンツの充実も必要」については、プロジェクト紹介のリーフレット作成や BioJapan や学会での展示など専門家向けの広報の充実を図ってきたものの、提言を受けサイエンスアゴラでの展示なども行った。

「商業的なベースの活用の検討」、「民間資金の導入」といった提言については、当面は、国家プロジェクトなど国の予算を使って生産したデータを、共有財としてアカデミアおよび産業界で自由に活用できる環境構築を最優先事項とした。受益者負担的な考え方は将来課題とさせていただいた。

5. 他機関との連携

プロジェクト全体の方向づけ、進捗のチェックについては、平成 18 年度より、ライフサイエンス、情報技術、ライフサイエンス DB の 3 分野の専門家からなる研究運営委員会、および具体的な情報収集や分析、戦略立案支援を行う研究運営委員会の実働部隊としての役割を担った研究運営委員会作業部会を組織した。平成 18 年度は、平成 19 年度以降も睨んだ統合データベース整備の考え方を中心に、上記二つの委員会の合同開催も含め、計 5 回の委員会を開催した。平成 19 年度からは、研究運営委員会に、分担機関に全体戦略を提示して連携する場としての役割も持たせるため、分担機関の研究代表者にも参加いただく形の構成とした。ただし、分担機関あるいは年度途中からの採択になった補完課題実施機関に実施いただく具体的な課題の推進については、後述のとおり、中核機関・分担機関研究連絡会議や個別の打ち合わせにより実施した。平成 19 年度は、研究運営委員会を 3 回、同作業部会を 4 回実施し、平成 19 年度の計画、補完課題の公募内容、実施状況と進め方、契約・著作権問題、個人情報問題、海外連携、3 年後の在り方などを討議した。平成 20 年度以降は、前項に記載したように、中間評価での指摘に対応すべく、研究運営委員会と作業部会構成を見直し、研究運営委員会では我が国の統合データベース整備に関わる大局的な議論を行い、作業部会では、特に作業部会分科会という場を中心に、外部の専門家と併せ、関連課題ごとのプロジェクト参加機関の代表者が集まって、進捗状況、進め方に関する議論を集中的に行う場とした。平成 20 年度は、研究運営委員会を 3 回、同作業部会を 1 回、作業部会分科会を延べ 4 回実施し、プロジェクトの進捗状況、中間評価への対応、プロジェクト終了後を見据えた統合データベースのあるべき姿、体制等に関して討議した。平成 21 年度は、本庶佑総合科学技術会議議員、井村裕夫京都大学名誉教授、平田正協和発酵キリン名誉相談役などの外部有識者を交えて開いた研究運営委員会拡大懇談会を含めて計 4 回の研究運営委員会と伸べ 7 回の作業部会分科会を開催し、プロジェクトの進捗状況と併せて、プロジェクト終了後を見据えた統合データベースのあるべき姿、体制等について、引き続き議論を行った。平成 22 年度は、現在まで 2 回の研究運営委員会と延べ 7 回の作業部会分科会を開催している。

分担機関、補完課題実施機関あるいは中核機関の参画機関との連携については、特に実質的に初年度であった平成 19 年度については、分担機関との間では全体会議を 3 回、個別会議を各 1 回開催し、実施内容と連携の進め方、実施状況のフォローなどを中心に討議した。補完課題実施機関との間では、計 12 回の打合せを実施し、特に蛋白立体構造ならびに植物関係の進め方に関する議論は、それぞれ中村春木教授、中村保一室長を中心に中核機関を含めたグループを構成して進めた。平成 20 年度は、分担課題、補完課題実施機関との個別の打合せを 16 回開催すると同時に、上述の通り、関連テーマごとに中

核機関の参加機関も含めて集まって議論する場である作業部会分科会を定期的を開催することにより、情報共有、相互の連携を図る場とした。また、個別機関の進捗については、書類ベースの管理も導入し、計画設定も含め年 4 回進捗状況を報告いただくと同時に、全体で情報を共有することとした。

平成 20 年度以降は、他省との連携も積極的に進めた。特に経済産業省の統合 DB プロジェクトとの間では、一月から二月に 1 回程度の定期的会合を持ち、メタデータや横断検索、アーカイブの共有化について討議した。その結果、メタデータ共有化案の作成と横断検索の仕様統一化を実現した。また、厚生労働省との間では、平成 20 年以来、医薬基盤研のトキシコゲノムプロジェクトの関係者との間で数回の打合せを持ち、同プロジェクトのデータベース公開時に本プロジェクトにもデータを提供いただく旨の合意ができています。一方、農林水産省統合 DB プロジェクトとの間でも、今後の連携のための会合を開催し、カタログや横断検索、アーカイブなどの連携のあり方や問題点について討議を行い、引き続き協議を続けることになっている。

また、本プロジェクトでは、医療に関わるデータもその枠内に入っていたため、当初より個人情報保護に関わる調査、検討を進めてきた。その結果をもとに本プロジェクトの分担機関である東京大学グループ（疾患解析データベースグループ）と連携して、平成 20 年度には、ゲノムワイドの多型情報および疾患関連遺伝子のリシーケンスデータの共有方針の検討を行い、共有方針文書およびデータアクセス同意書の起案やデータの漏えい等の危険に対するセキュリティ手順書の策定などを行った。併せて、検討委員会設置についても連携して行った。平成 21 年度にはこれらの文書について検討委員会での指摘等を反映させるとともに、登録・利用申請に関する書面の作成を担当した。一連の文書や書面は国外からの申請も考慮して日本語とともに英語版も作成し、「疾患解析から医療応用を実現する DB 開発」ウェブサイトから平成 21 年 12 月に公開した。

6. 今後の見通し、計画、展望

本プロジェクトでは、「1. 課題開始時における達成目標」の中で述べた、統合化への 5 段階の内、目標とした 4 段階については概ね実現したものと考える。ステップ 1 については、定量的な目標として掲げた DB カタログに加え、プロジェクト情報等のカタログを整備し、併せてプロジェクト成果全体への入り口となるポータルサイトを構築した。ステップ 2 については、特徴ある動画教材である TogoTV のコンテンツを 400 近く作成し、併せて年間 10 回近い講習会の実施等により、その使い方の普及を図った。ステップ 3 については、横断検索を軸に、検索対象コンテンツの充実を図ることにより、世の中のニーズに応えた。ステップ 4 については、当初より提供してきたデータベース便覧に加え、蛋白質分野の統合検索システムである TogoProt の構築、また DBCLS Galaxy によるワークフロー環境の構築を行った。今回開発した中の中心的サービスである、DB カタログ、横断検索、DB アーカイブについては、関連 4 省との連携の核として、JST ライフサイエンス DB 統合推進事業の中心的サービスになるものと期待している。

残る課題としては、ステップ 5 の目的、用途ごとのデータベース統合化、解析ワークフローの構築ということになるが、ライフサイエンスは生き物であり、新たなカテゴリーのデータが出てくると同時に、各分野の進展に伴い、統合的にみるべきデータの範囲、つながりもますます拡大している。それと同時にライフサイエンスを取り巻く社会的環境も変化を続けている。従って、中長期的には、個人ゲノム時代への対応、次世代 DNA シーケンサへの対応、オープンアクセス化への対応、医療、健康情報への対応、医薬品、化合物への対応、分子イメージング、画像、動画への対応、環境、生物多様性情報への対応、DB の視点からの国家プロジェクト管理、世界的な連携、役割分担への対応、そのためのさまざま

なガイドライン、法整備といった項目も、統合データベースとして検討、考慮する必要のある課題と考える。

また、ステップ 1 から 4 までの課題が一通り達成できたにしても、上記のとおりライフサイエンスで扱うべきデータの種類、量は拡大しつつあり、また、データベース構築と利用に関わる情報技術も日々進歩を続けている現状を考えると、統合データベースのあり方を常に分野ないし背景技術の進歩の中でアップデートさせ続けないと、すぐに時代遅れのものになってしまう。従って、確立したサービスを続けながら、そのバックグラウンドで、次の展開をにらんだ技術開発を継続することが重要と考える。

7. 全体総括

以下の 3 項目は、本プロジェクトの研究運営委員会の委員の方々の、統合データベース整備の在り方に関する議論から得られた統合化の理念をまとめたものである。

- ・ 公的資金を用いて得られたデータ、DB は広く公開し、その活用を図るべし
- ・ DB 統合により、生命科学の新たな展開とバイオ産業の国際的競争力向上に貢献すべし
- ・ 統合 DB は重要な基盤であり、省庁に囚われない存在であり、国が継続的支援すべき

第 1 項は、データベース統合が、単に存在するデータベースを集めてきて一体のものとして利用できるようにすればよいという、技術的な問題のみではなく、種々の公的プロジェクトで生み出されるデータをできるだけ制限なく、利用者が共有できる環境を構築するという制度的な側面を含んでいる。この点に関しては、次の特記事項で詳しく述べるが、文部科学省との密接な連携のもとで、平成 20 年度以降のライフサイエンス課管轄の委託事業の公募要領には、統合データベースへのデータおよび DB 提供条項を盛り込むことができた。また、受入れ DB のアーカイブサイト用にクリエイティブ・コモンズに基づく利用許諾テンプレートを作成するなど、DB 利用者が権利関係を心配することなく、安心して DB にアクセスし、また必要に応じて二次利用できる環境を構築した。

第 2 項については、生命科学の新たな展開とバイオ産業の国際的競争力向上のために、イノベーション、新たな知識発見、データベース生物学の構築を目標として、統合化に至る上記のステップ 4 までの機能を実現するために、本プロジェクトでは種々のサービス、技術を開発してきた。その具体的内容、有効性については、これまでの各項で記載した通りである。ほとんどなんの基盤もない状況から、種々のカタログの整備、横断検索を中心としたコンテンツを含めた検索システムの充実、これまで前例のなかった DB 受入れとそのアーカイブ化の実現、分散的な統合化も含めたシームレスな DB 統合を実現するための基盤技術の開発、DB やツールの使い方を動画で示すコンテンツの作成など、統合データベースを形作るための要素のシステム、技術の蓄積は相当程度進んだものとする。前項で述べたとおり、データベースは生き物であり、そのコンテンツも周辺技術も常に変化、発展している。従って、本プロジェクトで開発した技術を継続的に発展させることが、我が国における真の統合データベースを構築、維持するうえで重要と考える。

第 3 項については、文部科学省ライフサイエンス委員会の情報基盤整備作業部会、さらには総合科学技術会議ライフサイエンス PT の統合 DB タスクフォースの場において行われた検討結果に基づき、本プロジェクト終了後も、本プロジェクトの成果である統合データベースを恒久的に維持、発展させる必要があること、また、そのための具体策として、本プロジェクトと科学技術振興機構（JST）の BIRD 事業を一本化するための組織を JST 内に構築すべきである、との提言がなされ、本プロジェクト終了時に、JST 内にバイオサイエンスデータベースセンターが設置されることになった。最終的には 3 年後に見直しが行われるものの、本プロジェクトの成果が認められ、恒久化に向けて一歩進んだことは意

味のあることと考える。

8. 特記事項

本プロジェクトのきっかけの一つは、我が国のライフサイエンス関係の予算が年間 3000 億円以上にものぼり、また、多くのゲノム、ポストゲノム関係の大型プロジェクトが行われてきたにもかかわらず、その成果公開が不十分である、との産業界からの声であった。これに対しては、散在する DB の所在情報や使い方を明らかにする、あるいは更新や維持の止まった DB をアーカイブするといった技術的対応だけでは十分でなく、積極的に公的プロジェクトからのデータないし DB の提供を担保する仕組みが必要との結論になった。当初、国家プロジェクトの契約内容及び日本版バイドール法などその基礎となる法律に関して調査を行い、プロジェクトの契約書にデータないし DB の提供条項を盛り込むなどいくつかの方策を検討した。そうした中で、当面の現実的な対応として、文部科学省ライフサイエンス課管轄の委託プロジェクトの公募要領に、データないし DB の公開への努力、統合データベースへのデータないし DB 提供、非独占的な複製、改変権の担保、利用のために必要な情報の提供義務などを盛り込むことになり、平成 20 年度以降の委託業務に上記の条項が盛り込まれることとなった。平成 21 年度終了時において、脳科学研究戦略推進プログラムの平成 20 年度、21 年度分、革新的細胞解析研究プログラム、橋渡し研究支援プログラム、およびターゲットタンパク研究プログラム 2 機関分とナショナルバイオリソースプロジェクト 4 機関分に関わる平成 21 年度分が対象となっている。前例のない試みであることもあり、必ずしもスムーズに進んでいるわけではないが、データ提供をいただく手順の整理も含め、方向性が見えてきている。現在、ナショナルバイオリソースプログラムに関わる 2 機関からの入手データを公開準備中であるほか、その他数機関と交渉中である。また、開示いただいた成果報告書から提供いただけるデータないし DB の有無に関する評価も行っている。

また一方、いくら良質な DB でもその利用条件が明確でないと安心して DB を利用できない、との声も産業界から聞こえてきている。そこで、本プロジェクトでは、クリエイティブ・コモンズのライセンススキームをベースに利用許諾を行うこととした。クリエイティブ・コモンズとは、クリエイティブ・コモンズ・ライセンス（CC ライセンス）を提供している国際的非営利組織とそのプロジェクトの総称です。CC ライセンスはインターネット時代のための新しい著作権ルール of the 普及を目指す活動で、著作者自らが「この条件を守れば自らの著作物を自由に使って良いですよ」という意思表示をするためのツールである。本プロジェクトでは、クリエイティブ・コモンズ・ジャパンの提供する クリエイティブ・コモンズ・パブリック・ライセンス日本版の「表示・継承」を標準利用許諾として採用することとした。表示・継承とは、元の DB の構築者、著作者のクレジットを表示し、元の DB と同じ条件で公開することを二次利用者に課すものである。本件については、著作権に詳しい弁護士およびクリエイティブ・コモンズ・ジャパン関係者と相談しながら進めた。

9. 委託研究費一覧（カッコ内は JST 予算、ただし外数）

	18年度	19年度	20年度	21年度	22年度	計
設備備品費（千円）	870	240,714	31,590	60,000	(25,000)	333,174 (25,000)
人件費（千円）	10,631	142,878	143,681	158,104	(161,923)	455,294 (161,923)
業務実施費（千円）	162,412	305,499	374,374	174,659 (220,000)	(365,985)	1,016,944 (585,985)
一般管理費（千円）	26,087	68,909	54,964	39,276	(55,291)	189,236 (55,291)
合計（千円）	200,000	758,000	604,609	432,039 (220,000)	(608,199)	1,994,648 (828,199)

整備実績一覧

(1) データ（又はDB）の連結、統合化整備

通 番	データ（又はDB）の名称	公開／ 未公開	概要（データの種類（生物種）・数量（kB等）、本プロジェクトで実施した特徴点、進捗状況、今後の計画・課題などを簡潔にわかりやすく記述）
1	生命科学データベースカタログ http://lifesciencedb.jp/lbdb.cgi?pg=1 (英語版) http://lifesciencedb.jp/lbdb.cgi?gg=dbcatalog&lng=en	公開	生命科学分野のデータベースカタログ（注釈及びメタデータ）で、データベースサイトのトップページをサムネイルで表示。データベースの稼働状況をモニターして表示。利用者からのコメントを受け付ける。掲載データベース数を増やすことが課題で、海外DB数の増加と記載内容の充実化を図ることが必要。その他の課題はデータベースのIDの統一化など、他のデータベースに関する情報との対応を持たせること。平成22年度はDB情報の追加、更新、厚労省、農水省関係のDBの調査、補強を行った。 登録DB数：912件（平成22年10月現在）開始年月：平成20年2月。平成22年10月の訪問者数：295/訪問回数：381/ページビュー：500。
2	生命科学データベース横断検索 http://lifesciencedb.jp/dbsearch/ (英語版) http://lifesciencedb.jp/dbsearch/en/	公開	生命科学分野のデータベースや文献の横断検索サービスで、生命科学分野の国内外のデータベースと文献を網羅的に検索。平成20年3月にサービスを開始、平成20年度は対象データベースの拡大を図り、平成21年度はインデックスの更新管理並びに、関連キーワード表示機能やBlast検索機能の開発を行った。また経産省統合DBプロジェクトとの検索連携を実現した。H22年度は対象データベースの充実を図るとともに、検索キーワードの翻訳結果や関連誤表示機能の実装やデータ間のリンク情報の格納を進めている。平成22年9月に英語版を限定公開。平成22年度3月末に改良インターフェースの公開予定。検索対象DB数：270データベース（平成22年10月現在） 開始年月：平成20年3月。平成22年10月の訪問者数：11888/訪問回数：22460/ページビュー：34462。
3	DNAデータベース総覧と検索 http://lifesciencedb.jp/ddbj/ (英語版) http://lifesciencedb.jp/ddbj-e/	公開	巨大な国際DNAバンクを高速に検索し生物種、分子種、プロジェクトに分類して表示するシステム。従来巨大な計算機を必要としていたが、バンク内部を詳細に検討し1億300万レコードを58万プロジェクトに縮退させ実質上、計算機一台での検索を可能にした。さらに、ある研究機関から発表されたある生物のゲノムデータを一括してダウンロードするというような研究単位でのデータ取得を実現した。BLASTによる配列検索、核酸関連の特許公報へのダイレクトリンク、レコードの時系列展開も可能である。英語版サービスサイトの整備も完了。今後は、データベース受入れ、横断検索など他のサービスとの統合、検索の一層の高速化、ESTの臓器分類や発現データとの統合を図る。検索対象：1億300万レコード。開始年月：平成19年3月。平成22年10月の訪問者数：3914/訪問回数：16422/ページビュー：95984/ダウンロード：8.44Gbyte。
4	統合TV	公開	データベースやウェブツールの使い方などを動画で発信するウェブサイト。動画コンテンツ

	http://togotv.dbcls.jp/ (英語版) http://togotv.dbcls.jp/?category=English		数は 366 に達した (平成 22 年 10 月現在)。開始年月 : 平成 19 年 7 月。平成 22 年 10 月の訪問者数 : 4326/訪問回数 : 15425/ページビュー : 74288/ダウンロード : 150.47Gbyte。
5	生命科学学協会カタログ http://lifesciencedb.jp/lbdb.cgi?gg=tool_academy	公開	生命科学系の学協会のデータベースと検索サービス。国内の生命科学系学協会を網羅したデータベース。各学会の詳細情報やウェブサイト、学会誌などの公開情報を掲載。平成 19 年度は日本学術会議の資料等をもとに生命科学系の学協会をリストアップしそれぞれの情報を調査、DB カタログと同じユーザインターフェースを利用して公開している。平成 21 年以降は各学会が出版している学会誌や大会要旨の情報調査を行った。掲載学協会数 : 600 学会 開始年月 : 平成 20 年 4 月。平成 22 年 3 月の訪問者数 : 47/訪問回数 : 63/ページビュー : 171。
6	学会要旨統合検索 http://lifesciencedb.jp/lbdb.cgi?gg=tool_tproc	公開	学会要旨集の検索サービス。著者名や施設名の表記揺れを吸収する辞書を搭載し、研究者の研究履歴をたどることが可能で、キーワードのトレンドを表示させることもできる。要旨集の提供を受けた遺伝学会と進化学会については全文検索全文表示を行っている。非公開の分子生物学会については、索引に対する検索のみ行っており要旨本文は表示しない。H21 年度には生物物理学会から要旨データの提供を受けた。学協会カタログや横断検索との重複を避けるため現在は更新を中断している。検索対象 : 4 学会 15729 要旨。開始年月 : 平成 20 年 4 月。平成 22 年 10 月の訪問者数 : 30/訪問回数 : 38/ページビュー : 43。
7	蛋白質核酸酵素全文検索 http://lifesciencedb.jp/pne/	公開 (更新分一部未公開)	蛋白質核酸酵素バックナンバーの全文検索サービス。国内で出版社により発行されている総説誌は、研究者にとって重要な研究発表の場であるとともに、無くてはならない良質な情報源でもある。共立出版の協力により蛋白質核酸酵素のバックナンバー電子化と公開が可能となり、すでに 1985~2007 年分を公開中である。残念ながら本雑誌は平成 22 年 1 月号をもって休刊となったが、休刊号までの総説をアーカイブし今後も提供を続ける予定である。収録論文 : 1985 年~2007 年 (全文表示分)~2009 年 (要旨表示分)。開始年月 : 平成 20 年 6 月。平成 22 年 10 月の訪問者数 : 4643/訪問回数 : 11758/ページビュー : 22325。
8	MotDB http://MotDB.dbcls.jp/	公開	DBCLS を担っていく人材であるアノテータ、キュレータ、システムデータベース管理者向けの教材を提供するサイト (MotDB は Master of the database の略)。中核機関で独自に作成した教材 PDF ファイルもダウンロード可能。講習会開催のたびに、情報が追加されている。開始年月 : 平成 19 年 3 月。平成 22 年 10 月の訪問者数 : 967/訪問回数 : 1595/ページビュー : 8109/ダウンロード : 1.29Gbyte。
9	遺伝子発現バンク (GEO) 目次 http://okubolab.genes.nig.ac.jp/bodymap_i/ (英語版) http://lifesciencedb.jp/geo-	公開	ヒト遺伝子の解剖学的な発現パターンデータの統合サイトを構築した。発現パターンは、測定法毎に異なる場合があることが知られており、ここではできるだけ客観的な発現パターンの解釈を可能にするために、5 種類の発現データ、即ち iAFLP、GeneChip、EST、NCBI の SAGEmap によるタグマップ、SAGE データの独自タグマップに基づく発現データを表示可能にした。平成 22 年 10 月の訪問者数 : 1318/訪問回数 : 2852/ページビュー : 10598/ダウンロード :

	e/		885.65Mbyte。
10	生命科学系データベースアーカイブ http://dbarchive.lifesciencedb.jp/	公開	データベースのアーカイブを公開するサービス。各登録データベースに対して、データベースとデータのメタデータ、利用許諾が示され、アーカイブデータのダウンロードが可能である。データの一括ダウンロードと共に、検索して絞り込んだデータのダウンロード機能も備えている。また、データベースの寄託を公募するためのページを作成し、その中で「寄託データベース利用許諾作成マニュアル」、「データベース利用許諾用語定義解説」などの作成を行い公開している。現在国内の生命科学系データベースが37件登録されている。開始年月：平成21年3月。平成22年10月の利用状況（訪問者数：1063、訪問回数：1470、ページビュー：4456、ダウンロード：2.93GB）
11	生命科学系主要プロジェクト一覧 http://togodb.dbcls.jp/lbdb_project （英語版） http://togodb.dbcls.jp/lbdb_project_en	公開	日本国内の生命科学系関連の主要なプロジェクトに関して詳細な情報を収集しまとめたもの。1990年代以降の4省（文科省、経産省、厚労省、農林水産省）の生命科学に関連する80の主要プロジェクトについて、プロジェクトの期間、概要、予算、代表研究者、公開の状況、成果、文献等に関する情報が閲覧できる。開始年月：平成20年3月。平成20年度まで「ゲノム・ポストゲノム主要プロジェクト一覧」として公開していたものの後継版。平成22年10月の訪問者数：12/訪問回数：12/ページビュー：12
12	TogoProt http://lifesciencedb.jp/togoprot/	公開	蛋白質関連データベース統合検索システム。ユーザは日本語または英語のキーワードで蛋白質を検索し、得た蛋白質リストから蛋白質を選ぶと、その蛋白質に関連づけられる各種のデータをデータベース横断的に集めて表示することができる。蛋白質の代わりに蛋白質ファミリーを選べば、ファミリーのいずれかのメンバーに関連づけられる各種データを表示できる。41のデータベースを収録。開始年月日：平成22年6月。平成22年10月の訪問者数：75/訪問回数：123/ページビュー：482
13	日本の生命科学データベース政策 http://lifesciencedb.jp/sciencepolicy/	公開	これまでにライフサイエンスのデータベースに関わる検討がいろいろなところで行われており、一元的に俯瞰できる状態ではなかったため、本プロジェクトに関与しそうな報告書、会議資料（ウェブページ情報等含む）等を1箇所に集約し、時系列に整理した。開始年月：平成21年3月。4月から外部公開した。平成22年10月の訪問者数：86/訪問回数：103/ページビュー：113
14	文科省「ゲノム」研究報告書 全文検索 http://togodb.dbcls.jp/genome_sci_ryouiki1/	公開	ゲノム特定領域（2005年-2009年）の年次報告書データ 一覧サイト。全文検索可能で、全235の題目ごとのメタデータつきで、報告書のPDFファイルのダウンロードが可能。開始年月日：平成20年6月。平成22年10月の訪問者数：4106/訪問回数：6135/ページビュー：7273
15	Togoexp (RefEx) http://togoexp.dbcls.jp/	公開	RefEx (Reference Expression dataset) は、提供する4種類の異なる手法 (EST, GeneChip, iAFLP, CAGE) によるヒトおよびマウスの遺伝子発現データのリファレンスデータセット。個々の遺伝子については染色体領域や遺伝子ファミリー (InterPro)、Gene Ontologyに基づ

			く生物学的機能別に検索ができる他、発現パターンから探すことが可能。開始年月日：平成21年12月。平成22年10月の訪問者数：111/訪問回数：160/ページビュー：914
16	ライフサイエンス 新着論文レビュー http://first.lifesciencedb.jp/	公開	<i>Nature, Science, Cell</i> などに代表されるトップジャーナルに掲載された日本人を著者とする生命科学分野の論文について、論文の著者自身の執筆による、専門分野の異なる生命科学研究者にむけた日本語によるレビューを、だれでも自由に閲覧・利用できるよう Web 上にていち早く無料で公開。開始年月日：平成22年9月。平成22年10月の訪問者数：2756/訪問回数：11294/ページビュー：44484

(2) DB基盤システム、ツール等開発成果物の整備

通番	DB基盤システム、ツール等の名称	公開／未公開	概要（主な機能・特徴点、進捗状況、今後の計画などを簡潔にわかりやすく記述）
1	遺伝子名称シソーラス http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	分子情報から医学文献までの遺伝子や蛋白質名称の正確な同定のための遺伝子固有名（「遺伝子名」）と一般名称（「ファミリー名」）の辞書データの構築を目的として、様々なデータベースで利用されている名称の収集と専門的キュレータによる編集を行い、遺伝子を持つ多様な名称の関係を明示した遺伝子名称シソーラスを開発した。本シソーラスは、ヒトをはじめ9種類の生物をカバーしている。
2	生物学名羅日英対応辞書 http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	研究分野でよく使われる生物種の基準として、学名に日本語一般名を対応させた生物学名日本語一般名対応辞書を開発した。対応付けは、塩基配列データベース (DDBJ) の登録エントリ数が多い生物種から順番に行った。また、標準和名が存在しない場合、その生物を説明する一般的な名称を用いた。今年度は追加作業を継続。登録件数：学名（ラテン語）と他言語での種名の対応 - 77,625 件、学名と和名の対応 - 66,346 件、そのほか、学術用語の日英対応 - 42,823 件。
3	施設名称辞書 http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1 (http://lifesciencedb.jp:32888/agency/agency.cgi)	公開	日本のライフサイエンス研究を俯瞰するための重要な情報源として、各種関連学会の抄録などの報告文書があるが、これらをデータベース化する際に問題となるのが、例えば、大阪大、阪大、大阪大学大学院などといった、施設名称研究室名称の表記ゆれである。これに対応するために、施設名称辞書を開発した。すでに公開しているものは1625の名称についてまとめたものであるが、昨年度からその範囲を15,000件以上の研究機関・研究グループに広げ、ホームページ掲載情報や学会要旨データベースより収集した機関名別名情報を格納し、機関名の表記ゆれに対応した検索を可能とした。今年度も辞書の拡張を継続し、引き続き内部公開で試験運用中である。
4	動物解剖学自動分類タガー http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	動物の臓器、組織を、大きく10のグループに分類(大分類)し、さらにそれぞれのグループを細かく分類、合計40の小分類グループに分類する。基本的には、与えられた解剖用語に対して、解剖用語辞書、病理関連語彙の分類辞書、形容詞の解剖用語辞書、一般的な臓器名称の分類辞書、の4種類の辞書を順番に検索し、上記のカテゴリに分類する。

5	植物解剖学自動分類タガー http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	植物(維管束植物)の部位、組織を大きく6のグループに分類し、さらにそれぞれのグループを細かく分類、合計11の小分類グループに分類する。分類における検索対象としては、生物種に合わせて、種子を持たない維管束植物の解剖用語辞書、イネ科の解剖用語辞書、その他被子植物の解剖用語辞書、裸子植物の解剖用語辞書、トウモロコシ属の解剖用語辞書、フウチョウソウ目の解剖用語辞書を用いる。
6	都市名国名自動検出タガー http://lifesciencedb.mext.go.jp/result/tech.html#id2-1-1	公開	論文やデータベースレコードに見られる国の名称の未記載や国の名称にみられる表記ゆれを吸収することを目的に、国別に分類するための辞書を開発した。
7	学術用語集	公開	文部科学省発行の学術用語集のうち、生命科学分野に関連のあるもの医学編、農学編、植物学編、化学編、歯学編、薬学編、動物学編について、学術用語と対訳情報を取り出し対訳辞書としてまとめ公開した。開始年月：平成21年8月、更新年月：平成22年10月。登録用語数-150,592件
8	辞書の構築と公開 http://lifesciencedb.jp/lldb.cgi?gg=dic	公開	これまでに構築、集積してきた、辞書、シソーラス、オントロジーの公開サイト。学術用語、学名、学名と他言語での種名の対応、学名と和名の対応、学術用語の日本語と英語の対応、解剖学用語の和英ラテン対応辞書、遺伝子名称シソーラス、施設名辞書、解剖学辞書がダウンロード可能。開始年月：平成21年9月。平成22年10月の訪問者数：690/訪問回数：827/ページビュー：1850。
9	生物アイコン http://lifesciencedb.jp/lldb.cgi?gg=resource_icon (英語版) http://lifesciencedb.jp/lldb.cgi?gg=resource_icon&lng=en	公開	ユーザが自身のHPなどで自由に利用できるように提供されている生物のアイコン集。目的とするイラストをキーワード検索でき、生物に関する外部情報のリンクも付いている。ユーザはイラストに関するコメントを書き込むことができ、さらには同じ条件で配布可能なイラスト等を投稿することもできる。アイコン数：220。開始年月：平成20年6月。定期的にアイコン数を増やし、インターフェースを更新。ヘルプページを作製した。平成22年10月の訪問者数：249/訪問回数：307/ページビュー：452/ダウンロード：86.4Mbyte。
10	BodyParts3D http://lifesciencedb.jp/bp3d	公開	解剖学用語、すなわち臓器・器官・部位の名称やそれらをくくる概念をモデル人間中の3次元座標で定義した辞書。BodyParts3D内の空間関係で用語関係を表現すると、ツリー型表現(いわゆる解剖オントロジー)と違い、改変にロバストな表現が可能。今後も用語の追加を続ける。現在、全パーツ数が1314種である。生命科学系データベースアーカイブから全ポリゴンメッシュデータがダウンロード可能。今年度は脳モデルの改修を行った。アクセス情報はアナトモグラフィーと一緒にカウントしているため下記参照。
11	アナトモグラフィー http://lifesciencedb.jp/bp3d http://lifesciencedb.jp/bp3d?lng=en (英語)	公開	BodyParts3Dから部品を選択して自由に人体の部分のモデル図が書ける。視点固定の図譜や3Dアトラスと異なり透過度や着色、視点は自由に選択可能。今後は、解剖学教育や医学用語辞書の助けになる人体版グーグルマップを目指す。すでに公開したバージョンに加え、拡大・縮小、平行移動など自由度が高くかつ高速な図形処理が可能なバージョンを公開した。平成22年10月の訪問者数：3879/訪問回数：5432/ページビュー：143427/ダウンロード：

			27.93Gbyte。
12	TogoDB http://togodb.dbcls.jp/	公開	データベース受け入れシステム。さまざまな形式のさまざまな種類のデータを簡単に読み込み、自動的に共通のウェブ検索インターフェースを生成する。また、自動的にウェブサービス(SOAP/WSDL API)を生成する。これらの機能により、容易に維持困難なDBを移設できる。また、本ツールのダウンロードにより、誰でもDBを構築可能である。平成22年10月の訪問者数：1348/訪問回数：6181/ページビュー：114779/ダウンロード：5.27Gbyte。
13	TogoWS http://togows.dbcls.jp/	公開	国内外のウェブサービスを透過的に統合するシステム。国内各サービスの稼働状況を継続的に監視し、サービスの連携に必要なデータ形式変換機能等を提供する。これにより、ユーザは統一的命名規則と統一的データ構造で国内外のウェブサービスにアクセスすることができ、相互運用が容易となり、ワークフローの作成・蓄積を促進する。開始年月：平成20年4月。平成22年10月の訪問者数：433/訪問回数：973/ページビュー：36536/ダウンロード：2.25Gbyte。
14	OpenID 認証システム http://openid.dbcls.jp/ (英語版) https://openid.dbcls.jp/?lang=en	公開	一つのIDで複数のサイトを認証できるアカウントシステム。このシステムを利用することで、ユーザはひとつのIDとひとつのパスワードを保持すればよい。これにより、各サイトで認証サービスを用意する必要がなく、サイト間のユーザ情報の集約が容易になる。開始年月：平成20年4月。平成22年10月の訪問者数：300/訪問回数：663/ページビュー：7138。
15	OReFiL http://orefil.dbcls.jp/ (英語版) http://orefil.dbcls.jp/en	公開	オンライン上に存在する多数の生命科学系の資源(データベースやソフトウェアなど)を効率的に見つけるための検索システム。信頼性や概要を容易に取得するために、資源の所在だけでなく、その資源について書かれている論文情報や、他の利用者の評価を閲覧できる。アカウントを取得することで、自ら評価を行うことも可能である。また、検索に関連する機能を全てウェブサービスとして提供しているため、利用者のプログラムからOReFiLを利用できる。英語版ヘルプなどを整え、環境を整備した。開始年月：平成19年8月。平成22年10月の訪問者数：365/訪問回数：556/ページビュー：2227。
16	Allie http://allie.dbcls.jp/ (英語版) http://allie.dbcls.jp/	公開	医学生物系論文誌情報データベースMEDLINEを対象とし、出現する略語とその転回形のペアを検索するシステム。生命科学系の論文では非常に多くの略語が使われており、同じ表記でも全く違う意味を示していることが少なくない。そこで、利用者の興味のある略語を検索語として入力することで、その使われ方を一覧表示すると共に、論文の発表年を提示する。また、検索された各ペアについて、それが出現する論文中で共起する他の略語も提示する。検索に関連する機能を全てウェブサービスとして提供しているため、利用者のプログラムからAllieを利用できる。英語版ヘルプなどを整え、環境を整備した。開始年月：平成20年4月。平成22年10月の訪問者数：936/訪問回数：1625/ページビュー：4192。
17	DBCLS Galaxy http://galaxy.dbcls.jp/	公開	生命科学データに特化した、ウェブベースの対話的ツール組み合わせインターフェース。マウス操作と簡単な入力で、プログラムを書くことなく、途中結果を確認しながら、用意され

			たツールを対話的に組み合わせて使うことができる。DBCLS Galaxy では、TogoWS やテキストマイニングツールなど、DBCLS で開発されたツール群を組み込み、利用できるようにした。ログインは DBCLS OpenID で可能である。また、メニューの国際化対応へ協力し、メニューの日本語化を行い、DBCLS Galaxy のみならず、元の Galaxy にも反映させた。平成 22 年 10 月の訪問者数：190/訪問回数：260/ページビュー：7013。
18	ScrapParty	公開	ブラウザのプラグインによるキュレータ支援用アプリケーション。オンラインで文献やウェブサイトを閲覧しながら、重要な記述や画像、図表を集めコメントとともに簡単に DB 化するツール。平成 18 年度アノテーションの調査、試験に利用した。下記 Wired-Marker が後継アプリケーションであり、現在 ScrapParty 名での開発は終了。
19	Wired-Marker http://www.wired-marker.org	公開	情報共有ツール、アノテーション支援ツール。Firefox ブラウザのプラグインとして働く。データベース構築、アノテーションを支援するブラウザ用のプラグイン。オンラインで閲覧したページ、論文のテキストや図をマークしブックマークすることが可能。再訪問の際にも同じ箇所がハイライトされアノテーションのエビデンスとして詳細な情報が残せる。またフォルダを自由に階層化して情報を整理することが可能。新たにキャッシュ機能が加わり、ウェブページの書き換えが起きても過去のデータを参照できるようになった。 公開日：H19.12.19 総ダウンロード数：366,163 件。
20	MiGAP http://migap.lifesciencedb.jp/ (英語版) http://migap.lifesciencedb.jp/en/	公開	技術革新により大量に生産されるようになった微生物ゲノムの利用価値を高めるため、アノテーションを自動的に行うパイプラインを開発し、試験的にプロジェクト内部に公開、評価を受けた。評価の意見を反映した修正作業を進め、平成 21 年 6 月に一般公開。利用のためには、OpenID の取得が必要。開始年月：平成 21 年 3 月。平成 22 年 10 月の訪問者数：170/訪問回数：383/ページビュー：186383。
21	inMeXes http://docman.dbcls.jp/im/ (英語版) http://docman.dbcls.jp/im/	公開	医学生物系論文書誌情報データベース MEDLINE を対象とし、利用者が入力した文字列とマッチする表現を逐次的(1文字入力する度)に検索するシステム。生命科学系分野の学術文献を英語で執筆する際、既に発表されている論文中で実際に使われている表現を容易に検索出来るれば、特に非ネイティブスピーカーが多数の日本人研究者にとって有用であると考え、構築した。マッチした表現の全 MEDLINE 中(題目・要旨)での出現頻度も表示されるほか、当該表現が使われている文や関連情報を表示させることも可能。開始年月：平成 21 年 7 月。平成 22 年 10 月の訪問者数：508/訪問回数：1516/ページビュー：134342。
22	TogoDoc http://docman.dbcls.jp/pubmed_recom (英語版)	公開	文献情報及び論文 PDF を管理し、また、特定の文献情報群に関連する論文情報を提示するシステム。日々多くの論文が発表されている昨今においては自身が興味を持つ文献情報を効率良く取得することが容易では無くなってきている。また、多くの学術雑誌がオンラインジャーナルを発行していることから論文全文を PDF ファイルで取得出来る機会が増大し、研究者の PC に多くの論文 PDF ファイルが保存されている。以上の背景から文献情報の効率的な

	http://docman.dbcls.jp/pubmed_recom		取得、整理環境を実現すべく、そのために必要な情報は既に研究者自身の手元にある文献情報リストから得られるようなシステム TogoDoc を構築した。利用者 PC に保存されている PDF ファイルを解析し必要な書誌情報を自動的に付加するなどの機能を持つ文献情報管理ソフトウェア TogoDocClient と書誌情報のやり取りや PDF ファイルの授受を行うなどの連携機能を持つ。公開年月：平成 21 年 12 月。
23	Gendoo http://gendoo.dbcls.jp/	公開	Gendoo は、遺伝子や疾患について、関連する疾患や薬剤、生命現象、臓器などの情報をキーワードで表示するシステム。入力として、Gene ID、遺伝子名、OMIM ID、疾患名（OMIM のタイトル）、キーワード（MeSH terms）：疾患、薬剤、生命現象、臓器を受け付け、関連する語を表示する。開始年月：平成 19 年 12 月試験公開。平成 22 年 10 月の訪問者数：1639/訪問回数：2018/ページビュー：3445。

（３）その他の成果物（（１）、（２）に該当しないもの）

通番	名称	公開／未公開	概要
1	統合 PJ 成果公開サイト http://lifesciencedb.jp	公開	統合 DB プロジェクトに関する中核機関の成果公開サイト。生命科学データベースカタログ、DB 横断検索、INSD 全文検索、統合 TV、ORFeIL、生命科学学会検索、学会要旨統合俯瞰システム、MotDB、LSDB ブログ等を掲載している。また、プロジェクト参加機関の成果公開サイトへのリンクも張ってある。全掲載サービスは 60 件で、DBCLS 分 40 件、その他参加機関分 20 件である。開始年月：平成 20 年 6 月。DBCLS 提供サービスへのアクセス状況としては、平成 22 年 10 月の訪問者数：49639/訪問回数：118782/ページビュー：689712。
2	統合 PJ 成果公開サイト（英語版） http://lifesciencedb.jp/?lng=en	公開	国外で本プロジェクトの紹介をする機会も増えてきたことから、英語での利用が可能なサービスを集めて、英語版のポータルサイトを公開した。現在紹介されているデータ・サービスは、ゲノム・ポストゲノム主要プロジェクト一覧、生命科学データベースカタログ、生物アイコン、DNA データベース総覧と検索、遺伝子発現バンク (GEO) 目次、アナトモグラフィ/BodyParts3D、Wired-Marker、TogoWS、togoTV。そのほか、プロジェクト紹介ページ、参画機関の成果ページ（英語版があるもの）へのリンクを張った。開始年月：平成 21 年 3 月。
3	統合データベース間の連携調査 http://charles.kazusa.or.jp/lifesciencedb/	公開	代表的モデル研究植物である「イネ」ならびに「シロイヌナズナ」のゲノムアノテーション型公開データベース各 46 サイト、25 サイトの基本項目の調査と実験生物研究者 188 名のアンケートにより調査し、それぞれのデータベース間の連携のための現状の課題と将来の DB 統合にむけた課題の整理を行った。
4	検索アルゴリズムを含めた知 識情報技術の動向調査 http://lifesciencedb.mext.go.jp/result/strategy.html#id1	公開	次世代の生物情報データベース統合に必要な知識情報技術として、検索システム、データマイニング、Web 2.0 およびグリッドコンピューティングに焦点を絞り、聞き取り調査や文献調査によって動向を調べた。

	-4		
5	遺伝統計学分野における解析技術の基礎調査 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4	公開	遺伝統計学分野で用いられる連鎖解析、連鎖不平衡解析）、QTL 解析等の解析手法と、それぞれの手法における代表的なアルゴリズム計 8 種類の調査を行ない、その特徴および長所・短所を評価した。併せて、各手法の代表的プログラム計 15 種類に関して、実装されているアルゴリズム、動作環境などを調査し、その評価を行った。
6	臨床情報・疾患健康情報の調査 http://lifesciencedb.mext.go.jp/result/strategy.html#id1-4	公開	わが国の代表的コホート研究に関してその研究の背景と目的、対象地域、ターゲット疾患、特徴的な検査項目、対象人数、代表研究者、研究開始時期、資金源等について調査した。
7	データベース受入に関するアンケート調査 http://lifesciencedb.jp/lldb.cgi?pg=0&gg=ukeire_survey	試験的公開	受入による統合化の方針策定と受入対象の候補抽出を目的として、国内データベースを対象として、データベース内容と検索方法、経歴、維持・管理状況、提供可能性及び提供時の問題点などをアンケートにより調査した。送付した 104 件中、65 件の回答が得られ、半数以上の機関が、統合 DB 事業に対して全データもしくは一部データを開示できると回答した。結果の概要および詳細報告（PDF）をウェブ上で公開した。
8	BioHackathon2008 ホームページ http://hackathon.dbcls.jp/	公開	分散環境のままで、各種 DB をアルタイムかつ統合的に利用できる環境を構築すること目的に開催した、国内外のライフサイエンス DB のウェブサービスにおけるデータ構造や命名規則を標準化するための BioHackathon2008 の活動内容と経過を紹介するサイト。2008 年 2 月 11 日から 15 日まで都内で開催。海外参加者 34 名、国内参加者 37 名。平成 22 年 10 月の訪問者数：221/訪問回数：682/ページビュー：2278。
9	BioHackathon2009 ホームページ http://hackathon2.dbcls.jp/	公開	標準化技術および標準化を利用する技術をより安定的に使用可能にすることを目的として開催した BioHackathon2009 の活動内容を紹介するサイト。巨大なデータについてどうすべきか、セマンティックウェブ技術をどう利用すべきかなどの議論の報告が掲載されている。2009 年 3 月 16 日から 20 日まで、都内および沖縄で開催、海外参加者 30 名、国内参加者 28 名。平成 22 年 10 月の訪問者数：320/訪問回数：422/ページビュー：3348。
10	BioHackathon2010 ホームページ http://hackathon3.dbcls.jp/	公開	統合検索に向けたセマンティックウェブ技術の調査と開発を目的とした、BioHackathon 2010 の活動内容を紹介するサイト。2010/2/8 - 12 の期間、海外 32 名、国内 36 名が参加し、LinkedData を実現するための合意形成、セマンティックウェブ技術のフィージビリティ調査のための開発などが行われた。平成 22 年 10 月の訪問者数：309/訪問回数：1143/ページビュー：4203。
11	シンポジウム・講演会 http://symposium.lifesciencedb.jp/	公開	DBCLS が主催したシンポジウムサイトをリストしたサイト。現在は 2009 年 6 月 12 日および 10 月 5 日、2010 年 10 月 5 日開催の 3 つのシンポジウムについてリンクがあり、2009 年 6 月 12 日分のサイトでは、講演動画 9 件・スライド 7 件・発表ポスター PDF 68 件、2009 年 10 月 5 日分のサイトでは講演動画 8 件を掲載している。2010 年のサイトでは発表ポスター PDF 28

			件を公開、講演動画 9 件・スライド 10 件を公開準備中。平成 22 年 10 月の訪問者数：672/訪問回数：1292/ページビュー：3838。
12	外部評価 http://lifesciencedb.jp/lldb.cgi?gg=hyouka	公開	登録ユーザによるサービスへの評価内容を公開するサイト。平成 19 年から基本的に前年度の公開サービスへの評価を実施。平成 20 年度からはウェブサイト記入方式（5 段階評価とフリーコメント）に変更し、サービスについての理解を深めてもらう工夫も行った。なお、いただいたご意見への対応・回答も併せて掲載している。平成 21 年度には、前年度の要望に基づき参考資料として 20 年度の各機関の予算と目標/進捗情報を掲載した。22 年度はプロジェクト最終年度のためプロジェクト全期間について全ての参画機関についての評価を行うこととし、全期間の予算情報と 20 年度以降の目標/進捗情報、各サービスのアクセス情報（22 年度新規公開分は除く）も掲載した。評価者の人数：約 60 名（平成 19 年度）、58 名（平成 20 年度）、117 名（平成 21 年度）。平成 22 年 10 月の訪問者数：206/訪問回数：283/ページビュー：8675。
13	運営委員会等の資料 http://togodb.dbcls.jp/koukai	公開	平成 18 年度から行われてきた運営委員会、作業部会、作業部会分科会の会議資料と議事概要および委員名簿を掲載。開催日や資料名称の簡易検索機能付き。平成 22 年 10 月の訪問者数：244/訪問回数：288/ページビュー：3838。
14	統合データベース技術情報交換ワークショップ http://wiki.lifesciencedb.jp/mw/index.php/BH10.10	公開	次世代のライフサイエンスデータベースをより利便性の高いものとすることを目指して、これまでの BioHackathon で蓄積された知識を国内で共有する目的で行ったワークショップの活動を紹介するサイト。2010/10/18 - 22 の期間、修善寺で開催し、約 70 名の参加者があった。

外部発表実績一覧

学 会 等 発 表 実 績

委託業務題目： 「ライフサイエンス統合データベース開発運用」(戦略立案・実行評価／統合データベース開発／統合データベース支援)

機 関 名： 大学共同利用機関法人 情報・システム研究機構

1. 学会等における口頭・ポスター発表

発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
--------------------------	-------	--------------	--------	--------

平成 22 年度				
新たな日本語 Web コンテンツ「ライフサイエンス新着論文レビュー」[口頭]	飯田 啓介	第 6 回 SPARC Japan セミナー2010 「日本発オープンアクセス」	2010 年 10 月 20 日	国内(国立情報学研究所)
"Microbial Genome Annotation Pipeline (MiGAP)"	Sugawara H, Ohyama A, Mori H and Kurokawa K.	Biocuration 2010	2010 年 10 月 11-14 日	国内(台場)
Development of Cooperative Work Tools and Workflow for Multimodal Genome Annotation [ポスター]	Okamoto S, Nakao M	Biocuration 2010	2010 年 10 月 11-14 日	国内(台場)
TogoDB + TogoWS: a data integration platform for the Semantic Web [口頭]	Toshiaki Katayama and Mitsuteru Nakao	The Conference of International Society for Biocuration (Biocuration2010)	2010 年 10 月 11-14 日	国内(東京台場)
データ・情報の円滑な利用に向けた権利関係の整理の試み [ポスター]	箕輪真理、高祖歩美、中尾光輝、河野信、八塚茂、畠中秀樹、川本祥子、永井啓一、大久保公策、高木利久	シンポジウム 「ライフサイエンスの未来へ～10 年先のデータベースを考える～」	2010 年 10 月 5 日	国内(東京大学本郷キャンパス浅野地区武田ホール)
文部科学省「統合データベースプロジェクト」の広報・普及活動 [ポスター]	高祖歩美、箕輪真理	シンポジウム 「ライフサイエンスの未来へ～10 年先のデータベースを考える～」	2010 年 10 月 5 日	国内(東京大学本郷キャンパス浅野地区武田ホール)
統合データベースの教育・人材育成活動～統合 TV と MotDB～ [ポスター]	河野信、小野浩雅、坊農秀雅	シンポジウム 「ライフサイエンスの未来へ～10 年先のデータベースを考える～」	2010 年 10 月 5 日	国内(東京大学本郷キャンパス浅野地区武田ホール)
多様なゲノム注釈のためのツールと協働作業手法の開発 [ポスター]	岡本忍、中尾光輝、藤澤貴智、佐藤修正、金子貴一、吉村英尚、中平有香、阿久津智子、鐘ヶ江弘美、笠井誠、山本純子、作田千代子、眞板寛子、鹿島裕子、中村保一	シンポジウム 「ライフサイエンスの未来へ～10 年先のデータベースを考える～」	2010 年 10 月 5 日	国内(東京大学本郷キャンパス浅野地区武田ホール)
ウェブサービス相互運用、ワークフロー、データ共有のための基盤技術開発(2) [ポスター]	中尾光輝、山口敦子、山本泰智、片山俊明、川島秀一	シンポジウム 「ライフサイエンスの未来へ～10 年先のデータベースを考える～」	2010 年 10 月 5 日	国内(東京大学本郷キャンパス浅野地区)

				武田ホール)
ウェブサービス相互運用、ワークフロー、データ共有のための基盤技術開発(1) [ポスター]	中尾光輝、山口敦子、山本泰智、片山俊明、川島秀一	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
論文執筆の際に役立つサービスの紹介 [ポスター]	山本泰智、山口敦子、坊農秀雅、高木利久	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
生物学名和名対訳辞書の構築と公開 [ポスター]	森山丈士、栗原英輔、小笠原理、川本祥子、大久保公策	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
統合データベースプロジェクトの新しい日本語コンテンツ「新着論文レビュー」 [ポスター]	飯田啓介、中尾光輝、川本祥子、高木利久	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
「生命科学系データベースアーカイブ」の成果と今後 [ポスター]	坂東明日佳、熊谷禎洋、八塚 茂、西川哲夫、畠中秀樹、大久保公策、高木利久	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
生命科学系データベース情報の収集とポータルサイト構築 [ポスター]	川本祥子	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
ライフサイエンス統合データベースの横断検索サービスの現状 [ポスター]	杉崎太一朗、牧口大旭、高橋順子、平井信一、奥村利幸、谷口仁、畠中秀樹、川本祥子、高木利久	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
TogoProt による蛋白質関連データベースの統合検索 [ポスター]	畠中秀樹、平井信一、多門啓子、大野忠、牧口大旭、奥村利幸、川本祥子、中村春木、高木利久、	シンポジウム 「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)

遺伝子発現リファレンスデータセット『RefEx』 [ポスター]	小野浩雅、大久保公策、高木利久、坊農秀雅	シンポジウム「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
BodyParts3DとAnatomography: 3次元人体構造データベースと人体解剖図作成サービス [ポスター]	三橋信孝、藤枝香、今井紫緒、田村卓郎、川本祥子、高木利久、大久保公策	シンポジウム「ライフサイエンスの未来へ～10年先のデータベースを考える～」	2010年10月5日	国内(東京大学本郷キャンパス浅野地区武田ホール)
Reference Expression dataset (RefEx) for practical gene expression analysis of cancer-related molecular targets [口頭]	坊農秀雅	第69回日本癌学会学術総会	2010年9月22-24日	国内(大阪国際会議場)
データおよび知識の共有と統合化に向けて [口頭]	高木 利久	第48回日本生物物理学会年会 フォーラム「デジタル革命の本当のインパクト」	2010年9月20日	国内(東北大川内キャンパス)
デジタル革命の本当のインパクト[口頭]	大久保 公策	第48回日本生物物理学会年会 フォーラム「デジタル革命の本当のインパクト」	2010年9月20日	国内(東北大川内キャンパス)
RefEx: Reference expression dataset for practical use of gene expression data [ポスター]	Hidemasa Bono, Hiromasa Ono, Kousaku Okubo, Toshihisa Takagi	The 10th Cold Spring Harbor Laboratory/Wellcome Trust conference on Genome Informatics	2010年9月15-19日	国外 (Hinxton (UK))
Functional indexing and curation of next-generation sequencing data [ポスター]	Takeru Nakazato, Hidemasa Bono, Toshihisa Takagi	The 10th Cold Spring Harbor Laboratory/Wellcome Trust conference on Genome Informatics	2010年9月15-19日	国外 (Hinxton (UK))
"WFCC World Data Centre for Microorganisms (WDCM) hosted by the National Institute of Genetics: Lessons Learned" [招待講演]	Hideaki Sugawara	Research Network for Applied Microbes & International symposium on microbial resources management, utilization and fungal DNA barcode (Chinese Academy of Science Institute of Microbiology (Beijing))	2010年8月24日	国外 (北京)
統合データベースプロジェクト: ライフ・イノベーションによる知のめぐりのよいオープンな生命科学を目指して [口頭]	坊農秀雅	平成22年度脳科学研究戦略推進プログラム 課題A, B分科会プログラム	2010年7月27-28日	国内(ホテルさっぽろ芸文館)
ゲノムアノテーション協同作業を支援する手法と技術[口頭]	岡本 忍	ラン藻ゲノム研究会	2010年7月24日	

TogoProt: 蛋白 I D を用いたデータベース横断検索システム [ポスター]	畠中秀樹、平井信一、大野忠、多門啓子、高橋頼子、押田健寛、牧口大旭、奥村利幸、川本祥子、西川哲夫、中村春木、高木利久	第 10 回日本蛋白質科学会年会	2010 年 6 月 16-18 日	国内(札幌コンベンションセンター)
ゲノム注釈の民主化 [口頭]	岡本 忍	農業環境技術研究所公開セミナー	2010 年 6 月 14 日	国内 (つくば)
DBCLS Galaxy: A Galaxy Service to integrate databases in Japan [口頭]	Atsuko Yamaguchi	Galaxy Developer Conference 2010	2010 年 5 月 15-17 日	国外(NewYork)
Galaxy Internationalization (i18n) and Localization (L10n) [口頭]	Mitsuteru Nakao	Galaxy Developer Conference 2010	2010 年 5 月 15-17 日	国外(NewYork)
「科学技術コモンズと情報知識学の挑戦」ライフサイエンスにおける統合データベースの構築と課題 [招待講演]	高木 利久	情報知識学会 第 18 回 (2010 年度) 年次大会記念シンポジウム	2010 年 5 月 15 日	国内(東大本郷工学部 2 号館)
平成 21 年度				
生命科学系データベースアーカイブ [口頭]	畠中 秀樹	日本農芸化学会 2010 年度大会 ランチオンセミナー「情報整理術 2010 ～文献とデータを賢く管理する～」	2010 年 3 月 27-30 日	国内 (東京)
情報メディアとしてのゲノム [口頭]	岡本 忍	GC0E 若手研究異分野セミナー	2010 年 3 月 28-29 日	国内 (北海道)
内閣府統合データベース事業と理研への期待[口頭]	高木 利久、大石道夫	第 18 回理事長ファンドワークショップ「理研総合データベースを核にした研究基盤連携網の構築」	2010 年 3 月 12-13 日	国内 (葉山)
内閣府統合データベース事業と理研への期待、文科省ライフサイエンス統合データベースプロジェクトの目的、現状、課題[口頭]	高木 利久	第 18 回理事長ファンドワークショップ「理研総合データベースを核にした研究基盤連携網の構築」	2010 年 3 月 12-13 日	国内 (葉山)
微生物ゲノム基盤情報の高度化と統合のためのシステム	岡本 忍	第 4 回日本ゲノム微生物学会年会	2010 年 3 月 7-9 日	国内 (九州)
国際開発会議 DBCLS バイオハッカソン 2010 レポート[口頭]	中尾 光輝	人工知能学会第二種研究会 分子生物情報研究会 (SIG-MBI)	2010 年 3 月 5-6 日	国内 (能美)
ゲノムアノテーション情報の高度化統合化への取り組み [口頭]	岡本 忍	かずさ研究交流セミナー	2010 年 1 月 27 日	国内 (千葉)

BIOLOGICAL DOCUMENT CLUSTERING USING A GENE-GENE NETWORK[ポスター]	Hong-Woo Chun, Shinobu Okamoto, Yasunori Yamamoto, Atsuko Yamaguchi, Toshihisa Takagi	PSB(Pacific Symposium on Biocomputing) 2010	2010 年 1 月 4-8 日	国外 (Hawaii)
DDBJ Read Archive and DDBJ Read Annotation Pipeline:An Archive Database and an Analytical Tool for Next-Generation Sequence Data T003/P120[ポスター／口頭]	Eli Kaminuma, Yuichi Kodama, Satoshi Saruhashi, Takeshi Konno, Takako Mochizuki, Hidemasa Bono, Hideaki Sugawara, Kousaku Okubo, Toshihisa Takagi, Yasukazu Nakamura	The 20th International Conference on Genome Informatics (GIW2009)	2009 年 12 月 14-16 日	国内 (横浜)
なぜ今共有なのか？個人プレーから社会プレーへ[口頭]	大久保 公策	第 32 回日本分子生物学会年会、フォーラム「デジタル時代のジレンマ」”共有と秘匿のバランス”	2009 年 12 月 10 日	国内 (横浜)
データベース統合センターの役割[口頭]	高木 利久	第 32 回日本分子生物学会年会、フォーラム「デジタル時代のジレンマ」”共有と秘匿のバランス”	2009 年 12 月 10 日	国内 (横浜)
実験生物学者にとって有益な情報環境を提案する機会：「統合データベースプロジェクト」のユーザ評価[口頭]	高祖 歩美、箕輪 真理	第 32 回日本分子生物学会年会、ワークショップ	2009 年 12 月 10 日	国内 (横浜)
ブタ成熟脂肪細胞および顆粒層細胞における脱分化ならびに多能性獲得機構の統合トランスクリプトミクス[ポスター]、[口頭]	小野 浩雅、坊農 秀雅、加野 浩一郎	第 32 回日本分子生物学会年会、ワークショップ	2009 年 12 月 10 日	国内 (横浜)
かずさアノテーションを用いた分散型ゲノムアノテーションの実証実験[口頭]	岡本 忍、中尾 光輝、藤沢 貴智、佐藤 修正、金子 貴一、中村 保一	第 32 回日本分子生物学会年会、ワークショップ	2009 年 12 月 10 日	国内 (横浜)
統合データベースプロジェクト：オープンで知のめぐりの良い分子生物学を目指して[口頭]	坊農 秀雅	第 32 回日本分子生物学会年会、ワークショップ	2009 年 12 月 10 日	国内 (横浜)
DBCLS で提供中の文献関係サービス[ポスター]	山本 泰智、山口敦子、坊農 秀雅、高木利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内 (横浜)
KazusaMart:植物および関連微生物のゲノムデータ統合[ポスター]	藤澤 貴智、岡本 忍、中尾 光輝、中村 保一	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内 (横浜)

「DDBJ Read Archive」：次世代シーケンサからの出力データのためのアーカイブ[ポスター]	児玉 悠一、猿橋 智、五條 堀 孝、舘野 義男、神沼 英里、中村 保一、菅原	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
次世代シーケンサーのハイスループット・データ解析システム「DDBJ Short Read Annotation Pipeline」の構築[ポスター]	神沼 英里、児玉 悠一、大久保 公策、高木 利久、中村 保一	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
微生物ゲノムアノテーションパイプライン MiGAP[ポスター]	大山 彰、森 宙史、黒川 颯、菅原 秀明	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
DBCLS Galaxy: 生物データ処理の対話的ツール組み合わせインターフェース[ポスター]	山口 敦子、全 弘宇、中尾 光輝、大内田 賢太、山本 泰智、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
生命科学研究者のための統合文献情報管理システム[ポスター]	岩崎 渉、山本 泰智、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
Representation of the multiple roles of proteins from anatomical perspective by data analysis[ポスター]	坂東 明日佳、中村 桂子、大久保 公策	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
キーワードと蛋白 ID を用いた横断検索システム[ポスター]	畠中 秀樹、平井 信一、大野 忠、高橋 順子、牧口 大旭、奥村 利幸、熊谷 禎洋、八塚 茂、川本 祥子、西川 哲夫、中村 春木、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
生命情報データベースの横断検索システム：Web-API サービスの開発[ポスター]	牧口 大旭、高橋 順子、川崎 恭実子、川本 祥子、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
データバンク (INSD, GEO) 目次システムの開発 [ポスター]	小笠原 理、渡邊 康司、栗原 英輔、森山 丈士、大久保 公策	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
BodyParts3D と Anatomography: 3 次元人体構造データベースと人体解剖図作成サービス[ポスター]	三橋 信孝、藤枝 香、田村 卓郎、川本 祥子、高木 利久、大久保 公策	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）

生命科学系データベースアーカイブサービス[ポスター]	西川 哲夫、八塚 茂、熊谷 禎洋、畠中 秀樹、片山 俊明、川本 祥子、大久保 公策、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
ライフサイエンス統合データベース: アップデート 2009[ポスター]	河野 信、Lsdb 開発スタッフ、大久保 公策、高木 利久	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
かずさアノテーションを用いた分散型ゲノムアノテーションの実証実験 [口頭]	岡本 忍	第 32 回日本分子生物学会年会	2009 年 12 月 9-12 日	国内（横浜）
Functional analysis pipeline of transcript sequences as gene expression data[口頭]	Hidemasa Bono	第 7 回がんとハイポキシア研究会	2009 年 12 月 5-6 日	国内（京都）
シアノバクテリアのゲノム情報基盤[口頭]	岡本 忍	ラン藻ワークショップ	2009 年 12 月 4-5 日	国内（千葉）
GuideLink: A Corpus Annotation System that Integrates the Management of Annotation Guidelines[ポスター]	Kenta Oouchida, Jin-Dong Kim, Toshihisa Takagi and Jun'ichi Tsujii	The 23rd Pacific Asia Conference on Language, Information and Computation	2009 年 12 月 3-5 日	国内（香港）
デジタル科学における独占と共有の賢いバランス[口頭]	大久保 公策	産総研 生命情報工学研究センター2009 年度研究報告シンポジウム	2009 年 12 月 3-4 日	国内（お台場）
統合データベースプロジェクトのサービス紹介～複数のデータベースを繋ぐツールたち - [口頭]	山口敦子、中尾 光輝	産総研 生命情報工学研究センター2009 年度研究報告シンポジウム	2009 年 12 月 3-4 日	国内（お台場）
ライフサイエンス統合データベースの課題: 権利と法律、技術[口頭]	中尾 光輝	産総研 生命情報工学研究センター2009 年度研究報告シンポジウム	2009 年 12 月 3-4 日	国内（お台場）
「生命科学情報の共有を目指す」ライフサイエンス統合データベース」 [ポスター]	箕輪 真理、高祖 歩美、坂東 明日佳、川本 祥子	情報・システム研究機構シンポジウム 情報とシステム 2009	2009 年 11 月 30 日	国内（品川）
Integrated Database Project for Life Science in Japan[口頭]	Toshihisa Takagi	BIOINF02009, CBI-KSBSB Joint International Conference	2009 年 11 月 4-6 日	国外（釜山）
統合 DB プロジェクトのサービス紹介～データベースを役立てよう～[口頭]	川本 祥子	第 47 回日本生物物理学会年会 ランチオンセミナー「データベース統合の多面的展開」	2009 年 10 月 30 日-11 月 1 日	国内（徳島）

多様なデータが広く活用されるために[口頭]	畠中 秀樹	第 47 回日本生物物理学会年会 ランチ ンセミナー「データベース統合の多面的 展開」	2009 年 10 月 30 日-11 月 1 日	国内（徳島）
Systematic organization of gene expression data in Japan[口頭]	Hidemasa Bono, Eli Kaminuma, Yuichi Kodama, Yasukazu Nakamura, Kousaku Okubo, and Toshihisa Takagi	12th International MGED Meeting, focusing on translational genomics and high throughput sequencing	2009 年 10 月 5-8 日	国外(Phoenix, Arizona)
「デジタル科学への最後のステップ」[口頭]	大久保公策	シンポジウム「科学技術における情報の 上手な権利化と共有化」	2009 年 10 月 5 日	国内（東大弥 生）
ライフサイエンス統合データベースプロジェク ト[口頭]	高木 利久	日本人類遺伝学会 第 54 回大会 ワー クショップ「ゲノムインフマティクス」	2009 年 9 月 26 日	国内（品川）
ライフサイエンス分野の統合データベースの学 部や高校教育における活用[口頭]	坊農 秀雅	日本遺伝学会第 81 回大会 ミニ・シンポ ジウム「自分が興味を持つ遺伝子がどの 範囲の環境微生物に存在するのかを探 索する方法：環境微生物ゲノム配列から のお宝遺伝子発掘の学部や高校教育に おける活用」	2009 年 9 月 16 日	国内（長野）
RizoBase の新機能 [口頭]	岡本 忍	植物微生物研究会	2009 年 9 月 8-10 日	国内（長野）
統合データベースプロジェクト：オープンで知の めぐりの良いライフサイエンス研究を目指して [口頭]	坊農秀雅	文部科学省委託研究開発事業 統合デー タベースプロジェクト シンポジウム 2009「データベースが拓 くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本 郷）
BodyParts3D/Anatomography: 3 次元人体構造 DB と人体モデル図作成ツール[ポスター]	三橋 信孝	文部科学省委託研究開発事業 統合デー タベースプロジェクト シンポジウム 2009「データベースが拓 くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本 郷）
Gene Indexing: 植物と関連微生物におけるオー プンゲノム注釈実証実験[ポスター]	岡本 忍	文部科学省委託研究開発事業 統合デー タベースプロジェクト シンポジウム 2009「データベースが拓 くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本 郷）

生物種ごとの日本の分子生物学研究一覧の構築に向けて ～「生物アイコン」の現状と今後[ポスター]	坂東 明日佳	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
生命科学系データベースアーカイブサービス[ポスター]	西川 哲夫	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
DBCLS Galaxy: 生物データ処理の対話的ツール 組み合わせインターフェース[ポスター]	山口 敦子	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
発現統合: 配列データとしての発現データを扱う 解析パイプラインの開発[ポスター]	坊農 秀雅	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
BioHackathon 2009 レポート[ポスター]	中尾 光輝	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
MeSH キーワード集による疾患の特徴プロファイリング[ポスター]	仲里 猛留	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
テキスト処理して情報整理[ポスター]	山本 泰智	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
統合データベース講習会 AJACS[ポスター]	河野 信	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）

統合TV-動画によるバイオデータベースとウェブツールの活用法解説サイト[ポスター]	小野 浩雅	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
MeSH キーワード集による疾患の特徴プロファイリング[ポスター]	仲里猛留、坊農秀雅、松田秀雄、高木利久	文部科学省委託研究開発事業 統合データベースプロジェクト シンポジウム 2009「データベースが拓くこれからのライフサイエンス」	2009 年 6 月 12 日	国内（東大本郷）
統合データベースポータル「統合ホームページ」lifesciencedb.jp[ポスター]	仲里 猛留、坂東 明日佳、坊農 秀雅、全 弘宇、藤枝香、畠中 秀樹、平井 信一、川本 祥子、河野 信、高祖歩美、熊谷 禎洋、箕輪 真理、三橋 信孝、永井 啓一、中尾 光輝、西川 哲夫、岡本 忍、小野 浩雅、大内田賢太、谷口 仁、山口 敦子、山本 泰智、八塚 茂、大久保 公策、高木 利久	第 9 回日本分子生物学会春季シンポジウム	2009 年 5 月 12-13 日	国内（宮崎）
平成 20 年度				
Gene Indexing: Manual micro-annotation of plant-related organisms using KazusaAnnotation [ポスター]	岡本忍、中尾光輝	第 3 回国際バイオキュレーション会議 Biocuration 2009	2009 年 9 月 4 日, 11 月 19 日	国外（Berlin, Germany）
Semantic clustering of biomedical abbreviation definitions using bipartite graph [ポスター]	Atsuko Yamaguchi, Yasunori Yamamoto, Hong-Woo Chun and Toshihisa Takagi	The 14th Pacific Symposium on Biocomputing (PSB)	2009 年 1 月 5-9 日	国外（Hawaii）
Functional profiling of OMIM data using MeSH vocabulary [口頭及びポスター]	Takeru Nakazato, Hidemasa Bono, Hideo Matsuda, Toshihisa Takagi	2008 年日本バイオインフォマティクス学会年会（JSBi2008）	2008 年 12 月 15-16 日	国内（大阪）

統合データベースプロジェクトのサービスについて [口頭]	川本祥子	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008) シンポジウム「ライフサイエンス統合データベースプロジェクト」	2008 年 12 月 9-12 日	国内(神戸)
日本のライフサイエンス統合データベース構築・維持のための人材育成活動 [口頭]	坊農秀雅	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008) シンポジウム「ライフサイエンス統合データベースプロジェクト」	2008 年 12 月 9-12 日	国内(神戸)
科学データは誰のものか? [口頭]	大久保公策	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008) シンポジウム「ライフサイエンス統合データベースプロジェクト」	2008 年 12 月 9-12 日	国内(神戸)
KazusaWiki: ゲノムデータベースにおけるの Wiki の活用 [ポスター]	藤澤 貴智、岡本 忍、中尾 光輝、中村 保一	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008)	2008 年 12 月 9-12 日	国内(神戸)
Gene Indexing:KazusaAnnotation による遺伝子注釈実証実験 [ポスター]	岡本 忍、中尾 光輝、藤澤 貴智、中村 保一	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008)	2008 年 12 月 9-12 日	国内(神戸)
かずさアノテーション: 持続可能な遺伝子アノテーション整備のための新型システム [ポスター]	中尾 光輝、岡本 忍、藤澤 貴智、中村 保一	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008)	2008 年 12 月 9-12 日	国内(神戸)
ライフサイエンス統合データベースポータル lifesciencedb.jp [ポスター]	河野信、坊農秀雅、全弘宇、藤枝香、川本祥子、高祖歩美、箕輪真理、三橋信孝、永井啓一、仲里猛留、西川哲夫、大野忠、小野浩雅、山口敦子、山本泰智、八塚茂、大久保公策、高木利久	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008)	2008 年 12 月 9-12 日	国内(神戸)

MeSH を用いた生物学的解釈によるマイクロアレイデータの解析の実際 [ポスター]	仲里猛留、坊農秀雅、松田秀雄、高木利久	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008)	2008 年 12 月 9-12 日	国内(神戸)
統合 TV - 動画によるバイオデータベースとウェブツールの活用法解説サイト [ポスター]	小野 浩雅、河野 信、大村 蓉子、山口 瑤子、竹若 浩一、川本 祥子、高木 利久、坊農 秀雅	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008)	2008 年 12 月 9-12 日	国内(神戸)
BodyParts3D: 解剖学概念を表現した 3 次元人体構造データベース [ポスター]	三橋 信孝、藤枝 香、田村 卓郎、川本 祥子、高木 利久、大久保 公策	第 31 回日本分子生物学会年会 第 81 回日本生化学会大会 合同大会 (BMB2008)	2008 年 12 月 9-12 日	国内(神戸)
DDBJ, PDBj, KEGG, CBRC のウェブサービスによる連携 [口頭]	片山俊明	第 46 回日本生物物理学会年会 シンポジウム「動き出したライフサイエンス統合データベース」	2008 年 12 月 3-5 日	国内(福岡)
統合データベースプロジェクトは何の役に立つのか? [口頭]	高木利久	第 46 回日本生物物理学会年会 シンポジウム「動き出したライフサイエンス統合データベース」	2008 年 12 月 3-5 日	国内(福岡)
統合 DB プロジェクトのサービスについて[口頭]	川本祥子	第 46 回日本生物物理学会年会 シンポジウム「動き出したライフサイエンス統合データベース」	2008 年 12 月 3-5 日	国内(福岡)
A Literature Management System with A Recommendation Function at Database Center for Life Science. [ポスター]	Yaunori Yamamoto and Toshihisa Takagi	The 19th International Conference on Genome Informatics	2008 年 12 月 1-3 日	国外 (Gold Coast)
Life Science Database Portal Site: lifesciencedb.jp. [ポスター]	Shin Kawano, Hidemasa Bono, Shoko Kawamoto, Kousaku Okubo and Toshihisa Takagi	The 19th International Conference on Genome Informatics	2008 年 12 月 1-3 日	国外 (Gold Coast)
Functional organization of transcript sequences as gene expression data. [ポスター]	Hidemasa Bono, Shoko Kawamoto, Kousaku Okubo and Toshihisa Takagi	The 19th International Conference on Genome Informatics	2008 年 12 月 1-3 日	国外 (Gold Coast)
低酸素刺激における比較トランスクリプトーム解析 [口頭及びポスター発表]	坊農秀雅	第 6 回がんとハイポキシア研究会	2008 年 11 月 29-30 日	国内(広島)

生命科学データの共有に向けた「統合データベースプロジェクト」の試み [ポスター]	箕輪真理、坊農秀雅、全弘宇、藤枝香、川本祥子、河野信、高祖歩美、熊谷禎洋、三橋信孝、永井啓一、仲里猛留、西川哲夫、大村蓉子、大野忠、小野浩雅、山口敦子、山口瑤子、山本泰智、八塚茂、横山貴央、大久保公策、高木利久	情報とシステム 2008 「社会のイノベーションを誘発する情報システム 08	2008 年 11 月 5 日	国内(東京)
Text-related Services for Researchers in Life Sciences [ポスター]	Yaunori Yamamoto and Toshihisa Takagi	BIBM2008	2008 年 11 月 3-5 日	国外 (Phila. USA)
データの保護と利用のための情報セキュリティ [口頭]	箕輪真理	科学研究費補助金特定領域研究「ゲノム」4 領域シンポジウム「ゲノム医学のこれから」	2008 年 10 月 14 日	国内(東京)
Practical Organization of Sequence Data as Gene Expression Data [ポスター]	Hidemasa Bono, Shoko Kawamoto, Kousaku Okubo and Toshihisa Takagi	The 8th Cold Spring Harbor Laboratory/Wellcome Trust conference on Genome Informatics	2008 年 9 月 10-14 日	国外 (Hinxton UK)
Japan started data-sharing center for publicly-funded biomedical science[ポスター]	Kousaku Okubo, Shoko Kawamoto, Hidemasa Bono and Toshihisa Takagi	The 8th Cold Spring Harbor Laboratory/Wellcome Trust conference on Genome Informatics	2008 年 9 月 10-14 日	国外 (Hinxton UK)
ライフサイエンス統合データベースセンターにおける言語処理関係の活動 [口頭]	山本 泰智、大久保 公策、川本 祥子、全 弘宇、高木 利久	第 38 回 人工知能学会分子生物情報研究会	2008 年 8 月 28 日	国内(東京)
KazusaAnnotation Suite: 植物関連生物ゲノム情報統合の試み [口頭]	岡本 忍、中尾 光輝、藤澤貴智、中村 保一	日本進化学会第 10 回大会	2008 年 8 月 22-24 日	国内(東京)
統合 TV : 動画によるデータベースの使い方講座 [口頭]	河野 信	日本進化学会第 10 回大会	2008 年 8 月 22-24 日	国内(東京)
統合データベースプロジェクトとライフサイエンス統合データベースセンター [口頭]	坊農 秀雅	日本進化学会第 10 回大会	2008 年 8 月 22-24 日	国内(東京)

ライフサイエンス統合データベースプロジェクト [口頭]	坊農秀雅、全弘宇、藤枝香、川本祥子、河野信、高祖歩美、峰崎雄一、箕輪真理、三橋信孝、永井啓一、仲里猛留、西川哲夫、大野忠、小野浩雅、山口敦子、山本泰智、八塚茂、大久保公策、高木利久	第 28 回日本糖質学会年会	2008 年 8 月 18-20 日	国内(つくば)
統合データベースセンターの紹介 [ポスター]	河野 信、坊農 秀雅、全 弘宇、藤枝 香、川本 祥子、高祖 歩美、峰崎 雄一、箕輪 真理、三橋 信孝、永井啓一、仲里 猛留、西川 哲夫、大野 忠、小野 浩雅、山口 敦子、山本 泰智、八塚 茂、大久保 公策、高木利久	第 8 回日本分子生物学会春季シンポジウム	2008 年 5 月 26-27 日	国内(札幌)
平成 19 年度				
統合データベースセンターの紹介	河野 信	第 37 回人工知能学会分子生物情報研究会(SIG-MBI)・第 9 回オープンバイオ研究会	2008 年 3 月 7-8 日	国内(能美)
ライフサイエンス統合データベースプロジェクト	高木利久	「生命をはかる」研究会	2008 年 2 月 18 日	国内
TogoTV - a broadcast station of tutorial movies about bioinformatics resources	S. Kawano, H. Ono, H. Bono, S. Kawamoto, T. Takagi	日本バイオインフォマティクス学会年会(JSBi2007)	2007 年 12 月 17-19 日	国内(東京)
MeSH terms を用いた生物学的機能付与による遺伝子群の解析手法の開発	仲里猛留、滝中徹、坊農秀雅、麻生川稔、松田秀雄	第 30 回日本分子生物学会年会第 80 回日本生化学会大会 合同大会(BMB2007)	2007 年 12 月 11-15 日	国内(横浜)
ライフサイエンス統合データベースプロジェクトの課題と成果	川本祥子、坊農秀雅、池田佳代子、河野信、三橋信孝、峰崎雄一、永井啓一、西川哲夫、小野 浩雅、山口敦子、山本泰智、大久保公策、高木利久	第 30 回日本分子生物学会年会第 80 回日本生化学会大会 合同大会(BMB2007)	2007 年 12 月 11-15 日	国内(横浜)

ライフサイエンス統合データベースセンターにおけるデータベースポータル構築	河野信、小野浩雅、仲里猛留、坊農秀雅、岡本忍、中村保一、川本祥子、大久保公策、高木利久	第30回日本分子生物学会年会第80回日本生化学会大会 合同大会 (BMB2007)	2007 年 12 月 11-15 日	国内 (横浜)
次元人体モデルのセグメンテーションによる解剖学用語の形式表現	三橋信孝、藤枝香、川本祥子、高木利久、大久保公策	第30回日本分子生物学会年会第80回日本生化学会大会 合同大会 (BMB2007)	2007 年 12 月 11-15 日	国内 (横浜)
遺伝子発現統合データベースの開発	有川浩司、飯塚高康、渡邊康司、小笠原理、大久保公策	第30回日本分子生物学会年会第80回日本生化学会大会 合同大会 (BMB2007)	2007 年 12 月 11-15 日	国内 (横浜)
分子生物学の研究動向の俯瞰を目的としたデータバンク (INSDC, GEO) 目次の開発	小笠原理、飯塚高康、渡邊康司、有川浩司、大久保公策	第30回日本分子生物学会年会第80回日本生化学会大会 合同大会 (BMB2007)	2007 年 12 月 11-15 日	国内 (横浜)
ゲノムセントラルな遺伝子発現情報融通システム YUZ (柚子)	坊農秀雅	第30回日本分子生物学会年会第80回日本生化学会大会 合同大会 (BMB2007)	2007 年 12 月 11-15 日	国内 (横浜)
ライフサイエンス統合データベース基盤整備のためのデータベースポータル構築: WIINGpro	黒田 雅子, 小池 俊行, 安田 絵美, 佐藤 早苗, 藤田 晶子, 河村 明子, 川越 千晴, 坊農 秀雅, 川島 秀一, 谷口 丈晃, 川本 祥子, 高木利久	第30回日本分子生物学会年会第80回日本生化学会大会 合同大会 (BMB2007)	2007 年 12 月 11-15 日	国内 (横浜)
統合 TV によるがんとハイポキシア研究の推進	坊農秀雅	第 5 回 がんとハイポキシア研究会	2007 年 12 月 1 日	国内 (千葉)
YUZ: an environment for integration AND ANALYSIS of gene regulatory networks	H. Bono, S. Kawano, S. Kawamoto, and T. Takagi	GENOME INFORMATICS	2007 年 11 月 1-5	国外 (New York (USA))
An approach to decipher gene regulatory networks from the federation of databases in life science	H. Bono, S. Kawano, S. Kawamoto and T. Takagi	FUNCTIONAL GENOMICS & SYSTEMS BIOLOGY	2007 年 10 月 10-13 日	国外 (Hinxton (UK))
ゲノム研究のためのオントロジーとテキストマイニング [口頭、招待講演]	高木利久	第 16 回セマンティックウェブとオントロジー研究会	2007 年 7 月 23 日	国内
平成 18 年度				

使い倒し系バイオインフォマティクスによる知のめぐりのよい生物学的研究のすすめ	坊農秀雅	お茶の水女子大学「魅力ある大学院教育」第6回バイオインフォマティクスへの招待	2007年3月16日	国内
統合DBの構築に必要な情報技術	高木 利久	情報とシステム 2007	2007年3月1日	国内
ライフサイエンスのデータベースの現状と課題	大久保公策	情報とシステム 2007	2007年3月1日	国内
データベースから見たライフサイエンスプロジェクト	高木 利久	日本分子生物学会 2006 フォーラムシンポジウム[プロジェクト型研究時代の生命科学の課題]	2006年12月8日	国内
知的生産性向上のための情報処理	大久保 公策	日本分子生物学会 2006 フォーラムシンポジウム[プロジェクト型研究時代の生命科学の課題]	2006年12月8日	国内
ライフサイエンスDB その歴史とわが国の現状と課題	高木 利久	日本分子生物学会 2006 フォーラム バイオテクノロジーセミナー	2006年12月8日	国内
オントロジーや辞書は役に立つのか	大久保 公策	日本分子生物学会 2006 フォーラム バイテクノロジーセミナー	2006年12月8日	国内
ライフサイエンスにおけるゲノム情報の高度利用に向けた生命知識の構造化	大久保 公策	知の構造化ワークショップ 知の構造化ツールは、新しいサイエンスを開くのか	2006年12月4日	国内

2. 学会誌・雑誌等における論文掲載

掲載した論文（発表題目）	発表者氏名	発表した場所（学会誌・雑誌等名）	発表した時期	国内・外の別
平成 22 年度				
改変 OK? 許可は必要? 何を明示すればいい? いまさら聞けないクリエイティブコモンズ	永井 幸輔、中尾 光輝、平田 雄一、渡辺 智暁	技術評論社 59 巻、52-60	2010 年 10 月 23 日	国内

BioRuby: bioinformatics software for the Ruby programming language.	Naohisa Goto, Pjotr Prins, Mitsuteru Nakao, Raoul Bonnal, Jan Aerts and Toshiaki Katayama	BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 20 2010, pages 2617-2619	published: 2010 年 8 月 25 日	国外
The DBCLS BioHackathon: standardization and interoperability for bioinformatics web services and workflows.	Toshiaki Katayama, Kazuharu Arakawa, Mitsuteru Nakao et al.	Journal of Biomedical Semantics 2010, Vol. 1, 1-8	published: 2010 年 8 月 21 日	国外
文部科学省「統合データベースプロジェクト」と PubMed を中心とした関連データベース	仲里 猛留、坊農 秀雅	情報の科学と技術, 60 巻 7 号, 265-271	2010 年 7 月	国内
「セマンティック Web プログラミング」Toby Segaran, Colin Evans, Jamie Taylor 著	大向 一輝、加藤 文彦、中尾 光輝、山本 泰智 監訳、玉川 竜司 訳	オライリージャパン	2010 年 6 月	国内
TogoWS: integrated SOAP and REST APIs for interoperable bioinformatics Web services	Toshiaki Katayama, Mitsuteru Nakao, and Toshihisa Takagi	Nucl. Acids Res. (2010) 38 (suppl 2): W706-W711.	published online: 2010 年 5 月 14 日	国外
平成 21 年度				
GuideLink: A Corpus Annotation System that Integrates the Management of Annotation Guidelines	Kenta Oouchida, Jin-Dong Kim, Toshihisa Takagi and Jun'ichi Tsujii	Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation	2009 年 12 月 3-5 日	国外
CyanoBase: the cyanobacteria genome database update 2010	Mitsuteru Nakao, Shinobu Okamoto, Mitsuyo Kohara, Tsunakazu Fujishiro, Takatomo Fujisawa, Shusei Sato, Satoshi Tabata, Takakazu Kaneko and Yasukazu Nakamura	Nucleic Acids Research, 2010, Vol. 38, Database issue D379-D381	published online on 2009 年 10 月 30 日	国外 (Oxford University Press)
生命科学分野におけるテキストマイニング	山本 泰智	情報処理学会誌 2009, Vol. 50, No. 9, 853-858	2009 年 9 月	国内

生命情報学が直面する大規模ゲノムデータ時代の課題：1. 分散データの統合とセマンティックWeb	中尾 光輝、片山 俊明	情報処理学会誌 2009, Vol. 50, No. 9, 836-844	2009 年 9 月	国内
生命情報学が直面する大規模ゲノムデータ時代の課題：0. 編集にあたって	片山 俊明	情報処理学会誌 2009, Vol. 50, No. 9, 833-835	2009 年 9 月	国内
Gendoo: Functional profiling of gene and disease features using MeSH vocabulary	Takeru Nakazato, Hidemasa Bono, Hideo Matsuda and Toshihisa Takagi	Nucleic Acids Research, 2009, Vol. 37, Web Server issue W166-W169	2009 年 6 月	国外 (Oxford University Press)
ライフサイエンス統合データベースセンターと統合データベースプロジェクト	坊農 秀雅	情報の科学と技術, 59 巻 4 号, 165-170	2009 年 4 月	国内
平成 20 年度				
BodyParts3D: 3D structure database for anatomical concepts	Nobutaka Mitsuhashi, Kaori Fujieda, Takuro Tamura, Shoko Kawamoto, Toshihisa Takagi and Kousaku Okubo	Nucleic Acids Research, 2009, Vol. 37, Database issue D782-D785	2008 年 10 月 3 日	国外
蛋白質核酸酵素・連載 ライフサイエンス分野の統合データベース 「『蛋白質 核酸 酵素』バックナンバーの全文検索サービス公開にあたって」	川本祥子	蛋白質核酸酵素 Vol. 53, No. 9, pp. 1200-1205	2008 年 7 月	国内
ライフサイエンス分野のデータベースの現状と統合データベースプロジェクト	坊農秀雅、高木利久	日本化学会情報化部会誌 特集 Vol. 26, No. 1, pp. 5-7	2008 年 4 月	国内
平成 19 年度				
利用の立場からのコメント	高木利久	蛋白質核酸酵素(共立出版) 53 巻、5 号、686-691	2008 年	国内
統合データベースプロジェクトのサービスとその利用法	川本祥子、坊農秀雅	蛋白質核酸酵素(共立出版) 53 巻、3 号 281-287、	2008 年	国内
統合データベースセンターがめざすもの	高木利久	蛋白質核酸酵素(共立出版) 52 巻、11 号、1388-1393	2007 年	国内

生命科学系データベース統合化の背景	大久保公策	蛋白質核酸酵素(共立出版) 52 巻、9 号、1027-1031	2007 年	国内
OReFiL: an online resource finder for life sciences	Y. Yamamoto and T. Takagi	BMC Bioinformatics, 2007, 8, 287	2007 年	国外
BioCompass: A novel functional inference tool that utilizes MeSH hierarchy to analyze groups of genes	T. Nakazato, T. Takinaka, H. Mizuguchi, H. Matsuda, H. Bono and M. Asogawa	<i>In Silico</i> Biology, 2007, 8, 6	2007 年	国外
平成 18 年度				
生命科学データベースの現状と課題	大久保 公策	科学 77 巻、4 号、364-369	2007 年	国内

「統合データベースプロジェクト」事後評価参考資料
プロジェクト成果の詳細説明資料

大学共同利用機関法人 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

目 次

1. プロジェクトの目的	2
2. 戦略立案・実行評価	3
2. 1 戦略立案・実行評価の実施計画	3
2. 2 戦略立案・実行評価の実施内容	3
3. 統合データベース開発	21
3. 1 共通基盤技術開発	21
3. 1. 1 共通基盤技術開発の実施計画	21
3. 1. 2 共通基盤技術開発の実施内容	21
3. 2 ヒト統合 DB の開発・運用	30
3. 2. 1 ヒト統合 DB の開発・運用の実施計画	30
3. 2. 2 ヒト統合 DB の開発・運用の実施内容	31
3. 3 モデル生物・産業応用生物統合 DB の開発・運用	43
3. 3. 1 モデル生物・産業応用生物統合 DB の開発・運用の実施計画	43
3. 3. 2 モデル生物・産業応用生物統合 DB の開発・運用の実施内容	43
4. 統合データベース支援	44
4. 1 ポータル整備・運用、広報、普及啓発	44
4. 1. 1 ポータル整備・運用、広報、普及啓発の実施計画	44
4. 1. 2 ポータル整備・運用、広報、普及啓発の実施内容	45
(1) 統合 DB プロジェクトポータルサイトの整備・運用	45
(2) 広報・普及啓発	49
4. 2 データベース受入	52
4. 2. 1 データベース受入の実施計画	52
4. 2. 2 データベース受入の実施内容	53
5. プロジェクトの総合的推進（情報・システム研究機構）	64
5. 1 プロジェクトの総合的推進の実施計画	64
5. 2 プロジェクトの総合的推進の実施内容	64
(1) 研究運営委員会及び研究運営委員会作業部会	64
(2) プロジェクト実施機関並びに他省関連機関等との連絡会等の開催実績	70
(3) 外部有識者からの意見聴取	74
5. 3 その他シンポジウム、セミナー等	75
(1) シンポジウム	75
(2) 統合データベース講習会	79
(3) BioHackathon	92
6. 新聞発表等	100
7. 実施体制	102

1. プロジェクトの目的

ライフサイエンスやバイオ産業に従事する研究者や技術者がいわゆるゲノムプロジェクト・ポストゲノムプロジェクトの成果や多様な DB や解析ツールをストレスなく利用してより高度な研究開発が効率よくできる環境（統合 DB）を実現することを目的とする。このため、情報・システム研究機構、科学技術振興機構、産業技術総合研究所生命情報工学研究ユーザ、かずさ DNA 研究所、九州大学、奈良先端科学技術大学、東京大学、お茶の水女子大学、長浜バイオ大学は共同して、以下の 3 つの業務を行う。

（1）戦略立案・実行評価

ライフサイエンス（基礎生物学および応用生物学（医・薬・農））、情報処理技術、ライフサイエンス DB の 3 分野の専門家による研究運営委員会を組織し、各々の分野の動向に即した我が国の DB 整備戦略を立案する。また、3 分野の動向に関する情報の網羅的収集・分析を行い、戦略立案や実行評価を支援するチームを組織する。

（2）統合データベース開発

ライフサイエンス、バイオ産業にかかわる情報へのアクセスと利用に関する格段の利便性向上とそれによる研究開発の飛躍的な効率化と質的向上を目指して、ヒト統合化 DB およびモデル生物・産業応用生物統合 DB の開発・運用と、それに必要な統合 DB 構築技術、DB の標準化、文献等からの知識獲得技術等の共通基盤技術開発を実施する。

（3）統合データベース支援

個々の公開 DB や利用可能な DB 解析ツールなどを網羅的に調査し、提示するポータルサイトを構築する。運用にあたって専門家及び利用者の意見集約システムを開発し、意見をフィードバックして運用する。また日本語による研究情報の流通促進に注力する。プロジェクトデータベースを受入れ、相互運用可能な DB へ変換した上で公開し、運用する。また、広報・普及啓発の活動を行う。人材育成は実績のある機関と連携し階層別に、アノテータ・キュレータ教育、DB 構築者養成、及び DB 高度利用者養成を行う。

以上は、統合データベースプロジェクトの中核機関を構成する情報・システム研究機構以下の上記 9 機関が共同して実行すべき実施目的であるが、以下では、情報・システム研究機構の実施内容について記載する。

2. 戦略立案・実行評価

2. 1 戦略立案・実行評価の実施計画

ライフサイエンス（基礎生物学および応用生物学（医・薬・農））、情報処理技術、ライフサイエンス DB の 3 分野の専門家による研究運営委員会を組織し、各々の分野の動向に即した我が国の DB 整備戦略を立案する。上記委員会に対し、3 分野の動向に関する情報の網羅的収集・分析を行い、戦略立案や実行評価を支援するチームを組織する。このような体制のもとで、戦略立案、実行評価の支援業務とそのための情報収集・分析だけでなく、関係府省、利用者、産業界、学会・出版社、国内外の研究機関等関係者との連絡調整等の活動もあわせて行う。

具体的には、統合データベースプロジェクトの目標や年次計画の具体化、国内外の動向調査も踏まえたデータベース統合化のあるべき姿の検討を進める。併せてデータ共有化や公開促進の観点から、国家プロジェクト成果に関する権利関係や著作権、個人情報保護に関する調査を行い、プロジェクト成果であるデータの提供を促進するための取り組みや提供いただいたデータの運用方針の立案、個人情報に関わるデータの公開方針の策定などを行う。さらに、総説誌など日本語文献の公開に関する取り組みも進める。また、世界的に分散しているデータの統合利用に関する世界的なハブとしての役割を務める。

実行評価については、想定されるユースドメインから、有用な回答が期待できる有識者を選定したユーザ評価や、出展する学会等の展示会でのアンケートにより、ユーザの生の声の把握に努め、サービス内容のユースドメインへの適合に努める。併せて、プロジェクト成果のサービスサイトへのアクセスログを持続的に収集する体制を整え、ユーザからの利用状況の把握に努める。

2. 2 戦略立案・実行評価の実施内容

2. 2. 1 戦略立案

平成 18 年度は、ライフサイエンス、知識情報処理、ライフサイエンスデータベース (DB) の 3 分野の専門家による研究運営委員会を組織し、各々の分野の動向に即した DB 整備戦略について議論する場を構築し、上記委員会に対し、3 分野の動向に関する俯瞰データを提示し、適宜与えられる調査課題に対し答えることのできる体制（統合 DB 整備戦略作業部会）を構築した。この体制をベースに、平成 19 年度以降のプロジェクトを想定した戦略立案の基盤となる以下の情報の収集・分析を行った。

- (1) ゲノム注釈とデータベース間の連携における課題
- (2) 国内外の DB の俯瞰と質的量的比較
- (3) ライフサイエンス分野の研究の俯瞰調査
- (4) 検索アルゴリズムを含めた知識情報技術の動向調査
- (5) 臨床情報や医療統計の現状調査

(1) のゲノム注釈とデータベース間の連携における課題については、代表的モデル研究植物であり、全ゲノム塩基配列が決定済みである「イネ」ならびに「シロイヌナズナ」

のゲノムアノテーション型公開データベースを題材に、各データベースの基本項目を調査し、それぞれのデータベース間の連携と課題の整理を実施した。将来の統合化データベースの作成にあたっては、現在、利用頻度の高い個々のデータベースが保持している有用な情報を活かしつつ、齟齬を解消し連係させる形での統合化を考えていくべきであるとの結論を得た。（２）の国内外のDBの俯瞰と質的量的比較については、科学技術連携施策群の生命科学データベース統合に関する調査研究とJST-DB（WING）の調査をベースに、主に分子データに関する網羅性の高いデータベースカタログを生成した。さらに、DNA配列データバンクや遺伝子発現データベースを対象に、一般研究者でも分子レベルの生物学研究の現状を容易に俯瞰し、プロジェクト単位あるいは研究対象、生物種単位でデータをダウンロードすることが可能な目次システムを構築した。（３）のライフサイエンス分野の研究の俯瞰調査については、国内の研究を俯瞰するための情報源として各種学会の過去の抄録を統合し、施設別やテーマ別の再編成が可能なデータベース構築を目的として、分子生物学会８年分の書誌事項に加え一部要旨を電子化し、施設名称など基本的な用語の統一を行い、要旨の検索、ソート、可視化が可能なシステムを開発し、分野動向を逐次的に調査把握していくことを可能とした。（４）の検索アルゴリズムを含めた知識情報技術の動向調査については、次世代の生物情報データベース統合に必要な知識情報技術として、検索システム、データマイニング、Web 2.0 およびグリッドコンピューティングに焦点を絞り、聞き取り調査や文献調査によって動向を調べた。併せて、統合データベースに必要とされる計算機資源（CPU 数、ディスク容量など）の調査を行った。（５）の臨床情報や医療統計の現状調査については、臨床情報、コホート研究および医療統計学分野の現状調査と課題発掘を行った。これらの調査から、医学データ活用における課題として、病名の標準化、前向きコホート研究の推進、人類遺伝学基盤の充実が重要との結果を得た。病名の標準化に関しては、現状は死因統計などの主として保険行政統計用のものか保険診療用のものしかなく、臨床研究向けには使いにくく、抜けもあり、また必ずしも真の病名が記載されない、といった問題があり、また、前向きコホート研究については、既存のカルテの活用（後向き研究）は、検査値などを除くと難しく、しっかりデザインされた一定規模の前向き研究によって初めて有効なデータが得られることが分かった。また、米国では家系を集めるプロジェクトにも多額の投資がなされているのに対して、日本では家系データが軽視される傾向にあり、これは日本における人類遺伝学基盤の不十分さに起因することが分かった。

以上の調査活動の結果を踏まえ、さらに研究運営委員会等での議論を通じ、以下に示す統合データベース構築の際の目指すべき課題、取り組みの際の基本的考えをまとめ、これに基づく統合データベース構築の基本的アウトラインをまとめた。

- ・DB 構築者ではなく利用者の思考や意思決定を支援するDBを構築する。
- ・利用者の興味、知識に応じて必要な情報、判断材料をもれなく提示する。
- ・複数DBをつなぎ異種データ・知識の関係が俯瞰（仮説生成）できるようにする。
- ・いろいろなツールを簡便に組み合わせて解析（知識発見）できるようにする。
- ・できれば、上記のことが日本語で行えるようにする。
- ・DB化（構造化）されないもの（論文（テキスト、ポンチ絵、画像）、特許、教科書、報告書、解説記事、など）もうまく扱えるようにする。

- ・DB構築に最先端の研究開発は必要だが、あくまでも利用者の利便性向上のためのサービス事業であることを認識する。

平成19年度は、引き続きライフサイエンス、情報処理技術、ライフサイエンスDBの各分野の動向に即した我が国のDB整備戦略を立案するために、上記3分野の専門家、並びに関連府省の専門家、産業界の代表、分担機関の研究代表者を加えた研究運営委員会を組織した。また、研究運営委員会の下に研究運営委員会作業部会を組織し、個別具体的な課題について詳細な議論を行うこととした。また、上記研究運営委員会、作業部会に対して、3分野の動向に関する情報の網羅的な収集と分析を行い、戦略立案や実行評価を支援するためのチームを、大学利用機関法人 情報・システム機構（ROIS）内に設立したライフサイエンス統合データベースセンター（DBCLS）内に組織した。

このような体制の中で、本年度は特に、まずDB統合化のスキームの検討とこれに基づくサービスの実現方法を検討し、さらに統合化の年次計画を策定した。次に、これをベースに補完課題の公募方針およびこれに関連したデータベースの受入れ戦略を検討した。さらに、データベース提供を促進するための契約・著作権関係並びに個人情報関係の検討、日本語総説誌を中心とした日本語文献の公開のための方策の検討を行った。年度の後半には、本プロジェクト終了後の我が国のライフサイエンス統合DBの在り方に関する議論も行った。

統合化のスキームについては、主に検索のタイプから、DBカタログ、横断検索及び統合検索の三つのケースについて検討した。DBカタログはライフサイエンス分野のDBのメタデータを集め、これに注釈づけを施したDB情報に網羅的にアクセスできるものである。横断検索は、クロージングなどにより対象となるDBのインデックス情報を集め、これによりライフサイエンス分野のDBや文献の横断的なキーワード検索を実現するものである。また、統合検索は、外部にあるDBからAPIにより入手したデータ、あるいは統合データベースセンター内に受け入れたDBを処理して得たデータに対して検索を行うもので、横断検索より、より高度かつ詳細な情報を獲得できるものである。統合化の年次計画としては、DBカタログについては、平成19年度中に、我が国に存在するとされるゲノム、ポストゲノム関係の250のDBのほとんどを網羅することにした。また、横断検索については、平成19年度中に20程度の代表的なDBを掲載し、引き続き平成20年度中に我が国の関連DBの大部分と海外の主要DBを掲載することとした。さらに、統合検索については、国内DBの受入れと併せて、順次対象となるDB数を増やしていくこととした(図2.1参照)。

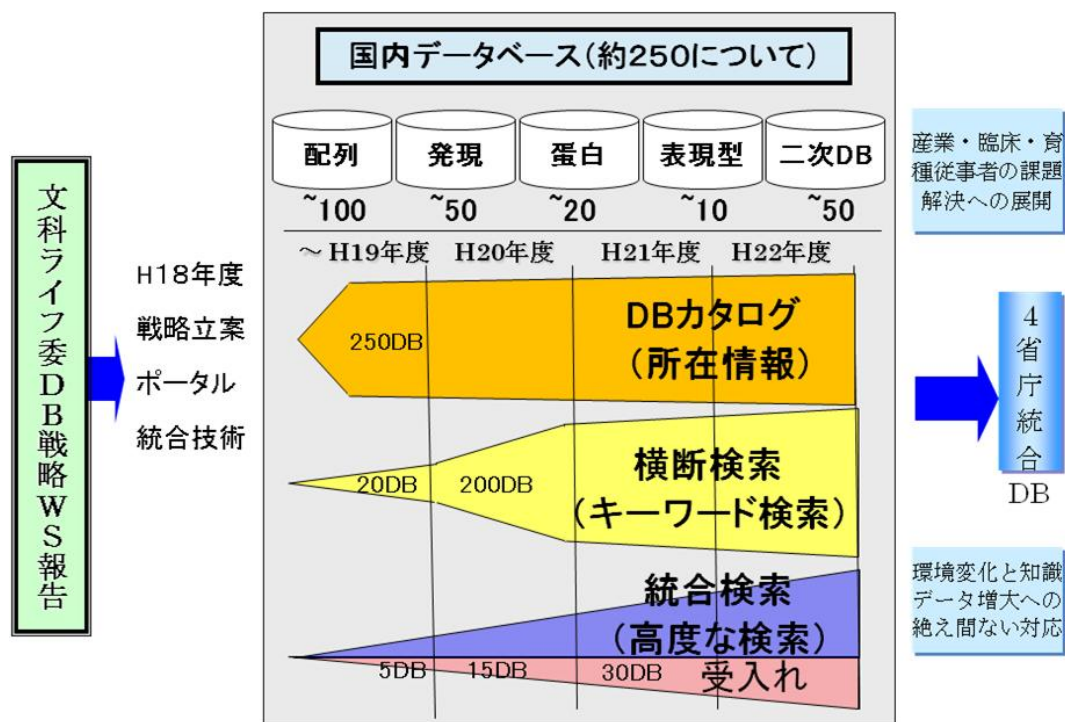


図 2.1 DB統合化の年次計画

補完課題については、上記作業部会で公募内容の検討を行い、大型プロジェクトや大量データ産生機関を対象として、ユーザの標準化仕様に従い統合化するための諸作業を課題化すると結論が得られた。この内容をベースに文部科学省が公募を行い、審査の結果、理化学研究所、産業技術総合研究所糖鎖工学センター、国立遺伝学研究所および九州工業大学の提案が採択された。この結果、図2.3に示すように、最終的に統合データベースプロジェクトは、中核機関、三つの分担機関、及び四つの補完課題実施機関で行うこととなった。

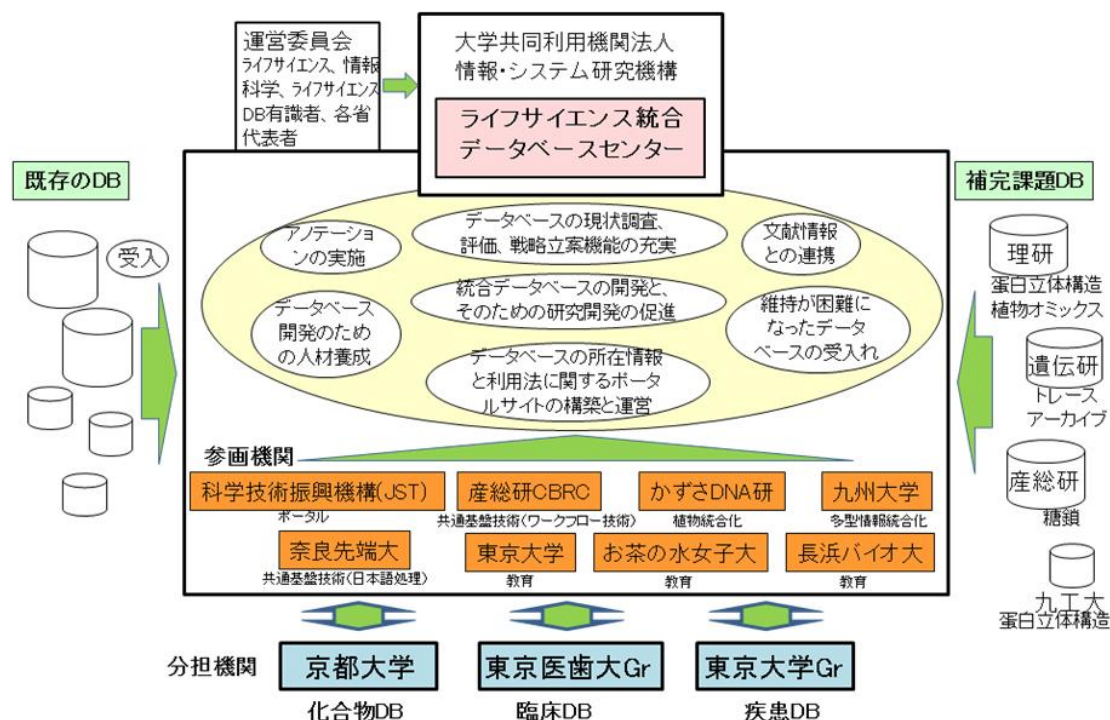


図 2.2 統合データベースプロジェクトの体制図

契約・著作権並びに個人情報関係については、まず前者について、国家プロジェクトの契約内容及びその基礎となる法律に関して調査を行った。また後者に関しては、国内の医療情報を取り扱う代表的プロジェクトに関して、個人情報の取り扱いに関する現状調査を行った。こうした検討結果を踏まえ、今後のライフサイエンス分野の国家プロジェクトからデータおよびデータベースの提供を受けるための方策について検討し、文部科学省との密接な連携の上で、プロジェクトの公募要領にデータおよびデータベース提供条項を盛り込むことができた。その結果、平成20年度以降のライフサイエンス課管轄の委託事業の公募要領には、図2.3に示すような条項が盛り込まれるようになった。また、国内外データベースの利用条件、免責事項、プライバシーポリシー等の整理、分析に基づき、本プロジェクトが運営・管理するサイトのサイトポリシーを作成した。また、日本語文献の公開については、特に日本語総説誌の公開のためのモデルの検討を行い、こうした検討結果をベースに出版社と交渉し、日本語総説誌公開ための契約に至った。また、日本語資料の公開の一環として総合科学技術会議、文科省等のライフサイエンス関連公開資料の収集も行った。なお、上記の法律、契約に関わる条項については、当該分野に詳しい弁護士のコンサルテーションを得ながら進めた。

(7) 統合データベースプロジェクトへのデータ提供

本プログラムはライフサイエンス分野のデータ産生を伴うプロジェクトであるため、本プログラムの実施者は、ライフサイエンス分野における研究に資するよう、本プログラムの実施の結果産生したデータ又はデータベースを、一般に利用可能な形で公開するべく努めるものとします。そのため、本プログラムの実施者は、毎事業年度にまとめる成果報告書の提出に合わせて、論文発表等で公表された成果に関わる生データの複製物、又は本プログラムで構築した公開用データベースの複製物を、ライフサイエンス統合データベースセンター※又は文部科学省の指定する機関に提供するものとします。提供された複製物については、ライフサイエンス分野のデータベース整備（統合データベース構築）に利用するため、これら複製物に関わる知的財産権を非独占的に複製・改変その他必要な形で利用できるものとします。なお、本プログラムの実施者は、複製物の提供を受けた機関の求めに応じ、複製物を利用するに当たって必要となる情報提供に協力するものとします。

※ライフサイエンス統合データベースセンター (<http://dbcls.rois.ac.jp/index.html>)

生命情報のデータ又はデータベースはライフサイエンス研究を支える基盤であり、その整備を推進するため、文部科学省では「統合データベースプロジェクト」を実施しています (<http://lifesciencedb.mext.go.jp/>)。ライフサイエンス統合データベースセンターは大学共同利用機関法人情報・システム研究機構に属し、当該プロジェクトを実施するための中核機関です。本センターでは、我が国のライフサイエンス関係データベースの利便性の向上を図るため、データベース整備戦略の立案・評価、データベース統合化の基盤技術開発、及び統合化データベース開発等を行い、ライフサイエンスデータベースの統合化を推進しています。

なお、本センターは、ライフサイエンスの基盤となるデータの保存と配布の役割を担うデータデポジトリのサイトであり、データ提供者と競合するような研究活動を行う機関ではありません。

(平成20年度以降の公募)

- ➡ データ/データベースの公開への努力
- ➡ データ/データベースの提供 (DBCLS or 指定機関へ)
- ➡ 非独占的な複製・改編権の担保
- ➡ 必要な情報の提供義務
- ➡ DBCLSと統合データベースプロジェクトの説明
- ➡ 提供者と競合しないことの説明

図 2.3 ライフサイエンス課委託業務公募要領のデータ提供条項

また、国内外の研究機関等関係者との連絡調整の一環として、平成20年2月11日～15日の5日間、ウェブサービスの統合を目指した現状調査・情報交換および技術開発を目的として、産業技術総合研究所生命情報工学研究ユーザの協力のもと、国際ワークショップ B

ioHackathon 2008 を開催した。ここでは、主にライフサイエンスの主要なウェブサービスプロバイダ、クライアント開発者、オープンバイオ開発者などから参加者を募った。その結果、海外から34名、国内から37名の参加者があり、トピックごとに熱心な議論が行われた。主なトピックはオントロジー、データ交換の標準化、非同期通信、大規模データ、セキュリティ、オープンバイオ、RESTvsSOAP、ワークフロー、BioMobyなどであった。

平成 20 年度は、年度前半に行われた本プロジェクトの中間評価結果や文部科学省ライフサイエンス情報基盤整備作業部会の報告書等の内容に対応すべく、これまで進めてきた統合データベースプロジェクトの目標や年次計画の具体化、国内外の動向調査をベースに、さらには、研究運営委員会の委員の意見も踏まえ、データベース統合化のあるべき姿の検討を進めた。併せてデータ共有化や公開促進の観点から、著作権、個人情報保護に関する調査、検討、並びに今後のデータ産出の中心となることが想定される次世代サーバの生み出すデータに関する調査を行い、特にヒトゲノム多型データの共有に関わる指針策定を行った。日本語文献の公開については、コンテンツを充実すべく学会要旨を提供いただく学協会との交渉を行い、コンテンツを充実した。また、国内外の研究機関等関係者との連絡調整という観点から、平成 19 年度に引き続き、世界的に分散しているデータの統合利用に関する国際ワークショップである BioHackathon を開催し（参加者は海外 30 名、国内 28 名）、国際的な標準化を実現する場、及び国内の関連研究者が海外の第一線の研究者、開発者と共同作業を通じて緊密な関係を築く場としての役割を果たした。その他、日本学術振興会ゲノムテクノロジー164 委員会と共催したシンポジウムでのパネルディスカッション等を通じた産業界及び研究者コミュニティの意見収集や、ヨーロッパの EMBL-EBI 主催の生物情報基盤構築プロジェクト「ELIXIR」の調査等によるニーズと動向把握を行った。

まず、データベース統合化のあるべき姿の検討については、本プロジェクトの周辺においても、平成 20 年度には、平成 20 年 7 月付で報告書が出された本プロジェクトの中間評価、平成 21 年度概算要求時の総合科学技術会議のヒアリング、及び文部科学省ライフサイエンス委員会の情報基盤整備作業部会（平成 20 年 10 月～平成 21 年 1 月）、さらには総合科学技術会議ライフサイエンス PT の統合 DB タスクフォース（平成 20 年 12 月～平成 21 年 4 月）の場において、我が国の関連分野の専門家による議論、検討が行われた。これらにおける共通点は、本プロジェクト終了後も、本プロジェクトの成果である統合データベースを恒久的に維持、発展させる必要があること、また、そのための具体策として、本プロジェクトと科学技術振興機構（JST）の BIRD 事業を一本化するための組織を JST 内に構築するべきである、との提言であった。これらの提言を受け、本プロジェクトでも松原謙一研究運営委員長のリーダーシップの下、運営委員の先生方から意見を収集して、誰でも理解できるライフサイエンス統合データベースのアウトラインを提示することになった。また、これと並行してプロジェクト側でも統合化のビジョンということで、改めて統合化の全体像とステップについて検討した。

運営委員の方々の意見として出てきたのは、まずデータベース統合化の理念として、
1. 公的資金を用いて得られたデータ、データベースは広く共有して、その活用を図るべきであること。2. データベース統合により、生命科学の新たな展開とバイオ産業の

国際的競争力向上に貢献するべきであること。3. 統合データベースは重要な基盤であり、省庁に囚われない存在として国が継続的支援すべきである、との3点が重要であり、また、これを実現するためには、公的資金で得られたデータをコミュニティで早期に共有するためのルールを作成すること、及び公的資金で得られたデータを保全、管理し育成するための中立的機関を設置する必要があるとの指摘である。次に統合化は、アカデミアの知識を集約し、日本で生産するデータの価値を最大化することにより、日本のユーザ、更には世界のユーザに貢献することを目標にするべきであり、具体的には、制約のないデータの共有、操作感の統一及び知識発見支援を提供することによる生命科学のブレークスルー、イノベーション、新たな研究スタイルの実現を目指すことが究極の目標になる。

統合データベースの持つべき機能として挙げられたなかの代表例は、サービスの側面として、

- ・個々のデータベース、解析ツールへの案内機能
- ・複数のデータベースの統合化による付加価値の向上
- ・標準パラメータ、フォーマット、オントロジーなどの共通基盤の提供
- ・維持困難なデータベースの保存によるサービスの継続
- ・一次データの登録、編集、標準化、品質管理、閲覧サービス
- ・企業アクセスも視野に入れたセキュアな環境の提供

基盤的側面として、

- ・使い勝手およびデータの付加価値向上のための研究開発機能
- ・データベース統合利用促進のためのファンディング機能
- ・世界的情報共有化のための国際的な連携機能
- ・データベースの構築や統合化のための人材育成機能と教材提供

などである。

また、今後の進め方としては、上記の提言の趣旨に従ってプロジェクト終了後の体制の構築を検討していくと同時に、関係4省の連携など全体としての統合を実現するための組織の重要性、及び統合データベースの整備と併せて基盤データベースの維持や個別データベース構築への支援にも留意する必要があるとの意見が寄せられた。その他、データの所有権、公開原則あるいはデータベースへの登録義務といった考え方を、省庁を越えて統一することや、個人情報保護、著作権、営利目的利用に関する考え方を整理して、関連法令、規則等の整備を推進する必要があるとの指摘があった。

以上のような統合データベースのアウトラインに関して寄せられた意見を受けて、中核機関でも、改めてデータベース統合推進に関する基本的考え方、及びデータベース統合化の全体像とそれに至るステップを整理した。基本的考え方は、

- ・データの共有化、価値の最大化が目的であり、必ずしも一カ所集中型を目指すものではない。
- ・データを統合化し、価値を高めるには、知識の集約、デジタル化が必要であり、これに積極的に取り組む。
- ・技術的には可能でも、制度的その他の理由により、共有化できていないものは、その障害の解消に積極的に取り組む。

- ・DB利用者だけでなく、DB構築提供者にも喜ばれる仕組みを作る。
- ・サービスの安定的、永続的な維持更新。

である。図2.4にデータベース統合化の全体像とそれに至るステップを示す。

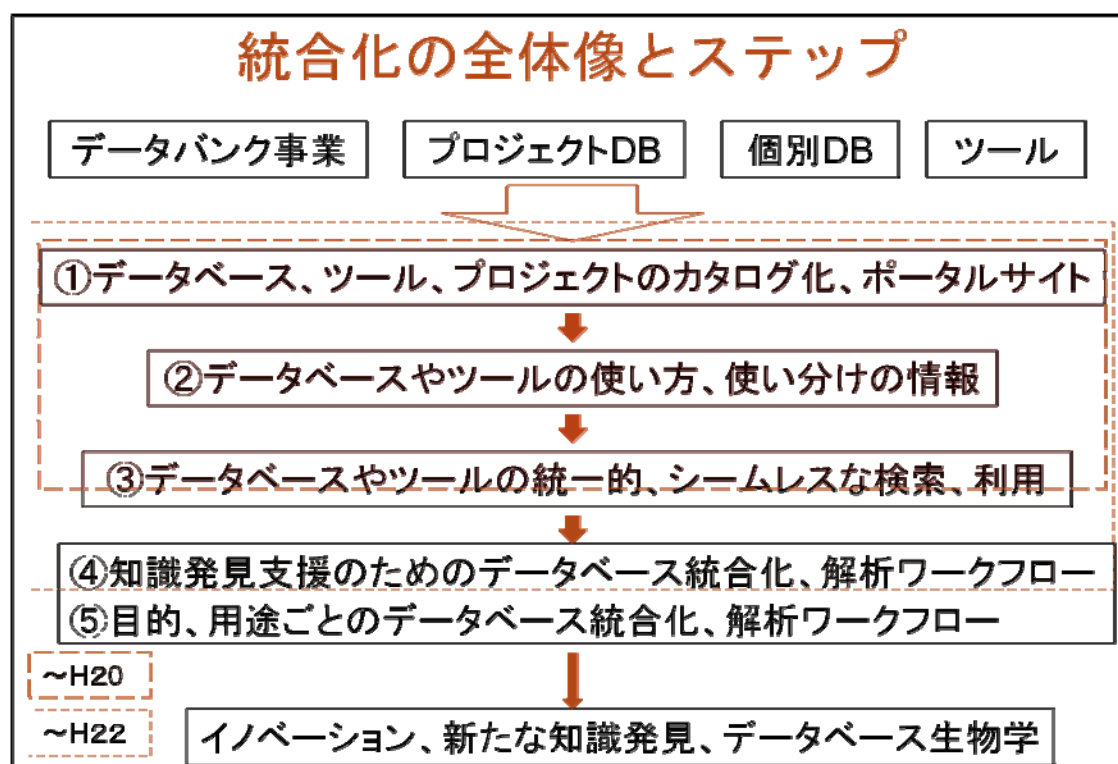


図 2.4 統合化の全体像とステップ

個人情報保護に関する調査、検討については、まず、次世代DNAサーバ技術の進展に伴い予想される個人ゲノム情報の蓄積とその医学への応用という局面で想定されるデータ量とデータのセキュリティの確保という観点での調査を行った。次に、こうした調査の結果をベースに、本プロジェクトの分担機関である東京大学の疾患解析データベースグループと連携して、同グループが提供するゲノムワイドな多型情報および関連解析情報および疾患関連遺伝子のリシーケンスデータのデータカテゴリー別の共有方針の検討を行った。併せて、データ提供に伴うユーザ同意書の起案や提供先でのデータの漏えい等の危険に対する安全性を確保するためのセキュリティ手順書の策定などを行った。これにより、疾患解析データベースグループの提供するデータの共有が進むことが期待できる。

平成21年度は、これまで進めてきた統合データベースプロジェクトの目標や年次計画の具体化、国内外の動向調査をベースに、データベース統合化のあるべき姿の検討を引き続き行った。併せてデータ共有化や公開促進の観点から、著作権、個人情報保護に関する調査、検討も引き続き行った。また、日本語文献の公開についても、コンテンツを充実すべ学協会等との交渉を積極的に行い、日本溶媒抽出学会 (http://lifesciencedb.jp/lldb.cgi?gg=tool_academy) や歯科基礎医学会 (http://lifesciencedb.jp/lldb.cgi?gg=tool_academy) の学協会カタログ、日本における生物多様性関連Webサイト一覧 (<http://life>

lifesciencedb.jp/lbdb.cgi?pg=1) や日本分類学連合 (<http://lifesciencedb.jp/lbdb.cgi?pg=1>) のカタログ、そして原生生物情報サーバ (<http://lifesciencedb.jp/lbdb.cgi?pg=1>)、日本産糸状菌類図鑑 (<http://lifesciencedb.jp/lbdb.cgi?pg=1>)、病原真菌データベース (<http://lifesciencedb.jp/lbdb.cgi?pg=1>) 等を追加し日本語コンテンツの充実を図った。さらに、国内外の研究機関等関係者との連絡調整という観点から、平成19年度、20年度に引き続き、世界的に分散しているデータの統合利用に関する国際ワークショップである BioHackathonを開催した(参加者は海外32名、国内36名)。今回はデータベースのより高度な収集と活用を目的にセマンティックウェブについて、国内外の主要な研究者が集まった。

データベース統合化の今後の進め方については、平成20年度に松原謙一研究運営委員長のリーダーシップの下、平成20年度にとりまとめた運営委員の先生方から収集したライフサイエンス統合データベースのあるべき姿に関するご意見をベースに、本庶 佑総合科学技術会議議員をはじめ、井村裕夫京都大学名誉教授、平田 正協和発酵キリン名誉相談役、水上政之科学技術振興機構理事をお招きして開いた研究運営委員会拡大懇談会や通常の研究運営委員会での議論などを通じて、上記のライフサイエンスPTタスクフォースの報告書で設立が提言されている統合データベース推進本部(仮称)への申し入れ事項の案文として、以下の9項目がまとめられた。

1. これからのライフサイエンスのブレークスルーは、存在する莫大なデータから仮説を創生してそれを検証する、データドリブン型の研究により達成される。
2. 広くデータを共有してこそイノベーションの加速化が実現できる。従って、我が国バイオ産業の発展には、網羅的に効率よくデータにアクセスできる環境が不可欠である。
3. データ共有できる環境を構築するためには、多岐にわたる対象や階層のデータベースの統合化とこれを構築、維持する機関の存在が必須である。
4. 統合データベースはライフサイエンス分野の国力の象徴であり、米国の NCBI、欧州の EBI に相当する機関の設立が望まれる。
5. 上記機関の運営は、データ生産者の協力や利用者の意見を集約しながら進めることが重要であり、そのため各機関の代表者も含めたオープンな運営に努める必要がある。
6. 統合データベースは質の高い個別データベースの上に成立するものであり、個別データベースを構築、維持するための仕組みは別途考慮されるべきである。
7. ライフサイエンス分野のデータは、種類、量とも急速に増大している。そのため、統合データベースの将来的な持続可能性を担保するための仕組みを検討する必要がある。
8. 統合データベースの有用性を維持するためには、恒常的に使い勝手やデータの付加価値向上を図る必要があり、そのための研究開発機能を備える必要がある。
9. ライフサイエンス分野のデータベース構築や統合化のための人材は不十分であり、統合データベース構築と連携した人材育成やキャリアパスの構築が必要である。

データ共有や個人情報保護については、分担機関の一つである東京大学の「疾患解析から医療応用を実現する DB 開発」において、共有方針の作成検討にオブザーバーとして参加し、共有方針や登録・利用申請に関する書面の作成を担当した。書面は国外からの申請

も考慮して、日本語とともに英語版も作製した。これらの文書は、12月に「疾患解析から医療応用を実現するDB開発」ウェブサイトから公開された。

当プロジェクトの実施期間も半ばを過ぎたことから、中間での成果発表の位置づけとして、また同時に、データベースを取り巻く様々なコミュニティメンバーが一堂に会する機会として、平成21年6月12日に東京大学本郷キャンパス浅野地区内武田ホールにおいて、統合データベースプロジェクトシンポジウム2009「データベースが拓くこれからのライフサイエンス」を開催した。講演7件、パネル討論、ポスター発表（プロジェクト内外問わず募集）85件からなるシンポジウムに、当日登録者も含め、のべ393名の参加者（含む関係者）があった。発表者に後日依頼してポスターPDFを提出してもらい、ウェブサイトで公開した（<https://symposium.lifesciencedb.jp/2009/>）。



図 2.5 平成 21 年 6 月シンポジウムのポスターと会場の様子

データベースの受入・公開にともなう著作権問題の解決方法の一つとして当プロジェクトではクリエイティブ・コモンズ・ライセンスを採用している。同ライセンスの創始者である米国の法律学者が来日する機会をとらえ、この問題を広く一般に問いかけるために、国内の著作権法関係者、欧州での動向に詳しい関係者らを集めて、平成21年10月5日に東京大学本郷キャンパス弥生地区内弥生講堂において、統合データベースプロジェクトシンポジウム「科学における情報の上手な権利化と共有化」を開催した。参加者は257名（含む関係者）で、テーマ内容に関連してライフサイエンス関係者以外に法学や出版関係者も集うシンポジウムとなった。講演の内容は動画としてウェブサイトで公開した（<http://symposium.lifesciencedb.jp/IPDS/>）。



図 2.6 平成 21 年 10 月シンポジウムのウェブサイトと動画配信

平成 22 年度は、平成 21 年 12 月に公開した、上記「疾患解析から医療応用を実現する DB 開発」ウェブサイトを通じた実際のデータ受け渡しに伴って顕在化した手続き上の不便を解消する手順を再整備し、書面もそれに合わせて改訂した。また、「ゲノム支援」のデータベース構築に関する打ち合わせにオブザーバーとして参加し、ヒトゲノム情報の共有化について関係者と意見を交わした。倫理の問題に関しては、先行している国外のゲノムやコホートのプロジェクト等において関連する部門でリーダーシップを取っている研究者が一堂に会する初めての国際会議が英国で開催されたので、DBCLS から参加して各国の具体的な状況を知るとともに日本の状況についても議論する機会を得た。一方で、本プロジェクトサイトを利用した録画公開についての外部からの要請が増加していることに伴い、録画公開についてのポリシーを新たに検討して公開した。また、平成 22 年 10 月には 4 省（文科、厚労、農水、経産）間でのデータ共有に関して積極的な関係機関が共同で合同シンポジウム「ライフサイエンスの未来へ～10 年先のデータベースを考える～」を開催した。平成 21 年度、22 年度に開催したシンポジウムの発表資料等(要旨、ポスターPDF、講演録画など)はそれぞれのウェブサイトから検索・閲覧可能である(表 2.1 参照)。

表 2.1 シンポジウムウェブサイト

シンポジウム名	開催日	URL
「データベースが拓く これからのライフサイエンス」	平成 21 年 6 月 12 日	https://symposium.lifesciencedb.jp/2009/
「科学における情報の 上手な権利化と共有化」	平成 21 年 10 月 5 日	http://symposium.lifesciencedb.jp/IPDS/
「ライフサイエンスの未来へ	平成 22 年	https://symposium.lifesciencedb

～10 年先のデータベースを考える～	10 月 5 日	. jp/godo/
--------------------	----------	------------

また、平成 20 年度以降のライフサイエンス課管轄の委託事業の公募要領には、図 2.3 で示したように、統合データベースへのデータ、データベース提供条項が盛り込まれている。平成 22 年度より、該当プロジェクトへのデータ提供の働きかけを開始した。前例のない試みであることもあり、必ずしもスムーズに進んでいるわけではないが、データ提供をいただく手順の整理も含め、方向性が見えてきている。表 2.2 に、対象となる 5 つのプロジェクトに関する状況をまとめた。

表 2.2 公募要領記載の「統合データベースへのデータ提供」に関するフォロー状況

事業名	対象年度	対象機関	状況(平成 22 年 10 月現在)
脳科学研究戦略推進プログラム	20 年度 21 年度	全機関	対象となるデータがあるかどうか事務局にて調査いただき、1 件データ提供いただける機関を紹介いただいた。現在先方機関との交渉準備中。
ターゲットタンパク研究プログラム	21 年度	阪大、理研	成果報告書の確定後、DBCLS が成果報告書を閲覧する旨、文部科学省から各機関に了解をとっていただいたうえで、対象データの存非を確認することになった。 ライフサイエンス課取りまとめで成果報告書を入手予定。
革新的細胞解析研究プログラム	21 年度	全機関	事務局に確認、現在のところ対象データがないとの回答を得た。
ナショナルバイオリソースプロジェクト	21 年度	理研、NINS、横浜市、筑波大	現在データ取りまとめ機関と相談後、個別機関へ DBCLS から対応することになった。2 機関からデータを入手済みで、現在公開準備中。残り 2 機関とは現在交渉中。
橋渡し研究支援プログラム	21 年度	全機関	成果報告書の確定後、DBCLS が成果報告書を閲覧する旨、文部科学省から各機関に了解をとっていただいたうえで、対象データの存非を確認することになった。 現在、一部成果報告書を入手済み、内容確認中。

2. 2. 2 実行評価

平成 18 年度に先行着手した本プロジェクトであるが、年度末時点で公開されていたいくつかのサービスや取り組みについての説明などを対象として H19 年度以降の統合 DB の進め方について製薬企業を中心に約 60 名の生命科学関係者から意見をうかがう第 1 回の評価を平成 19 年 3 月に行い、19 年度の開発の参考にした。

平成 19 年度以降は年度が終了してその成果がウェブサイト反映されてから実施することとし、取り組み内容の説明に加えてテスト検索が実行できるページや実施内容を表す模式図を見せるなどの工夫をしてサービス内容の紹介も兼ねたウェブサイト記入方式のユーザ評価のシステムを構築した(図 2.7 21 年度評価記入サイトの一部)。評価の内容は各サービス・取り組みについては 5 段階評価とフリーコメントで記入する方式とし、プロジェクト全体にわたる内容(例:プロジェクトが提供すべきサービス、ユーザ評価の方法/

形式など）についてはフリーコメントを記入してもらうようにした。また、評価者の中で氏名や所属を公表してもかまわない場合は評価期間終了後に一括して公開することとした。

サービスのステップごとの分類説明

データベース統合化への道のり (総合データベースの考え方)

データベース構築 → プロジェクト化 → 統合化 → ツール

ステップ1. データベース、ツール、プロジェクトのカタログ化、ポータルサイトの構築
ステップ2. データベースやツールの使い方と教材の提供
ステップ3. 目的、用途ごとのデータベース統合化、解析ワークフロー
ステップ4. データベースやツールの統一、シームレスな検索、利用の支援

評価サイトトップ

選択数	得票数	得票率
かなり役に立ちそうなので、どんどん進めてほしい	33	54.8%
それなりに役に立ちそうなので、できる範囲で進めてほしい	28	42.4%
責任をもって評価できない	5	7.6%
あまり役に立たなそうなので、進めなくてもよい	0	0.0%
まったく役に立たなそうなので、やめたほうがよい	0	0.0%
投票総数	66	

このサービスについて以下のコメント欄にご意見をお寄せ下さい。(前年のご意見へのリンクあり)
また、未登録のデータベースがありましたら以下にお知らせ下さい。

コメント:

個別サービス評価ページ

図 2.7 21 年度評価記入サイトの一部

平成 20 年度には 19 年度に公開されたものを含め代表的な 17 種類のサービスを 5 つのカテゴリ（生命科学研究の情報へ案内する／国内外の様々な情報の検索サービスを提供する／情報整理や抽出の道具を提供する／DB 統合に必要な新しい基盤技術を創る／DB の未来を担う人材を育成する）に分類し、それらに対する 5 段階評価とフリーコメントの記入を 92 名の有識者に依頼してのべ 58 名から回答を得た。平成 21 年度には 20 年度に新たに公開されたサービスを中心に中核機関以外のサービスについてもほぼすべてを対象としたため、多数となった評価対象サービスをデータベース統合化のステップごと（ステップ 1. データベース、ツール、プロジェクトのカタログ化、ポータルサイトの構築／ステップ 2. データベースやツールの使い方と教材の提供／ステップ 3. 目的、用途ごとのデータベース統合化、解析ワークフロー／ステップ 4. データベースやツールの統一的、シームレスな検索、利用の支援）に分類して、ステップごとに評価ページを作成し、サービスごとに新たに開発された内容を記述するなどして、2 度以上評価に参加しているユーザにも新規部分がわかりやすい表記を心掛けた。また、前年の評価の際に要望のあった、成果目標や予算配分の資料も併せて掲載した。記入方式は前年と同様で、学会の展示会でのアンケートやプロジェクト関係のイベント等で募集した評価者も加えて 291 名に依頼してのべ 117 名からの回答を得た。いずれの評価においても、コメントのうちすぐに実現可能なものは実施し、次年度以降の開発に盛り込めるものは盛り込んだ（実施例：パラメータの改善点の意見→データベースカタログ等の結果表示の方式変更、英語化推進の意見→英語での紹介が有効と考えられるサービスを中心にウェブサイトの英語化を実施、など）。評価者の分類情報や評価結果およびフリーコメントの内容、コメントに対する開発サイドから

の回答内容は全てウェブサイト上に公開している。平成 22 年度は、プロジェクトの最終年度であるため、全ての関連機関に対する評価を実行できるよう各機関にウェブサイトでの何らかの情報やサービスの提供を依頼し、機関ごと（DBCLS のみ実施内容が多岐にわたるため、サービス内容による 5 つの分類ごと）に評価ページを用意した。評価対象の内容は、プロジェクトの全期間に実施した内容の全てとし、参考資料としてプロジェクト全期間にわたる各機関の予算配分結果と平成 20 年度以降の各機関の年度目標と進捗のまとめ表を掲載し、前年の評価の際に要望のあった各サービスのアクセス数について（22 年度に新規に公開したサービスを除く）も参考情報として提供した。評価作業は 470 名余に依頼し、のべ 138 名からの回答を得ている。この結果については暫定的に評価記入サイトを閲覧可能にすることで暫定公開している（11 月現在）。

表 2.3 評価結果公開サイト

対象年度	実施年月	のべ 138 名からの回答を得ている。この結果については暫定的に評価記入サイトを閲覧可能にすることで暫定公開している
平成 18	H19 年 3 月	http://lifesciencedb.jp/feedback/hyouka18/hyouka_H18.html
平成 19	H20 年 8 月	http://lifesciencedb.jp/feedback/hyouka19/?result
平成 20	H21 年 10 月	http://lifesciencedb.jp/feedback/hyouka20/?result
平成 18-22	H22 年 11 月	http://lifesciencedb.jp/feedback/hyouka21/?FrontPage （評価記入サイトにて評価内容の未公開）

参考）外部評価のページ <http://lifesciencedb.jp/lldb.cgi?gg=hyouka>

一方、平成 19 年 10 月のサービス開始以来収集していたアクセスログの解析については平成 20 年度から開始した。本センター以外のプロジェクト参加機関にもログの提供を依頼した。解析方針は以下の通りで、基本的には現在も変わっていない。

- 解析ソフトは AWStats ver.6.6 (<http://awstats.sourceforge.net/>)を採用し、サービス開発者の IP アドレスを除外する以外は、デフォルト設定のまま利用した。AWStats が出力する主な集計値は以下の 3 つである。
 - 一意な訪問者 IP：アクセス元の IP アドレス数。同一 IP からの複数回のアクセスは 1 と計数
 - 訪問回数（セッション数）：アクセス元の IP アドレス数。ただし 1 時間以上の間隔をおいての同一 IP からのアクセスは別訪問と計数
 - ページ数（いわゆるページビュー）：閲覧ページ数
- 統合 HP(<http://lifesciencedb.jp>)のポータルにリストされている項目（サービス）毎に月次集計。
- 毎日自動更新し、本センターの関係者が常時 AWStats の結果を閲覧できるようにする。

平成 21 年度は、本センターのみならずプロジェクト全体としての利用状況の把握を効率的に行うため、参加機関からのアクセスログの自動収集も検討したが、セキュリティの問題を考慮して実施は見送った。しかし各機関との協力体制は確立され、現在まで順調に運用されている。

平成 22 年度は、プロジェクト全期間にわたる成果についてのユーザアンケートとアクセス解析等により提供サービスの評価を行う段階に入った。10 月に実施したユーザアンケート資料にはサービス単位の訪問回数とページ数を表示し、評価指標のひとつとなった。作業部会や運営委員会の報告、研究者による学会発表にもこのアクセスログ解析結果は多数引用されている。利用目的が明確化しつつあり、それに沿った利用やすい形式で解析結果を提供するのが今後の課題である。

表 2.4 ライフサイエンス統合データベースセンター全体の集計結果

	平成 20 年度	平成 21 年度	平成 22 年度 (10 月末現在)
一意な訪問者 IP	231,305	439,119	310,736
訪問回数	436,203	942,174	721,858
ページ数	1,765,823	6,350,496	4,891,269

本センター全体の集計値は表 2.4 のとおりである。サービス別訪問回数の合計であるため、例えば 1 ユーザが 1 回の訪問で 2 つのサービスを訪問した場合は 2 と計数している。また、一意な訪問者 IP 数も月次の合計のため、同一の IP から異なる月に 2 回訪問した場合は 2 となる。

図 2.8 のヒートマップは、本センターの提供するサービス全体を時系列かつサービス別に俯瞰するためのものである。

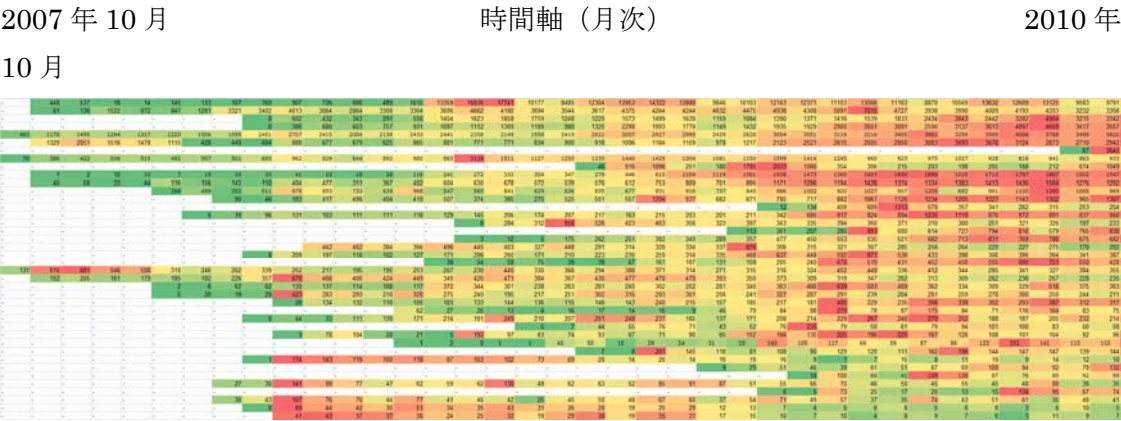


図 2.8 ライフサイエンス統合データベースセンターが提供するサービス全体のヒートマップ

横軸は時間軸、縦軸はサービス軸で最大月次アクセス数の降順にソートしている。1 行が 1 サービスに対応しており、赤いセルはそのサービスでのアクセス数の月次最大値、緑は最小値、黄は中間の値を示している。ヒートマップの右上が赤いことから、アクセス数の多いサービスほど順調にアクセス数を伸ばしているといえる。反対に、右下の一部のみが緑であることから、アクセス数が減少しているサービスは少数であり、総アクセス数への影響も少ないことがわかる。しかも、更新が終了した過去のコンテンツとして掲載しているものであるため、更新を継続しているサービスはアクセス数が増加している。なお上

サービス（アクセス数降順）

のマップでは一意な訪問者 IP 数をアクセス数としたが、訪問回数、ページ数の場合もほぼ同様の傾向であった。

アクセス数上位のサービスは表 2.5 の通りである。サービス開始時期の違いや月毎の変動を考慮して、サービス開始以降の月次最大値を代表値とした。一意な訪問者 IP 数をユーザ数とみなすと、表 2.5 からは独自のコンテンツを提供しているサービスがユーザを獲得しているといえる。トップの生命科学データベース横断検索は独自コンテンツではないが、内訳は蛋白質核酸酵素バックナンバー日本語 PDF ファイルを出力するリクエストが大部分を占めている。表 2.6 もほぼ同様の傾向を示している。

表 2.5 一意な訪問者 IP 数の月次最大値上位 10 サービス

サービス名	開始時期	月次最大値
生命科学データベース横断検索	2007/10～	17,741
DNA データベース総覧と検索 (DDBJ/EMBL/GenBank)	2007/10～	7,015
蛋白質核酸酵素 全文検索	2008/5～	4,904
文科省「ゲノム」研究報告書 全文検索	2008/5～	4,609
統合 TV(DB やツールの動画教材)	2007/7～	4,326
アナトモグラフィー/BodyParts3D	2007/10～	3,879
ライフサイエンス新着論文レビュー	2010/8～	3,645
統合ぐらし	2007/9～	3,128
シンポジウム・講演会	2009/4～	2,033
Gendoo	2007/10～	1,888

表 2.6 訪問回数の月次最大値上位 10 サービス

サービス名	開始時期	月次最大値
生命科学データベース横断検索	2007/10～	32,474
DNA データベース総覧と検索(DDBJ/EMBL/GenBank)	2007/10～	17,890
統合 TV(DB やツールの動画教材)	2007/7～	15,425
蛋白質核酸酵素 全文検索	2008/5～	11,758
ライフサイエンス新着論文レビュー	2010/8～	11,294
文科省「ゲノム」研究報告書 全文検索	2008/5～	6,786
誰でもデータベースが構築できる TogoDB	2008/4～	6,181
Gendoo	2007/10～	6,082
統合ぐらし	2007/9～	5,941
アナトモグラフィー/BodyParts3D	2007/10～	5,449

表 2.7 のページ数は、ユーザ数というよりはむしろリクエストを処理した回数とみなせるため、表 2.5、2.6 とは多少異なる傾向を示している。ページ数の多少はシステムの構築

方法にも依存するとはいえ、上位にはコンテンツ提供サービスだけでなく、解析処理を行う TogoWS や MiGAP のようなサービスも現れる。バッチ処理等で 1 度の訪問で多数のリクエストを処理しているものと思われる。

表 2.7 ページ数の月次最大値上位 10 サービス

サービス名	開始時期	月次最大値
国内ウェブサービスの統合 TogoWS	2008/2～	399,873
DNA データベース総覧と検索(DDBJ/EMBL/GenBank)	2007/10～	380,834
アナトモグラフィ/BodyParts3D	2007/10～	248,853
誰でもデータベースが構築できる TogoDB	2008/4～	114,779
統合 TV(DB やツールの動画教材)	2007/7～	77,700
生命科学データベース横断検索	2007/10～	52,946
遺伝子発現バンク(GEO)目次)	2007/10～	51,250
MiGAP	2009/4～	46,700
ライフサイエンス新着論文レビュー	2010/8～	44,484
Gendoo	2007/10～	25,003

表 2.8 はアクセス元のドメインの内訳である。プロバイダなど組織を特定できないドメイン(net/ne/com)からのアクセスが大きく増加している。教育機関(ac/edu)、政府系機関(go/gov)については実数が増えているにもかかわらず、割合は横ばいもしくはやや減少している。民間企業(co)についても減少しているが、その他に分類している com ドメインにも民間企業が含まれておりやや低めの数字となっている。

表 2.8 アクセス元ドメインの内訳

ドメイン種別	平成 20 年度	平成 21 年度	平成 22 年度 (10 月末まで)
教育機関(ac/edu)	18%	10%	13%
民間企業(co)	13%	5%	2%
政府系機関(go/gov)	4%	1%	2%
インターネットプロバイダ等(ne/net)	23%	57%	54%
その他(com など)	42%	27%	29%

3. 統合データベース開発

3. 1 共通基盤技術開発

3. 1. 1 共通基盤技術開発の実施計画

生命科学研究者の興味を反映した文献情報の検索、整理が容易にできる環境の構築を行うことを目的に、言語横断的な文献検索システムや質問応答システム、情報抽出シス

テムを実現するための技術開発および必要な調査を行う。さらに、統合検索に向けた基盤技術として、セマンティックウェブ技術に関する調査を行い、必要となる技術を開発する。

3. 1. 2 共通基盤技術開発の実施内容

本課題では、研究者が多様な生命科学 DB や解析ツールをストレスなく利用できるための基盤技術を提供することを目標とし、様々なサービスを開発してきた(図 3.1)。それらは、文献執筆支援や文献情報からの情報抽出を行うテキストマイニングサービス群、簡単に誰でもデータベースが構築可能となるサービス TogoDB (<http://togodb.dbcls.jp/>)、国内外の主要な DB や TogoDB のウェブサービスを統一的インターフェイスで利用できる TogoWS (<http://togows.dbcls.jp/>)、ウェブサービスやツールを組み合わせで解析ワークフローを作成できるサービス DBCLS Galaxy (<http://galaxy.dbcls.jp/>)、シングルサインオンサービス DBCLS OpenID (<http://openid.dbcls.jp/>) から成り、互いに連携して、様々な研究スタイルをもつ生命科学研究者が DB を利用するための基盤サービスとなっている。また、これらのサービスを開発するにあたり、国際的動向の調査、国際的連携と標準化、それに伴う開発が必要であり、それらを行うために、国際開発会議 DBCLS BioHackathon を平成 19 年度 (<http://hackathon.dbcls.jp/>)、20 年度 (<http://hackathon2.dbcls.jp/>)、21 年度 (<http://hackathon3.dbcls.jp/>) の三回開催した。以下、サービスと会議それぞれについて、年次ごとの成果について述べる。



図 3.1 共通基盤技術開発によって提供されるサービスと開催した会議

現在一般公開しているテキストマイニングサービスは、OReFiL (<http://orefil.dbcls.jp/>), Allie (<http://allie.dbcls.jp/>), inMeXes (<http://docman.dbcls.jp/im/>), TogoDoc (https://docman.dbcls.jp/pubmed_recom) の 4 つからなる。OReFiL はオンライン上に公開されている生命科学系のデータベースやソフトウェアなど（オンライン資源）のうち、学術雑誌に引用のあるものを検索するシステムであり(図 3.2)、東京大学で開発されたオンライン資源検索システムをベースとして、平成 19 年度にユーザの利便性と検索精度を高めるための資源評価書き込み機能、資源評価閲覧機能などの追加を行ったものである。



図 3.2 OReFiL: オンライン上の資源を文献情報と共に検索し提示

Allie, inMeXes, TogoDoc の三つのサービスは文献執筆支援を目的に開発され、具体的には、Allie は生命科学における略語とその展開形の検索システム、inMexes は生命科学文献における英語表現検索システム、TogoDoc は生命科学分野の書誌情報と論文(PDF ファイル)を管理するためのサービスとツールである。Allie は略語として同じ表記であっても、異なった展開形をもつ例が多数存在することを踏まえ、平成 19 年度に、遺伝子名やたん

ばく質名等の略語とその展開形のペアを MEDLINE から抽出してデータベース化し、頻度順などで表示する検索システムとして構築した(図 3.3)。inMeXes は平成 21 年度に文献 (PubMed/MEDLINE に含まれるタイトルとアブストラクト) に頻出する英語表現を、1 文字の入力ごとに高速に検索するシステムとして開発した。

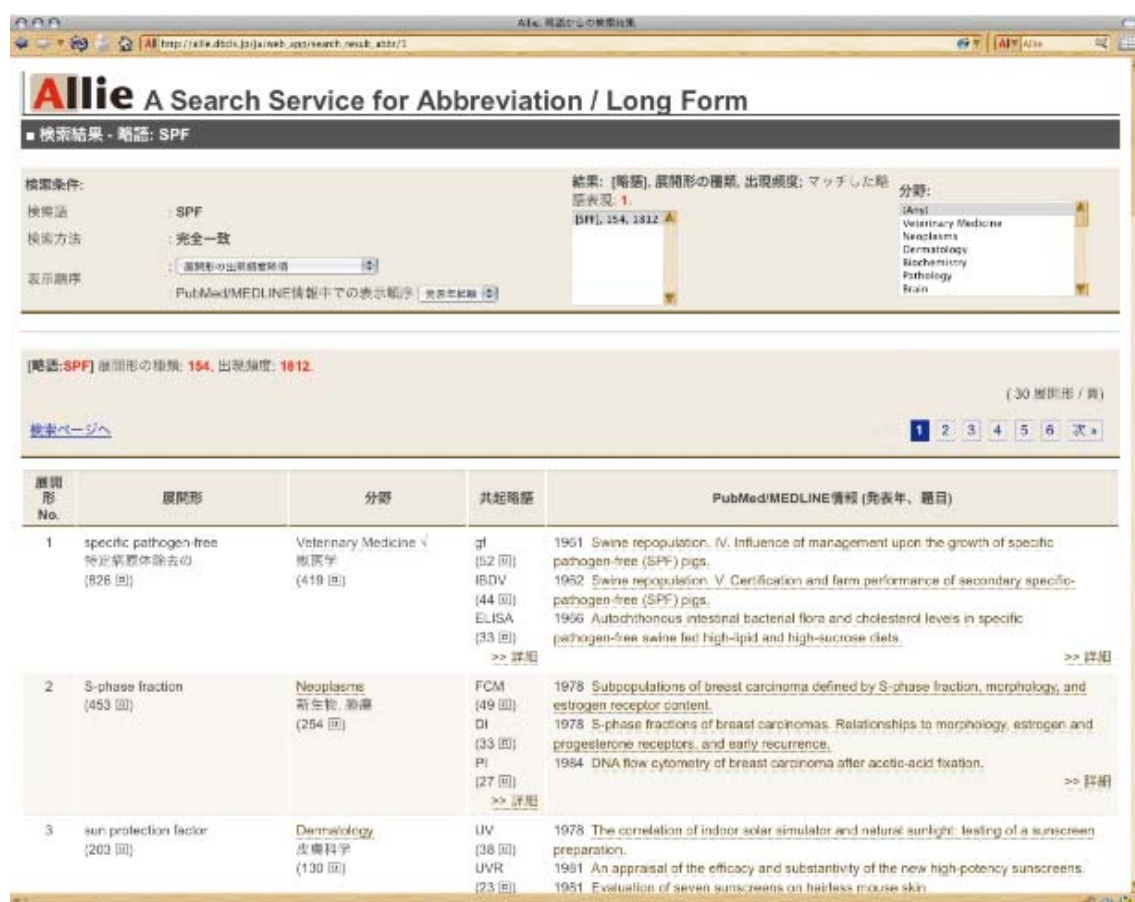


図 3.3 Allie: 論文に出現する略字とその展開系を検索

TogoDoc の開発は平成 20 年度より 3 年かけて行われた(図 3.4)。平成 20 年度には、研究者が既に保持している文献情報を基に MEDLINE データ全体を対象として関連度の高いものを高速に検索し、提示するシステムとして開発し、内部向けに公開した。さらに、当時、このシステムでは、文献管理機能やパラメータの検討を行うために、文献情報の推薦は暫定的に NLM の提供する PubMed Related Articles 機能を用いていたが、当該機能は利用規程に起因する処理速度の遅さと仕様の不安定性、検索の硬直性などが問題であるため、DBCLS 独自に運用可能な文献推薦システムが必要と考え、一般的な文献推薦を行う既存システム Hotate (株式会社プリファードインフラストラクチャー社製品) を、MEDLINE の部分集合(2,753,786 件)に適用する試みを行った。その結果、全 MEDLINE データを対象としても動作可能である見通しを得ることができた。平成 21 年度は、利用者が自身の文献情報や文献 PDF ファイルを効率的に整理出来る環境を提供するために開発されている TogoDocClient (東京大学による)との連携機能の開発を中心に行い、システムを公開した。TogoDocClient は利用者の PC にインストールして動作させるソフトウェアで、利用者の PC 中にある文献 PDF ファイルから自動的に書誌情報を抽出する機能を持つが、本機能を支援するシステムや、得られた文献情報、更には文献 PDF ファイルを

TogoDoc に保存したり取り出したりする機能を開発した。また TogoDoc に保存されている文献情報を基にした推薦機能を TogoDocClient において利用出来る仕組みを構築した。平成 21 年度は、TogoDoc については DBCLS で推薦機能の開発を行って、動作確認を行い、TogoDocClient についても安定して動作する確認ができたことから、これらを公開した。この独自推薦機能の採用に伴い、システムが推薦時に重要視すべきキーワードの編集を可能とした。

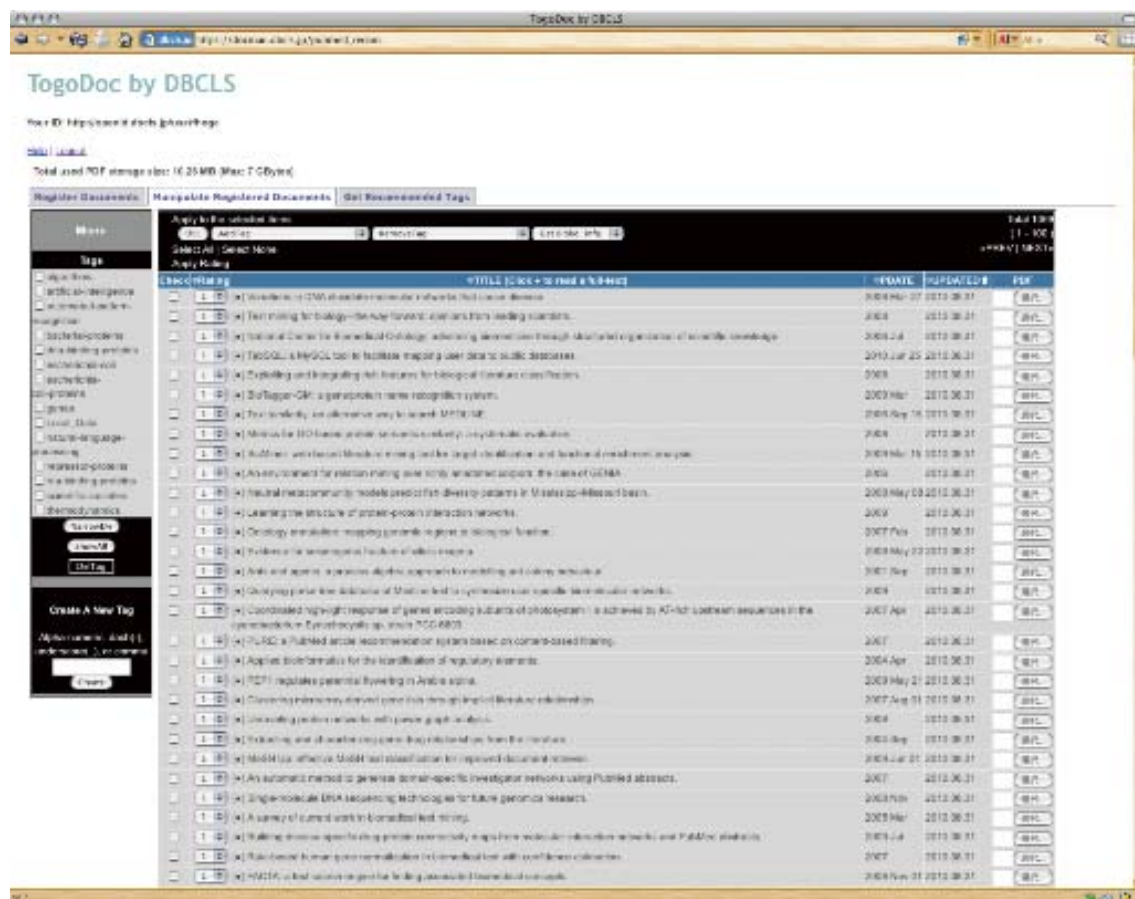


図 3.4 TogoDoc の文献情報操作画面

その他、テキストマイニングツールに関しては、以下のような内部向けの技術開発も行ってきた。平成 19 年度、遺伝子の対応関係や文献の引用関係などの二項関係をマトリックス上で視覚的に表示し、対応関係の編集を手助けする技術を開発し、プロトタイプ IDEditor を実装した。対応関係は読み込み時に自動的にクラスタリングされ、また、削除の候補としてクラスターの外れ値となる対応関係が示される。これを用いることで、ユーザは対応関係を視覚的に確認しながら編集を進めることが可能となった。平成 20 年度は、文献を始めとするテキストデータや図表データ等の非構造化データから各種生命科学情報を、より詳細に、精度良く、幅広い関連資源から抽出する要素技術開発を目標とし、フルペーパー内の藍藻遺伝子名を認識するツールを開発した。この検索システムでは、藍藻遺伝子名からの文献検索、PubMed ID からその論文に含まれる藍藻遺伝子名頻度情報取得、藍藻遺伝子名と PubMed ID からその遺伝子が論文内のどこに出現するかについての情報を取得することができる。平成 21 年度は言語横断的な質問応答システムを実現するための技術開発を行った。具体的には、日本語自然文を入力とし、MEDLINE の関連文献を

関係のある文と共に表示するシステムのプロトタイプを作成した。このシステムは主に次の三つのステップからなる：(1) 日本語自然文を解析して、主語・述語・目的語を抽出する、(2) ステップ(1)で抽出した主語・述語・目的語から、MEDIE(テキストマイニング技術を用いた知的 MEDLINE 検索システム)のクエリを生成する、(3) ステップ(2)で作成したクエリを MEDIE 上で実行し、その結果を表示する。

TogoDB は表形式のデータを CSV 形式でインポートすることで、レコード表示、リスト表示、検索機能をもったウェブデータベースの公開・管理を誰でも手軽に行えるシステムである(図 3.5)。TogoDB の基本的な機能は平成 19 年度に、DB 受入のためのテンプレートシステムとして開発された。具体的には、ウェブブラウザからエクセルなどで作成した CSV ファイルを読み込み、関係データベースを簡単に構築することができる機能、検索パラメータなどのウェブ公開に必要なインターフェイスが自動生成される機能、データ読み込み後すぐウェブ公開することができ、また、ウェブサービス(SOAP/WSDL)による検索 API も自動生成される機能などである。平成 20 年度における利用ユーザの利用実態の調査やフィードバックをもとに、平成 21 年度はすでにある機能をユーザにとって使い易くなるような拡張を主に行った。たとえば、カラム追加機能やデータインストールを繰り返すときの手間を省く機能、表示・編集項目の選択機能、リンク支援機能などである。また、他サービスとの連携も進め、TogoWS REST API 経由でのデータ取得機能や DBCLS Galaxy からのデータの入力機能を開発した。これらの連携機能は、ユーザにデータ共有の効率化のためのデータ中心のワークフローを提供する基盤となる。平成 22 年度はユーザのデータ利用の利便性向上のため、データビュー機能とデータリリース機能の新規開発をおこなった。データビュー機能を利用すれば、TogoDB のデータを任意のウェブサイト に埋め込むことができ、データリリース機能を利用すれば、データセットを作成し、BLAST 検索や RDF リリースが可能となる。現在、およそ 1000 ユーザが TogoDB を利用し、月間 5GB のデータ転送が行われている。



図 3.5 TogoDB の基本機能

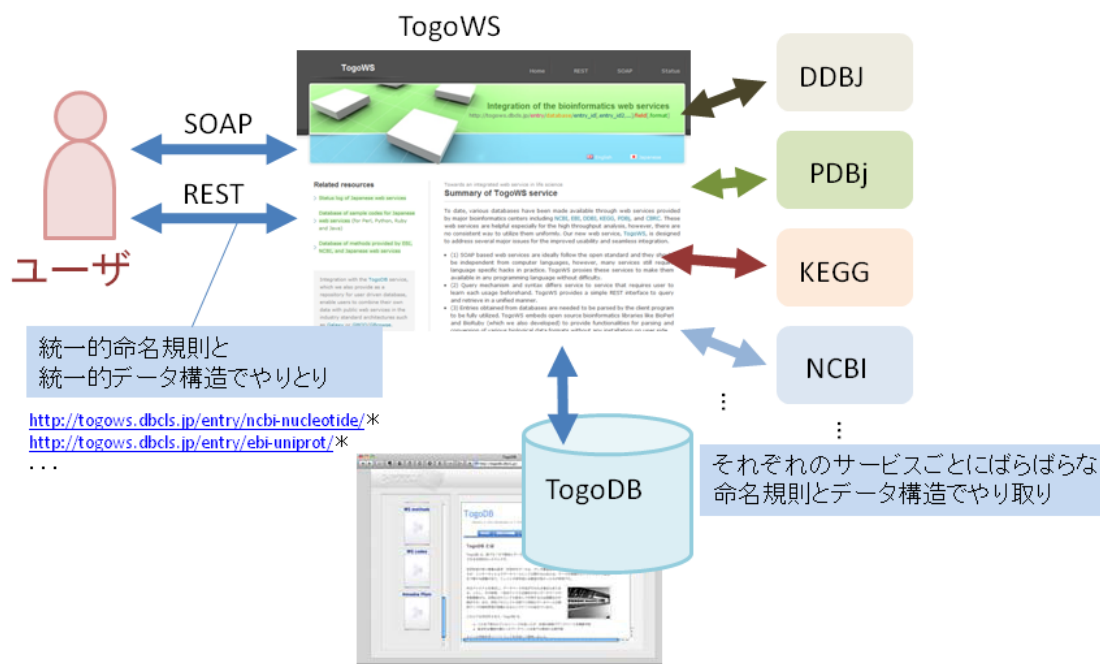


図 3.6 TogoWS: ウェブサービスの統一的なパラメータ提供

TogoWS は国内主要拠点のデータベース（DDBJ、PDBj、KEGG）や海外の主要拠点データベース（NCBI、EBI）を仮想的につないで、統一的なパラメータを提供するサービスである(図 3.6)。平成 19 年度は、国内外でどのようなウェブサービスが用意されており、また、その稼働状況がどうなっているかを調査し、現時点のウェブサービスの問題点を洗い出した。そして、各サービスでやりとりされる生物学データは、フォーマットの違いなどにより簡単には相互利用できない状況であること、さらにサービス毎にメソッドの命名規則が異なることやドキュメントが不足しているといった問題を解決するため、ウェブサービスの変換サーバとして TogoWS を開発した。平成 20 年度は、生物系のウェブサービスの相互連携を容易にするため、各サービスではどのような機能を最低限用意すべきであるか、またエラー処理の扱いをどのように統一すべきかなどの設計指針を示したウェブサービスのガイドラインを作成し、TogoWS をこのガイドラインに準拠させた。平成 21 年度および平成 21 年度は対応データベースと対応フォーマット形式の拡充を進めた。現在、およそ 400 ユーザが TogoWS を利用し、月間 2GB のデータ転送が行われている。

DBCLS Galaxy はウェブブラウザ上で複数のツールを組み合わせで解析ワークフローが容易に作成可能なサービスである(図 3.7)。平成 20 年度にはツールを組み合わせるパラメータとして、ペンシルバニア州立大学が開発したプラットフォーム Galaxy を中心にツール組み込みパラメータの調査を行い、その結果をもとに、TogoWS のサービスの一部を Galaxy に組み込んだプロトタイプを開発した。平成 21 年度には、DBCLS Galaxy での解析結果を TogoDB へ出力するツールが組み込まれ、上記の TogoDB の DBCLS Galaxy からのデータの入力機能と繋げて、DBCLS Galaxy で解析した結果をスムーズに TogoDB

から公開できるようにした。また、セマンティックウェブ技術を用いた知識ベース freebase (<http://www.freebase.com/>)からの RDF データ取得など、セマンティックウェブ技術関連のツールの組み込み開発を行った。平成 22 年度には、統合 DB で開発した横断検索の組み込みや、前年度に引き続いてセマンティックウェブ技術関連のツール組み込みを拡充し、さらに、日本語化および日本語ドキュメントの作成をすすめた。

DBCLS OpenID は一つの ID で複数のサイトの認証が可能なシングルサインオンサービスである(図 3.8)。平成 19 年度に当時の標準規格 OpenID 1.1 に準拠し、その認証プロトコル仕様に沿って認証サーバを構築した。この開発により、統合データベースプロジェクトに関連する認証をこのサーバで一元的に行うことが可能となり、複数のサイト間を行き来するユーザの負担を減らすことができるようになった。平成 20 年度には OpenID 2.0 にも対応し、より幅広いユーザの利用を可能とした。現在、DBCLS OpenID は、統合データベースプロジェクトのページをはじめ、TogoDoc、TogoDB、DBCLS Galaxy、TogoWiki、かずさアノテーション、かずさナビゲーション、MiGAP など多くのサービスで採用されており、ユーザはこれらのサービスを一つの DBCLS OpenID アカウントで利用可能となる。

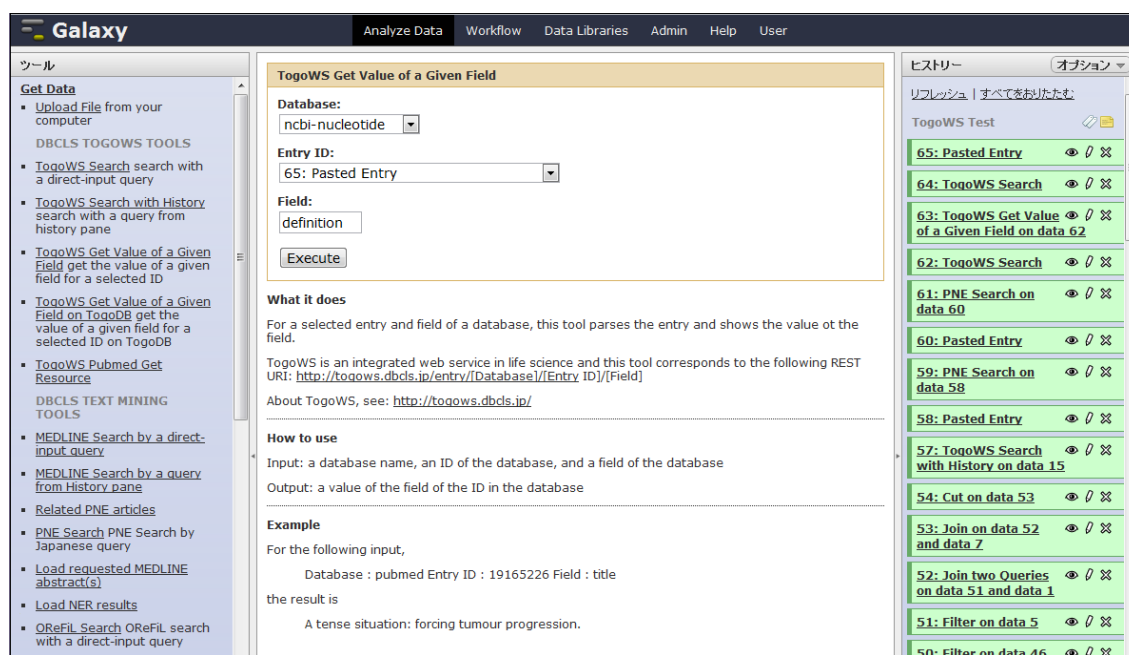


図 3.7 DBCLS Galaxy: 生命科学データ解析のためのワークフロー作成・共有サービス



図 3.8 DBCLS OpenID: 複数のサイトにログイン可能な認証サービス

上記の諸サービスの開発と並行して、国際的な動向・現状調査と連携した技術開発を目的として、国際開発会議 DBCLS BioHackathon をこれまでに三回開催してきた。第一回は平成 19 年度に「ウェブサービスの相互運用性」をテーマとして開催した。主に生命科学の主要なウェブサービスプロバイダ、クライアント開発者、オープンバイオ開発者などから参加者を募り、その結果、海外から 34 名、国内から 37 名の参加者があり、トピックごとに熱心な議論が行われた。主なトピックはオントロジー、データ交換の標準化、非同期通信、大規模データ、セキュリティ、オープンバイオ、RESTvsSOAP、ワークフロー、BioMoby などであった。その成果は上記 TogoWS の平成 19 年度の調査結果にも反映され、続く平成 20 年度のウェブサービスの設計指針作成の基ともなった。第二回は平成 20 年度に標準化技術をよりユーザ側から利用可能とする目的で「ワークフロー」をテーマとして開催した。ウェブサービス設計指針を改良するグループ、Galaxy の国際化や今後の展開について議論・開発するグループ、次世代サーバからの大量データをどう扱うかを議論するグループ、大量データの可視化について議論・開発するグループ、マイクロアレイデータ解析について議論・開発するグループ、実際の生物系データを持ち寄ってユースケースについて議論するグループ、セマンティックウェブをどう実装すべきか議論するグループ、文献情報のツール群について議論するグループなどがあり、これらの議論は、共通基盤技術開発の諸サービス、特に TogoWS と DBCLS Galaxy の開発に大きな影響を与えた。第三回は平成 21 年度に統合検索に向けた基盤技術調査を目指し「セマンティックウェブ技術」をテーマに開催した。ここでは、国内外のデータプロバイダーの実務者レベルでの利用技術の観点における合意形成がおこなわれた。それに基づいて、今後も国際協力的に、データ統合化やそのサービスの開発をすすめる予定である。さらに、データの応用部分での議論に基づいて開発を行い、その結果、萌芽的なものではあるが、ユーザインタ

ーフェイス部分やデータ統合化についてのいくつかのあたらしい技術的開発的取り組みを実施した。ここでの議論は平成 21 年度から 22 年度にかけて行われた TogoDB、TogoWS、DBCLS Galaxy でのセマンティックウェブ技術関連の開発を大きく促進した。平成 22 年度は、これまで得てきた国際技術動向や知見を国内で共有する必要があると考え、国内のデータベース構築者・利用者を対象に、統合データベース技術情報交換ワークショップ (<http://wiki.lifesciencedb.jp/mw/index.php/BH10.10>)を行った。国内から 70 名弱の参加者があり、活発な議論と情報交換がおこなわれた。三回の DBCLS BioHackathon および統合データベース技術情報交換ワークショップでの議論と成果はそれぞれウェブ上でまとめられており閲覧が可能となっている。

3. 2 ヒト統合 DB の開発・運用

3. 2. 1 ヒト統合 DB の開発・運用の実施計画

医学研究や創薬開発研究に携わる利用者がゲノム情報、分子データベース、文献情報等の、研究開発に必要な情報を容易に統合利用できるように、背景知識の整理を行い、辞書やオントロジーの形で公開する。またこれらを用いて データベースの整理、統合を進める。

3. 2. 2 ヒト統合 DB の開発・運用の実施内容

より多種類の分子データや文献データを統合利用するにはデータの種類の依存しない上位のレベルでの共通の索引付けや分類整理が必要である。「文書のメタデータ」である。

日付著者アドレス引用関係等を一階層医学生物学に下ると、生物種、解剖、疾患、分子遺伝子名称、化合物薬品、操作検査手技、研究分野などがこの上位レベルである。これは文献整理のためのキーワード群と同じであるが米国以外ではこれらの概念（用語）の収集が未発達であり、また膨大な文献や実験データのメタデータ部分に自由に記載された内容へのマッピング技術も未整備である。日本語を置き去りにせずこの課題の解決に向けて実証的サービスも構築と基礎資源の作成をおこなった。フェーズの違う産物が散在するが、ゴールは未統制用語による表現を操作可能にできる十分な網羅性の用語集、その意味による整理を行うシソーラス、であり、さらに分類表現を理解や利用に繋げるための情報ハブ表現の開発である。

実証実験 既存の分子データベース（DDBJ および GEO, BodyMap, H-Angel）を統合し年度、施設所在国、生物種、解剖、研究目的、など共通の視点から俯瞰可能にする統合

「遺伝子発現バンク(GEO)目次システム」(以下、GEO 目次 <http://lifesciencedb.jp/geo/>)

「DNA データベース総覧と検索システム」(以下、INSD 目次 <http://lifesciencedb.jp/ddbj/>)

材料データとして DDBJ リリースファイル (<ftp://ftp.ddbj.nig.ac.jp/>、NCBI GEO メタ

データ ftp://ftp.ncbi.nih.gov/pub/geo/ をおよび GenBank 生物分類ルールを利用し
研究分類のためのフィールドの値および埋まり具合に基づく分類ルール、国別分類のため
の頭文字表記を含む世界の都市名の政治的国家名対応ルール、動物解剖分類のための正規
表現による頻出臓器細胞名称と臓器分類の対応ルール、同様な植物解剖分類ルール、およ
び別項で報告する生物名称日本語辞書を開発利用し提供してきた。

その結果、1) 配列データ、遺伝子発現データともに各種の分類ルールによる多元独立
自動分類を行い、シソーラス的に上位からたどることで自分が排除している対象を意識し
ながら関心プロジェクトに到達しデータを取得することを可能となった。

2) キーワード検索や配列検索の後で得られるレコードを同様に検索後分類が可能であり、
多数のヒットを上位概念から絞り込むことが可能となった。

3) さらに自動分類に従った各種の統計によるデータ全体の俯瞰的表現を可能にした。
俯瞰表現には表以外に googlemap および anatomography, (内部開発別項)、生物アイコン
(別項) を用いている。

遺伝子発現バンク(GEO)目次 バージョン: 2010-11-13 [English page](#)

NCBI Gene Expression Omnibus (GEO) に登録されているデータを、測定技術と材料の異性に基づいて整理しました。

登録データリスト										国別登録データ分布	登録データ検索	登録データ全容	ヘルプ	
データ単位: [データセット / サンプル / プラットフォーム] 単位の説明>>														
各タブ内に表示される数値は、そのタブ分類に属するデータ数です。														
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
ヒト	霊長	動物	哺乳	脊椎	無脊椎	植物	バクテリア	ウイルス	ファージ	未分類	すべて			
(260,433)	(2,464)	(108,539)	(7,457)	(9,424)	(20,616)	(54,135)	(26,811)	(994)	(1,992)	(983,839)				
すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて	すべて			
(983,839)														
SAE NGS	SAE RNA	SAE SRA	PPS	GeneChip	タイリングアレイ	OMAアレイ	オリコアレイ	ヒースアレイ	タンパク質アレイ	抗体アレイ	HT-PCR	HT-Seq	その他	すべて
(37)	(0)	(0)	(0)	(2,143)	(123)	(3,646)	(2,467)	(63)	(0)	(0)	(178)	(1)	(7,887)	
RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)				

用語集 日本語対応メタ用語集の作成

英語圏のシソーラスやオントロジーにマップ可能な形で出典情報を明確にして日本語を集めた初の転用自由の日本語専門用語辞書

解剖学 7411 概念 http://togodb.dbcls.jp/terminologia_anatomica_japonica13

生物種 62274 種名称 http://togodb.dbcls.jp/species_names_latin_vs_japanese

学術用語 150592 概念 http://togodb.dbcls.jp/scientific_term_ja_vs_en

我が国では昭和初期より継続的に学術用語や生物学名を個別の学会が主体となって収集更新しているがそのほとんどが書籍として販売され統合利用や転用が困難である。また JICST (JST) および NII はそれぞれの事業の為に保有していると思われるシソーラスなどはその契約上や事業モデル上外部に利用許諾を出すことが困難で当然そのサイズや機能を客観的に評価した例も見当たらず個別の用語の出典情報も記載がなく不整合時に正当性を検証する手段もない。統合データベース事業では上記の例に学びおもに書籍を情報源に現行法制度下において誰もが自由に転用利用できることを命題に、「転用可能な巨大言語資源」としてメタ用語集を編纂した。メタ用語集では英語、ラテン語などの科学基本用語（基本概念）との対で日本語表記が収集されているのでオントロジーや分類学などの構造化は外部資源に依存して自由に行うことができる。

材料: 文部科学省学術用語 1 1 篇を含む書籍 3 1 巻および wikipedia を含む 3 2 の web サイトを情報源として用い、電子化したのちに創意著作権のないファクトデータのみを下記の一定形式で抽出する手作業およびパースを行った。

(<http://lifesciencedb.jp/bdls/90326357>)

開発: 単文形式による電子化統合

辞書や用語集およびウェブサイトはカラム構造を用いた表形式で語彙およびその意味が提示されており、人間には容易に意味が理解できる。しかし複数の書籍の間では同じカラム見出しの意味やその種類、行であらわされる集積単位 (概念) がそれぞれに異なっている。本プロジェクトでは、編纂中に意味粒度のまるめや不整合の断りない修正が行われないように、第一段階として電子化と形式の統一のみを考え、可逆的にオリジナルデータを再現可能な形式に転載することを行った。その形式とはテーブルの一つのセル中の値を文献 X の表 Y の行 N では列 M の値は Term・T であるという一文であらわす単文形式であり、容易にオリジナルデータを再現することが可能である。複数のテーブルを一つの単文集形式で集めたのちに「表 Y の列 N と表 Z の列 NN は同じ意味カラムである」などのルール集を別途明示的に作成適用することで転用を一層容易に出来る。

@Aartsenia japonica@	is @学名@	for @Aartsenia japonica@	in @東海大学出版会@	@日本近海産貝類図鑑@	version @第 1 版第 1 刷@
@コシタカドマルクチキレ@	is @和名@	for @Aartsenia japonica@	in @東海大学出版会@	@日本近海産貝類図鑑@	version @第 1 版第 1 刷@
@Abacetus leucotelus@	is @学名@	for @Abacetus leucotelus@	in @北隆館@	@新訂 原色昆虫大図鑑 第 II 巻 (甲虫 篇)@	version @平成 19 年 5 月 1 日@
@ホソツヤナガコミムシ@	is @和名@	for @Abacetus leucotelus@	in @北隆館@	@新訂 原色昆虫大図鑑 第 II 巻 (甲虫 篇)@	version @平成 19 年 5 月 1 日@
@Abacetus tanakai@	is @学名@	for @Abacetus tanakai@	in @北隆館@	@新訂 原色昆虫大図鑑 第 II 巻 (甲虫 篇)@	version @平成 19 年 5 月 1 日@
@ヨツボシツヤナガコミムシ@	is @和名@	for @Abacetus tanakai@	in @北隆館@	@新訂 原色昆虫大図鑑 第 II 巻 (甲虫 篇)@	version @平成 19 年 5 月 1 日@

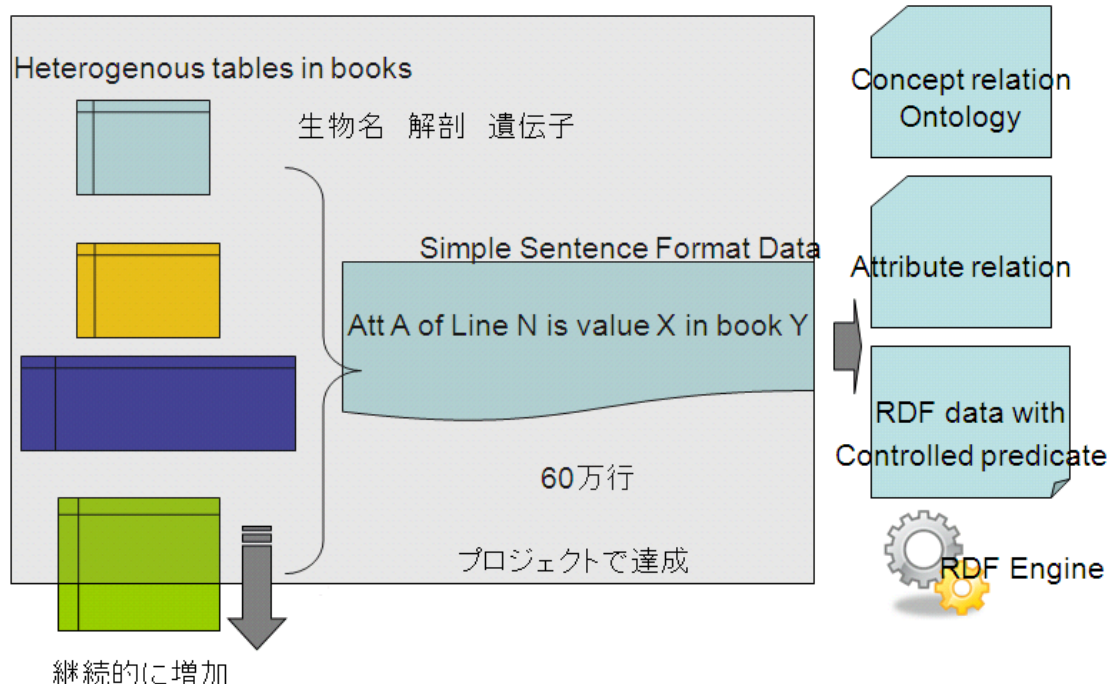


図 3.11 単文表現の例とメタ用語集プロジェクトの概要

これまでに 100 万行を超える単純文データを集め、Species2000, NCBItaxonomy, Nomina anatomica, Foundation Model of Anatomy など各種の世界規模での用語集とマップすることで常に概念単位の数とカバー率を測定している。生物名称に関して日本語書籍で与えられたラテン語と Species2000 や NCBItaxonomy との対応が予想外に悪く、学名の更新によるものやラテン語表記の語尾変化によるものなどの原因が多く、これら世界資源へのマップは未着の技術的課題として残された。マップの結果まとめは以下の URL を参照されたい。

<http://lifesciencedb.jp/bdls/8abf67fb72b66cc153ca7d50679c>

ごみ取りと簡単な整理を行った 2 次データ（用語対訳辞書）は全項の URI から検索ダウンロードが可能である。

以下の外部サイトおよび内部サービスではこの資源による日本語化を行っている。

- ・ Species Dictionary - 学名和名検索。

(http://bm2.genes.nig.ac.jp/Integrated_BodyMap/species.php)

- ・ 地球規模生物多様性情報機構 日本ポータルサイト 検索 (<http://gbif.ddbj.nig.ac.jp/>)。

- ・ DNA データベース総覧と検索における生物種名の和訳表示 (<http://lifesciencedb.jp/ddbj/>)。

辞書作成支援ツール

商業雑誌の一部分など web ページの特定部分へのリンクを著作権違反なく作成し共有可

能な引用プラグイン “Wired-Marker”

解説＋ダウンロード <https://addons.mozilla.org/ja/firefox/addon/6219/>

非専門家が用語の混乱を統制したり、用語に定義を与える場合にはクレジットおよび正当性を確保する為に常に情報源へのリンクを残す必要がある。しかしながら電子化オンライン化情報を直接コピーペーストする引用は厳密には「違法」である。したがって論文同様に DOI レベルのリンク付けを行うのが通例である。これは情報源のサイズによっては当該部分の再発見の負担が大きい。商用オンライン教科書を含む電子文献へのラインマーカの解像度でリンク付けが法的に行えれば本事業のみならず広く利用可能な道具である。ページ全体のコピーを利用した著作権違反を誘発する類似ツールは散見するが、位置情報のみを保存する本手法は他に例を見ない。

機能 1. 責任箇所（引用部分）への相対座標の自動付与

Wired-Marker は X-path 形式でページ内の位置を表現し、解釈する Hyper-Anchor エンジンを利用して URL に追記してやり取りする Firefox プラグインである。アノテータも利用してもブラウザに Firefox+Wired-Marker を用いればカーソルで選択した箇所をページ情報と独立にやりとり可能である（図 3.16）。

機能 2. ブックマークの編集機能

マーク領域のメタ情報は Book マークのようにフォルダ整理や日付順自動整理が行われる

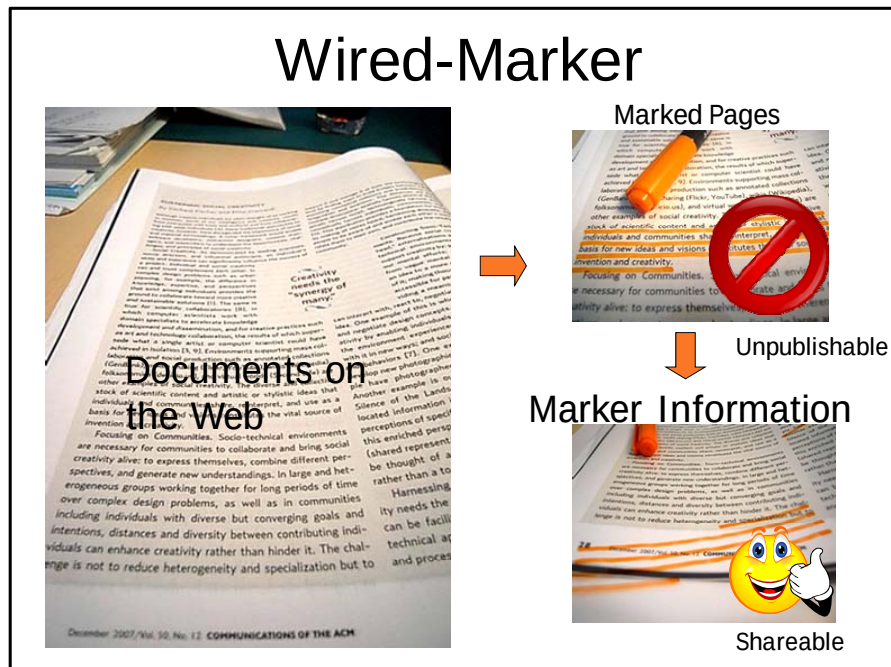


図 3.12 Wired-Marker 概念図

機能 3. 共有、レポート機能

マーカーメタ情報はローカルのブラウザ内に DB として保管されており、利用者の意志で該当部分のコピーもオリジナルページ内容も加えることが可能であり、オフラインでも利用できる。さらにこれらは項目別に自由に形式を指定して xml テキストや表形式で出力することが可能である。

機能 4. 特殊な情報源への対応

しばしば利用される情報源の中には特別なページ表現を行っておりひとつのURLとして特定のページが与えられない場合がある。PDF文献、オンライン教科書、PubMedCentral、商用教科書等がある。サイト別のパースを行いPDFやAjax + JavaScriptにより取得されるコンテンツを含むWebページからの情報収集を可能とした。またPubMedCentralではマーカー付与した文献についてはPubMed経由で自動的に書誌情報を収集しマーカーのメタデータに加えマーカータイトルに文献タイトルが与えられる。Gray's Anatomy(40版)は書籍を購入すれば出版社が提供する特殊な電子教科書サイトを利用することが可能であり、特にBodyParts作成グループや利用者の間で相互に関心領域を引用することを可能にした。

学術文献での言及引用

1) Design and development of SNS system and user profile module for connecting learning research and teaching and classroom practice. S. Yuta *Educational Information and Technology Digital Library*. 2010

2) バルセロナ大学教育シラバス インターネット利用案内

“Mi computadora desde el llavero USB” by Carlos Bravo Reyes. PhD

<http://ddd.uab.cat/pub/dim/16993748n17a7.pdf>

3) イタリアfreeSoftware conference “Atti della IV Conferenza Italiana sul Software Libero, 11 12 Giugno 2010, Cagliari” by D'Andrea, Vincenzo, Dipartimento di Ingegneria e Scienze dell'Informazione, Università degli Studi di Trento,

4) 「真実のための運動? ウィキペディアでの現実社会の建設」オーストリアBielefeld大社会学部 René König博士論文

http://bieson.ub.uni-bielefeld.de/volltexte/2010/1763/pdf/Koenig_09_-_Eine_Bewegung_fuer_die_Wahrheit_-_Diplomarbeit.pdf

その他、雑誌およびネットでの紹介 44件

<http://www.wired-marker.org/article.html>

情報ハブの開発 医学系情報全体を統合利用可能にする情報ハブとして機能する可能性
がある BodyParts3D/Anatomography

Web site: <http://lifesciencedb.jp/bp3d/>

全モデルのポリゴンメッシュデータのダウンロード

<http://dbarchive.lifesciencedb.jp/jp/bodyparts3d/download.html>

解剖学用語は医学系の文献、分子データ、医療記録等を統合整理するためのもっともカバー率の高い索引用語である。しかしながら用語の関係に空間的な関係を多く持つためにオントロジー等のシンボル関係への表現が難しく、またシンボルからヒトの理解への整理表現手法としての能力も劣る問題がある。BodyParts/Anatomography はシンボル表現と利用者のパラメータとして計画された。解剖学概念集である Nomina Anatomica (7700 語)

に対してすべて個別に人体モデル中でのセグメントを与えさらに米国で発達中のシンボル表現 **Foundation Model of Anatomy** を利用してこれらの立体を概念的に離散会合可能にすることを目標にした医学系情報のハブである。

関連分野

画像診断領域ではCTやMRIなどの断面像の濃度差で与えられる臓器輪郭の抽出課題の解決レベルが飽和し、臨床現場でも個人のCTやMRI画像の輪郭を3次元合成した3Dデータは容易に得られる。最近はその延長として **Computer Asisted Diagnosis** と呼ばれる課題が注目されており、輪郭抽出 臓器同定から異常画像（器質的疾患）の検出や同定を目指した研究が盛んに行われている。診断以外の目的で解剖学的に精緻な3Dモデルの構築に関しては **VisibleHumanProject**(NIH 1990-) による特定の検体の詳細な固定断面カラー画像に由来するものが有名であるが、主にこれも輪郭抽出の課題としての研究が盛んに行われた。そのほか教育コミュニケーション用メディカルアートとしてのカノニカル画像集や3D人体モデル表示サービスもいくつか存在する。これらは教育目的や手術シミュレーションの背景として主に商業ドメインで応用が行われている。

一方、脳に限定すれば関心領域(ROI)の共通表現や操作の為に標準座標系を導入する考えは1950年代に始まる定位脳手術に源が求められる。この目的のモデルは標本脳に規準点とグリッドをはじめて持ち込んだ **Talailach** アトラスに始まる (J. Talairach and P. Tournoux, "Co-planar Stereotaxic Atlas of the Human Brain 1988)。この70才の女性の切片アトラスに基づく座標系は後にfMRIなどを使う機能画像診断分野でも広く使われるようになる。**Talailach** アトラスは電子化共有され **BrainMap** などのニューロイメージングデータの統合サイトでも座標系として利用されている。このような展開は **BodyParts3D** プロジェクトの意図するところと共通である。最近では詳細なMRI画像を用いた **Harvard Surgical Planning Laboratory (SPL)** の分節脳MRI3Dデータとそのイメージ視覚化スイート "**Slicer**"(2008)などが多目的に利用されている。しかしながら、解剖学全体をカバーする精緻で分節数の大きなデータはパブリックドメインには存在せず、情報の整理統合共有目的には利用できない。

情報ハブとして3Dモデルを用いる例は最近急に海外で散見されるようになった。出版されている例では "Spatiotemporal integration of molecular and anatomical data in virtual reality using semantic mapping" Sun Center of Excellence for Visual Genomics, University of Calgary, Calgary, AB, Canada; Int J Nanomedicine. 2009; 4: 79-89 のみであるがIBMのプレスリリースは広く知られている。

"3D Avatar to Help Doctors Visualize Patient Records"

<http://www-03.ibm.com/press/us/en/pressrelease/22375.wss>

BodyParts: 解剖学オントロジー Foundation Model of Anatomy (FMA 75,000 concepts)のうち解剖学会が認める7000あまりの概念 (Terminologia Anatomica) を3次元人体モデル中のセグメントとして定義します。シンボル表現とモデル表現間の対応を行うことで以下の利点があります。

- オントロジーが立体データに与える便益
 - 1) 点線面空間穴などを区別して取り扱える
 - 2) セグメントの集合としてカテゴリ概念も表現できる (ex. 副腎=右副腎+左副腎)
 - 3) 未定義の概念も定義済み概念の部分として取り扱える
- 立体定義がオントロジーに与える便益
 - 1) ほとんど説明的な定義を与えられていないFMAやTAに対して見える定義を与えられる
 - 2) トップレベルが難解で階層が深く探索しづらいオントロジーからの概念探索のインターフェイスを与える

Anatomography: web-レンダリング serverでAvatarをmush-up素材として提供します

2007年の開始時には同じ目的のプロジェクトは存在していましたが、2010年現在はIBM's Zurich Research LabのAnatomic-Symbolic Mapping EngineとCalgary大のSun center of excellence visual genomics laboratoryのCavemanが競合目的のプロジェクトとして進行中です。比較すると特徴は以下のようになります。

DBCLS:	分散型統合	FMA	カノニカル	CC-BY
Sun micro CEO	集中型統合	SNOMED-CT	Visible Human	独占
IBM's Zurich :	集中型統合	SNOMED-CT	Visible Human ?	独占

図 3.13 BodyParts/Anatomography の特徴と競合プロジェクトとの比較

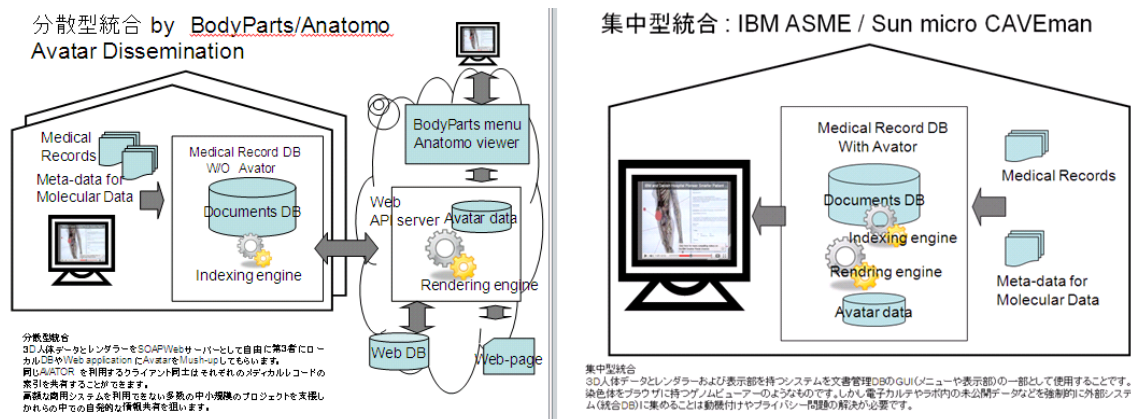


図 3.14 医療医学情報の分散統合と集中統合の比較

BodyParts ポリゴンデータ

神経	感覚	循環	呼吸	消化	内分泌	リンパ	泌尿器	生殖系	骨格	関節	筋肉	外皮	その他	合計
54	4	29	11	47	5	5	9	18	272	7	749	6	3	1,219

独立 FMA/NominaAnatomica 概念数として上記の数だけのデータを作成した。分子データの表現と一般教育用レベルの解像度である。これらはすべて3次元 CAD を用いて DBCLS の研究員が作成してきた。作成指導の為に教科書や web 上のアトラス等の著作を自在に参照できる環境を構築したが、著作権上の問題で一般には公開できない。意識して作者がデータに込めた特徴はリリースノート

<http://lifesciencedb.jp/bp3d/info/userGuide/releaseNotes/release-2.0.html>

およびモデリングノート

<http://lifesciencedb.jp/bp3d/info/userGuide/releaseNotes/modelingNote-2.0.pdf>

として公開した。

また参照用立体データとして各種の人体模型のスキャニングデータおよび Talailach データや複数の連続切片アトラスを取り込んで立体再構成したデータを作成して、研究員のモデリング環境に取り込めるようにしているがこれも著作権情報源とコンタクト中で公開できない内部データである。(評価時にご連絡いただければ内部用の URL と一次パスワードを発行します)

Anatomography

Anatomography は BodyParts データを呼び出してユーザの指示によって著作権のないイラスト画像(Avatar 像) を作るためのレンダリングサーバである。分散型 Avatar の特別な特徴としておもに API 経由での利用によるオンザフライでの描画を企図したサーバであり、メニュー画面を利用して作成した Avatar はパーツ ID や設定パラメータを記載した RESTfulURL として出力され保存または埋め込みされることを前提としている。

Web テキスト、個人ブログ、メール添付、データベースとして外部でも利用されることを前提とし、負荷分散装置を備えたレンダリングサーバを構築した。

Avatars in Web textbook



Avatars in Web DB



図 3.15 オンザフライで Avatar を表示した例 これらはhtml 文であり画像データは持たない。 http://221.186.138.155/project/anatomo5/demo_graysanatomy2/index.html

学術誌での引用言及

“A Framework for Constructing Coordinated Multiple 3D Visualizations on Excel”
Itoh, M. et al. Information Visualisation, 2009 13th International Conference
Issue Date: 15-17 July 2009 pp162 - 170

“Component-based networking for simulations in medical education “von Mammen, Sebastian et.al. IEEE Symposium on Computers and Communications (ISCC), 2010
Issue Date: 22-25 June 2010 On page(s): 975 - 979

“Modular, semantics-based composition of biosimulation models” Maxwell Lewis

Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2009 Michael Y. Galperin et.al, Nucleic Acids Res (2009) 37(suppl_1): D1-D4

次世代サーバの登場により、新たな大量配列解析時代を迎え、遺伝子発現データにも新たな解析手法が必要になって来ている。そこでそのためのプラットフォームとなるシステムの検討も行った。

遺伝子発現を測定する手法としてはさまざまな実験が試みられ、次第にノザンブロットイングなどの定性的な実験から大規模シーケンシングによる発現配列タグ (EST: Expressed Sequence Tag) や DNA マイクロアレイによる定量的な実験手法が用いられるようになり、その大量さゆえにデータを生物学的に解釈することが困難となってきている。次世代高速 DNA サーバの登場によりさらに大量の塩基配列データが発現データとして産生されることが決定的となっている今日、まずは (1) それらの発現データを生物学的に解釈するためのツールが必要であり、さらに (2) それらのために比較解析対象となる信頼できる参照発現データセットを構築することが必要不可欠である。

そこで、次世代サーバより出力された生データを EST 発現頻度のデータとして可視化し、ウェブブラウザで閲覧、一般公開に耐えうるシステムの開発を行った。EST 発現頻度のデータとして加工するまでの流れは Buncher と呼ばれる解析パイプライン(ゲノム配列へのマッピング、複数箇所にヒットするものの除去、塩基座標ごとのヒット数の集計を一度に集計)を独自に構築し、遺伝子ごとに各位置でのヒット数をヒストグラム表示する Buncher viewer を開発した (図 3.16、<http://togoexp.dbcls.jp/buncher/>)。さらに、様々な形態の遺伝子発現データを統合して可視化し、ウェブブラウザで閲覧、一般公開に耐えうるシステムの開発を行った (図 3.17、<http://togoexp.dbcls.jp/RefEx/human/>)。

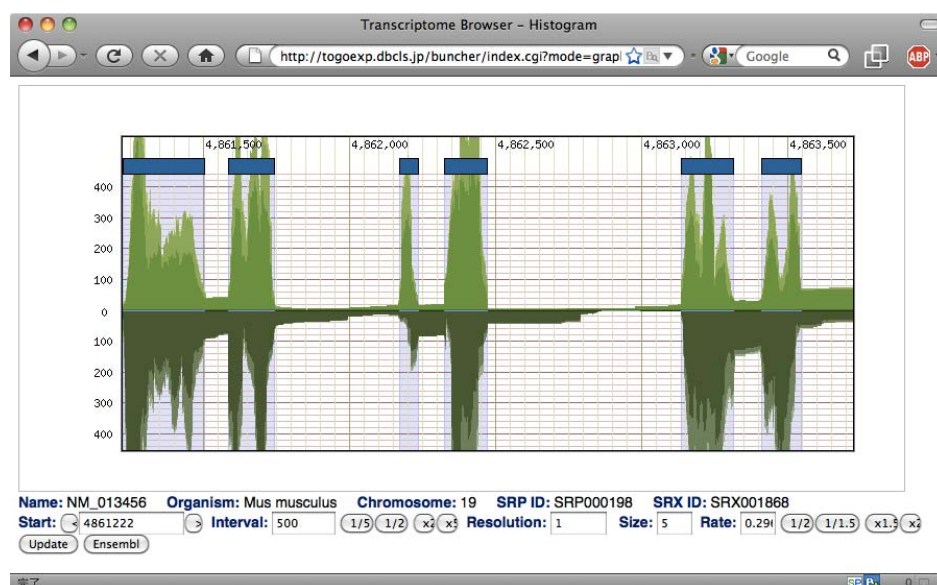


図 3.16 EST ゲノムマッピング表示アプリケーションの Web 画面

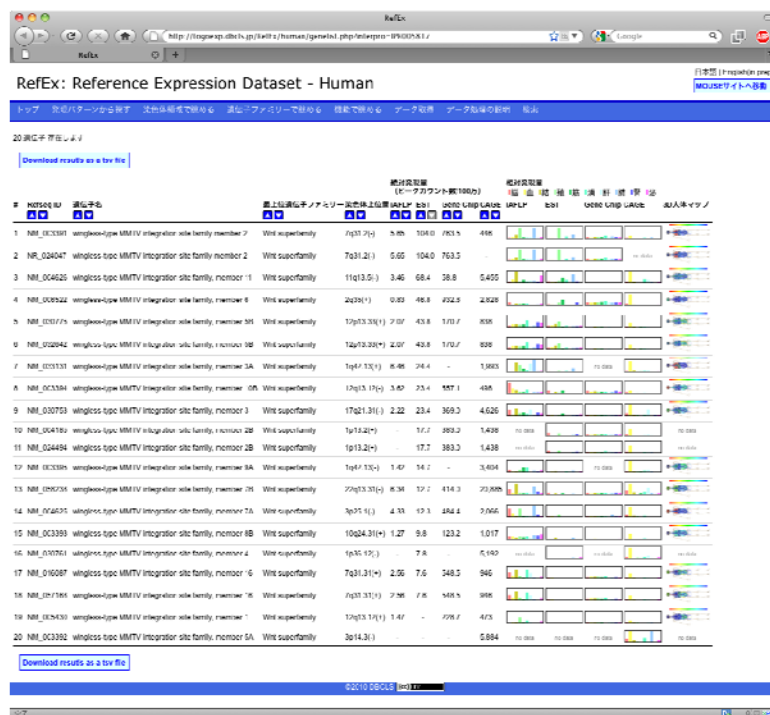


図 3.17 遺伝子発現データの統合可視化システムの Web 画面

現在、ヒトとマウスの正常組織での遺伝子発現データを iAFLP, EST のカウント数、NCBI GEO で公開されている GeneChip によるマイクロアレイに加えて、FANTOM3 で公開されている CAGE データを利用した CAGE タグのカウント数で作成し、可視化を行っている。かつては UniGene の ID で紐付けられることが多かった遺伝子発現データであるが、現状では NCBI RefSeq がよく用いられていることもあり、RefSeq ID ですべてのデータが整理されている。RefSeq のアノテーションとして公開されている InterPro や Gene Ontology の ID やキーワードで検索することも実装されているほか、染色体の位置情報も物理的な位置だけでなく遺伝的距離に基づく染色体領域からも検索できるようになっている。また、BodyParts3D/Anatomography システムを利用した発現部位の「3D 人体マップ」も統合されている。

3. 3. モデル生物・産業応用生物統合DBの開発・運用

3. 3. 1 モデル生物・産業応用生物統合 DB の開発・運用の実施計画

主として微生物を対象とする基礎生物学研究および生物の産業応用研究に携わる利用者がゲノム情報、分子データベース、文献情報等の、研究開発に必要な情報を容易に統合利用可能とする環境を整備する。

3. 3. 2 モデル生物・産業応用生物統合 DB の開発・運用の実施内容

平成19年度

ゲノム解析が進んでいる多様な微生物菌株を当面のターゲットとして、統合DB の開発・

運用に着手した。微生物菌株については、1) 微生物ゲノム情報の統合利用・解析環境の調査、2) 微生物情報資源の機能、ならびに3) 微生物ゲノム論文からの知識抽出について調査して、微生物に関して、系統保存の情報からゲノム解析の結果の情報までを統合利用可能とする環境構築の要素を整備した。特に、次世代サーバの開発と普及がもたらす膨大なゲノム情報から有用な生命情報を効率よく抽出するために必須と考えられる自動アノテーションシステムについて調査した。

平成20年度

平成19年度の調査以来、次世代サーバの普及によって比較的小規模な研究室やグループにもゲノム解析への道が広がってきた。一方で、シーケンシングセンターにおいてさえ時間を要していたアノテーションは、依然として、微生物ゲノム解析のボトルネックになることが明らかであった。そこで、アノテーションの基準にもなりうる微生物ゲノム自動アノテーションのパイプラインMicrobial Genome Annotation Pipeline (MiGAP)の開発に着手した。海外の類似サービスに比べて、ゲノム解析の初心者から専門家まで活用可能なように予め、3つのレベルの利用者パラメータを用意した点と、国際塩基配列データベースへの登録ファイルの自動生成機能がMiGAPの特徴である。

平成21年度

MiGAPには、デフォルト解析の初級用のブロンズ (b-MiGAP)、パラメータを設定可能とする中級用シルバー (s-MiGAP)、さらにデータベースやアルゴリズムを付加することができる上級用のゴールド (g-MiGAP) の3つのレベルを実装したが、2009年6月にMiGAPの初心者用バージョンを公開し、一定の質が担保されたアノテーションの普及を目指した。公開後、利用者からの要望を受けて、MiGAPに20回を超す改訂を加えた。また、b-MiGAPの利用状況を見ながら、b-MiGAPでの利用経験を積んだ利用者にs-MiGAPを公開した。G-MiGAPについても、ゲノム解析の経験が深い利用者に公開した。

平成22年度

MiGAPの利用実績が高まってきたため、解析サーバを国立遺伝学研究所の大規模PCクラスターの一部に移行して、複数ゲノムの並行解析を可能とした。また、既知の情報が蓄積されているデータベースを分散させることによって、解析対象配列からの高速な参照を可能とした。解析の高速化に加えて、真核微生物ゲノムのアノテーション機能も加えた。これらによって、利用者の利便性が大幅に高まった。

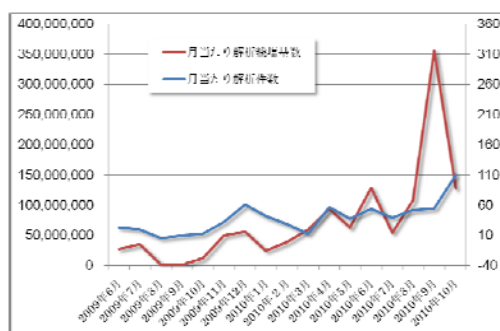


図 3.19 MiGAPの利用実績

公開後 17 カ月間の解析総件数が 658 件、解析総塩基数が 1.2Gbases 余り、予測した CDS 総数が 110 万個余りであった。解析件数と解析塩基数については月ごとの変動を図 3.19 にまとめた。比較的大きなゲノムを集中的に投入されることもあり、月ごとの変動が大きい。解析件数および解析データ量ともに着実に増加してきている。今後、近縁種ゲノムの知識の重みづけや既知遺伝子のアミノ酸配列マッピングの機能拡充といった利用者からの要望に応じていく。

4. 統合データベース支援

4. 1 ポータル整備・運用、広報、普及啓発

4. 1. 1 ポータル整備・運用、広報、普及啓発の実施計画

統合データベース支援として業務計画書に定められた、ポータルサイトの構築運用、日本語による研究情報の流通促進、広報・普及啓発を実施するため、下記のように各年度の実施計画を設定した。

平成 19 年度

- ・ 統合DBプロジェクトポータルサイトの開設(<http://lifesciencedb.jp>)。
- ・ 横断検索システムの開発。
- ・ ライフサイエンス関連情報の自動収集システム（Webクロール）の開発。
- ・ 日本語ポータルシステム開発に必要な言語情報資源の電子化。
- ・ 事業内容や趣旨を解説したパンフレットを発行し学会等で普及する。
- ・ 意見や質問に対応する窓口の設置。
- ・ 教育コンテンツの作成と講習会等の開催。

平成 20 年度

- ・ ポータルサイトの整備運用の継続。
- ・ 日本語総説や学会要旨、報告書等、生命科学に関する文献や関連情報の収集。
- ・ 横断検索の対象DB数の増加とパラメータの改良。
- ・ 動画教育コンテンツ「統合TV」の構築と運用。
- ・ データベースやツールに関する講習会の開催。

平成 21 年度

- ・ ポータルサイトの整備運用の継続。
- ・ 横断検索対象DBの充実化とキーワードの日英翻訳機能の強化。
- ・ 統合検索に向けてメタデータやキーワードを利用した検索手法の検討。
- ・ 参加機関や関係各省の統合プロジェクトとの検索機能の連携。
- ・ 各地での講習会を開催する。
- ・ 教材コンテンツの作成継続と内容の更新。
- ・ 国内の主要学会での広報活動並びに、種々の媒体を利用した統合DBの普及
- ・ アンケートの実施と意見の反映。

平成 22 年度

- ・ ポータルサイトの整備運用の継続。
- ・ 最終年度にあわせた成果のとりまとめと公開。
- ・ 横断検索の運用と、プロジェクト成果の統合を中心とした統合検索プロトタイプ開発。
- ・ 学会要旨等の日本語コンテンツの充実化。
- ・ 動画教材の製作継続と英語版の作成
- ・ 講習会や講演会のプレゼンテーションの公開
- ・ 主要学会に加え特定の分野の学会でのブース出展や、各種媒体を利用した広報活動、公開シンポジウムの開催。
- ・ コンテンツや広報資料の英語化による海外への普及。
- ・ アンケートの実施と意見の反映。

4. 1. 2 ポータル整備・運用、広報、普及啓発の実施内容

(1) 統合 DB プロジェクトポータルサイトの整備・運用

1) ポータルサイトの構築、整備運用

統合 DB プロジェクトの成果を公開するウェブサイトとして平成 19 年 10 月より <http://lifesciencedb.jp> のドメイン名を取得しライフサイエンス統合データベースセンター

のサーバより運用開始した(図 4.1)。このサイトは、中核機関だけでなく分担機関、補完課題すべての機関の成果を集約したサイトである。平成 22 年 9 月現在で、58 件のツールやサービス、成果物、関連サイトへのリンクがあり、そのうち 32 件が中核機関からのサービスやコンテンツである。各成果物はポータル、検索、データベース、ツール、というように分類表示している。



図 4.1 統合データベースプロジェクトサービス公開サイトの変遷

中核機関のポータル系のサービスについては、パネルやツリーを利用したユーザインターフェイスと、TogoDB によるデータベース化によって、使用感の統一を図っている。平成 20 年には英語版のポータルサイトを開設した。

2) データベースカタログ

ポータルサイトの重要な機能として、どこにどのようなデータベースがあるのか、利用者が目的のデータベースを簡単に探すことができることが必要である。データベースカタログは国内のデータベースに関する情報を、各データベースの内容をもとに、分子種や生物種、運営機関ごとに調べることでできるサービスである。H19年度はパラメータを開発するとともに、科学技術連携施策群 生命科学の基礎・基盤 補完的課題 生命科学データベース統合に関する調査研究の調査データをもとに、国内350件、海外50のデータベースに関する情報を公開した。その後、データベースについての調査を継続し、平成22年9月現在で912件のデータが公開されている(図4.2)。平成23年度からは4省庁連携の枠組みで調査データを作成していく予定である。またパラメータについては検索エンジン対策のために詳細データの表示にMediaWikiを利用し、JSTで構築されてきたWINGproとの統合を行う。



図 4.2 データベースカタログのパラメータ

データベースカタログと同一のパラメータを利用して、生命科学系の学協会情報と生物アイコンの公開も同時に行っている。

3) 生命科学データベース横断検索

横断検索は統合の第一段階として国内 DB を中心に、各 DB はそれぞれのサイトに置いたまま、キーワード検索を可能にした検索サービスである。平成 19 年度 13 件のデータベースを検索対象にスタートしたが、平成 20 年度は 200 件に、平成 22 年 9 月現在で 270 件の DB が、一括して検索可能である。そのパラメータを図 4.3 に示す。入力した検索キーワードは自動的に日英翻訳されるため、相手のデータベースが日本語か英語かを気にする必要は無い。

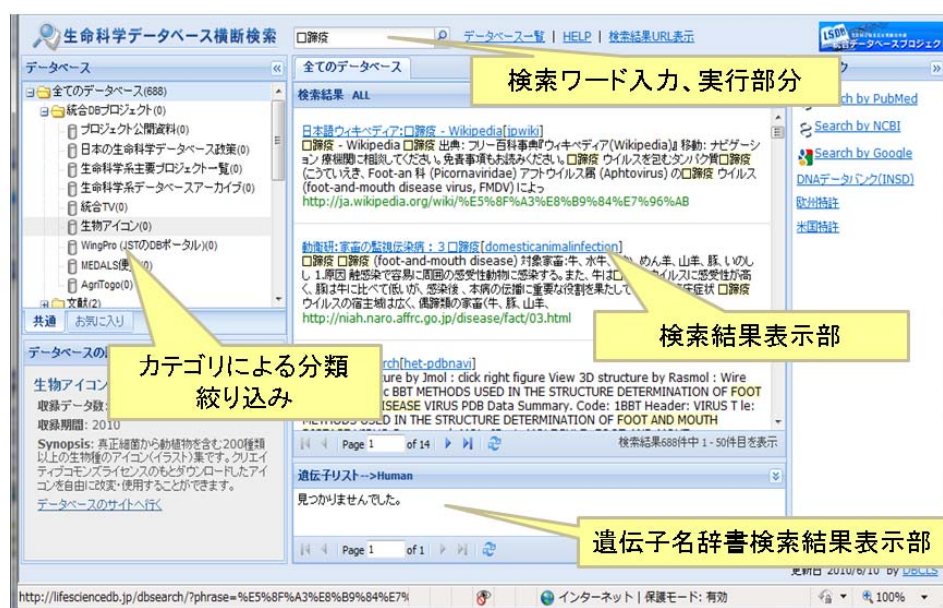


図 4.3 生命科学データベース横断検索のパラメータ

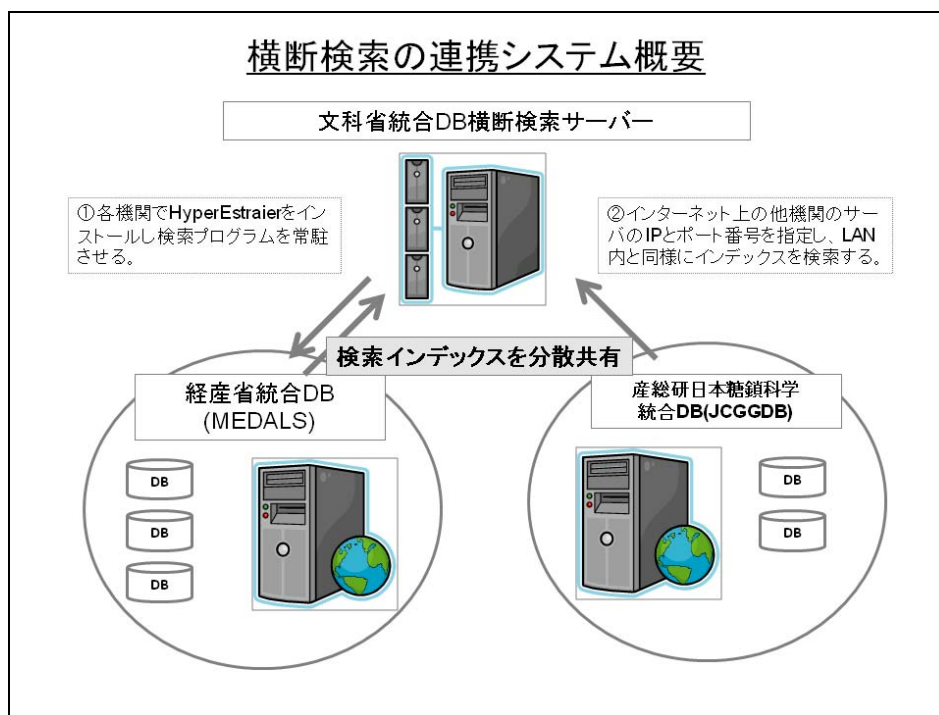


図 4.4 横断検索における連携システムの概要

左側のパネルには対象としているデータベースをカテゴリに分類してツリー表示し、利用者が目的分野の検索結果にたどりつきやすいようになっている。カテゴリの下に各データベースの概要とリンクを表示するパネルを設置した。中央のパネルに検索結果が表示され、カテゴリーツリーをクリックすることによって該当するカテゴリの検索結果が新規タブとして表示されるようにし、利用者が前の絞り込み結果に戻れるように配慮した。平成 20 年度には補完課題の糖鎖科学統合 DB とは共通の検索エンジンで開発を進め、インデックスの共有による検索連携を行った。平成 22 年度には経済産業省統合データベースプロジェクトにも技術情報を提供し検索連携を実現、利用者は省庁の区別なくデータベースを横断的に検索可能となった（図 4.4）。

4) 日本語オリジナルコンテンツの作成並びに英語サイトの作成

ポータルサイト構築にあたって、データベースの検索だけでなく、独自の日本語情報資源やオリジナルコンテンツの制作も行った。H19 年度から生命科学分野の日本語総説誌である蛋白質核酸酵素（共立出版）のバックナンバー電子化並びに検索サービスを開始した。また、文部科学省特定領域研究「ゲノム」報告書を公開し、横断検索の対象として分子データベースや特許公報と合わせて検索可能にした。報告書に関しては平成 22 年度よりウェブ上で公開されていた CREST やさがけの報告書についても横断検索対象とした。

データベースの利用促進を目的として、平成 19 年 7 月にはデータベースやウェブツールの使い方を動画で解説する統合 TV を開始した。ウェブで公開されているものの中にはヘルプが不足していて使用法がわからないものも多くあるが、本サービスは実験系の研究者にあわせて使用方法を解説しており、大変よく利用されるサービスとなった。書籍で出版される解説文書はすぐに古くなってしまふなどの欠点があったが、本サービスでは動画の更新も定期的に行なっている。平成 22 年 9 月現在、366 件の動画を公開しており、公報普及の場ともなっている。

(2) 広報、普及啓発

統合データベースプロジェクトの普及のために、実際にインターネットにつながった端末を前にして主として初心者（大学後半から大学院前半が中心、ただし企業等からの参加も OK で、特に制限は設けない）向けに行う統合データベース講習会および単独講師による短いバージョンあるいは実習なしの講演会を平成 19 年度から開催することを計画し、19 年度末に第 1 回の講習会を開催、平成 20 年度からは年間を通じてコンスタントに開催してきた(表 4.1)。

講習会の内容は、統合データベースプロジェクトそのものあるいは本プロジェクトで開発されたサービスの解説を主体とした初心者向けの内容が中心であるが、広くデータ解析に利用されているウェブサイトやツールの解説なども盛り込み、研究の初心者にも役立つ内容を幅広く提供できるように努めた。一方で、平成 21 年度からは 2 日間にわたる講習会を数回企画し、全般的な通常の講習会形式（1 日目）と質問への回答形式（2 日目）によってやや高度な内容にも対応できるような方法も試みたほか、Nucleic Acids Research 誌データベース特集号およびウェブサーバ特集号の解説を中心としたやや専門的な勉強会も実施し、中上級者向けのニーズにマッチした構成の会も設けた。平成 21 年には、前年の展示会等で要望が多かった、大学等でのバイオインフォマティクスの課程の講義としての講演会も数回実施した。これ以外にも、DDBJ や PDBj、産総研主催講習会へ講師としての要請等、機会があれば積極的に参加し、先方の要望に沿った専門的な内容についてはプロジェクト参画機関の協力も得ながら、様々な形式の普及活動に努めた。

会場としては、DBCLS の教室（東京大学本郷キャンパス浅野地区）以外にも、ご協力いただける共催機関（大学や、公的研究機関など）の PC 施設や教室をお借りし、可能であれば学外・機関外からの参加者も受け入れて実施した。参加者の募集は、講習会用のウェブサイトでの告知、プロジェクト成果公開サイトでのニュース配信のほか、共催機関での告知やバイオインフォマティクス関連のメーリングリストなどを利用した。また講習会用のポスターも作成し、DBCLS のある東大キャンパスや共催機関での掲示をお願いした。講習会等の参加者の中には統合データベースプロジェクトに興味を持った学生もいて、その後アシスタントとしてプロジェクト業務に従事した者も複数名現れた。

表 4.1 統合データベース講習会等の実施内容

年 度	統合 DB 講習会（回）	講習会のべ参加者数（人） （ustream による参加者数）	統合 DB 講演会（回） [#]	統合 DB 勉強会（回）
19	1	20	2	
20	6	232		2
21	7	210	5	1
22	7 《予定含む》	131 (413) 《実施分のみ》	2	

[#]他機関からの要請分含まず

（講習会実施共催機関等：科学技術振興機構、日本大学生物資源科学部、長浜バイオ大、北海道大学、九州大学、東京工業大学、基礎生物学研究所、大阪府立大学、福岡県バイオ産業拠点推進会議、金沢大学、東北大学、筑波大学[予定]）

統合データベース講習会等で使用したテキストやその様子を録画した動画は、教育/人材育成の各参画機関において作成された教育プログラムを集約して平成 18 年度より公開しているウェブサイト「MotDB」から利用可能となるよう維持管理し、常に最新の情報が提供されるようにした。動画の一部は「統合 TV」のコンテンツとしてあとからいつでも視聴できるようにしたほか、ustream を利用して開催と同時の配信（中継）も行った。

プロジェクトの成果を集めたウェブサイトの本格的な公開(2008 年 6 月)に合わせて、平成 20 年度から、プロジェクトで開発したデータベースやサービスの紹介やその他データベース統合化に必要な活動を紹介するために、学会年会におけるワークショップやフォーラムの企画・運営、ブースの出展やランチョンセミナーの開催といった広報活動を開始した(表 4.2)。平成 20 年度には、ユーザとなる研究者が多く出席するであろう、BioJapan2008（商用のバイオ関連の大型展示会の一つ、日経主催）、生物物理学会年会、BMB2008（分子生物学会と生化学会の合同年会）、農芸化学学会年会等での活動を行った。平成 21 年度は、BioJapan に替えて国際バイオ EXPO（もう一つの商用の大型展示会、イベント会社主催）へ出展し、関連すると思われるいくつかの学会へのブース出展を追加した。平成 22 年度は、効果の点から考えて再度 BioJapan2010 へ出展し、個別のサービスに関連の深い中規模の学会へのブース出展を新たに試みた（平成 22 年度はプロジェクト最終年度のため、12 月で出展等の活動を終えるように計画した）。これらの展示会等においてはアンケートを実施し、サービス改善のための情報を得るとともに、ユーザ評価者（2.2.2 項参照）の募集を行った。なお、22 年度の BMB2010 では、プロジェクト内外のライフサイエンスデータベース関連機関と相談のうえ、通路を挟んだ 6 つのブースを並べて出展することを試みる。また、多数のユーザが参加すると思われる分子生物学会年会（あるいは BMB[生化学会との合同年会]）ではブース出展のほかに、サービスの紹介などのために DBCLS 研究員や参画機関が多数のポスター発表も行ってきており、平成 20、21 年度大会では同一会場で 20 件以上の連続発表となるようにアレンジしてアピールしてきたが、22 年度大会（2010 年 12 月実施）ではテーマごとに分散する形で発表することとし、会期中様々な場面で本プロジェクトの成果がユーザの目に触れるようにする。

表 4.2 広報活動まとめ

年度	内容	回数（対象学会・展示会名称）
20	学会等企画提案	3（進化学会、生物物理学会、BMB2008）
	ブース出展	5（BioJapan2008、情報とシステム 2008、サイエンスアゴラ、BMB2008、農芸化学会）
	ランチョンセミナー	1（農芸化学会）
21	学会等企画提案	2（分子生物学会/2 件）
	ブース出展	9（国際バイオ EXPO、セルロース学会、脂質生化学会、癌学会、生化学会、生物物理学会、情報とシステム 2009、分子生物学会、農芸化学会）
	ランチョンセミナー	2（生物物理学会、農芸化学会）
22	学会等企画提案	3（生物物理学会、BMB2010 [#] /2 件）

	ブース出展	8（脂質生化学会、蛋白質科学会、CBRC2010、BioJapan2010、Biocuration2010、生物工学会、人類遺伝学会、BMB2010#）
	ランチョンセミナー	2（脂質生化学会、生物物理若手夏の学校）

#の BMB2010 は 2010 年 12 月開催のため予定

プロジェクトに関する情報発信（表 4.3）としては、これまで文科省の記者会室を通じたプレスリリースを 3 回（プロジェクト成果公開サイトの本格開始、経産省統合データベースサイトとの相互検索の開始、4 省関連機関での合同開催シンポジウムの案内）実施した。また、プロジェクト成果公開サイト、DBCLS のサイトおよび JST の「文科省統合データベース整備事業サイト」を通じたニュース配信を行うとともに、そのうちの一部は日経の *Biotechnology Japan* のサイトを利用した PR も実施した。22 年の 4 月には、DBCLS のウェブサイト进行全面改定したが、その際に DBCLS サイト、プロジェクト成果公開サイト、さらに DBCLS 用を取得した Twitter アカウントへ同時にニュースが流れるように工夫し、情報発信の効率化を図った。Twitter アカウントでは、このほか、DBCLS メンバーが発表する外部の講演会等の情報や、DBCLS 主催のイベントの実況中継なども配信している。プロジェクトに関するイベント等の案内や人材育成プログラムへの参加者募集には、展示会等でのアンケートの際承諾をいただいた方のメーリングリスト、製薬企業を中心としたメーリングリスト、「ゲノム」特定領域研究の班員のメーリングリスト等を利用した。本プロジェクトのサービスを紹介した総合パンフレットについては平成 20 年 7 月に初版を作成し、その後毎年度の 6 月ごろまでには内容を改訂して、展示会や講習会等での配布を行った。また、個別のサービスに関するリーフレット（A4 サイズ 1 枚もの）を参画機関の協力も得て現在までに 43 種類[うち英語 9 種類]作成した。リーフレットについては DBCLS で編集・印刷するものとし、展示会等大きなイベントの前には新サービスの追加や内容の改訂を行い、情報を随時アップデートできるようにしている。また、これら総合パンフレットやリーフレットの最新版は、プロジェクト成果公開サイト「統合ホームページ」から自由に閲覧・ダウンロードできるようにしている。平成 20 年度 3 月に公開したプロジェクト成果公開サイトの英語版においては、プロジェクト自体の説明とともに主にプロジェクトで独自に開発したサービスのうち英語でも利用が可能なサービスについて紹介しているが、サイトでの案内機能の充実のために、説明資料（リーフレットや説明動画統合 TV）の英語化を行い、これらも英語版のプロジェクト成果公開サイトより閲覧・ダウンロードできる。

表 4.3 情報発信について

年度	プレス リリース（回）	ニュース 配信（回）	Biotech. Japan（回）	パンフレット 配布数（部）	リーフレットの 種類(英語版) 累計
20	1	62	13	3800	27
21		84	9	4300	35（9）
22	2	57 [#]	6 [#]	2400 [#]	43（9） [#]

[#]2010 年 11 月現在の数値

4. 2 データベース受入

4. 2. 1 データベース受入の実施計画

我が国のプロジェクトデータベースについて、データベースのタイプに応じた方法で、データベース受け入れを実施する。

平成19年度は、我が国のプロジェクトデータベースの現状を調査し、それらをそれぞれの形態に従って分類し、それに基づいた受け入れ方法を検討する。また、どのデータベースから優先的に受け入れるべきかその優先順位付けの方法を立案し、それに基づきデータベースの受け入れと運用を試行する。

平成20年度は、平成19年度の調査・検討をさらに進めるとともに、実際にデータベースのタイプに応じた方法で、データベース受け入れを実施する。さらに、受け入れたデータベースの運用システムをポータルサイトに構築し、運用を行う。

平成 21 年度は、プロジェクト参加機関のデータベース受け入れと連携を進めると同時に、より広い範囲の機関からのデータベース受け入れを実施する。また、前年度開発した受け入れデータベースの公開システムに、検索機能・横断検索との連携機能等の拡張を行い、運用を行う。さらに、我が国で実施された主要な生命科学系プロジェクトの調査を実施し、その結果を公開する。

平成 22 年度は、受け入れ対象をさらに多様・複雑なデータベースに広げるとともに、公開システムについて検索機能・他サービスとの連携機能等の拡張を行いながら運用を行う。また、主要な生命科学系プロジェクトの調査を広げるとともに、データベース受け入れに関連する他機関との連携を進め、これらのコンテンツについてデータ統合に向けた整備を進める。

4. 2. 2 データベース受入の実施内容

1) ライフサイエンスプロジェクト及びデータベースの調査

データベース受け入れの対象や方法を検討するために、我が国のプロジェクトデータベースの現状を調査した。まずデータベースカタログデータの補充を行う形で、平成19年度に内閣府の連携施策群成果リストに理研、ゲノム特定領域(新旧)、JST SNPポータル、JBIRC、CBRC、かずさDNA研究所、農業生物資源研究所等のリストを追加し、平成20年度には分担機関である東京医科歯科大学が実施したアンケートのデータベース情報からデータベースカタログに未収録のデータベース情報(15件)を登録した。

また、受け入れ対象データベースの把握のために、データベース自体の調査に並行して、ライフサイエンス国家プロジェクトの事業や成果に対する調査を行い、「生命科学系主要プロジェクト一覧」(図4.5、http://togodb.dbcls.jp/lbdb_project)として公開した。まず平成19年度に、また、多くの予算が投入されてきた約10年間のライフサイエンス国家プロジェクトの事業や成果に対する調査を行い、プロジェクトの期間、概要、予算、代表研究者、公開の状況、文献等に関する情報を収集した。平成20年度は、この「主要プロジェクト一覧」の英語化を実施し、各項目による検索やソートが可能なパラメータを前年度構築したTogoDBを用いて作成し公開するとともに、文部科学省のプロジェクトについてのデータ内

容の追加(10件)を行った。平成21年度は、主要プロジェクト一覧のデータ内容の更新と22件の追加（文科省18件、経産省4件）、及びそれらの英語化を実施した。平成22年度は文科省所管のプロジェクトデータ19件を更新、2件を追加し、これらのデータを前年度公開分22件とともに英語化した。

The screenshot shows the L5DB (Life Science Data Base) project overview page. It features a search bar, a table of projects, and navigation links. The table lists projects such as 'Genome Network Project', 'Genome Network Project', 'Genome Network Project', and 'Genome Network Project'.

プロジェクト名	開始年度	終了年度	概要	データベース公開状況	公開データベース	データダウンロードサイト	公開ホームページ	実施者
ゲノムネットワークプロジェクト	2004	2008	遺伝子の発現調節機構に関する網羅的な解析	公開	ゲノムネットワークプロジェクトデータベース(http://genomeweb.nig.ac.jp/)	ゲノムネットワークプロジェクトデータベース(http://genomeweb.nig.ac.jp/public/download/dataset.html)	ゲノムネットワークプロジェクトデータベース(http://genomeweb.nig.ac.jp/public/contents/release.html)	国立研究開発法人理化学研究所
動物系統生物ゲノム情報データベース	2006	2010	動物系統生物ゲノム情報データベースの構築	公開	AgriTGOデータベース(http://tgo.dna.affrc.go.jp/)	AgriTGOデータベース(http://tgo.dna.affrc.go.jp/download.html)	-	農研機構
カイコゲノム解析プロジェクト	1994	-	カイコゲノム解析プロジェクトの推進	公開	カイコゲノムデータベース(http://kikocodb.dna.affrc.go.jp/)	カイコゲノムデータベース(http://kikocodb.dna.affrc.go.jp/cgi-bin/search_L5DB.cgi)	カイコゲノム解析プロジェクトデータベース(http://kikopass.dna.affrc.go.jp/)	農研機構
家畜ゲノム解析プロジェクト	1999	2001	家畜ゲノム解析プロジェクトの推進	公開	Animal Genome Database(http://animal.dna.affrc.go.jp/)	-	家畜ゲノム解析プロジェクトデータベース(http://animal.dna.affrc.go.jp/)	農研機構

図4.5 生命科学系主要プロジェクト一覧

また平成19年度には、本事業の分担機関や補完課題実施機関以外で整備・構築された国内データベースを対象として、データベースの内容、維持・更新状況、提供可能性及び提供時の問題点などを調査し、受入による統合化の方針策定に役立てるとともに、受入対象の候補の抽出を目的としてアンケート調査を実施した。データベースコンテンツと検索方法、データベースの経歴・管理状況、データベース維持の将来計画、提供可能性及び条件、提供方法、受入時の連絡先について質問し、65/104 (62.5%) の回収数(率)を得て、以下のようなアンケート結果を得た。データベースの利用条件では、大半は1次利用に対して制限を設けていないが、2次利用や再配布に対しては半数近くが制限を設けている。今後については、大半が現状維持・データの増強を検討しているが、平成20年度以降もプロジェクトが継続しているのは全体の30%程度にすぎない。また、半数以上が、統合データベース事業に対して全データもしくは一部データを開示できると回答している。提携条件としては、データ更新権限や著作権などの留保、利用条件の継承、統合データベース事業による維持・管理・提携費用の負担を挙げる回答が多かった。

さらに平成20年度には、Nucleic Acids Research誌 (NAR) に収録された国内のライフサイエンスデータベース (77件) について、利用許諾やダウンロードの可否等の調査を行った。その結果、データダウンロード可能なデータベースは約6割、何らかの利用許諾条件が明記されているものは約3割であることがわかった。NARに収録されていないデータベースでは、ダウンロードダブルなものの割合はさらに減少すると考えられ、データの再利用がしにくい状況であることがわかった。

2) ライフサイエンスデータベース受入方法の検討

上記の調査に基づいて、データベースをそれぞれの形態に従って分類し、それに基づいた

受け入れ方法を検討した。

平成19年度は、データベースの分類に基づく受入優先順位の検討と受入方法の検討を行った。まず、収集したデータベースリストに基づき、データベースの対象を分類し、さらにDDBJなどの公開バンクに登録されていないオリジナルデータがあるかないかに基づいて、受入の優先順位の検討を行った。その結果、DNA配列や発現のデータが最もオリジナルデータの割合が多いことがわかった。ここでは、DNA配列の中でも日本が優位性を持つcDNAに着目し受入の試行を行うこととした。次に、データベース作成者との間の様々な状況に応じた受入方法の検討を行った。受入や連係の方法として、一括してデータを集める場合とデータベースが分散している場合に分けて検討した。データを集める場合は、1)データだけをFTPでダウンロード可能にする場合とデータベース自体を移設する場合、2)データベースの型を揃え、管理はそれぞれのデータベース構築者が運用する場合（データベース受入テンプレートの利用）3)データベースの型を揃え、統合した形で当ユーザが構築管理を行う場合、いわゆる集中型統合の場合を検討した。また、データベースが分散している場合の連係方法としては、1)横断検索、及び2)API等を用いる分散型の統合検索を検討した。

平成 20 年度はこの検討を受けて、集中型統合で対応が可能なデータベースに関しては後述するアーカイブの形で受け入れを開始した。その過程で、ほとんどのデータベースはアーカイブの形で受け入れて公開することができると考えられたが、今後データがさらに多様化するとアーカイブ以外の受け入れ方法を検討することが必要であると考え、平成 21 年度、クラウドを用いる方法も含めて費用対効果などの面から新たに調査と検討を行った。その結果、現在主に用いている簡易検索機能を併用した形に比べると、自前サーバもしくはクラウドを用いたミラーサーバの構築は、個別データベース独自の機能を移管できデータベース寄託者側からの支持は得易いが、データベース間で利用法に統一感がないためユーザからの支持は得難く、またコストが高いことがわかった。しかし今後受け入れるデータベースの内容によっては、クラウドを用いたミラーサーバ構築も検討の余地があると考えられた。

3) 受け入れデータベースの運用システムの構築

以上の調査と検討に基づいて、受け入れデータベースの運用のために、受け入れたデータベースをダウンロード可能な形（アーカイブ）で公開するサービス「生命科学系データベースアーカイブ」（図 4.6、<http://dbarchive.lifesciencedb.jp/>）を構築、公開した。また、運用の過程で各種の機能拡張を進めた。



図 4.7 データベースのメタデータの例



図 4.8 データのメタデータの例



図 4.9 利用許諾のページの例



図 4.10 簡易検索ページの例

平成 21 年度は、サイトの名称を現在の「生命科学系データベースアーカイブ」に改めるとともに、各種の機能拡張を行った。

i) 簡易検索ページの機能の拡張と応用：簡易検索ページで用いているデータベーステンプレートシステム TogoDB に関して、まずユーザ側の利便性向上のために、テーブル間の 1 対多リンク、リスト表示での件数の設定、詳細表示でのセルの高さの切替などを設けた。図 4.11 に、Dicty_cDB（粘菌の cDNA データベース）の例を示す。Contig テーブル（図 4.11-(a)）から Clone テーブル（図 4.11-(b)）へと 1 対多リンクを張ることにより、例えば cDNA ライブラリプロフィールによりコンティグを検索し、その中に含まれる各クローンの配列等の情報へと辿ることが可能になっている。また Contig 詳細情報ウィンドウ（図 4.11-(c)）では、情報量の多い項目のセルの高さを切替可能にして見易くしている。

このようにして、複雑で多様なコンテンツを有するデータベースについても、利便性をあまり落とすことなく TogoDB の枠組みに収めることができるようになった。さらに、データベース管理者側の利便性向上のために、テーブルや詳細ウィンドウでの非表示やダウンロード非対象を項目毎に設定できるようにした。また、テーブルに新たに項目を追加しその項目のデータのみをアップロードできるようにした。これらの拡張によって寄託ごとに異なる様々な状況に効率よく柔軟に対処できるようになった。

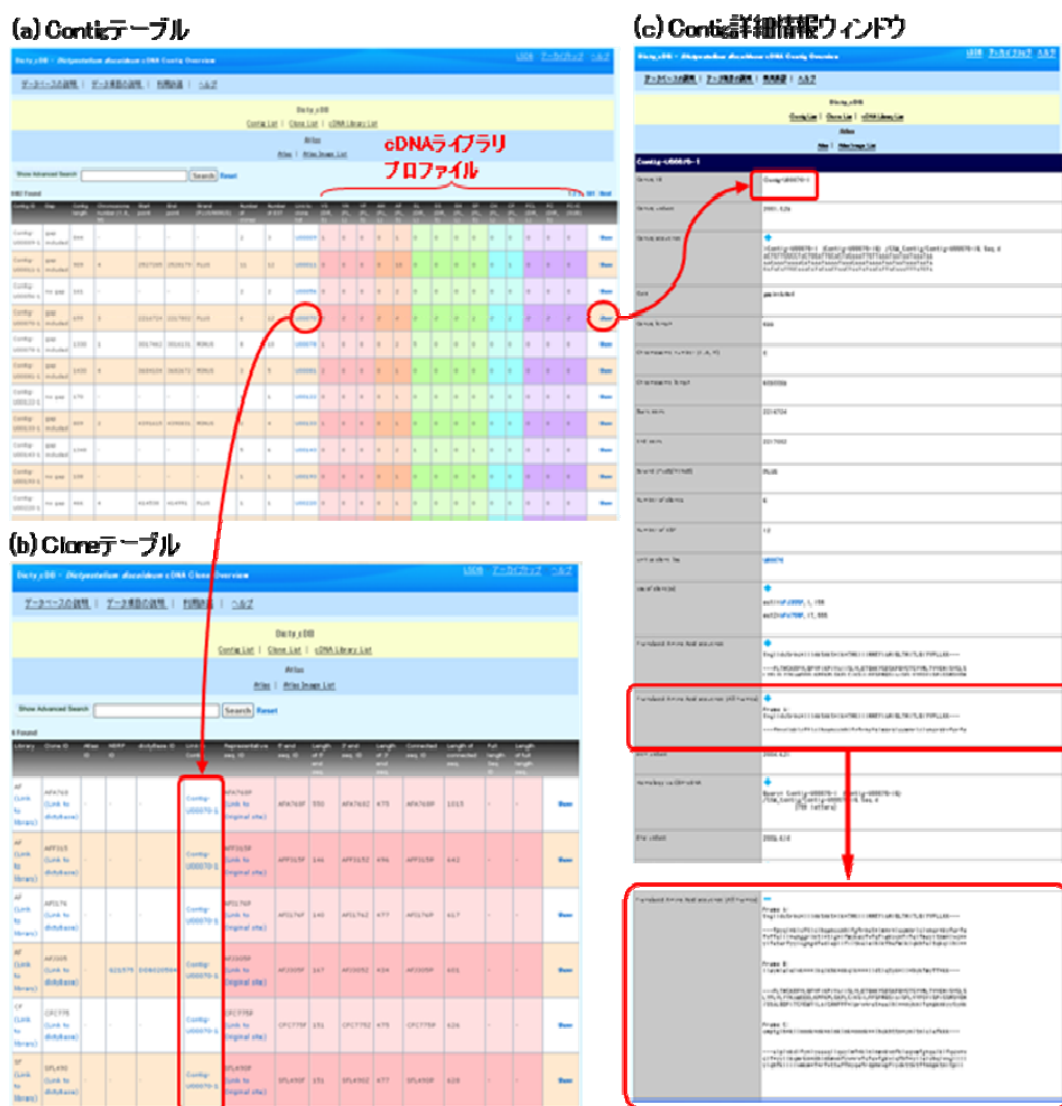


図 4.11 TogoDB のパラメータの機能拡張 (Dicty_cDB の例)

ii) アーカイブサイトの改定：データベースアーカイブサイトのユーザ向けの改定としては、サイトのトップページにアーカイブ一覧を設置し、関連項目を隣に配置し、オリジナルデータベースの Top ページの画像を左端に置くなどの工夫により、見やすく使いやすいページになるよう改定した。利用許諾について、寄託者の手間を低減するために、一般的に採用されやすいと思われる標準利用許諾と追加利用許諾の組み合わせを利用許諾テンプレートとして作成し公開した (図 4.12)。また寄託者との間の交渉をスムーズに進めるために、覚書の文言を簡素化した寄託規約を作成し (図 4.13)、寄託者の了解を取り付ける形の契約締結を行うこととした。これら利用許諾テンプレートと寄託規約の採用により、

寄託者との間の寄託契約の締結を、迅速にかつ一定の様式で進めることができるようになった。また、アーカイブサイトの構築と運用を管理するために、コンテンツマネジメントシステムの構築を行った。これまで、各データベースのメタデータや利用許諾等のページ情報を個別に管理していたため、枠組みや共通の文言の修正を行うときにも個別に修正を行う必要があった。そのため、次第に寄託データベースの数が多くなってきて管理運用に非常に手間がかかるようになってきた。そこで、各データベースのメタデータやその他の情報を一元的に管理する仕組みを構築した。このコンテンツマネジメントシステムにより、今後ますます増加するアーカイブ化作業において、重複する手間を減らすことによる効率化と間違いの低減を実現することができるようになった。

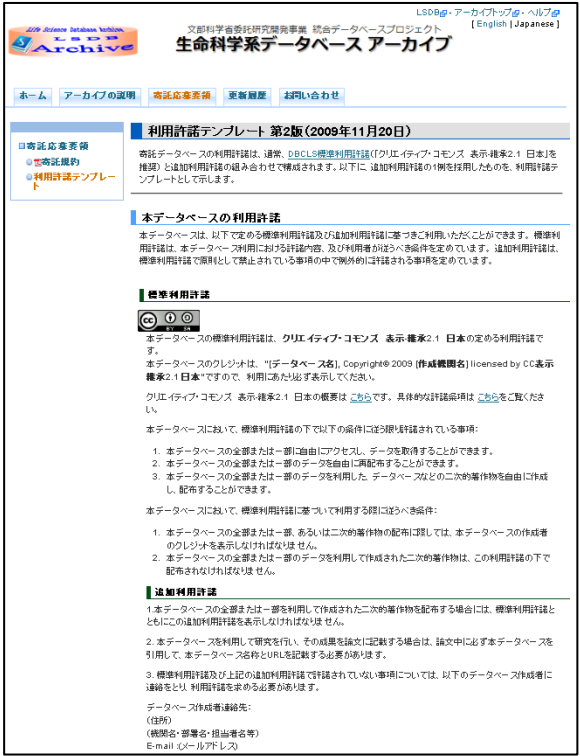


図 4.12 利用許諾テンプレート

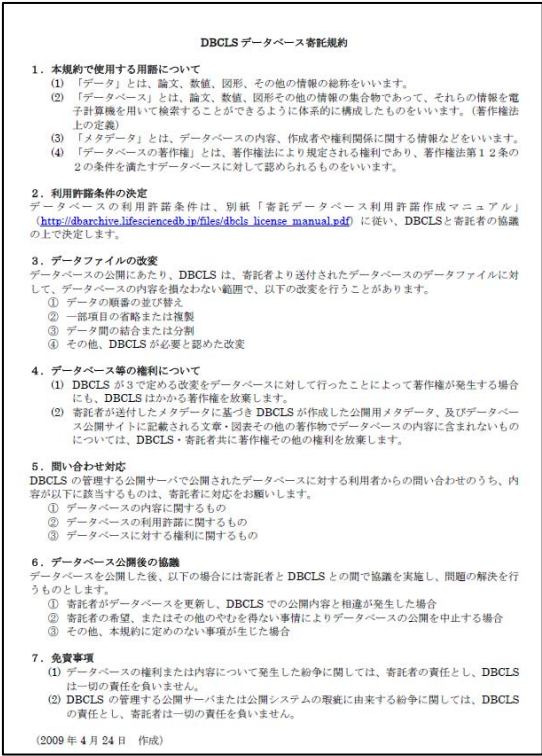


図 4.13 データベース寄託規約

iii) 横断検索への導入：データベースアーカイブのメタデータをインデックス化することにより、横断検索への導入を行った。これによりキーワード検索でアーカイブを探すことが可能になり、またキーワード検索を主なデータベース利用法とするユーザにもアーカイブ利用を促せるようになった。

平成 22 年度は、登録データベース件数の増加に対応するために、データベース一覧においてインクリメンタルサーチによる絞り込みを可能にして、前年度導入したメタデータ横断検索と合わせてユーザが適切なデータベースを選択しやすくなるように配置した。また簡易検索ページでは、検索において AND 検索、OR 検索、フレーズ検索やデータ型に合わせたフォームでの検索をユーザが適切に行えるよう整備した他、一覧における表示件数の変更や小数の表示桁数の管理など、多様なデータに対応するための機能の拡張を続けている。

4) データベースの受け入れ

①各種機関からの受け入れ

平成19年度は、ライフサイエンスデータベース受入の試行を行った。cDNA関連のデータベースとして、BodyMap、HGPD (Human Gene Protein Database)、FLJ Human cDNA DB (スプライシングバリエーションデータベース) と FLJ-DB (ヒト完全長cDNA)、DBTSS (Database of Transcriptional Start Sites)、及びMiBASE (トマトEST データベース) と KafTom (トマト全長cDNA データベース) の受入と関係方法を検討した。また、理研補完課題のテーマ「植物オミックス情報および蛋白質構造情報」に関連したデータベースとして、KATANA (シロイヌナズナデータベース) の受入と関係方法を検討した。

平成 20 年度は、補完課題の各機関、及びその他の機関との間で、3 つの観点 (メタデータ、横断検索、データベースアーカイブサービス) で連携を行うという方針のもとに、それぞれ連携作業を実施し、産総研糖鎖医工学センター (1 件)、九工大 (2 件)、理研 (4 件)、JST (1 件)、かずさ DNA 研 (4 件) の総計 12 件のデータベースについてメタデータとデータを受け取り、これらのメタデータとデータの一部をアーカイブデータベース一覧システムのテストサイト内に登録し、公開の準備を進めた。

平成 21 年度は、前年度にアーカイブデータベース公開システムが構築されたことを受けて、本格的にデータベース受け入れを行った。

i) 参加機関からの受け入れと公開:平成 20 年度に受け入れていたデータベースのうち、理研 1 件 (理研播磨 重原子データベース) についてアーカイブ公開を行った。また新たに長浜バイオ大学より 1 件のデータベース (tRNADB-CE) の受け入れとアーカイブ公開を行った。

ii) データベース構築者へのアンケート回答者からの受け入れと公開:平成 19 年度に実施したデータベース構築者へのアンケートで協力的な返答を寄せていた機関に対して、寄託への賛同を呼びかけた。その結果 4 件のデータベース (Fluorome (東京大学、現早稲田大学)、Gclust Server (東京大学)、Dicty_cDB - Dictyostelium cDNA Database (筑波大学)、FAMSBASE (北里大学)) を受け入れ、3 件のデータベース (Fluorome、Gclust Server、Dicty_cDB - Dictyostelium cDNA Database) のアーカイブ公開を行った。

iii) 文科省ゲノム特定研究の課題機関からの受け入れと公開:ゲノム特定研究でデータベースを構築している課題 135 件を対象にして、データの種類によって分類を行った結果、公共バンクが存在する 1 次データ、公共バンクが存在しない 1 次データ、2 次データがそれぞれ、51 件、22 件、62 件あるという調査結果を得た。これらのうち公共バンクが存在しない 1 次データがアーカイブ化の意義がより高いものと考え、それらのデータとして変異体と表現型、及び cDNA 等の発現画像データベースをアーカイブ化の優先的なターゲットに選び、寄託への賛同を呼びかけた。その結果、Atlas (ISH Data Base) (筑波大学)、EGTC - Database for the Exchangeable Gene Trap Clones (熊本大学)、KTmAbDB - C. elegans Monoclonal Antibody Collection (理化学研究所)、Transposon insertion site database - germline (大阪大学)、WorTS - Worm TS mutant Database (大阪大学)、GETDB - Gal4 Enhancer Trap Insertion Database (京都工芸繊維大学)、The NAISTrap database (奈良先端科学技術大学院大学)、DGSP Database - Drosophila Gene Search Project Database (首都大学東京) の 8 件のデータベースについて受け入れを実施しアー

カイブ公開を行った（括弧内は運用場所）。

iv) その他の受け入れと公開：文科省ゲノムネットワークプロジェクトの成果である DMPD (Dynamic Macrophage Pathway CSML Database)、Yeast Interacting Proteins Database、及び Budding yeast cDNA sequencing project を東京大学より、AcEST を首都大学東京より、DIAM-安全に関する文献情報をバイオインダストリー協会より受け入れた。これらのうち、JBA からのデータベースを除いてアーカイブ公開を行った。

平成 22 年度は、10 月末日までに 8 件のデータベースを受け入れるとともに、9 件のアーカイブ公開を行った結果、受け入れデータベースは累積 38 件、公開アーカイブは累積 37 件に達した（図 4.14）。新たに公開したアーカイブには RA target genes in the Ciona embryo（高知大学）、TMPDB（弘前大学）、Heterozygosity of the Microsatellite Markers in the Japanese Population（九州大学）、Medaka EST Database（東京大学）、MAGEST（東京大学で運用）の他、バイオインダストリー協会からの 4 つのアーカイブの 1 つ DIAM-微生物情報のように、ダウンロードできるデータのファイル数が 40 を超える複雑なアーカイブも現れている。前年度まではファイル数 3~4 の場合が多く、その多くは TogoDB のテーブルとして表現しておりテーブル間のリンクによって効率的に情報の閲覧をできるようにしていたが、今後は巨大もしくは複雑なデータベースの扱いが増えることが想定され、対応の検討を進めている。

データベース	データベース運用場所	代表者
 MAGEST ダウンロード 簡易検索 お礼状	東京大学 医科学研究所	高堂 和裕
 Medaka EST Database ダウンロード 簡易検索 お礼状	東京大学大学院理学系研究科 生物科学専攻 動物科学大講座動物発生学研究室	武田 洋幸
 Heterozygosity of the Microsatellite Markers in the Japanese Population ダウンロード 簡易検索 お礼状	九州大学 生体防衛医学研究所付属遺伝情報学センター グノム種遺伝学分野	山本 健
 TMPDB ダウンロード 簡易検索 お礼状	私立大学大学院 理工学研究科 電子情報工学科	清水 俊夫
 DIAM - 微生物情報 ダウンロード 簡易検索 お礼状	バイオインダストリー協会	-
 RA target genes in the Ciona embryo ダウンロード 簡易検索 お礼状	高知大学 理学部 応用理学科	森原 道樹
 DIAM - 安全性に関する文献 ダウンロード 簡易検索 お礼状	バイオインダストリー協会	-
 DIAM - バイオテクノロジーの基礎知識 ダウンロード 簡易検索 お礼状	バイオインダストリー協会	-
 DIAM - バイオセーフティ関連の法規制 ダウンロード 簡易検索 お礼状	バイオインダストリー協会	-
 The NAITrap database ダウンロード 簡易検索 お礼状	奈良先端科学技術大学院大学 バイオサイエンス研究科 動物遺伝子機能学講座	石田 靖雅
 DGSP Database - Drosophila Gene Search Project Database ダウンロード 簡易検索 お礼状	首都大学東京 大学院理工学研究科 生命科学専攻 細胞遺伝学研究室	相塚 敏郎
 EGTC - Database for the Exchangeable Gene Trap Clones ダウンロード 簡易検索 お礼状	熊本大学 生命資源研究・支援センター	寛木 正徳
 Transposon insertion site database - germline ダウンロード 簡易検索 お礼状	大阪大学大学院 医学系研究科 環境・生体機能学	堀江 熱二
 HATODAS - 重原子データベース ダウンロード 簡易検索 お礼状	独立行政法人理化学研究所 構造研究所 放射光科学総合研究センター	国島 道樹
 KTMABDB - C.elegans Monoclonal Antibody Collection ダウンロード 簡易検索 お礼状	理化学研究所 発生・再生科学総合研究センター 発生ゲノミクス研究チーム	杉本 聖砂子
 Atlas (ISH Data Base) ダウンロード 簡易検索 お礼状	筑波大学 大学院生命環境科学研究所 グノム情報	遠原 秀子
 GETDB - Gal4 Enhancer Trap Insertion Database ダウンロード 簡易検索 お礼状	京都工芸繊維大学	特 茂生
 WorTS - Worm TS mutant Database ダウンロード 簡易検索 お礼状	大阪大学 微生物研究所 細胞機能学	中村 邦明
 AcEST ダウンロード 簡易検索 お礼状	京都大学 京大 大学院理工学研究科 生命科学専攻 植物資源学研究室	鎌ヶ江 健
 Dicty cDB - Dictyostelium cDNA Database ダウンロード 簡易検索 お礼状	筑波大学 大学院生命環境科学研究所 グノム情報	遠原 秀子
 Gclust Server ダウンロード 簡易検索 お礼状	東京大学大学院 総合文化研究科 広域科学専攻 生命環境科学系	佐藤 直樹
 Budding yeast cDNA sequencing project ダウンロード 簡易検索 お礼状	東京大学 大学院理学系研究科 生物化学専攻	伊藤 隆司
 Yeast Interacting Proteins Database ダウンロード 簡易検索 お礼状	東京大学 大学院理学系研究科 生物化学専攻	伊藤 隆司
 tRNADB-CE ダウンロード 簡易検索 お礼状	長浜バイオ大学 生体分子情報学研究室	阿部 良志
 DMPD (Dynamic Macrophage Pathway CSML Database) ダウンロード 簡易検索 お礼状	東京大学 医科学研究所	宮野 悟
 Fluorome ダウンロード 簡易検索 お礼状	早稲田大学 教育学部理学科	園池 公繁
 GlycoProtDB ダウンロード 簡易検索 お礼状	独立行政法人 産業技術総合研究所 糖鎖工学研究センター グライコデータベース解析チーム	堀 裕之
 Arabidopsis thaliana EST index ダウンロード 簡易検索 お礼状	かずさDNA研究所 植物ゲノム研究部	中村 保一
 Chlamydomonas reinhardtii EST index ダウンロード 簡易検索 お礼状	かずさDNA研究所 植物ゲノム研究部	中村 保一
 Lotus japonicus EST index ダウンロード 簡易検索 お礼状	かずさDNA研究所 植物ゲノム研究部	中村 保一
 Trifolium pratense EST index ダウンロード 簡易検索 お礼状	かずさDNA研究所 植物ゲノム研究部	中村 保一
 BodyParts3D ダウンロード 簡易検索 お礼状	ライフサイエンス統合データベースセンター	大久保 公繁
 生命科学系データベースカタログ ダウンロード 簡易検索 お礼状	ライフサイエンス統合データベースセンター	川本 祥子
 生命科学系学協会カタログ ダウンロード 簡易検索 お礼状	ライフサイエンス統合データベースセンター	川本 祥子
 生命科学系主要プロジェクト一覧 ダウンロード 簡易検索 お礼状	ライフサイエンス統合データベースセンター	西川 智夫
 施設名辞書 ダウンロード 簡易検索 お礼状	ライフサイエンス統合データベースセンター	大久保 公繁
 遺伝子名称シンボラス ダウンロード 簡易検索 お礼状	ライフサイエンス統合データベースセンター	大久保 公繁

図 4.14 生命科学系データベースアーカイブのデータベース一覧

②蛋白質関連データベース統合検索システムの開発

平成 20 年度に、理研と九工大の蛋白質関連情報の整理の一環として、蛋白質に関する国内外のデータベース統合化のための内容検討を行った。蛋白質に関する国内外のデータベース情報を収集し、その中で約 40 件のインデックス化が可能なデータベースについて、

横断検索システムのプロトタイプを構築した。また、蛋白 ID による統合を目的として、上記の蛋白質データベースから ID 情報が抽出できるかどうかの調査を行った。

平成 21 年度は、これらの検討を元に蛋白質関連データベース統合検索システム TogoProt (図 4.15、<http://lifesciencedb.jp/togoprot/>) を開発した。このシステムでは、ユーザは望みの蛋白質に関する多様なデータを一覧することができる。ユーザがまず日本語または英語のキーワードを入力すると、検索の結果、蛋白質のリストが PIRSF、Pfam、UniRef50 などのファミリーに分類された形で表示される。ここで 1 つの蛋白質を選択すると、その蛋白質に関連づけられる各種データが一覧表示される。ここで蛋白質の代わりにファミリーを選択すると、ファミリーのいずれかのメンバーに関連づけられる各種データが一覧表示される。12 月に TogoProt のプロトタイプを公開し、年度末までにパラメータの開発を行うとともに、配列・分類・構造・機能・発現・相互作用・辞典など多種多様なデータベースから 40 種のデータベースを選んで UniProt accession や PDBID などの蛋白 ID の付加を行い TogoProt に収録した。

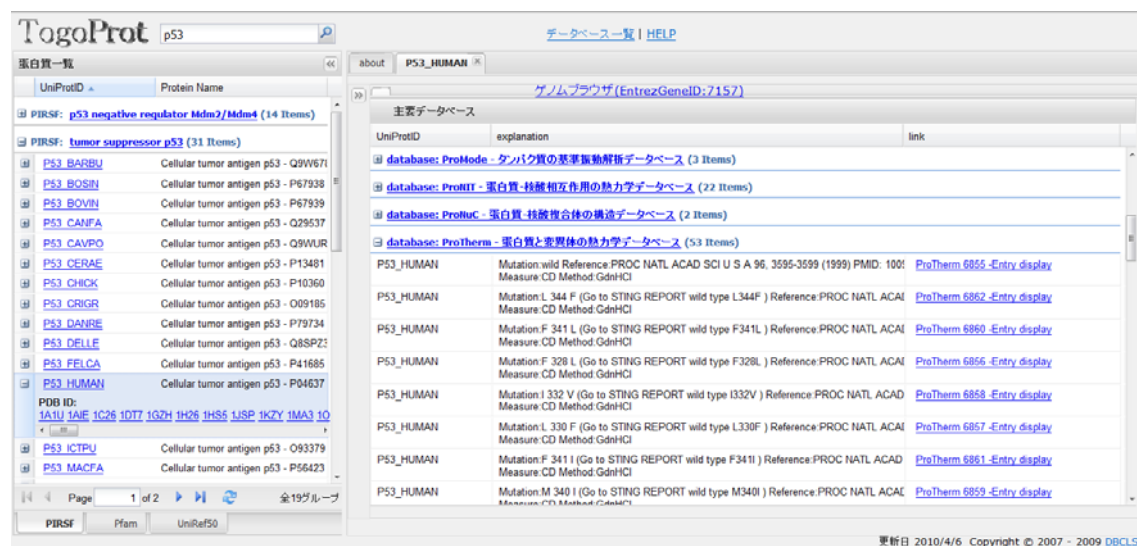


図 4.15 TogoProt の検索画面

平成 22 年度は TogoProt の正式公開を始めるとともに、対象データベース更新への対応を定期的にできるように、スクリプト作成によりシステム化を進めている。また、ユーザの便宜のために、データベース以外の有用なリンク、例えば本プロジェクトで開発された生命情報工学研究ユーザの各種 Web ワークフローや、経産省統合データベースプロジェクトのリンク自動解析システムなどへのリンクをまとめたフレームを含むパラメータを作成し、試験的に公開している。

③他省庁との連携

経産省統合データベースプロジェクトとの間では平成 20 年度、メタデータや横断検索、利用許諾、ウェブサービスなどでの共通化をはかることで合意するなどの連携を図った。メタデータについては、27 件のデータベースについて、双方のデータベースカタログの記述を統一化した。平成 21 年度はデータベース受け入れに関連して、経産省統合データベースプロジェクトとの連携のため 9 回の会合を開催し、メタデータや横断検索、アーカイ

ブの共有化について討議した。その結果、メタデータ共有化案の作成と横断検索の仕様統一化を実現した。さらに、データベースの法的な問題への当ユーザの取り組みを経産省統合データベース側関係者に説明し討議する会を、コンサルテーションをいただいている弁護士を含めて実施した。これらの活動により、経産省統合データベースプロジェクトとの間でメタデータや横断検索は共通仕様に、アーカイブについては文科省統合データベースプロジェクトのアーカイブサイトに集約する方向でコンセンサスを得ることができた。平成 22 年度も会合を続けるとともに、上記コンセンサスに沿って協力し複数のデータベースに対するアーカイブ作成を進めている。

一方、農水省統合データベースプロジェクトとの間では、今後の連携のための会合を平成 21 年度より開催し、カタログや横断検索、アーカイブなどの連携のあり方や問題点について討議を行っている。

5. プロジェクトの総合的推進

5. 1 プロジェクトの総合的推進の実施計画

プロジェクト全体の連携を密としつつ円滑に推進・運営していくため、ライフサイエンス、知識情報処理、ライフサイエンス DB の 3 分野の専門家による研究運営委員会を組織すると共に、研究運営委員会の実働部隊（情報の収集・分析、動向調査、戦略立案支援、など）として統合 DB 整備戦略作業部会を設置する。これにより統合化 DB の整備戦略を議論する。また、本センターはプロジェクトの中核機関として他分担機関、補完機関と密に連携しつつ、共通の課題やターゲットを抱えた機関を集め、運営委員会や分科会の開催、あるいは外部有識者からの意見聴取の機会を増やすなど進捗を適宜確認しプロジェクトの推進に資する。

5. 2 プロジェクトの総合的推進の実施内容

(1) 研究運営委員会及び研究運営委員会作業部会

平成18年度は、プロジェクト全体の連携を密としつつ円滑に推進・運営していくため、ライフサイエンス、知識情報処理、ライフサイエンスDBの3分野の専門家による研究運営委員会を組織し、統合化DBの整備戦略を議論した。また、研究運営委員会の実働部隊（情報の収集・分析、動向調査、戦略立案支援、など）として統合DB整備戦略作業部会を設けた。

研究運営委員会のメンバーは下記の通りである（敬称略、順不同）。

堀田 凱樹	情報・システム研究機構長
小原 雄治	情報・システム研究機構国立遺伝学研究所長
五條堀 孝	情報・システム研究機構国立遺伝学研究所副所長
大久保公策	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ 研究ユーザ教授
勝木 元也	自然科学研究機構理事

高木 利久	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長/ 東京大学教授
大倉 克美	(独) 科学技術振興機構研究基盤情報部長
吉田 光昭	東京大学大学院新領域創成科学研究科客員教授
辻井 潤一	東京大学大学院情報理工学系研究科教授
金久 實	京都大学化学研究所バイオインフォマティクスセンター長
中村 春木	大阪大学蛋白質研究所蛋白質情報科学研究系教授
榊 佳之	(独) 理化学研究所横浜研究所ゲノム科学総合研究ユーザ長
秋山 泰	(独) 産業技術総合研究所生命情報科学研究センター長
増保 安彦	東京理科大学薬学部生命創薬科学科教授
中村 桂子	JT 生命誌研究館館長
松原 謙一	(株) DNA チップ研究所代表取締役社長
田畑 哲之	(財) かずさ DNA 研究所副所長
久原 哲	九州大学大学院農学研究院遺伝子資源工学部門教授

統合 DB 整備戦略作業部会のメンバーは下記の通りである(敬称略、順不同)。

五條堀 孝	情報・システム研究機構国立遺伝学研究所副所長
田畑 哲之	(財) かずさ DNA 研究所副所長
大久保公策	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ 研究ユーザ教授
菅原 秀明	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ 研究ユーザ長
高木 利久	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長/ 東京大学教授
高野 明彦	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/ 連想情報学研究開発ユーザ長
黒田 雅子	(独) 科学技術振興機構研究基盤情報部バイオインフォマティクス課長
藤山秋佐夫	情報・システム研究機構国立情報学研究所情報プリンスプル研究系教授
久原 哲	九州大学大学院農学研究院遺伝子資源工学部門教授
中村 桂子	JT 生命誌研究館館長
増保 安彦	東京理科大学薬学部生命創薬科学科教授

以下に示す研究運営委員会、統合 DB 整備戦略作業部会の開催を行った。

第一回研究運営委員会・統合 DB 整備戦略作業部会(合同会議、2006 年 11 月 8 日)

第二回統合 DB 整備戦略作業部会(2006 年 12 月 18 日)

第二回研究運営委員会(2006 年 12 月 25 日)

第三回研究運営委員会・統合 DB 整備戦略作業部会(合同会議、2007 年 2 月 1 日)

第四回研究運営委員会・統合 DB 整備戦略作業部会(合同会議、2007 年 3 月 19 日)

平成 19 年度は、プロジェクトの本格的なスタートに対応し、分担機関の研究代表者に

研究運営委員会に加わっていただくなど、新たな体制を構築し、以下に示したメンバーからなる研究運営委員会及び研究運営委員会作業部会を組織して、本プロジェクトの円滑な運営に努めた。

研究運営委員会のメンバーは下記の通りである（敬称略、順不同）。

秋山 泰	東京工業大学大学院情報理工学研究科計算工学専攻教授
浅井 潔	（独）産業技術総合研究所生命情報工学研究ユーザ長
大倉 克美	（独）科学技術振興機構研究基盤情報部長
勝木 元也	自然科学研究機構理事
金岡 昌治	（株）大日本住友製薬研究本部副本部長
金久 實	京都大学化学研究所バイオインフォマティクスセンター長
久原 哲	九州大学大学院農学研究院遺伝子資源工学部門教授
榊 佳之	（独）理化学研究所横浜研究所ゲノム科学総合研究ユーザ長
田中 博	東京医科歯科大学情報医科学センター長
田畑 哲之	（財）かずさ DNA 研究所副所長
徳永 勝士	東京大学大学院医学系研究科教授
中村 桂子	JT 生命誌研究館館長
中村 春木	大阪大学蛋白質研究所蛋白質情報科学研究系教授
長村 吉晃	（独）農業生物資源研究所基盤研究領域ゲノムリソースセンター長
松原 謙一（委員長）	（株）DNA チップ研究所代表取締役社長
吉田 輝彦	国立がんユーザ研究所腫瘍ゲノム解析・情報研究部長
吉田 光昭	東京大学大学院新領域創成科学研究科客員教授
大久保公策	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ 研究ユーザ教授
小原 雄治	情報・システム研究機構国立遺伝学研究所長
五條堀 孝	情報・システム研究機構国立遺伝学研究所副所長
坂内 正夫	情報・システム研究機構国立情報学研究所長
菅原 秀明	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ 研究ユーザ長
高木 利久	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長/ 東京大学教授
堀田 凱樹（副委員長）	情報・システム研究機構長

また、研究運営委員会作業部会のメンバーは下記の通りである（敬称略、順不同）。

浅井 潔	（独）産業技術総合研究所生命情報工学研究ユーザ長
金岡 昌治	大日本住友製薬（株）研究本部副本部長
黒田 雅子	（独）科学技術振興機構研究基盤情報部バイオインフォマティクス課長
田畑 哲之	（財）かずさ DNA 研究所副所長
松原 謙一	（株）DNA チップ研究所代表取締役社長
大久保公策	情報・システム研究機構国立遺伝学研究所

生命情報・DDBJ 研究ユーザ教授

高木 利久(主査) 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長
/東京大学教授

永井 啓一 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
特任教授

以下に示す研究運営委員会及び研究運営委員会作業部会を開催し、平成19年度の計画、補完課題の公募内容、実施状況と進め方、契約・著作権問題、個人情報問題、海外連携、3年後の在り方などを討議した。

第一回研究運営委員会（平成19年4月27日）

第一回研究運営委員会作業部会（平成19年6月15日）

第二回研究運営委員会作業部会（平成19年10月1日）

第二回研究運営委員会・DBCLSお披露目会（平成19年10月5日）

第三回研究運営委員会作業部会（平成19年12月6日）

第三回研究運営委員会（平成20年1月8日）

第四回研究運営委員会作業部会（平成20年3月27日）

平成 20 年度は、中間評価結果を踏まえて研究運営委員会と作業部会の在り方を見直し、研究運営委員会ではより広い視点からの議論を行い、具体的なプロジェクトの進捗に関わる議論は作業部会で行うこととした。このため、委員会の構成を見直し、平成 20 年 8 月開催の第五回研究運営委員会で承認を得たのち、以下に示す新たな体制で進めた。また、新たに作業部会分科会を設け、全ての課題の研究代表者にも参加いただき、関連テーマごとに計画、進捗状況を議論する場を設けた。

第六回研究運営委員会以降の運営委員は下記の通りである（敬称略、順不同）。

大倉 克美	（独）科学技術振興機構研究基盤情報部長
勝木 元也	自然科学研究機構理事
金岡 昌治	（株）大日本住友製薬研究本部副本部長
金久 實	京都大学化学研究所バイオインフォマティクスセンター長
榊 佳之	豊橋技術科学大学長
豊田 哲郎	（独）理化学研究所横浜研究所生命情報基盤研究部門長
長洲 毅志	エーザイ（株）理事・研究開発担当付部長
中村 桂子	JT生命誌研究館館長
中村 春木	大阪大学蛋白質研究所蛋白質情報科学研究系教授
長村 吉晃	（独）農業生物資源研究所基盤研究領域ゲノムリソースセンター長
松尾 泰樹	（独）理化学研究所横浜研究所研究推進部長
松原 謙一(委員長)	（株）DNAチップ研究所代表取締役社長
湯元 昇	（独）産業技術総合研究所理事
吉田 輝彦	国立がん研究センター腫瘍ゲノム解析・情報研究部長
吉田 光昭	東京大学大学院新領域創成科学研究科客員教授

大久保公策	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ研究ユーザ教授
五條堀 孝	情報・システム研究機構国立遺伝学研究所副所長
高木 利久	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長
堀田 凱樹(副委員長)	情報・システム研究機構長

また、新たな研究運営委員会作業部会のメンバーは下記の通りである(敬称略、順不同)。

浅井 潔	(独) 産業技術総合研究所生命情報工学研究ユーザ長
黒田 雅子	(独) 科学技術振興機構研究基盤情報部バイオインフォマティクス課長
五斗 進	京都大学化学研究所バイオインフォマティクスセンター准教授
田中 博	東京医科歯科大学情報医科学センター長
田畑 哲之	(財) かずさ DNA 研究所副所長
徳永 勝士	東京大学大学院医学系研究科教授
豊田 哲郎	(独) 理化学研究所横浜研究所生命情報基盤研究部門長
成松 久	(独) 産業技術総合研究所糖鎖医工学研究ユーザ長
松原 謙一	(株) DNA チップ研究所代表取締役社長
森下 真一	東京大学新領域創成科学研究科教授
五條堀 孝	情報・システム研究機構国立遺伝学研究所副所長
菅原 秀明	情報・システム研究機構国立遺伝学研究所 生命情報・DDBJ 研究ユーザ特任教授
高木 利久(主査)	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長

上記の体制で、以下に示す研究運営委員会及び研究運営委員会作業部会、同分科会を開催し、プロジェクトの進捗状況、中間評価への対応、プロジェクト終了後を見据えた統合データベースのあるべき姿、体制等に関して討議した。

第四回研究運営委員会 (平成20年4月3日)

第五回研究運営委員会 (平成20年8月11日)

第五回研究運営委員会作業部会 (平成20年10月10日)

第六回研究運営委員会 (平成21年1月19日)

研究運営委員会作業部会分科会－医学・化合物・糖鎖関係 (平成20年1月21日)

研究運営委員会作業部会分科会

－遺伝研・JST・中核機関の参画機関(基盤技術関係機関) (平成20年1月27日)

研究運営委員会作業部会分科会－立体構造・植物関係機関 (平成20年1月29日)

研究運営委員会作業部会分科会－人材育成 (平成20年1月30日)

平成21年度も、平成20年度と同じメンバーに研究運営委員、研究運営委員会作業部会委員をお願いし、併せて共通の課題やターゲットを抱えた機関を集めた分科会等の開催数を増やすなど、引き続き、研究運営委員会作業部会その他の会議(下記)の開催等を通してプロジェクト参加機関の連携・調整、及び進捗の確認を行った。また、翌平成22年度が最終年度に当たるため、当初立てた計画通りに進んでいるのか、あるいは修正が必要なのか

等の議論も積極的に行った。

開催した研究運営委員会及び研究運営委員会作業部会は次の通りである。

研究運営委員会拡大懇談会	平成 21 年 4 月 13 日
第七回研究運営委員会	平成 21 年 6 月 25 日
第八回研究運営委員会	平成 21 年 10 月 23 日
第九回研究運営委員会	平成 22 年 1 月 25 日

第五回作業部会分科会	平成 21 年 5 月 20 日
第六回作業部会分科会	平成 21 年 5 月 25 日
第七回作業部会分科会	平成 21 年 5 月 27 日
第八回作業部会分科会	平成 21 年 5 月 29 日
第九回作業部会分科会	平成 22 年 1 月 22 日
第十回作業部会分科会	平成 22 年 1 月 29 日
第十一回作業部会分科会	平成 22 年 2 月 2 日

平成 22 年度は、プロジェクトの最終年度に当たるため、開発関係は 9 月末までには終了させ、年度末までにはその効果を検証することを中核機関はじめ関係機関においては徹底することとした。さらに、平成 23 年度には新センターが JST 内に発足し、本プロジェクトで得られた成果を引継ぐことがほぼ決定的となっているため、今年度は当初立てた計画通りに進んでいるのか、あるいは修正や終了させるべき項目があるのか否か等、そして新センターを取り巻く状況も踏まえ研究運営委員会及び作業部会分科会その他の会議を通してプロジェクト参加機関の連携・調整、及び進捗の確認を行った。なお、(独)科学技術振興機構内の体制変更に伴い、大倉克己委員から、同機構のイノベーション推進本部バイオサイエンスデータベースセンター準備室長の白木澤佳子委員への変更があった。

開催した研究運営委員会及び研究運営委員会作業部会は次の通りである。

第十回研究運営委員会	平成 22 年 4 月 26 日
第十一回研究運営委員会	平成 22 年 11 月 12 日
第十二回作業部会分科会	平成 22 年 6 月 1 日
第十三回作業部会分科会	平成 22 年 6 月 2 日
第十四回作業部会分科会	平成 22 年 6 月 3 日
第十五回作業部会分科会	平成 22 年 10 月 21 日
第十六回作業部会分科会	平成 22 年 11 月 2 日
第十七回作業部会分科会	平成 22 年 11 月 2 日

(2) プロジェクト実施機関並びに他省関連機関等との連絡会等の開催実績

平成 19 年度は、中核機関と分担機関の実施内容と連携の進め方、実施状況のフォローなどを中心に討議した。全体会議を 3 回と、個別会議を各 1 回開催した。また、補完課題

の事業に関しては、各機関と密接な打合せを行いながら進めた。また、特に蛋白立体構造ならびに植物関係の進め方に関する議論は、それぞれ中村春木教授、中村保一室長を中心に中核機関を含めたグループを構成して進めた。会合の一覧は以下に示すとおりである。

日付	時間	内容	場所
H19.06.22	10:00-12:00	第1回中核機関・分担機関研究連絡会	ROIS 会議室
H19.08.09	10:00-12:00	第2回中核機関・分担機関研究連絡会	ROIS 会議室
H19.08.29	15:00-16:00	東大グループとの打合せ	DBCLS
H19.09.28	15:30-17:30	第3回中核機関・分担機関研究連絡会	東大理会議室
H19.10.11	11:00-12:30	産総研糖鎖医工学センターとの打合せ	DBCLS
H19.10.18	18:00-19:00	分担機関との個別打合せ(東大)	東大医学部
H19.10.19	12:30-14:00	九工大との打合せ	DBCLS
H19.10.19	14:00-15:30	理研との打合せ	DBCLS
H19.11.07	10:00-12:00	補完課題(蛋白構造関係)打合せ	DBCLS
H19.11.15	10:00-11:00	補完課題(植物関係)打合せ	DBCLS
H19.11.30	10:00-12:00	分担機関との個別打合せ(京大)	DBCLS
H20.01.07	10:00-11:00	分担機関との個別打合せ(東京医科歯科大)	東京医科歯科大
H20.01.31	13:00-15:00	理研との打合せ	理研
H20.02.06	13:00-15:00	理研(蛋白構造関係)打合せ	播磨理研
H20.02.18	18:00-21:00	補完課題(蛋白構造関係)打合せ	東京三菱ビル
H20.02.22	13:00-18:00	第一回日本糖鎖統合 DB 会議	東京コンファレンス センター品川
H20.03.06	13:00-15:00	遺伝研との打合せ	遺伝研
H20.03.13	14:00-17:00	産総研糖鎖医工学センターとの打合せ	産総研
H20.03.14	15:00-16:00	九工大との打合せ	DBCLS
H20.03.14	17:00-19:00	東大グループとの打合せ	DBCLS

平成19年度はさらに、各機関の具体的課題の解決と全体的な調整、意識合わせを図るために、各機関と年数回の打合せを行った。会合の一覧は以下に示すとおりである。

日付	時間	内容	場所
H19.05.24	15:30-17:00	講習会計画立案、人材育成計画に関する打合せ	DBCLS
H19.06.06	17:00-19:00	JST との打合せ (DB 受入れ関係)	JST
H19.06.07	15:30-18:00	CBRC、奈良先端大との打合せ(共通基盤技術)	DBCLS
H19.07.04	15:30-17:00	モデル生物有用微生物チーム打合せ	DBCLS
H19.07.10	15:00-18:00	統合データベース支援拡大会議	DBCLS
H19.07.23	16:00-18:00	バイオポータル合同会議	DBCLS
H19.08.21	14:00-16:00	JST との打合せ (DB 受入れ関係)	DBCLS
H19.08.29	18:00-20:00	CBRC との打合せ(共通基盤技術)	DBCLS
H19.09.14	15:00-17:00	奈良先端大との打合せ	奈良先端大
H19.11.05	18:30-20:30	奈良先端大との打合せ	DBCLS

H19.11.15	16:00-18:00	モデル生物有用微生物チーム打合せ	DBCLS
H19.11.27	15:00-17:30	CBRC、かずさ DNA 研との打合せ	DBCLS
H19.12.03	17:00-18:30	JST との打合せ（文献検索）	JST
H19.12.07	10:00-11:00	JST との打合せ（文献検索）	DBCLS
H19.12.13	13:00-15:00	九州大学との打合せ	DBCLS
H19.12.18	14:00-16:00	かずさ DNA 研との打合せ	DBCLS
H19.12.27	13:30-16:30	人材育成関係合同打合せ	DBCLS
H20.01.09	18:40-21:00	かずさ DNA 研との打合せ	DBCLS
H20.01.10	13:00-15:00	奈良先端大浅原先生講演と打合せ	DBCLS
H20.01.22	14:00-15:00	CBRC との打合せ	CBRC
H20.02.04	15:30-18:00	モデル生物有用微生物チーム打合せ	DBCLS
H20.02.25	16:00-17:00	かずさ DNA 研との打合せ	DBCLS
H20.02.26	15:00-17:00	農業生物資源研との打合せ	農資研
H20.03.21	13:00-17:00	大阪府立成人病ユーザ研究所との打合せ	成人病ユーザ

なお、共通基盤技術に関わる東大医科研、かずさ DNA 研との打合せは週一回定例で実施した。

平成 20 年度は、中核機関とその参画機関、及び分担機関、補完課題実施機関等のプロジェクト参加機関とは、適宜打合せを行うことにより、全体戦略を共有して、連携して本事業を推進するべく努めた。データベースの受け入れという観点で中核機関と連携の可能性のあるプロジェクト外の機関とも、必要に応じて打合せを行った。会合の一覧は以下に示すとおりである。

また、プロジェクトの参加機関が、科学技術振興機構(JST)も含めて、全体で 16 機関と多岐にわたるため、プロジェクト参加者全員が、全体としての目標を理解して、進捗状況を把握できるようにするために、参加機関ごとの平成 20 年度の具体的目標と進捗状況を示す全体テーブルを作成し、これを定期的にリバイズすることにより、プロジェクト全体での情報共有に努めた。併せて、JST が構築した「参画研究者の情報交換サイト」の活用による情報共有の促進にも努めた。

日付	時間	内容	場所
H20.04.25	13:00-14:40	九工大ミーティング第 1 回	DBCLS
H20.04.30	10:00-12:00	経産省統合 DB とのミーティング	DBCLS
H20.05.09	16:30-17:30	東大グループ DB 移植	DBCLS
H20.05.19	15:00-16:00	SVDB 等移植	DBCLS
H20.05.20	17:00-18:30	経産省統合 DB とのミーティング	DBCLS
H20.05.28	13:30-15:10	九工大ミーティング第 2 回	DBCLS
H20.06.17	11:00-15:30	産総研糖鎖ユーザとのミーティング第 2 回	糖鎖ユーザ
H20.06.27	15:00-16:00	九工大ミーティング第 3 回	DBCLS
H20.07.02	10:00-12:00	JST との打ち合わせ	DBCLS
H20.07.29	10:00-13:00	微生物グループ打合せ	DBCLS
H20.07.31	14:00-17:00	かずさ DNA 研(植物統合関連)打合せ	DBCLS

H20.08.01	13:00-15:00	理研とのミーティング	理研
H20.08.26	16:00-19:00	経産省統合 DB とのミーティング	DBCLS
H20.08.29	14:00-16:00	東大・磯貝先生とのミーティング(SVDB 関係)	東大薬学部
H20.09.16	15:00-18:30	FAMSBASE 受け入れヒアリング	北里大
H20.09.26	14:00-16:00	DDBJ とのミーティング(トレースアーカイブ関係)	遺伝研
H20.10.22	13:30-16:30	理研とのミーティング	理研
H20.10.30	14:00-18:00	HGPD ミラー化のミーティング	産総研 BIRC
H20.10.31	17:00-18:00	東大グループとの打合せ	東大医学部
H20.11.07	10:30-13:00	トキシコゲノミクスプロジェクトとのミーティング	医薬基盤研
H20.11.07	19:00-20:00	東大グループとの打合せ	DBCLS
H20.11.14	15:00-17:00	人材育成関係打合せ (お茶大、長浜バイオ大)	DBCLS
H20.11.19	15:30-18:30	経産省統合 DB とのミーティング	DBCLS
H20.11.20	13:00-15:00	医科歯科大グループとの打合せ	DBCLS
H20.11.28	14:00-16:30	HGPD ミラー化のミーティング	産総研 BIRC
H20.12.25	13:30-15:30	かずさ DNA 研(植物統合関連)打合せ	かずさ DNA 研
H21.01.08	14:30-17:00	DDBJ とのミーティング(トレースアーカイブ関係)	遺伝研
H21.01.16	14:00-17:00	糖鎖ユーザとのミーティング	DBCLS
H21.03.24	13:00-14:00	かずさ DNA 研(植物統合関連)打合せ	DBCLS

平成 21 年度は 4 省連携を中心に進めるために関係機関との打ち合わせを積極的に設定した。会合の一覧は以下に示す通りである。

日付	時間	内容	場所
H21.04.10	14:00-16:00	倫理検討委員会委員長(京大位田先生)打合せ	京大法学部
H21.04.21	15:30-18:00	経産省統合 DB とのミーティング	DBCLS
H21.04.28	10:30-12:00	東大・磯貝先生とのミーティング	東大薬学部
H21.05.08	17:00-18:00	分担機関 (東大 G・東大病院との打合せ)	東大病院
H21.05.28	16:00-19:00	経産省統合 DB とのミーティング	DBCLS
H21.06.05	13:00-15:30	第 1 回疾患解析 DB 開発「倫理検討委員会」	外部会議室
H21.06.15	13:00-15:00	医薬基盤研打ち合わせ	医薬基盤研
H21.06.23	10:00-12:00	経産省統合 DB とのミーティング	DBCLS
H21.06.30	14:00-15:00	東大医科研・宮野研とのミーティング	東大医科研
H21.07.22	10:00-12:00	経産省統合 DB とのミーティング	DBCLS
H21.07.27	16:00-18:00	第 2 回疾患解析 DB 開発「倫理検討委員会」	外部会議室
H21.08.06	15:00-18:00	経産省統合 DB とのミーティング	JBIC
H21.11.04	10:00-12:00	JBA とのミーティング	DBCLS
H21.12.16	16:00-18:00	医用画像データ関係打合せ	精神・神経ユーザ
H21.12.21	10:00-12:30	JBA とのミーティング	DBCLS
H21.12.25	13:30-15:30	農水省とのミーティング	DBCLS

H22.01.06	15:30-17:30	FAMSBASE 寄託・郷先生とのミーティング	ROIS
H22.01.21	14:00-15:00	東大・磯貝先生とのミーティング	東大薬学部
H22.01.21	17:30-19:00	FAMSBASE 寄託・梅山先生とのミーティング	北里大

平成 22 年度は、21 年度に引き続き各省との連携を進める打ち合わせを設定すると共に、参画機関間の連携の見直しと今後重要になると思われる分野に関する会合を設定した。会合の一覧は以下の通りである。

日付	時間	内容	場所
H22.04.20	16:30-18:00	分担機関（東大 G との打合せ）	東大医学部
H22.04.28	18:00-20:00	ヒトゲノム variation database 打合せ	外部会議室
H22.05.12	17:00-19:00	分担機関（東大 G との打ち合わせ）	東大医学部
H22.05.24	16:00-16:30	トキシコゲノミクスプロジェクトとの打合せ	テレビ会議
H22.05.26	14:00-15:30	理研 OSC とのミーティング	DBCLS
H22.05.28	15:00-18:00	経産省統合 DB とのミーティング	産総研
H22.06.07	13:00-14:30	トキシコゲノミクスプロジェクトとの打合せ	医薬基盤研
H22.06.02	18:00-20:00	ヒトゲノム variation database に関する打合せ	外部会議室
H22.07.08	14:00-16:00	分担機関（医科歯科大、阪大）打合せ	DBCLS
H22.07.27	14:00-15:30	SEVENS アーカイブについての打合せ	産総研
H22.08.06	16:00-18:00	経産省統合 DB とのミーティング	DBCLS
H22.08.18	14:00-18:00	医用画像データ関係打合せ	DBCLS
H22.09.09	13:00-14:30	トキシコゲノミクスプロジェクトとの打合せ	DBCLS
H22.09.21	16:00-17:00	経産省統合 DB とのミーティング	DBCLS
H22.11.10	10:00-12:00	分担機関（東大 G との打ち合わせ）	東大医学部

（３）外部有識者からの意見聴取

平成 18 年度、統合 DB の人材の育成プロジェクト（教育プロジェクト）に関して、外部有識者を招聘して意見を聞くための以下の会を開催した。①の外部有識者としては、長浜バイオ大学の池村教授を招いて、ゲノムアノテーションを通じたアノテーション教育のあり方と今後の方針についての討論を実施した。また、②の「シニア世代と学部教育」第一回ミーティングにおいては、池村教授が提唱されている「シニア研究者の高度な知識の活用・継承によるアノテーション教育」について、シニア世代 6 名と現役世代 3 名の研究者を招き討論を行った。これらの討論の結果は、「ライフサイエンス分野の統合データベース整備事業」の全体計画策定において反映された。

①統合 DB の教育プロジェクトに関する打ち合わせ（2006 年 10 月 17 日）

②「シニア世代と学部教育」第一回ミーティング（2007 年 1 月 29 日）

平成 19 年度以降、プロジェクトの推進に資するため、以下に示す先生方との打ち合わせやセミナーを開催した。

Martin Yuille 先生 (Manchester 大 教授)、隅蔵康一先生 (政策研究大 准教授)

金子周司先生(京大薬学部 教授)、 簗島伸生先生(浜松医大 教授)
 浅原正幸先生(奈良先端大 助教)、 増井徹先生(医薬基盤研 主任研究員)
 加藤和人先生(京大生命科学 准教授)、 神寫敏弘先生(産総研脳神経情報研究部門)
 市瀬龍太郎先生(国立情報研 准教授)、 向井秀幸先生(神戸大 准教授)
 権藤洋一先生(理研GSC PJ副ディレクター)、西村邦裕先生(東大情報理工 助教)
 星山大介先生(東大新領域 助教)、 田中英夫氏(トミーディジタルバイオロジー)
 野口 祐子氏(弁護士)、 末吉互氏(弁護士)
 John Wilbanks(Science Commons)、 Alan Ruttenberg(Neuro Commons)
 鳥澤健太郎(情報通信研究機構 グループリーダー)、新里圭司(京都大学 特任助教)
 荒木次郎(三菱総合研究所 主任研究員)、 牧野貴樹(東京大学 助教)
 木村宏一(日立製作所 主任研究員)、 岩崎渉(東京大学)
 畠中秀樹(九州大学 特任准教授)、 坂東明日佳(国立遺伝研)
 山田智之(株式会社RNAi) 神沼英理(国立遺伝研 助教)
 新井紀子(国立情報研 教授) 荒木啓充(セルイノベータ)
 有田正規(東大新領域 准教授)

5. 3 その他シンポジウム、セミナー等

(1) シンポジウム

1. ゲノムテクノロジー第164委員会との共催による研究会

「生命科学データの共有に向けて：統合データベースプロジェクトの挑戦」

日時：平成20年7月22日(木) 13:30 – 17:30

場所：東京大学山上会館

参加者：約50名

プログラム：

13:30-13:45 統合データベースプロジェクト

ライフサイエンス統合データベースセンター 高木利久

13:45-14:15 統合データベースプロジェクトの提供サービスについて

ライフサイエンス統合データベースセンター 川本祥子

14:15-14:45 疾患解析から医療応用を目指すDB開発：ゲノムワイド関連解析DBと
リシーケンシングDB

東京大学大学院医学系研究科 徳永勝士

14:45-15:15 日本糖鎖科学統合DBの取り組みについて

産総研糖鎖医工学センター 成松久

15:15-15:45 休憩(公開サービスのデモ)

15:45-16:15 科学を社会に役立てるには？

国立遺伝学研究所 大久保公策

16:15-17:30 パネルディスカッション「統合DBに期待するもの」

エーザイ(株) 長洲毅志

協和発酵工業(株) 太田紀夫

理化学研究所 権藤洋一

東京工業大学 秋山 泰
 国立遺伝学研究所 大久保公策
 (司会) ライフサイエンス統合データベースセンター 高木利久

2. 統合データベースプロジェクト シンポジウム 2009

「データベースが拓くこれからのライフサイエンス」

開催日：2009 年 6 月 12 日

開催場所：東京大学本郷キャンパス浅野地区武田ホール

参加人数：393 名

実施内容：

《講演》

時間	演者	タイトル他
10:00～10:20	文部科学省/五條堀孝(国立遺伝学研究所)	開会の挨拶
セッション 1 データベースを統合する (座長：高木利久、ライフサイエンス統合データベースセンター)		
10:20～10:40	平田正(協和発酵キリン株式会社)	「わが国ライフサイエンスデータベースの現状と統合データベースセンターへの要望」
10:40～11:10	坊農秀雅(ライフサイエンス統合データベースセンター)	「統合データベースプロジェクト：オープンで知のめぐりの良いライフサイエンス研究を目指して」
11:10～11:40	村上勝彦(バイオ産業情報化コンソーシアム)	「経済産業省ライフサイエンス統合 DB プロジェクトのポータルサイト MEDALS」
セッション 2 データベース勢ぞろい (11:40～13:30) ポスター発表		
セッション 3 データベースを作る (座長：浅井潔、産業技術総合研究所 生命情報工学研究ユーザ)		
13:30～14:00	森下真一(東京大学)	「次世代シーケンサーを用いたゲノムインフォマティクス」
14:00～14:30	中村春木(大阪大学)	「蛋白質構造データバンク事業と国際的な展開」
セッション 4 データベースを使う(座長：徳永勝士、東京大学)		
14:30～15:00	山田弘(医薬基盤研究所)	「トキシコゲノミクスプロジェクトで構築したデータベースとその応用」
15:00～15:30	斉藤和季(理化学研究所、千葉大学)	「データベースの有効利用による植物メタボロミクス研究の推進」
セッション 5 これからの統合データベース		
15:55～17:15	司会：堀田凱樹(情報・シ	パネル討論

	ステム研究機構)	パネリスト：大久保公策、金岡昌治(大日本住友製薬株式会社)、高木利久、中村春木、菱山豊(文部科学省 研究振興局 ライフサイエンス課)
17:15～17:25	堀田凱樹(情報・システム研究機構)	閉会の挨拶

《ポスター発表》 85 件

3. 「科学における情報の上手な権利化と共有化」 シンポジウム

開催日：2009 年 10 月 5 日

開催場所：東京大学本郷キャンパス弥生地区一条ホール

参加人数：257 名

実施内容：

《講演》

時間	演者	タイトル他
10:00-10:10	高木利久(ライフサイエンス統合データベースセンター、センター長)	開会挨拶・開催趣旨説明
<第 1 部> (同時通訳付) 座長：野口祐子(CCJP 専務理事、国立情報学研究所 客員准教授)		
10:10-10:30	大久保公策(国立遺伝学研究所生命情報・DDBJ 研究センター、センター長)	「デジタル科学への最後のステップ」
10:30-11:20	Lawrence Lessig (Harvard Law School, Professor)	“ Copyright in the Digital Age and Its Impact on Scientific Data Sharing”
11:20-11:40	中山信弘(明治大学特任教授、東京大学名誉教授、弁護士)	「デジタル時代の著作権とイノベーション」
11:40-12:10		質疑応答
<第 2 部> 座長：大久保公策		
13:20-13:50	長尾真(国立国会図書館長)	「公表したものは共有財産」 (Published Materials Are Common Property)
13:50-14:20	末吉互(末吉綜合法律事務所、弁護士)	「情報の共有と独占について」
14:20-14:50	深作裕喜子(コンサルタ	「OECD データ・アクセス・ガイドラインお

	ント業 INNOVMOND 主宰)	よび欧州におけるユーザ推進政策動向」
15:00-16:30	司会：大久保公策、高木 利久	パネル討論 パネリスト：末吉亙、長洲毅志（エーザイ株式 会社）、永山裕二（文化庁）、深作裕喜子
16:30-16:40	堀田凱樹(情報・システム 研究機構、機構長)	閉会挨拶

4. 合同シンポジウム「ライフサイエンスの未来へ～10年先のデータベースを考える～」

開催日：2010年10月5日

開催場所：東京大学本郷キャンパス浅野地区武田ホール

参加人数：会場参加者 230 名、同時配信視聴者 70 名以上

実施内容：

《講演》

時間	演者	タイトル他
10:00～10:20	郷 通子（長浜バイオ大 学、情報・システム研究 機構）	開会の挨拶・統合に関する現状報告
10:20～10:45	伊藤 隆司（東京大学大学 院理学系研究科）	「オーミクス研究者から見たデータベース：デ ータ生産者の目線」
10:45～11:10	伊藤 啓（東京大学分子細 胞生物学研究所）	「ショウジョウバエ脳の神経画像データベー ス Flybrain：15 年の蓄積と将来への課題」
11:10～12:30	ポスター発表	
13:30～14:10	堀田 凱樹（情報・システ ム研究機構）	基調講演「私はなぜ統合データベースを立ち上 げたのか ～ 2050 年に向けて」
14:10～14:35	青島 健（エーザイ株式会 社）	「バイオマーカー・薬剤作用機序研究における データベースの必要性和課題」
14:35～15:00	石川 裕（東京大学情報基 盤センター）	「東京大学情報基盤ユーザ 大規模データ解析 基盤運用に向けて」
15:30～15:55	辻井 潤一（東京大学大学 院情報学環、マンチェス ター大学計算機科）	「生命科学のための情報統合とテキストマイ ニング」
15:55～16:20	山崎 由紀子（国立遺伝学 研究所生物遺伝資源情報 総合ユーザ）	「バイオリソースデータベース：15 年の歩みと 将来展望」
16:20～16:45	佐藤 雄一郎（神戸学院大 学法学部）	「医療情報の研究利用について：外国から学べ ること」
16:45～17:05	高木 利久（東京大学大学	まとめ・閉会の挨拶

	院新領域創成科学研究 科、科学技術振興機構)	
--	---------------------------	--

《ポスター発表》 44 件

(2) 統合データベース講習会

4. 1. 1. 項で記載した統合データベース講習会の詳細は以下のとおりである。

平成 19 年度

①統合データベース講演会

日時：平成 20 年 1 月 25 日 午後 1 時半～午後 5 時半

場所：長浜バイオ大学

参加者：外部研究機関の研究者と学部学生（21 名）

対象：長浜バイオ大学学生並びに興味のある外部研究機関の研究者

内容：

1. ライフサイエンス分野の統合データベースプロジェクト（講師 川本祥子）
2. バイオデータベース構築（講師 中谷洋一郎）
3. ゲノムアノテーションとは何か：その付けられかた／活用方法の紹介（講師 中村保一）
4. バイオデータマイニング入門（講師 瀬々 潤）
5. 使い倒し系バイオインフォマティクスによる遺伝子発現情報解析（講師 坊農秀雅）
6. テキストを対象とした生命科学研究（講師 山本泰智）

②統合データベース講習会：AJACS 東京

日時：平成 20 年 3 月 5 日 午前 10 時～午後 5 時半

場所：科学技術振興機構 東京本部

参加者：20 名

対象：学部 3 年生や修士 1 年生、博士後期課程 2 年生を始めとした統合データベースセンターに就職する興味のある人。統合データベースが実際にどのように作成、維持管理されているかに興味のある人

内容：

1. 統合データベースプロジェクトとライフサイエンス統合データベースセンター
(講師 坊農秀雅)
2. かずさにおける新しいアノテーションの試み（講師 岡本忍）
3. ゲノムアノテーションとは何か：その付けられかた／活用方法の紹介（講師 中村保一）
4. 統合データベースへの期待（講師 樋口千洋）
5. 講習：ゲノム情報リテラシー入門（講師 中村保一、その他ライフサイエンス統合データベースセンター員）

平成 20 年度

① 統合データベース講習会：AJACS 湘南

日時：平成 20 年 7 月 3 日（木）10:30-17:30

場所：日本大学生物資源科学部

参加者：64 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに（就職する）興味のある方

内容：

10:30-12:00 中村保一 自己紹介 【実習】まずは肩慣らし「検索使い倒し」

13:30-14:40 中村保一 ゲノム塩基配列と類似検索 【実習】配列情報へのアクセス、BLAST
あれこれ

14:50-16:00 中村保一 塩基配列の注釈入門と遺伝子構造予測 【実習】遺伝子発見ツール・
スプライスサイト予測・Martでデータ取得

16:10-17:30 中村保一 遺伝子機能予測とDB高度化 【実習】Interpro・GO・CyanoGenes・
KazusaAnnotation

② 統合データベース講習会：AJACS 本郷 1

日時：平成 20 年 7 月 3 日（木）10:30-17:30

場所：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

参加者：21 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに（就職する）興味のある方

内容：

午前：講演 10:30-12:00

10:30-10:50 高木利久「統合データベースプロジェクトとは」

10:50-11:10 岡本忍, 中尾光輝, 藤澤貴智, 中村保一

「Kazusa Annotation Suite によるライフサイエンスデータ統合」

11:10-11:30 上原啓史、阿部貴志、池村淑道

「学部学生による持続可能型社会への貢献遺伝子データベースの構築」

11:30-11:50 大久保公策（国立遺伝学研究所）「科学データは誰のものか？」

午後：講習 13:30-17:30（統合データベースプロジェクトで提供しているサービスの使い方）
13:30-14:00 坊農秀雅 使いこなすための最低限のコンピュータ使いこなし術

14:00-14:30 川本祥子 生命科学横断検索の利用法

14:30-15:00 休憩

15:00-15:30 三橋信孝 アナトモグラフィ

15:30-16:00 山本泰智 OReFiL と Allie

16:00-16:30 休憩

16:30-17:00 坊農秀雅 DBCLS 独自サービス「目次」、「カタログ」系の使いこなし術

17:00-17:30 坊農秀雅 教育・人材育成のリソース紹介他

③ 統合データベース講習会：AJACS 蝦夷

日時：平成 20 年 8 月 29 日（金）10:00-17:30

場所：北海道大学

参加者：32名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに（就職する）興味のある方

内容：

11:10 - 12:00 中村保一 インTRODakション【実習】まずは肩慣らし「検索使い倒し」

13:30 - 14:40 中村保一 ゲノム塩基配列と類似検索【実習】配列情報へのアクセス、BLAST
あれこれ

14:50 - 16:00 中尾光輝 ゲノム配列の注釈と構造解析～【実習】ゲノムブラウザの利用、
アノテーションの閲覧

16:10 - 17:30 中尾光輝 遺伝子機能予測と DB 高度化～【実習】機能アノテーション付け、
アノテーションの利用

④ 統合データベース講習会：AJACS 博多

日時：平成 20 年 10 月 18 日（土）10:00-17:30

場所：九州大学情報基盤研究開発ユーザ

参加者：43 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに（就職する）興味のある方

内容：

午前：講演 10:00-12:00

10:00-10:30 高木利久「統合データベースプロジェクトとは」

10:30-10:50 佐藤賢二（金沢大学大学院）「ライフサイエンス用語の意味推定」

10:50-11:00 休憩

11:00-11:20 五斗進（京都大学化学研究所）「KEGG Atlas と KAAS アノテーション ～
KEGG における最近の開発から」

11:20-11:50 大久保公策「科学データは誰のものか」

午後：講習 13:30-17:30（統合データベースプロジェクトで提供しているサービスの使い方）

13:30-14:00 河野信「データベースを使いこなすための最低限のコンピュータ使いこなし
術」

14:00-15:00 川本祥子・仲里猛留「生命科学横断検索の利用法」

15:00-15:15 休憩

15:15-15:45 山本泰智「OReFiL と Allie」

15:45-16:15 藤枝香「アナトモグラフィ」

16:15-16:30 休憩

16:30-17:00 中村保一「Kazusa Annotation Suite の活用」

17:00-17:30 坊農秀雅「その他 DBCLS 提供サービスの活用術」

⑤ 統合データベース講習会：AJACS 近江

日時：平成 20 年 10 月 31 日（金）13:30-16:50

場所：長浜バイオ大学

参加者：45 名

対象：長浜バイオ大学生命情報コース配属の 3 回生及び一般参加者

内容：

講演 13:30-14:20

13:30-14:00 坊農秀雅「はじめに」

14:00-14:20 鈴木健介（タカラバイオ株式会社ドラゴンジェノミクスセンター）

「次世代型高速サーバによるゲノム解析」

講習 14:30-17:00 統合データベースプロジェクトで提供しているサービスの使い方実習

14:30-15:10 川本祥子「生命科学横断検索の利用法」

15:10-15:30 金子聡子「統合データベースプロジェクトにおける人材育成活動の紹介」

15:30-15:50 小野浩雅「統合 TV による DBCLS 提供サービスなどのバイオデータベース
とウェブツールの活用術」

15:50-16:00 休憩

16:00-16:30 藤枝香「アナトモグラフィ」

16:30-17:00 中村保一「Kazusa Annotation Suite の活用」

⑥ 統合データベース講習会：AJACS 長津田

日時：平成 21 年 1 月 23 日（金）10:30-17:30

場所：東京工業大学すずかけ台キャンパス ものづくり教育研究支援センター

参加者：30 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統
合データベースセンターに（就職する）興味のある方

内容：

午前：講演＋講習 10:30-12:30

10:30-11:00 坊農秀雅「統合データベースプロジェクトとは？」

11:00-11:30 黒川顕「バイオ大航海時代」

11:30-12:30 河野信「データベースを使いこなすための最低限のコンピュータ使いこなし
術」

午後：講習 14:00-17:30（統合データベースプロジェクトで提供しているサービスの使い
方）

14:00-15:00 川本祥子「生命科学横断検索の利用法」

15:00-15:30 藤枝香「アナトモグラフィ/BodyParts3D の利用法」

15:30-16:00 休憩

16:00-16:30 仲里猛留「自然言語処理技術の活用実例」

16:30-17:00 中村保一「Kazusa Annotation Suite の活用」

17:00-17:30 小野浩雅 「遺伝子発現データの活用実例」

平成 21 年度

① 統合データベース講習会: AJACS 本郷 4

日時: 平成 21 年 4 月 17 日 (金) 13:00-17:30

場所: DBCLS 教室

参加者: 30 名

対象: ライフサイエンス分野のデータベースを効率よく利用することに興味のある方。

内容:

13:00-13:15 坊農秀雅「はじめに: 統合データベースプロジェクトとは？」

13:15-13:45 坊農秀雅「データベースを使いこなすための前準備」

13:45-14:30 川本祥子「生命科学横断検索の利用法」講習で利用する URL のリスト

14:30-14:50 休憩

14:50-15:20 坊農秀雅「DNA データベース総覧と検索、遺伝子発現バンク(GEO)目次の使い方」

15:20-16:05 大久保公策「アナトモグラフィー/BodyParts3D の利用法」
<http://lifesciencedb.jp/ag/>

16:05-16:35 仲里猛留「文献検索にまつわる各種ツールの使いこなし術」

16:35-16:50 休憩

16:50-17:30 大喜利「明日から使えるデータベース・ウェブツール」(河野信、小野浩雅、他)

② 統合データベース講習会: AJACS 湘南 2

日時: 平成 21 年 5 月 15 日 (金) 10:00-17:30

場所: 日本大学生物資源科学部

参加者: 24 名

対象: 日本大学生物資源科学部の学部 3、4 年生や大学院生、教員を始めとしたライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容:

10:00-10:15 「はじめに: 統合データベースプロジェクトとは？」(坊農秀雅)

10:15-10:45 「データベースを使いこなすための前準備」(坊農秀雅)

10:45-11:30 「生命科学横断検索の利用法」(川本祥子)

11:30-13:00 休憩

13:00-13:30 「DBCLS 目次サービス (DNA データベース総覧と検索、遺伝子発現バンク(GEO)目次) の利用法」(坊農秀雅)

13:30-14:00 「アナトモグラフィー/BodyParts3D の利用法」(三橋信孝・藤枝香)

14:00-14:30 「文献検索にまつわる各種ツールの使いこなし術」(仲里猛留)

14:30-14:45 休憩

14:45-15:45 「統合 TV を駆使した遺伝子発現データの使い倒し術」(小野浩雅)

15:45-16:15 「Genome Annotation と KazusaAnnotation」(岡本忍)

16:15-16:30 休憩

16:30-17:30 大喜利「明日から使えるデータベース・ウェブツール」

③統合データベース講習会：AJACS 三河

日時：平成 21 年 7 月 10 日（金）13:00-18:00

場所：自然科学研究機構 基礎生物学研究所

参加者：17 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

13:00-13:15 「はじめに：統合データベースプロジェクトとは？」

13:15-14:15 「生命科学横断検索の利用法」

14:15-14:45 「DNA データベース総覧と検索の使い方」

14:45-14:55 休憩

14:55-15:35 招待講演

○ 内山郁夫「微生物ゲノム比較のためのデータベース MBGD」

○ 成瀬清「メダカ完全長 cDNA リソースの整備とデータベース構築」

15:35-16:20 「遺伝子発現データベースを使い倒す」

16:20-16:30 休憩

16:30-17:00 「アナトモグラフィー/BodyParts3D の利用法」

17:00-17:45 『Genome Annotation と KazusaAnnotation』

17:45-18:00 「おわりに」

④ 統合データベース講習会：AJACS 蝦夷 2

日時：平成 21 年 9 月 7 日（月）09:45-17:00、8 日（火）10:00-15:00

場所：北海道大学札幌キャンパス

参加者：1 日目 20 名、2 日目 11 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

1 日目：ライフサイエンス統合データベースセンターが提供するサービスの使い方

9:45-10:05 「はじめに：統合データベースプロジェクトとは？」

10:05-10:15 「統合 TV」

10:15-11:15 「生命科学横断検索の利用法」

11:15-12:00 「DNA データベースの使い方」

12:00-13:30 昼休憩

13:30-14:00 招待講演：瀬々潤（お茶の水女子大学）「データベースからの生命科学：タンパク質相互作用と遺伝子発現を例に」

14:00-14:45 「遺伝子発現データベースを使い倒す」

14:45-15:15 休憩

15:15-16:00 「アナトモグラフィー/BodyParts3D の利用法」

16:00-16:45 「Genome Annotation と KazusaAnnotation」

16:45-17:00 「おわりに」

2 日目：一歩進んだバイオデータベース・ツールの使い方（中上級者向け）

10:00-12:00

続・遺伝子発現データベースを使い倒す

次世代サーバから得られる遺伝子配列データの解析手法：配列データとしての発現データを扱う解析パイプラインの開発から

12:00-13:00 昼休憩

13:00-15:00

TogoWS の使い方

Galaxy を使った解析ワークフローの作り方

他、質問など

⑤ 統合データベース講習会：AJACS 近江 2

日時：平成 21 年 10 月 16 日（金）13:30-16:50

場所：長浜バイオ大学

参加者：43 名

対象：長浜バイオ大学生命情報コース配属の 3 回生及び一般参加者

内容：

13:30-13:45 「はじめに」

13:45-14:00 「統合 TV」

14:00-14:45 「生命科学横断検索の利用法」

14:45-15:15 「遺伝子配列・発現データベースの利用法」

15:15-15:25 休憩

15:25-15:45 招待講演

15:45-16:15 「アナトモグラフィー/BodyParts3D の利用法」

16:15-16:45 「Genome Annotation と KazusaAnnotation」

16:45-16:50 「おわりに」

⑥ 統合データベース講習会：AJACS りんくう

日時：平成 21 年 11 月 6 日（金）09:45-17:00

場所：大阪府立大学りんくうキャンパス

参加者：33 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

9:45-10:05 「はじめに：統合データベースプロジェクトとは？」

10:05-10:15 「統合 TV」
 10:15-11:15 「生命科学横断検索の利用法」、番外編/生物アイコン
 11:15-11:45 特別講演「科学データは誰のものか」国立遺伝学研究所 大久保公策 先生
 11:45-13:00 昼休憩
 13:00-13:30 招待講演「FANTOM4 プロジェクトに見る次世代サーバの可能性とデータベース」理化学研究所 オミックス基盤領域 川路英哉 先生
 13:30-14:00 「DNA データベースの使い方」
 14:00-14:45 「遺伝子発現データベースを使い倒す」
 14:45-15:15 休憩
 15:15-16:00 「アナトモグラフィー/BodyParts3D の利用法」
 16:00-16:45 「Genome Annotation と KazusaAnnotation 」
 16:45-17:00 「おわりに」

⑦ 統合データベース講習会：AJACS 筑後

日時：平成 22 年 1 月 18 日（月）09:30-17:30、19 日（火）10:00-15:00

場所：久留米ビジネスプラザ

参加者：1 日目 20 名、2 日目 12 名

対象：ライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースセンターに就職する興味のある方、統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

1 日目：ライフサイエンス統合データベースセンターが提供するサービスの使い方

9:30-10:00 「はじめに：統合データベースプロジェクトとは？」

10:00-10:30 「統合 TV」

10:30-12:00 「生命科学横断検索、生命科学系 DB カタログの利用法」

「概要」(30mins)

「実習」(1hr)

12:00-13:00 昼休憩

13:00-15:00 「遺伝子発現データベースを使い倒す」

15:00-15:30 休憩

15:30-17:00 「DNA データベース総覧と検索、MiGAP、アナトモグラフィー／BodyParts3D の使い方」

2 日目：一步進んだバイオデータベース・ツールの使い方（中上級者向け）

10:00-12:00 「遺伝子発現データを使い倒す～R/BioConductor の導入と解析例～」

12:00-13:00 昼休憩

13:00-15:00 午後の部

Reactome を使い倒す http://lifesciencedb.jp/image/small_video_icon.png

Biomart を使ってマイクロアレイ probeID の対応表を作る

Biomart を使って遺伝子の上流配列を取得する

Galaxy を使って転写因子予測結合領域と遺伝子上流領域の「交差点」をリストアッ

プする。

平成 22 年度

①統合データベース講習会：AJACS 湘南 3

日時：2010 年 6 月 4 日（金）午前 10 時～午後 5 時

場所：[日本大学生物資源科学部](#) 2 号館 231 教室

対象：学部 3、4 年生や大学院生、教員を始めとしたライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容：

10:00-10:20 「はじめに：統合データベースプロジェクトとは？」

10:20-10:40 [「統合TV」](#)

10:40-10:50 休憩

10:50-11:40 [「生命科学横断検索の利用法」](#)

11:40-12:20 [「文献情報を効率的に活用する」](#)

12:20-13:20 昼休憩

13:20-14:20 [「遺伝子発現データの使い倒し術」](#)

14:20-15:20 「次世代サーバの活用法」

15:20-15:40 休憩

15:40-16:20 「[BodyParts3D/Anatomography](#)の利用法」

16:20-17:00 「[KazusaMart](#)の使い方」

②AJACS & 第 22 回 DDBJing 講習会

日時：2010 年 6 月 23 日（水）10 時から 17 時まで

2010 年 6 月 24 日（木）10 時から 17 時まで

場所：ライフサイエンス統合データベースセンター（[東京大学本郷キャンパス](#) 工学部 12 号館 4 階）

主催：大学共同利用機関法人 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) 国立遺伝学研究所 生命情報・DDBJ 研究センター 日本 DNA データバンク(DDBJ)

対象：DBCLS・DDBJ を利用されている方

内容：

第 1 日目 6 月 23 日（水） 10:00～17:00

9:30-10:00 受付

10:00-10:15 統合データベースプロジェクトの紹介

高祖 歩美(ライフサイエンス統合データベースセンター 特任研究員)

10:15-11:30 カタログサービス・横断検索の利用法

川本 祥子(ライフサイエンス統合データベースセンター 特任准教授)

11:30-12:20 文献関連サービスの活用法

山本 泰智 (ライフサイエンス統合データベースセンター 特任研究員)

12:20-13:30 昼食休憩

13:30-15:00 遺伝子発現情報を使い倒す
坊農 秀雅 (ライフサイエンス統合データベースセンター 特任准教授)

15:00-15:20 休憩

15:20-16:10 BodyParts3D/Anatomography の利用法
三橋 信孝 (ライフサイエンス統合データベースセンター 特任研究員)

16:10-17:00 DBCLS Galaxy の利用法
山口 敦子 (ライフサイエンス統合データベースセンター 特任准教授)

第 2 日目 6 月 24 日 (木) 10:00～17:00

10:00-11:30 DDBJ の紹介と配列データの検索
中村 保一 (国立遺伝学研究所 教授)

11:30-12:30 次世代サーバ配列の登録・データ解析
次世代サーバのクラウド型解析パイプライン (15 分)
神沼 英里 (国立遺伝学研究所 助教)
次世代シーケンサアーカイブ DB (20 分)
児玉 悠一 (DDBJ 構築チーム アノテータ)
クラウド型解析パイプライン・実習 assembly/mapping (25 分)
望月 孝子 (国立遺伝学研究所 エンジニア)

12:30-13:30 昼食休憩

13:30-15:10 SAKURA を用いた塩基配列登録の方法・実習
小菅 武英 (DDBJ 構築チーム アノテータ)

15:10-15:30 休憩

15:30-17:00 相同性検索と系統解析: BLAST と ClustalW
中村 保一 (国立遺伝学研究所 教授)

③統合データベース講習会: AJACS 加賀

日時: 2010 年 7 月 6 日 (火) 午前 9 時 30 分～午後 4 時

場所: 金沢大学 角間キャンパス 自然研本館 2 階 計算機実習室 1

対象: 学部 3、4 年生や大学院生、教員を始めとしたライフサイエンス分野のデータベースを効率よく利用することに興味のある方や統合データベースが実際にどのように作成、維持管理されているかに興味のある方

内容:

9:30 - 09:45 [「はじめに: 統合データベースプロジェクトとは？」](#)

9:45 - 10:45 [「生命科学横断検索の利用法」](#)

10:45 - 11:45 [「タンパク質データベースの活用法」](#)

11:45 - 13:00 昼休憩

13:00 - 14:50 [「遺伝子発現データの活用法/次世代サーバの活用法」](#)

15:00 - 16:00 [「文献情報を効率的に活用する」](#)

④統合データベース講習会：AJACS みちのく

日時：2010年8月4日（水）12：30～18：00

2010年8月5日（木）10：00～17：00

場所：東北大学 星陵地区 IT 教育施設（医学部4号館6階）

対象：ライフサイエンス分野のデータベースを効率よく利用し、研究等に役立てたい方

内容：

8月4日（水）

12:30～15:00 統合データベースの利用法

統合データベースプロジェクトとは

[統合TVを使い倒す](#)

[統合データベースプロジェクトのサービスを使い倒す](#)

15:30～18:00 次世代サーバの活用法

[DDBJのNGSデータアーカイブ: DRA \(DDBJ Sequence Read Archive\)](#)

[おまけ：第三世代サーバ](#)

[データの解析法](#)

8月5日（木）

10:00～12:30 情報の可視化

[ゲノムブラウザ、Cytoscape](#)

[BodyParts3D/Anatomography](#)

14:00～17:00 遺伝子発現情報を使い倒す

[遺伝子発現データベースの活用法](#)

[R/Bioconductorを使った遺伝子発現解析](#)

⑤統合データベース講習会：AJACS 本郷7

日時：2010年10月4日（月）13：00～18：00

場所：ライフサイエンス統合データベースセンター（東京大学本郷キャンパス浅野地区工学部12号館4階）

対象：ライフサイエンス分野のデータベースを効率よく利用し、研究等に役立てたい方

内容：

13:00～13:20 [統合データベースプロジェクト](#)について

[河野信](#)（[ライフサイエンス統合データベースセンター](#)）

13:20～14:00 [TogoProt](#)とデータ統合

畠中秀樹（[ライフサイエンス統合データベースセンター](#)）

14:00～14:40 [ワークフロー](#)による配列及び構造解析

野口保、田代俊行（[産業技術総合研究所](#) [生命情報工学研究ユーザ情報統合班](#)）

15:00～15:40 [糖鎖関連データベース](#)の紹介

鹿内俊秀（[産業技術総合研究所](#) [糖鎖医工学研究ユーザ](#)）

15:40～16:20 [tRNADB-CE](#)：エキスパートがキュレートしたtRNA遺伝子データベースの紹介

阿部貴志 (長浜バイオ大学)

16:40 ~ 17:20 微生物ゲノム自動アノテーションパイプライン [MiGAP](#)の使い方・使われ方

菅原秀明 (国立遺伝学研究所)、兼崎友 (東京農業大学)、大山彰 (インシリコバイオロジー)、黒川顕 (東京工業大学)

17:20 ~ 18:00 [GWAS/CNV-DB](#)とMutation DBの紹介

小池麻子 (日立製作所 中央研究所)

⑥統合データベース講習会：AJACS 近江 3

日時：2010年11月5日(金) 13:30~17:00

場所：長浜バイオ大学命江館2階 情報実習室1

対象：長浜バイオ大学生命情報科学コース3回生 + ライフサイエンス分野のデータベースを効率よく利用し、研究等に役立てたい方

内容：

13:30-13:45 はじめに～「統合データベースプロジェクトとは」 担当・川本祥子
(DBCLS) [統合ホームページ\(http://lifesciencedb.jp/\)](http://lifesciencedb.jp/)から データベース
カタログ&横断検索の簡単な実習

13:45-14:25 「[Kazusa Annotation Suiteによるゲノムアノテーション](#)」

14:25-15:05 「[遺伝子発現データベースの活用法](#)」

15:05-15:20 休憩

15:20-16:00 「[TogoProtから探るタンパク質](#)」

16:00-17:00 「[Galaxy でできるゲノムスケール共同研究](#)」

(3) BioHackathon

平成19年度は、国内外の研究機関等関係者との連絡調整の一環として、平成20年2月11日～15日の5日間、ウェブサービスの統合を目指した現状調査・情報交換および技術開発を目的として、BioHackathon2008を開催した。本会議は、情報・システム研究機構と生命情報工学研究ユーザの協力のもとに開催した。

以下に、初日に開催されたの講演会の内容と各分科会の議論の内容を示す。

1) 講演会

- Objectives of this hackathon and current status of Japanese web services - Toshiaki Katayama
- From Web API for Biology (WABI) to Semantic Web API for Biology (SABI) - Hideaki Sugawara
- Current status of the BioMOBY project and vision for the future directions - Mark Wilkinson
- Moby, Legacy Apps and the Semantic Web - Paul Gordon
- Data, Services and Computational Resources Integration at the INB - Oswaldo Trelles
- The EMBRACE project and WS-I standard - Jan Christian Bryne

- Soaplab2 project to wrap up command line packages - Martin Senger
- Taverna (part of myGrid project) - Tom Oinn
- Generation Challenge Program effort at building interoperability - Richard Bruskiewich

2) 分科会の構成と議論の内容

• OpenBio、Web サービスのための Semantics と Ontology

語義、プロトコル、ドメインモデル、データタイプ、サービスタイプ（含 BioMoby 等）、API 等。特に語義の重要性が再確認され、異なるサービス間での同一語義の使用に関して議論された。

• 標準化

Glycoinformatics、相互作用ネットワーク、テキストマイニング、Phyloinformatics の各分野で取り扱うデータが異なるため、それぞれの分野にて標準化のための議論がなされた。

• 非同期 Web サービス

同期 Web サービスのみならず、非同期のサービスも必要であるため、そのためのアプリケーション、環境、方法に関して議論された。

• ラージデータ

バイオでは頻繁に巨大データを扱うが、SOAP では扱えない等の問題がある。その取り扱い方法に関して議論された。

• セキュリティ

Web サービス提供では当然セキュリティを確保する必要がある。認証方法、セキュリティレベル、計算機資源保護等に関して議論された。

• サービス記述

Web サービスに対する適切な説明や、ユーザがそのサービスを迅速に探し当てるためにはどうすればよいか等に関して議論された。

• OpenBio

バイオデータを RDBMS に格納し検索するための標準モデル（BioSQL）、既存のバイオプロジェクトに関して議論された。

平成 20 年度はライフサイエンスにおける標準化技術および標準を利用する技術をより安定化させるため、国内外の生命科学研究機関との情報交換および共同での技術基盤開発を目的として、平成 21 年 3 月 16 日～20 日の 5 日間、ライフサイエンス統合データベースセンターと沖縄科学技術研究基盤整備機構の協力の下に BioHackathon 2009 を開催した。初日の 3 月 16 日に一般向けに東京大学武田ホールにて公開シンポジウムを開催した。17 日から 20 日の 4 日間、沖縄科学技術研究基盤整備機構 Seaside House にて、情報技術開発者や生物系研究者が標準化技術のよりよい利用の議論や開発のための開発合宿を行った。シンポジウム参加者は 80 名、17 日からの開発合宿参加者は 58 名（うち海外参加者 30 名）である。

また、3 月 16 日東京大学武田ホールにて、BioHackathon 2009 のオープニングとして、

公開シンポジウムを行った。シンポジウムのプログラムは以下の通りである。

9:00 - 9:20	Introduction to the BioHackathon 2009 and integration efforts in DBCLS	Toshiaki Katayama
9:20 - 9:40	Integrated reproducible analysis with Galaxy	James Taylor
9:40 - 10:00	Biomolecular Function Discovery with the ANNOTATOR	Tobias Gattermayer
10:00 - 10:20	Accelerating experimentation and biological discovery within RIKEN OSC LifeScienceAccelerator -- eeDB gLyphs	Jessica Severin
10:20 - 10:40	BioMart and third party applications: Working towards better interoperability	Arek Kasprzyk
10:40 - 11:00	The Taverna 2 Workflow Engine API : Extensible, High Performance and Easy to Integrate	Tom Oinn
11:00 - 11:20	Multi-repository architecture for service integration	Oswaldo Trelles
11:20 - 11:40	Integrating literature and annotation resources: the Elsevier Grand Challenge experience	Alberto Labarga
11:40 - 12:00	Looking back on a WABI life for tomorrow	Hideaki Sugawara

更に、3 月 17 日から 20 日の 4 日間、沖縄科学技術研究基盤整備機構の施設である Seaside House にて、開発合宿を開催した。開発合宿では、初日の午前中に参加者全員で共通の問題意識について話し合い、それをもとにいくつかのワーキンググループに分かれ、初日午後以降、グループごとに活発な議論と開発を行った。参加者が多かったグループとして、次世代サーバのデータをはじめとした巨大なデータについてどのようにデータを溜めどのように利用するかを議論するグループ、セマンティックウェブ技術をどのように生物系のデータに利用していくか議論するグループ、タンパク質相互作用やゲノムなどの大きなデータをいかに可視化するかについて議論するグループがあった。その他、具体的なユースケースについて議論するグループ、文献データの処理について議論するグループ、ゲノム解析プラットフォーム Galaxy の国際化や今後の展開について議論・開発するグループなど、十数のグループで活発な議論が行われ、開発に反映された。またグループ間の議論の交換も頻繁に行われた。例えば、巨大なデータをどのように蓄積し、どの利用するかに関する話題は、それをどのように可視化すべきかという話題に密接に関わってくる。そのため、関連するグループの参加者間では意見交換や情報交換が行われ、結果として、横断的に開発に生かされることになった。

この合宿の成果物は大きく二つに分けることができる。すなわち、開発による成果物、および、情報共有による成果物である。開発による成果は、セマンティック自動発見統合フレームワーク SADI に対するプログラミング言語 Perl API (アプリケーション・プログラミング・インターフェイス) 開発、Galaxy 国際化のための開発、Galaxy とゲノムブラウザ gbrowse の間でのデータ入出力機能の開発、ゲノム解析環境 g-language の新しいデータ交換機能や解析パラメータの開発、癌ゲノムデータ閲覧ソフトウェア PARP

に対する 3D 可視化ツールの開発、可視化ツール Processing でネットワーク解析ツール Cytoscape の結果を表示するための開発、DDBJ-KEGG-PDB ワークフロー連携のための開発、糖鎖情報解析のためのワークフローの開発などが挙げられる。これだけの量のシステム開発が 5 日間で可能であったのは、関係する開発者同士が顔を合わせて議論をすすめお互いを刺激しながら開発を行う Hackathon スタイルならではのといえる。情報共有による成果は、参加者同士が問題意識の共有を行ったこと、あるいは、顔を合わせて密接に議論したことで、今後の共同研究などにつながっている。実際に、この会議をきっかけに転写制御のグループのメンバーは共同研究を始めており、将来的には潜在的な転写制御を検索するサービスを開始する予定である。また、遺伝子アノテーションパイプライン ANNOTATOR の大規模解析に東京工業大学の スーパーコンピュータ TSUBAME を利用するなどの提案、Galaxy の開発について開発者とライフサイエンス統合データベースセンターの要望についての情報交換、ウェブサービス設計ガイドラインについての議論コミュニティの作成など、BioHackathon 2009 での情報共有からはじまる共同研究の輪は確実に広がっている。

会議開催後におこなった参加者アンケートでは、他の参加者と議論する十分な時間があつたこと、システムを開発するのに非常に向いた環境であつたことなどの理由で BioHackathon 2009 は高い評価を得ている。参加者からは「セマンティックウェブや巨大データなどのホットなトピックについて議論ができてよかった」「生物系研究者から直接ニーズをきけてよかった」「普段なら数か月かかる開発を 5 日で終わらせることができた」など、他の機会では得られない成果が得られたという声があがっており、これらの感想はこの会議が主催者の目論見どおりの形で成功裏に終わったことを示唆していると思われる。

平成 21 年度は統合検索に向けた基盤技術の調査と開発を目的とし、国内外の研究者を招聘等した国際開発会議（BioHackathon 2010）を開催した。統合検索は、統合化されたデータの上に構築される高度なデータ検索と閲覧支援によって構成され実現するものである。内外の研究者とともに、医学生物学分野における、いま現在の最先端のデータ統合技術やその応用についての調査検討をおこない、当ユーザが今後の統合検索実現化に向けて必要なリソースや技術、開発すべきことについての情報を得ることができた。初日の 2 月 8 日は独立行政法人産業技術総合研究所生命情報工学研究ユーザでシンポジウムを開催した。プログラム等スケジュールは下記の通りである。

2/8 (Mon)

Symposium 10:00-18:00

Program

Open 9:00

Talks 10:00-12:00

10:00-10:05 Toshihisa Takagi (DBCLS/DDBJ)

Opening remarks [video](#)

10:05-10:10 Toshiaki Katayama (HGC, University of Tokyo)

Introduction to the BioHackathon 2010 for Semantic Web

10:10-10:40 Erick Antezana

[Towards a Semantic Systems Biology: Biological Knowledge Management Using Semantic Web Technologies](#)

10:40-11:00 Matthias Samwald

[High-level knowledge representation on the Semantic Web: the Concept Web Alliance and related efforts](#)

11:00-11:20 Thomas Kappler and Jerven Bolleman (Swiss Institute of Bioinformatics)

[UniProt in RDF](#)

11:20-11:40 François Belleau (Bio2RDF)

[Bio2RDF cognoscope : A killer app for the life science.](#)

11:40-12:00 Heiko Horn (Novo Nordisk Foundation Centre for Protein Research)

[Reflect - text mining in Semantic Web](#)

Lunch break

Talks 13:30-14:50

13:30-13:50 Tetsuro Toyoda (RIKEN)

RIKEN SciNeS

13:50-14:10 Mark Wilkinson (Genome Canada Bioinformatics Platform)

[SADI - Semantic Web Services](#)

14:10-14:30 Andrea Splendiani (Rothamsted Research)

[Visualization and analysis of biological networks on the Semantic Web](#)

14:30-14:50 Mitsuteru Nakao (DBCLS) and Toshiaki Katayama (HGC, University of Tokyo)

[Extending TogoWS and Open Bio* libraries for Linked Data](#)

Break

Poster session 15:00-16:30

All invited participants

Open space discussion 16:30-18:00

Reception 19:00-21:00 2/9 - 2/11 9:00-18:00

Day 1 (Tuesday)

• The first activity of the group was to watch [Tim Berner Lee presentation about Linked Data at TED](#) The whole group also shouted RAW DATA NOW for the the life science.

• [RDF primer lecture by François Belleau](#)

• [SadiDayOne](#)

• [TextMiningDayOne](#)

• [OpenBio](#)

• (ongoing: [FAQ](#))

Day 2 (Wednesday)

- [SPARQL primer lecture by François Belleau](#)
- [TextMiningDayTwo](#)
- 16:30 [Data provider session](#)
- [NextGenerationSequencing](#)
- (ongoing: [FAQ](#))

Day 3 (Thursday)

- Predicate
- [DDBJ-KEGG-PDBi](#)
- [TextMiningDayThree](#)
- (ongoing: [FAQ](#))

Day 4 (Friday)

- [TextMiningDayFour](#)
- (ongoing: [FAQ](#))

Hacks

- [G-languageProject](#)
- [SystemsBiologySBML](#)
- [VirutosoQuickPrimer](#)
- [VisualizationAndInteraction](#)
- [GFF3toRDF](#)

Hack, hack, hack

2/12 9:00-18:00

Hack, hack, hack 9:00-16:00

[SummarySession](#) 16:00-18:00

Farewell party 18:00-

平成 22 年度は、これまで得てきた国際技術動向や知見を国内で共有する必要があると考え、国内のデータベース構築者・利用者を対象に、統合データベース技術情報交換ワークショップ(<http://wiki.lifesciencedb.jp/mw/index.php/BH10.10>)を行った。国内から 70 名弱の参加者があり、活発な議論と情報交換がおこなわれた。プログラム等スケジュールは下記の通りである。

会期：2010 年 10 月 18 日（月）～22 日（金）

場所：静岡県修善寺 ラフォーレ修善寺

内容（プログラム）：

2010-10-18（月）

13:30-14:00 オープニング

中尾 光輝（DBCLS）：開催趣旨説明

大久保 公策（NIG DDBJ センター長）：ウェルカムメッセージ

14:00-15:00 シンポジウム・セッション 1：統合データベースプロジェクト

川本 祥子 (DBCLS) : 文部科学省ライフサイエンス統合データベースプロジェクトと DBCLS のサービス概要とポータル・横断検索

村上勝彦、山崎千里 (AIST) : 経済産業省ライフサイエンス統合データベースプロジェクト

15:00-17:00 ライトニングトーク 参加者全員、一人 3 分の [4 コマプレゼンテーション](#)
17:00-18:00 ポスターセッション

[4 コマプレゼンテーション](#)をA4 四枚に印刷したものを掲示。

18:00 懇親会

2010-10-19 (火)

9:30-12:00 シンポジウム・セッション 2 : データベース

中村保一 (DDBJ) : DDBJ ってみんなの役にたつんだろうか

飯田啓介 (DBCLS) : データベースと編集 (仮)

熊谷禎洋 (日立ソリューションズ) : データベース受け入れ事業

岡本忍 (DBCLS) : KazusaAnnotation によるゲノムアノテーション協働作業手法の開発

五十嵐芳暢 (医薬基盤研究所) : トキシコゲノミクスインフォマティクスプロジェクト

河野信 (DBCLS) : 統合 TV とスクリーンキャスト (仮)

荒川和晴 (慶應義塾大学先端生命科学研究所) : ウェブサービス統合のためのユーザインターフェイス

櫻井望 (かずさ DNA 研究所) : プロジェクト専用データベースの事後公開にあたって

13:00-15:00 シンポジウム・セッション 3 : サービス・技術

西山智明 (金沢大学/ERATO 長谷部分化全能性) : SOLiD シーケンスデータの解析

川路英哉 (RIKEN) : 公共 DB へのデータ登録とフォーマット

粕川雄也 (RIKEN CDB) : Brain Star の API と可視化 (仮)

木下聖子 (Soka University) : 糖鎖情報の統合を目指して

金 進東 (DBCLS) : バイオ文献からの情報抽出ー世界の動向

15:00-19:00 チュートリアル

牧口大旭 (三井情報)、杉崎太一朗 (三井情報) : [横断検索のP2P型データベース統合](#)

片山俊明 (HGC) : TogoDB/TogoWS

神沼英里 (DDBJ) : DDBJ パイプライン

三橋信孝 (DBCLS) : [BodyParts3D/Anatomography](#)

西田孝三 (NAIST) : [CytoscapeへのKEGGデータベース情報の取り込みと実験情報のマッピング](#)

大林武 (東北大学) : 共発現データベース ATTED-II が広く利用されるようになった訳

荒川和晴（慶應義塾大学先端生命科学研究所）：[Dashcode 3 による次世代ウェブアプリケーション開発](#)

2010-10-20（水）

9:30-11:00 ワークショップ 1

西田孝三：[Cytoscapeを用いたデータベースクライアント、可視化](#)

後藤直久：[BioRuby](#) と [ソースコードレポジトリ](#)

11:00-12:00 オープンスペース

これまでの話題から、掘り下げて議論や調査、開発するトピックを列挙する。

13:00-18:00 ハッカソン 1

ハッカソンでは、調査や開発、議論を小グループに別れて実施する。

- ・ [データベースのRDF化](#) 金城
- ・ [実用的なバイオオントロジー関連技術の調査](#) 川島
- ・ [ウェブサービスマニフェストの開発とBioDBCore対応](#) 山崎
- ・ データベースのライセンス問題
- ・ [糖鎖API標準化部分の実装](#) 木下
- ・ [TogoWS, TogoDB の開発](#) 片山
- ・ [BioRuby](#) 後藤
- ・ [iPhone アプリ](#)
- ・ ブックマークレット開発
- ・ バイオデータ可視化
- ・ [Cytoscapeプラグインを作る](#)
- ・ [Galaxy ツールを作る](#) 神田
- ・ R/Bioconductor パッケージを作る
- ・ [OSQAを利用したあたらしいQAサイトもしくはサイトホスティング](#) 中尾、神田
- ・ [オルソログクラスターへの自動命名手法の開発](#) 岡本
- ・ [ウェブサービスAPI とCytoscapeプラグイン](#) 大林、笠原、西田

18:00 から今日のまとめ。

2010-10-21（木）

9:30-12:00 ハッカソン 3

11:45 午前のまとめ

13:00-19:00 ハッカソン 4

18:30 今日のまとめ

2010-10-22（金）

9:30-12:00 ハッカソン 5

13:00-15:00 ハッカソン 6

15:00-16:00 まとめ

6. 新聞発表等

平成18年度の新聞、雑誌等における本プロジェクトに関する記事は以下の通りである。

タイトル	メディア名	掲載日
「ライフサイエンス分野の統合データベース整備事業」の成果公開についてーライフサイエンス研究の発展に向かいけてー	—	—

平成 19 年度の新聞、雑誌等における本プロジェクトに関する記事は以下の通りである。

タイトル	メディア名	掲載日
生命科学 DB 統合へ	朝日新聞	H20.3.21
「BioHackathon 2008」が六本木で開幕、13 カ国から 80 人参加、12 日から 4 日間お台場で「Hack、hack、hack」	日経バイオテク	H20.2.12
生命科学の情報を網羅 今年度から統合データベース作成	読売新聞	H19.11.7
大学共同利用機関法人情報・システム研究機構、ライフサイエンス統合データベース・ポータルサイト開設	日経バイオテク	H19.10.5
統合 DB プロジェクトが本格始動 仮説の枯渇と蛸壺化を改善	BTJ ジャーナル (日経 BP)	H19.8

平成 20 年度の新聞、雑誌等における本プロジェクトに関する記事は以下の通りである。

タイトル	メディア名	掲載日
ライフサイエンス統合 DB センターがポータルサイトで各 DB を公開、[蛋白質核酸酵素] の総説記事は 6 月 2 日	Biotechnology Japan (日経 BP)	H20.5.28
ライフサイエンス分野 データベースを 11 年度から一元化(1 面)・ライフサイエンスの発展 真に必要なデータベースとは	科学新聞	H20.12.5
情報の循環を改善する統合 DB プロジェクト本格化	BTJ ジャーナル	H21.3.25
生命科学分野の DB から文献まで扱うわが国初のサービスを目指す	Medical Tribune	H21.3.26

平成 21 年度の新聞、雑誌等における本プロジェクトに関する記事は以下の通りである。

タイトル	メディア名	掲載日
文科省、バイオリフ DB を JST 事業と一本化	化学工業日報	H22.1.28
記者発表、情報・システム研究機構、 H20 年度「統合データベースプロジェクト」事業成果に対するユーザ評価の結果	Biotechnology Japan (日経 BP)	H22.2.3

平成 22 年度の新聞、雑誌等における本プロジェクトに関する記事は以下の通りである。

タイトル	メディア名	掲載日
情報・システム研究機構(ROIS)、産総研など 生命科学系 DB 相互検索一省の垣根超え効率化ー	日刊工業新聞	H22.5.25

情報・システム研究機構(DBCLS)、産総研など DB 相互検索可能に—文科・経産関連を統合—	化学工業日報	H22.5.25
---	--------	----------

7. 実施体制

役割 ◎：業務主任者、●：サブテーマ代表、○：実施担当

氏名	所属		研究項目	年度毎の登録				
	部門	役職		H18	H19	H20	H21	H22
堀田 凱樹	情報・システム研究機構	機構長	サブテーマ（１）	◎	—	—	—	—
高木 利久	ライフサイエンス統合データベースセンター	特任教授	サブテーマ（１），（２），（３）	●	◎ ●	◎ ●	◎ ●	◎ ●
大久保 公策	国立遺伝学研究所生命情報・DBJ 研究センター	教授・センター長	サブテーマ（２）ヒト統合 DB の開発・運用	●	●	●	●	●
小笠原 理	国立遺伝学研究所生命情報・DBJ 研究センター	助教	サブテーマ（２）ヒト統合 DB の開発・運用	—	○	○	○	○
川本 祥子	ライフサイエンス統合データベースセンター	特任准教授	サブテーマ（３）ポータル整備・運用、広報、普及啓発	○	● ○	● ○	● ○	● ○
菅原 秀明	国立遺伝学研究所生命情報・DBJ 研究センター	特任教授	サブテーマ（２）モデル生物・産業応用生物統合 DB の開発・運用	○	○	○	○	○
高野 明彦	国立情報学研究所連想情報学研究開発センター	教授・センター長	サブテーマ（１）	○	—	—	—	—
武田 英明	国立情報学研究所情報学プリンシプル研究系	教授	サブテーマ（２）共通基盤技術開発	○	—	—	—	—
五條堀 孝	国立遺伝学研究所生命情報・DBJ 研究センター	教授	サブテーマ（１）	○	—	—	—	—
西川 建	国立遺伝学研究所生命情報・DBJ 研究センター	教授	サブテーマ（２）共通基盤技術開発	○	—	—	—	—
藤山 秋佐夫	国立情報学研究所情報学プリンシプル研究系	教授・研究主幹	サブテーマ（１）	○	—	—	—	—
市瀬 龍太郎	国立情報学研究所プリンシプル研究系	准教授	サブテーマ（２）共通基盤技術開発	—	—	—	—	○
岡本 忍	ライフサイエンス統合データベースセンター	特任准教授	サブテーマ（２）モデル生物・産業応用生物統合 DB の開発・運用	○	—	—	○	○
水田 洋子	ライフサイエンス統合データベースセンター	プロジェクト研究員Ⅱ	サブテーマ（２）共通基盤技術開発	○	—	—	—	—
三橋 信孝	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（２）ヒト統合 DB の開発・運用	○	○	○	○	○
永井 啓一	ライフサイエンス統合データベースセンター	特任教授	サブテーマ（１），（３）	—	○	○	○	○
箕輪 真理	ライフサイエンス統合データベースセンター	特任准教授	サブテーマ（１），（３）広報・普及啓発	—	—	○	○	○
西川 哲夫	ライフサイエンス統合データベースセンター	特任教授	サブテーマ（１），（３）	—	○	○	○	—
吉羽 洋周	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（１）	—	—	—	○	○
山口 敦子	ライフサイエンス統合データベースセンター	特任准教授	サブテーマ（２）共通基盤技術開発	—	○	○	○	○
坊農 秀雅	ライフサイエンス統合データベースセンター	特任准教授	サブテーマ（２），（３）	—	○	○	○	○
山本泰智	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（２）共通基盤技術開発	—	○	○	○	○

河野 信	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（３）ポータル整備・運用、広報・普及啓発	－	○	○	○	○
仲里猛留	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（３）ポータル整備・運用	－	○	○	○	○
全 弘宇	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（２）共通基盤技術開発	－	○	○	○	－
藤枝 香	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（２）ヒト統合 DB の開発・運用	－	○	○	○	○
坂東 明日香	ライフサイエンス統合データベースセンター	特任技術専門員	サブテーマ（３）データベースの受入と運用	－	－	－	○	○
S.Riedel	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（２）共通基盤技術開発	－	－	○	－	－
中尾 光輝	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（２）共通基盤技術開発	－	－	－	○	○
大内田 賢太	ライフサイエンス統合データベースセンター	特任技術専門員	サブテーマ（２）共通基盤技術開発	－	－	○	○	○
高祖 歩美	ライフサイエンス統合データベースセンター	特任研究員	サブテーマ（３）広報・普及啓発	－	－	○	○	○
小野 浩雅	ライフサイエンス統合データベースセンター	RA（→特任技術専門員）	サブテーマ（３）ポータル整備・運用	－	○	○	○	○
中村 春木	ライフサイエンス統合データベースセンター	客員教授	サブテーマ（３）データベースの受入と運用	－	○	○	○	○
金城 玲	ライフサイエンス統合データベースセンター	客員准教授	サブテーマ（３）データベースの受入と運用	－	○	○	○	○
畠中 秀樹	ライフサイエンス統合データベースセンター	特任准教授	サブテーマ（３）データベースの受入と運用	－	－	－	○	○
金 進東	ライフサイエンス統合データベースセンター	特任准教授	サブテーマ（２）共通基盤技術開発	－	－	－	－	○
飯田 啓介	ライフサイエンス統合データベースセンター	特任技術専門員	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
王 悦	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（２）共通基盤技術開発	－	－	－	－	○
大村 蓉子	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	○	○	－
千葉 温子	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	○	○	－
山口 瑶子	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	○	○	－
横山 貴央	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	○	○	○
大石晶一郎	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
大田 達郎	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
奥田 裕樹	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
小野 智央	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
熊谷 雄介	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
高橋 公海	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	－
寺田 愛花	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
並木 洋平	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○

根来 哲也	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
橋本 拓也	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	－
林 和弘	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
山田 恵	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	○
吉田 康久	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	○	－
飯田 宗玄	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
伊藤 真和吏	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
久保 奈月	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
坂田 隼人	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
水津 珠世	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
高橋 文緒	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
武田 英之	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
竹馬 光太郎	ライフサイエンス統合データベースセンター	技術補佐員	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
趙 漢哲	ライフサイエンス統合データベースセンター	RA	サブテーマ（２）共通基盤技術開発	－	－	－	－	○
比留川 治子	ライフサイエンス統合データベースセンター	RA	サブテーマ（３）ポータル整備・運用	－	－	－	－	○
加藤 菊也	大阪府立病院機構大阪府立成人病センター研究所	所長	サブテーマ（２）ヒト統合 DB の開発・運用	○	○	○	○	○
黒川 顕	東京工業大学大学院生命理工学研究科生命情報専攻	教授	サブテーマ（２）モデル生物・産業応用生物統合 DB の開発・運用	－	○	○	○	○
内山 郁夫	自然科学研究機構基礎生物学研究所	助教	サブテーマ（２）モデル生物・産業応用生物統合 DB の開発・運用	－	○	○	○	○
伊藤 剛	（独）農業生物資源研究所基礎研究領域ゲノム情報研究ユニット	ユニット長	サブテーマ（２）モデル生物・産業応用生物統合 DB の開発・運用	－	○	○	○	○
片山 俊明	東京大学医科学研究所ヒトゲノム解析センター	助教	サブテーマ（２）共通基盤技術開発	○	○	○	○	○
川島 秀一	東京大学医科学研究所ヒトゲノム解析センター	助教	サブテーマ（２）共通基盤技術開発	○	○	○	○	○
PAPAZIAN Frederic	ライフサイエンス統合データベースセンター	外来研究員	サブテーマ（２）共通基盤技術開発	－	－	－	－	○
岩崎 渉	東京大学大学院新領域創成科学研究科	助教	サブテーマ（２）共通基盤技術開発	－	－	－	－	○
簗島 伸生	浜松医科大学光量子医学研究センター	教授	サブテーマ（２）ヒト統合 DB の開発・運用	－	－	－	－	○
後藤 直久	大阪大学微生物病研究所	助教	サブテーマ（２）共通基盤技術開発	－	－	－	－	○
内藤 雄樹	東京大学大学院理学系研究科	助教	サブテーマ（３）ポータル整備・運用	－	－	－	○	○

なお、サブテーマ（１）は「戦略立案・実行評価」、サブテーマ（２）は「統合データベース開発」、サブテーマ（３）は「統合データベース支援」である。