平成20年度産業技術研究開発

「統合データベースプロジェクト」
に関する報告書

平成21年3月

（委託先）社団法人　バイオ産業情報化コンソーシアム

目次

1. 総括

ヒトゲノム配列決定計画に象徴されるように今日ではライフサイエンス研究は大規模事業となり、"BIG DATA"と呼ばれる大量データの生産が行われている。そして、研究者は大量データの中から有益な情報を抽出する作業により多くの時間と労力をかけるようになり、データベースの構築とその有効利用の重要性が以前にも増して高くなっている。こうした中、2005 年頃から総合科学技術会議等の場では、ライフサイエンス分野の「統合データベース」の必要性が議論され始めた。特に、基盤的なデータベースの整備拡充の遅れや大型プロジェクト研究の成果のデータベース化が不十分であることとともに、研究予算終了後にその維持管理責任の不明確等により成果物が散逸すること等が問題視され、対策の検討が始められた。これを受けて文科省では2006 年秋から「統合データベースプロジェクト」を開始し、2007 年 4 月には情報・システム研究機構にライフサイエンス統合データベースセンター（DBCLS）が設立された。一方経済産業省では、ミレニアムプロジェクト「統合データベース」の中で、環境対応技術開発「バイオインフォマティクス関連データベース整備」（2000-2005）を実施し、さらにモデル事業「ゲノム情報統合プロジェクト」（2005-2008）を実施することにより、ヒト全遺伝子のアノテーション統合データベース H-Invitational Database（H-InvDB; http://www.h-invitational.jp/）という基盤データベースを筆頭として種々の有用なデータベース群を開発してきた。これらの成果は、今後の統合データベースの基盤となると期待されている。

こうした背景の下で、経済産業省では 2008 年 4 月から「経済産業省統合データベースプロジェクト」（2008-2011 予定）を開始することとし、社団法人・バイオ産業情報化コンソーシアム（JBiC）と独立行政法人産業技術総合研究所（産総研）・バイオメディシナル情報研究センター（BIRC）の共同チームが 2008 年度事業を実施することとなった。本プロジェクトの目的は、第一に、経済産業省関連機関により実施されたライフサイエンス分野の研究開発プロジェクトの成果であるデータベースに関する情報提供サイトを作成することである。第二に、ヒト遺伝子に関連した各種の研究成果に関して、ゲノム情報統合プロジェクトにおいて構築したヒト遺伝子の統合データベース H-InvDB を基礎として、経済産業省関連の研究成果を連携して利用できるシステムを構築することである。以上を通じて、これまでに経済産業省関連機関により実施された研究開発プロジェクトの成果等を整備し、ライフサイエンス分野における研究開発の促進に資するデータベースを構築することを目的としている。

ライフサイエンス分野は、自身の研究成果を既に蓄積されている公知の情報や他の研究データと対比することにより、自身の研究の仮説を考案する手がかりを得たり、その検証を行ったりすることが極めて重要である。それによって例えば効果的な治療薬など新しい実用化の発想が得られたりする可能性がある。このため、国家プロジェクト等により産生された研究データを基盤データベースとして一括して活用できるデータベースが、産業界・学術界や社会から必須の知的基盤として強く要望されている。本プロジェクトではこうした強い要望に応えることを至上命題にしている。

こうして開始された本プロジェクトは、本報告書で詳細を述べるように計画通りに順調に成果を出すことができた。2008 年 4 月からデータベース便覧作成のための調査を開始し、2008 年 10 月にポータルサイト MEDALS（http://medals.jp/）を先行公開した。この時点では便覧で公開したデータベースの情報は限られていたが、その後も調査と情報整備の作業を進め、2009 年 2 月には件数を大幅に充実させたデータベース便覧を MEDALS から公開した。また、国際的な利便性を図って英語版のポータルサイトを新規に公開した。一方、経済産業省

プロジェクトの成果物の利用促進のための各種サービスの提供については、リンク自動管理システム（Hyperlink Management System; http://biodb.jp/index.cgi?lang=jp）、ID 一括変換システム（ID Converter System; http://biodb.jp/index.cgi?lang=jp#ids）、新規関連文献お知らせツール（PubMedScan）というオリジナリティのある有用なツール群を公開した。

さらに、ヒトの分子データの実質的な統合についての作業も多くの成果を挙げた。ヒト全遺伝子のアノテーション統合データベース H-InvDB と独立行政法人新エネルギー・産業技術総合開発機構（NEDO）「機能性 RNA プロジェクト」の成果である機能性 RNA データベース fRNAdb および機能性 RNA 用 UCSC ゲノムブラウザ（http://www.ncrna.org/）を連携させるため、双方のシステムに共通仕様のウェブサービス（Web サービス）を開発して導入し、ユーザーが双方の情報をひとつの画面の中で閲覧することができるしくみを実現した。同様に、産総研・糖鎖医工学研究センターで測定された糖転移反応に関する情報を集めた糖鎖関連遺伝子データベース（GlycoGene Database （GGDB）; http://riodb.ibase.aist.go.jp/rcmg/ggdb/）と H-InvDB の連携についても、ウェブサービスの導入によって情報の統合に成功した。以上の成果は研究開発者やその他の利用者の利便性を高め、データベースからの知識の取得を促進・効率化すると期待される。

このほか、産業界からのニーズをアンケートやインタビュー等により調査したこと、成果の広報活動を積極的に行いマスメディア等にも取り上げられたこと、論文投稿や学会発表を行ったことなど、データベース構築を取囲む諸活動についても多くの成果を挙げることができた。今後も産業界等の研究者の意見を反映させつつ、本プロジェクトを継続的に実施していくことが、きわめて重要であると考える。

経済産業省統合データベースプロジェクト・プロジェクトリーダー　五條堀　孝

サブリーダー　今西　規

サブリーダー　村上　勝彦

2. 研究開発成果

2. 1　経済産業省ライフサイエンスデータベース・ポータルサイトの構築と運営

- 目的

ライフサイエンス分野は、自身の研究成果を既に蓄積されている研究成果や研究データと対比することにより、自身の研究成果の仮説を考案する手がかりが得られたり、効果的な治療薬など新しい実用化の発想が得られたりする可能性がある。国家プロジェクト等により産生された研究データを一括して活用できるデータベースが、産業界や社会から要望されている。本事業では、経済産業省関連機関により実施されたライフサイエンス分野の研究開発プロジェクトの成果等（データベース、解析ソフト等）に関する情報提供ポータルサイトを構築した。これまでに経済産業省関連機関により実施された研究開発プロジェクトの成果やそれらに関する様々な情報を集めて整備することが目的である。ポータルサイトの名称は、MEDALS とした。 これは、METI Database portal for Life Science の略であり、METIとは経済産業省（Ministry of Economy, Trade and Industry）である。

- 開発方針

本事業にあたり、製薬企業へのインタビューやインターネットによるアンケートを行い、ユーザーニーズを調査した（詳細は 2.3.2 を参照）。この調査から分かった「ユーザーがデータベースやツールを探している時の悩み」の一部を表 2.1-1 に示す。

表 2.1-1　ユーザーの悩み

項番	悩み
1	種類が多すぎてどれが良いのか良く分からない
2	特徴や操作方法が分からない
3	いつのデータなのか分からない
4	外部リソースが何なのか分からない
5	商用利用していいのか分からない
6	関連論文があるのか分からない

これより、ユーザーが成果物の利用に至るまでの悩みには、成果物の基本的な情報から商用利用まで幅があることがわかった。さらに、開発者の手元にはあるような情報が簡単に入手できないために生じる悩みが多いことがわかった。そこで、MEDALS の開発では 3 つの方針を立てて、ユーザーが成果物を利用しやすくすることで、データベースを有効活用してもらうため以下の工夫を行った（工夫の詳細については、2.1.1 節にて述べる）。

1. 成果物を一括して簡単に見られるサイトを提供する
2. 関連情報も含めて詳しい内容を提供する
3. 記述内容の品質を保つ

- MEDALS が提供するコンテンツ概要

MEDALS は、検索機能を備えたポータルサイトとして 10 月に公開を行った。その後、2 回のデータ更新を経て 2 月には計 71 件の成果物に関する便覧を提供するに至り、ユーザーからのアクセスも更新毎に増加している。また、産業技術総合研究所（産総研）の研究成果を活用した「新規関連文献お知らせツール」をダウンロードコンテンツとして、ポータルサイトより公開し、ユーザーニーズを満たす関心の高いツールになっている。ポータルサイトでは、内容を 6 つのカテゴリー（図 2.1-1 （ア）～（カ））に分類して提供している。各カテゴリーで提供する情報について概要を以下に記載する。

図 2.1-1 6 つのカテゴリー

（ア）データベース便覧では、様々なデータベースに関する情報や説明を便覧形式で閲覧することができる。特に、格納されている分子種によって小カテゴリーを作り、専門が大きく異なるユーザーでも利用し易くしている。具体的には、DNA・ゲノム、RNA、蛋白質、その他である。

（イ）解析ツール（ソフト）便覧では、データ解析に役立つツールに関する情報や説明を便覧形式で閲覧することができる。データベース便覧と同じように、格納されている分子種によって小カテゴリーを作り、専門が大きく異なるユーザーでも利用し易くしている。具体的には、DNA・ゲノム、RNA、蛋白質、その他である。

（ウ）ダウンロードデータ（アーカイブ）では、成果物をダウンロードできる。現在のアーカイブは医薬品開発支援用分子シミュレーション計算ソフトおよびデータベースの集合体である myPresto、新規関連文献お知らせツールの 2 件がダウンロードできる。myPresto は、国等から委託を受けて実施した「生体高

分子立体構造情報解析」等の NEDO プロジェクトの中で開発した分子シミュレーションを構成するプログラム群等である。

（エ）統合データベースへのリンク集では、本事業の分子データのメインデータベースであるヒト全遺伝子のアノテーション統合データベース H-InvDB と文部科学省が実施している統合データベースプロジェクトへリンクを設置している。分子情報の統合と他省庁との連携については、2.2 項と 2.4 項で述べる。

（オ）サイト内検索では、便覧に記載した情報を英語・日本語による検索（全文検索）することができる。この機能により、ユーザーは欲しい情報をキーワードのみで探し出すことができるだけではなく、関連する新たな成果を知ることもできる。

（カ）MEDALS ツールでは、本事業にて開発・応用したツールを提供している。例えばリンク自動管理システムは、多くの外部サイト間リンクを一元管理し、自動更新を行い、データベースの運用・管理のコストを大幅に削減するシステムである。詳しくは、2.2.1 節で述べる。また、新規関連文献お知らせツールは 2.1.4 節でも述べるが、ライフサイエンス分野のユーザーが日常業務の中で削減したい作業の上位に挙がった文献収集の手間を削減するシステムである。

- MEDALS の開発・運用スケジュール概要

2008 年度の開発・運用のスケジュールを図 2.1-2 に示す。10 月のサイト先行公開、2 回のデータ更新を行い、2 月に英語版公開とともに本格公開を行った。

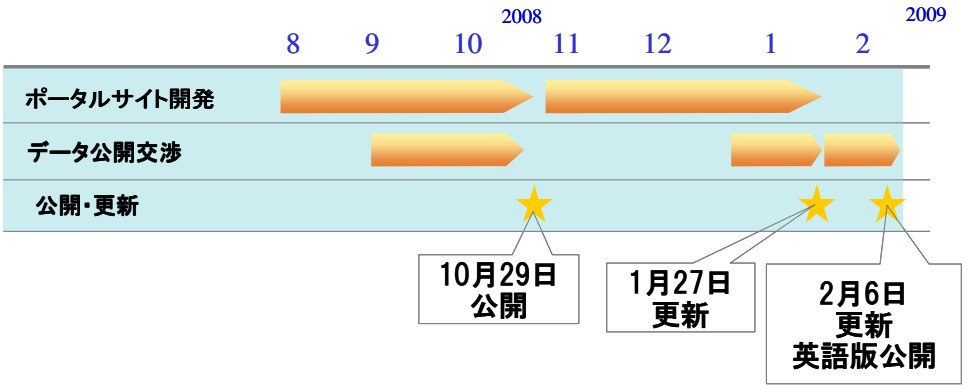


図 2.1-2 開発スケジュール概要

2．1．1　データベース・解析ツール便覧の公開

経済産業省関連の公的資金により産生される、ライフサイエンス分野の研究開発プロジェクトの成果である各種データベースやソフトウエア等の便覧を作成して公開した。便覧には、対象データベースを知らない人や、専門が大きく異なるユーザーでも理解しやすいものとなるように説明、表示を工夫した。文部科学省統合データベースプロジェクトとは、作成したカタログ等の情報提供、データベースの説明項目として何を載せるかなどの議論等を行い、連携を図った。

（1） 便覧に掲載する成果物の調査

始めに、ライフサイエンスに関するプロジェクト・成果物のリストを作成するための調査を行った。初期の調査によって、経済産業省関連のライフサイエンスに関するデータベース等の成果物の調査対象範囲は、バイオ産業情報化コンソーシアム（JBiC）、産業技術総合研究所（産総研）、製品評価基盤機構(NITE)、および独立行政法人新エネルギー・産業技術総合開発機構（NEDO）の 4 機関に絞れば十分であることが判明した。このため、データベース等の成果物の調査対象を上記 4 機関の成果物に定めた。

各機関に共通して行った調査内容は、各機関のウェブサイト（Web サイト）に記載された情報やパンフレットの調査、各機関の関係者へのインタビューなどである。但し、産総研の成果物は、研究成果発表データベース（RRPDB, URL: http://www.aist.go.jp/RRPDB/system/Koukai.Top）と研究情報公開データベース（RIO-DB, URL: http://riodb.ibase.aist.go.jp/riohomej.html）も加えて調査をした。また、NEDO については、ホームページ（http://www.nedo.go.jp/index.html）の成果報告書データベースに膨大な資料があった。そのため上記に加えて、以下の a) NEDO のデータベースを利用したプロジェクトリストの作成、b) プロジェクト報告書による成果物リストの作成、に示す 2 つの特別な調査を行った。

a)　NEDO のデータベースを利用したプロジェクトリストの作成

プロジェクト調査対象を 1981 年～2008 年のプロジェクト、成果報告書データベース検索サービスにおける対象抽出条件を以下の 2 つとして、NEDO が実施したライフサイエンス関連プロジェクトのリストを作成した。

抽出条件 1　特定技術分類（大分類）に分類されたプロジェクトであること

- バイオテクノロジー
- 医療福祉

抽出条件 2　特定担当部が担当するプロジェクトであること

- バイオ部
- 環境部
- 応技
- 研業
- エネ対部
- 基技
- 国際部

- バイオ健康福祉
- 研開推
- 企画調整部

検索項目を「技術分類（大分類）」と「担当部」で非排他的に抽出し、事業・プロジェクト紹介ページにてプロジェクトの分野を確認した。例えば、バイオテクノロジーで検索した場合は、NEDO の検索サービスからは図 2.1.1-1 に示すプロジェクトのリストを手に入れることができた。尚、現在当該検索システムは改修中の為、2.1.1-1 に示すリストは手に入れることはできない。

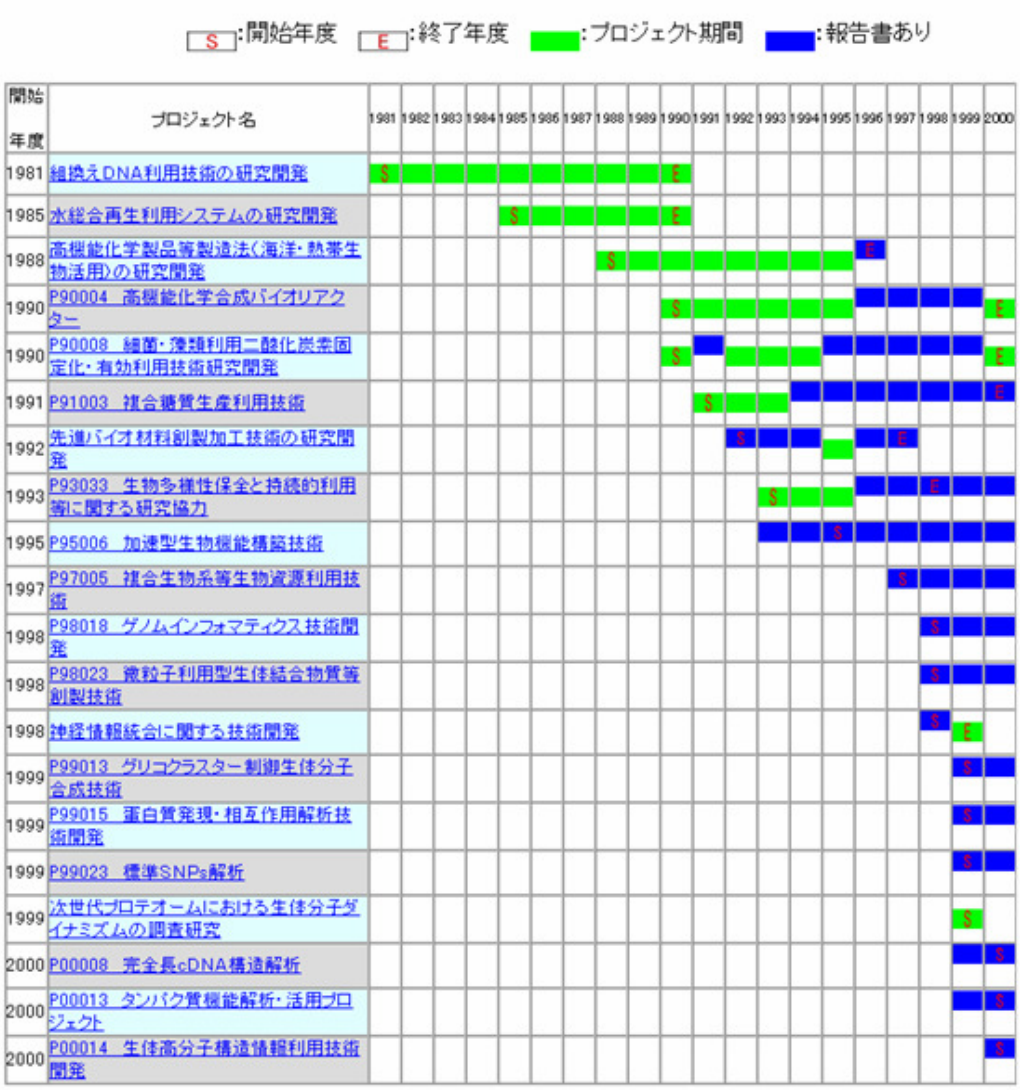


図 2.1.1-1　NEDO プロジェクトのリストの例

b) NEDO プロジェクト報告書による成果物リストの作成

収集したプロジェクトの報告書を NEDO から取得し、「データベース」、「ツール」、「システム」、「データセット」と書かれた成果物を拾いながら、全体を実際に読んで成果物リストを作成した。その結果、NEDO 関連成果物として、表 2.1.1-1 にあるリストを作成した（本リストでは NEDO から JBiC と産総研に委託され、2008 年度内に便覧にて公開された成果物は除いている）。

表 2.1.1-1 NEDO プロジェクト・成果物リスト

プロジェクト項番	プロジェクト名	成果名
1	化学物質のリスク評価及びリスク評価手法の開発	モニタリングDB PRTR DB 大気中の濃度マップ 初期リスク評価書 排出経路データシート 有害性評価書 PRTR対象化学物質総括管理表 詳細リスク評価書
2	標準SNPs解析	JSNP 解析工程支援システム コール作業支援システム 二次元散布図描画システム データ集計システム 公開用データチェックシステム アレル頻度集計システム 解析事業のアレル頻度データ
3	タンパク質機能解析・活用プロジェクト	ヒトの遺伝子全てに対するノックダウンsiRNA ライブラリーを作製し、ガンやHIV など、様々な疾患に関連する遺伝子、あるいは細胞分化等の様々な生物現象に関わる機能遺伝子を網羅的に同定するシステム BIOPREDsi off-target 排除システム Intris
4	ゲノムインフォマティクス技術開発	発現頻度情報収集用ソフトウェア 転写因子データベース（TFDB） 遺伝子同定支援ワークベンチシステム 遺伝子発現プロファイル計測実験支援システム
5	生体高分子構造情報利用技術開発	低分子処理のための支援ツール GamessViewer トポロジージェネレーター たんぱく質・核酸など生体分子を対象に計算するためのMD シミュレーション計算プログラprestoX トポロジージェネレーター:分子トポロジー作成と力場パラメータの割当をおこなうプログラム「topologene（ト
6	生物機能を活用した生産プロセスの基盤技術開発	MGF 研究支援システム 代謝物質標準液の測定データベース 代謝物質標準液のMS/MS スペクトルデータベース 生命分子ネットワークシステム設計支援システムCADLIVE
7	植物の物質生産プロセス制御基盤技術開発	KATANA(ANNOME, Metabolic Pathway Search, Gene Ontology Search, マメ科植物糖の代謝化合物データ DAGViz KaPPa-View(Kazusa Pathway viewer) MapDrawer(ツール)
8	複合生物系等生物資源利用技術	転写因子破壊株から得られた遺伝子発現プロファイルからの遺伝子制御配列解析結果（CD-ROM） 77の遺伝子発現プロファイルデータ 出芽酵母を構成する生体高分子間の相互作用66、668件とタンパク質間相互作用11、102件を納めた再編成データベース
9	多種遺伝子導入技術開発	DNAアレイによる遺伝子発現プロファイルデータを公開するデータベースシステム 遺伝子発現プロファイルと代謝産物プロファイルの統合データベースとしての代謝経路データベース kazusa Pathway
10	抗がん剤等の副作用及び効果予測システム等の評価・検証及び成果普及事業	SNP解析による副作用予測システム 遺伝子発現解析による治療に対する感受性や転移の予測システム
11	構造活性相関手法による有害性評価手法開発	毒性知識情報データベース 代謝知識情報データベース 有害性評価支援システム統合プラットフォーム
12	医用化合物スクリーニング支援システム	化合物データベース 医用化合物スクリーニング支援システム
13	蛋白質発現・相互作用解析技術開発	組織別遺伝子発現データベース 癌組織遺伝子発現データベース
14	完全長cDNA構造解析	cDNA機能アノテーション統合データベースシステム（cDNA Integrated Database System） 完全長cDNA発現プロファイルデータベース
15	染色体解析技術開発	染色体異常関連データベースの構築 個別化医療実現のために必要な日本人ゲノム多型のデータ収集と情報データベース
16	機能性RNAプロジェクト	ターゲットmRNA候補の発現変動を加味したターゲット予測システム
17	糖鎖構造解析技術開発（糖鎖エンジニアリングプロジェクト）	オリゴ糖データベース
18	定位的がん治療装置	治療計画データベース
19	微粒子利用型生体結合物質等創製技術	各種ライブラリデータ
20	グリコクラスター制御生体分子合成技術	バーチャルスクリーニングシステム
21	糖鎖合成関連遺伝子ライブラリーの構築	糖転移酵素データベース
22	エネルギー使用合理化生物触媒等技術開発	セルロース分解酵素、発現遺伝子のデータベース
23	環境中微生物の高精度・高感度モニタリング技術の開発	画像解析モジュールの開発Bacteria Automatic Counting System (BACS)
24	細胞内ネットワークのダイナミズム解析技術開発	リアルタイム・マルチカラー・デジタル録画システム
25	ゲノム情報に基づいた未知微生物遺伝資源ライブラリーの	RDP0Ⅱ
26	機能性糖鎖複合材料創製技術開発	糖鎖構造解析並びに機能予測システム
27	高機能簡易型有害性評価手法の開発	遺伝子情報、ラット毒性データ及び公開データを含めた統合データベース
28	化学物質総合リスク評価管理システムの開発	化学物質総合リスク評価管理システム
29	高精度・簡易有害性（ハザード）評価システムの開発	バイオインフォマティクスデータベースシステム
30	植物利用エネルギー使用合理化工業原料生産技術開発（植物の物質生産プロセス制御基盤技術開発）（P02001の継続プログラム）	遺伝子発現および代謝産物のプロファイルデータの統合データベース
31	細胞組織工学（ティッシュエンジニアリング）の研究開発	継代培養中に見られる種々の生物学的パラメータ（細胞接着、増殖、分化、細胞寿命のパラメータなど）を用いて整理したデータベース（観察システム）
32	タンパク質－汎用低分子医薬品相互作用の重点的解析による創薬研究のための基盤技術開発 [基盤技術研究促進事業（民間基盤技術研究支援制度）の一部]	相互作用解析用に開発した可視化ツールの相互作用データベースシステムとの連携機能 （汎用医薬品の構造、物性、薬理作用、適応疾患、副作用およびタンパク質の構造・機能の情報に関するデータベース）
33	糖鎖の極微量構造解析技術開発研究 [基盤技術研究促進事業（民間基盤技術研究支援制度）の一部]	構造単位データベース・検索インタフェース
34	自己集合性タンパク質に基づくバイオマテリアル創成基盤整備事業 [基盤技術研究促進事業（民間基盤技術研究支援制度）の一部]	塩濃度応答性自己集合ペプチドを最適化するためのペプチドライブラリーとデータベース
35	二本鎖RNA発現ベクターを用いた変異マウス作製法の開発 [基盤技術研究促進事業（民間基盤技術研究支援制度）の	組織特異的に二本鎖RNA発現ベクター
36	siRNA発現ライブラリーを用いた迅速な標的探索と医薬品	siRNA発現ライブラリー
37	開発成果標準化フォローアップ等標準化調査研究事業 平成19年度成果報告書 開発成果標準化フォローアップ等標準化調査研究事業 遺伝型ー表現型データベースのデータ記述形式標準化事業	XMLによるデータ定義書
38	開発成果標準化フォローアップ等標準化調査研究事業 平成18年度成果報告書 遺伝型-表現型データベースのデータ記述形式標準化事業	XMLによるデータ定義書

4 機関の成果物調査の初期段階（2008 年 9 月時点）では、145 件の候補があった。その後、ライフサイエンスの範囲を見直したり、プロジェクトの進捗で新たにデータベースが開発されたために加えたり、開発者へのインタビューによる統合・削除などを行い、その数は増減し、2009 年 2 月時点で 158 件となった。各機関が産生した

成果物の内訳を図 2.1.1-2 に示す。ただし、2 機関以上が関わった成果物の場合は、JBiC、産総研、NITE および NEDO の順で初めに書いた機関の成果物として分類している。以下に述べるように、これらの成果物リストから、便覧に掲載する対象とアーカイブ対象を決定した。

（2） 便覧に掲載した成果物

上記の調査で得られた成果物のうち、本年度便覧に掲載するものは、JBiC と産総研の成果物で公開されているものとした。件数は、JBiC が担当したプロジェクトの成果物が 20 件、産総研の成果物が 51 件である。本年度は、JBiC と産総研の成果物（データベースと解析ツール）を全て便覧に掲載、公開出来た。これらは、図 2.2.1-2 にあるグレイ部分（JBiC と産総研）を合計した 71 件に該当する。但し NEDO の 78 件と独立行政法人製品評価技術基盤機構（NITE）の 9 件は、本年度では便覧に掲載していない。

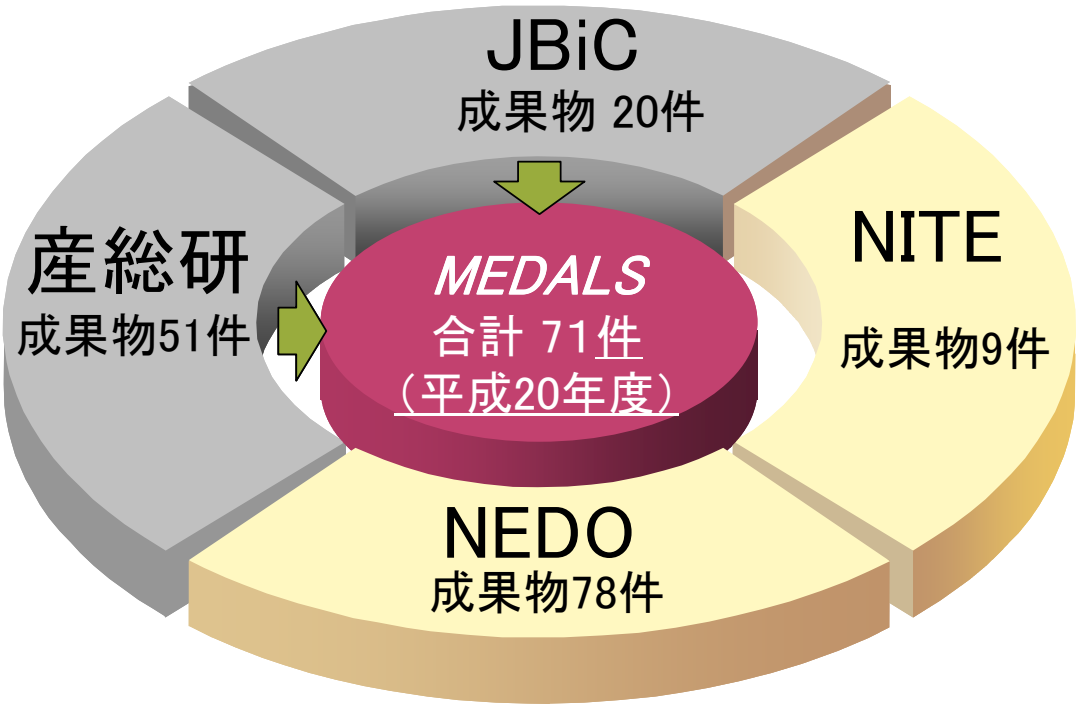


図 2.1.1-2 成果物とそれを産生した機関の内訳

（3） 便覧に掲載しているデータタイプによる内訳

掲載中の成果物のデータタイプによる内訳を表 2.1.1-2 に示す。これは、排他的分類であるので、1 つの成果物が複数のタイプに数えられている。

データベースと解析ツールの割合をデータタイプ別にみてみると、DNA・ゲノムと蛋白質については約 2:1 の割合であった。RNA については約 1:1 であり、比較的解析ツールが多いという傾向であった。その他（多型、発現またはパスウエイなど）については、3:1 であり比較的データベースが多い傾向にあった。これらの割合の違いは、現代の産業や学問のフェーズ、ないしは国や産総研の重点活動を反映しているとみられる。

表 2.1.1-2 成果物のデータタイプ内訳

	データタイプ	件数
データベース 40 件	DNA・ゲノム	17
	RNA	14
	蛋白質	22
	その他	22
解析ツール 31 件	DNA・ゲノム	9
	RNA	13
	蛋白質	12
	その他	7

（4） 便覧に掲載している説明項目

便覧では、成果物とプロジェクトという 2 つの情報を提供している（表 2.1.1-3, 4）。成果に関する項目は計 21 項目、プロジェクトに関する項目は 13 項目ある。項目の選定は、アンケートやインタビューから得られた産業界のニーズに準拠しており、例えばデータ更新頻度や論文や特許の情報である。また、将来的な他省庁との連携成果（項目の共通化など）を鑑みて、成果やプロジェクトの説明項目は、DBCLS や WingPro を参考にした。以下に具体的な項目リストを示す。

表 2.1.1-3 成果物の項目

項番	名前
1	成果物の名前
2	成果物の別名
3	成果物に関する説明（400文字程度）
4	クラス
5	運用機関（メインで運用・提供している機関）
6	機関所在国
7	キーワード
8	サイトURL
9	インタフェース([GUI Webサービスその他])
10	入力例（使い方）
11	データー括取得方法・ダウンロードデータ総量（Mbyte）
12	使っている外部リソース（データベース）
13	カテゴリor分子種
14	生物種
15	利用条件
16	データ更新頻度
17	最新更新記録（調査日）
18	データベース内での主な利用ID
19	利用IDを使った個別データへのアクセス方法（URL）
20	外部リンク
21	論文（学術雑誌、和雑誌）、特許

表 2.1.1-4 プロジェクトの項目

項番	名前
1	プロジェクト名称
2	分野
3	目的
4	プロジェクト紹介
5	キーワード
6	開始年度・終了年度
7	代表者
8	代表者所属組織
9	予算額（￥総額）
10	代表受託機関
11	参加機関
12	論文[PMID] （学術雑誌、和雑誌：投稿中は除く）
13	特許（日本、海外：出願中は除く）

（5） 画面構成

MEDALS は、トップ画面からデータベースやツール便覧、検索、ツールダウンロードなど全てのサービスを利用できる構成である。システムは全部で 21 画面から構成されており、各画面間の関係を図 2.1.1-3 に示す。各サービスは最大で 4 操作で詳細或いは、ダウンロードまで完了できるシンプルなサイト設計をしている。

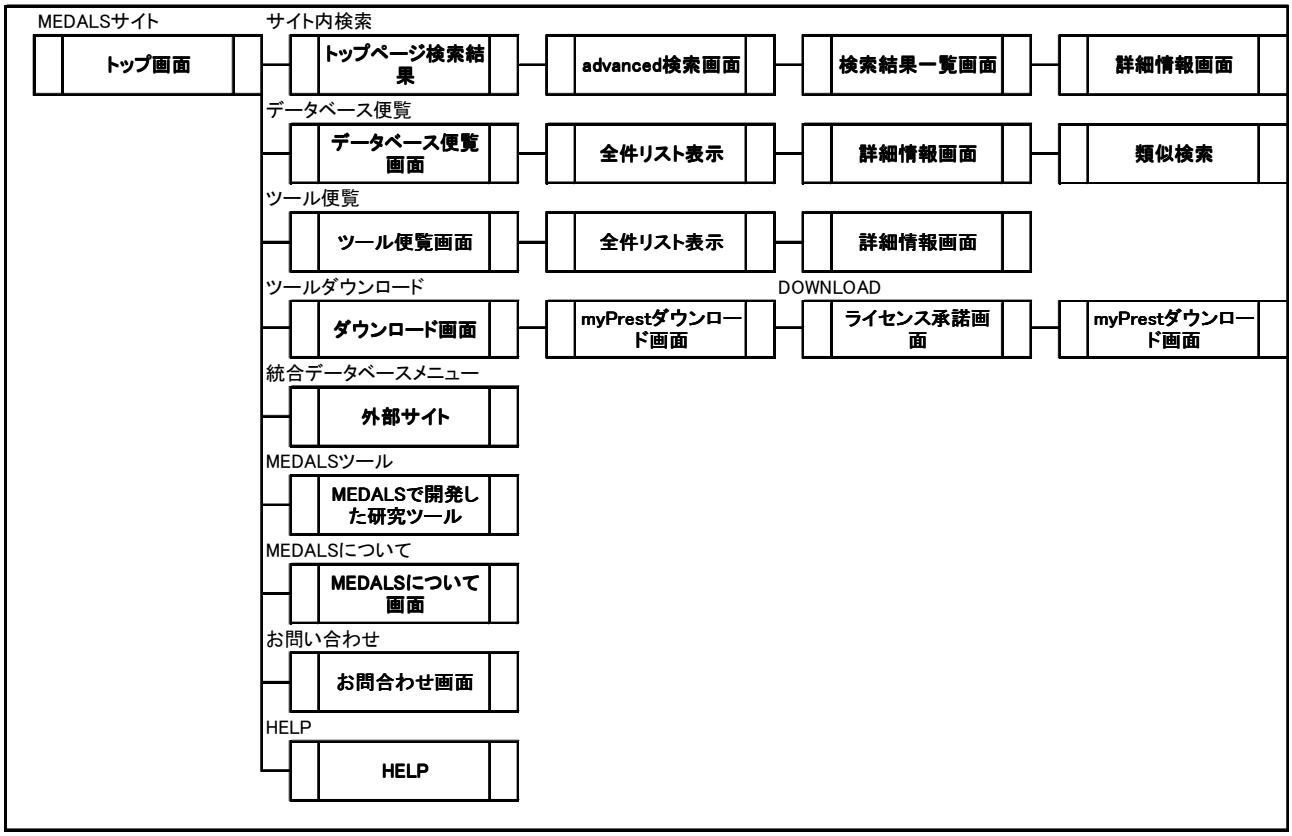


図 2.1.1-3 サイト構成

画面は全部で 21 画面あるが、そのうちよく利用されるトップ画面、全リスト、データタイプ別のリスト、詳細画面、ダウンロード画面の仕様について以下に述べる。

- トップ画面

トップ画面（図 2.1.1-4）は、上段から「ナビゲーションメニュー」、最新情報を表示する「What's New」、MEDALSの「概要」、便覧・ダウンロードなど各種メニュー、最下段にはMEDALSの詳細情報やお問い合わせページなどの付記情報へのリンクを設置している。

図 2.1.1-4 トップ画面

- データベース及びツール便覧画面（全リスト表示）

データベース及びツール便覧画面の全リスト表示（図 2.1.1-5）では、成果物の概要をリスト形式で表示する。表示する項目は、成果物の名称、解説へのリンク、主な対象データ、生物種、運用機関である。また、画面上段には件数（総数と現在の表示件数）を表示している。標準の設定では、1 ページ当りに 10 件をリスト表示するが、次の 10 件を見る場合や全件を表示する場合は画面最下段にあるページ、件数表示切り替えを利用できる。

図 2.1.1-5 便覧画面の全リスト表示

- データベース及びツール便覧画面（データタイプ別表示）

データタイプ別のリスト表示（図 2.1.1-6）では、成果物を利用・配布している公式サイトのイメージ画像や要約、主な対象データ、運用機関を表示する。また、公式サイトや詳細情報へのリンクを設置している。

成果物名	ProSeg
要約	PrpSegは、タンパク質分子の部分構造を網羅的に分類、整理したデータベースです。7万以上のセグメントが主鎖の構造類似性にしたがって数千個程度のクラスタに分類されています。[出典] 産総研プレスリリース(2008.4.5) http://unit.aist.go.jp/brf/brf-breed/ci/outputs/
主な対象データ	タンパク質-モチーフ
運用機関	産業技術総合研究所（AIST） 生命情報工学研究センター（CBRC）
詳細情報	データベースやツールの詳細情報を参照
リンク	公式 サイト

成果物名	SEVENS
要約	GPCR遺伝子の網羅的データベースです。7本膜貫通ヘリックス型タンパク質のGPCR遺伝子を網羅的に収めたデータベースです。32の真核生物のゲノムからバイオインフォマティクス手法で、遺伝子を高精度に同定しています。
主な対象データ	タンパク質-配列 アミノ酸配列
運用機関	産業技術総合研究所（AIST） 生命情報工学研究センター（CBRC）
詳細情報	データベースやツールの詳細情報を参照
リンク	公式 サイト

図 2.1.1-6 便覧画面のデータタイプ別のリスト表示

- 詳細情報画面

個別成果物に関する詳細な情報を表示するのが、「詳細情報画面」である。この画面では、収集した成果物に関する 21 項目と成果に関係するプロジェクトの情報 13 項目を表示する（図 2.1.1-7, 8）。

アンケート調査から分かったニーズの高い項目も記載している（項目の詳細は 2.1.2 節）。

成果物名	SEVENS
成果物の別名	なし
成果物に関する説明	GPCR遺伝子の網羅的データベースです。7本膜貫通ヘリックス型タンパク質のGPCR遺伝子を網羅的に収めたデータベースです。32の真核生物のゲノムからバイオインフォマティクス手法で、遺伝子を高精度に同定しています。
成果物のタイプ	DB
運用機関	産業技術総合研究所（AIST） 生命情報工学研究センター（CBRC）
機関所在国	日本
サイトURL	http://sevens.cbrc.jp/
インターフェイス	GUI
入力例	生物種を選択し、キーワードを検索テキストボックスに入力、[submit]ボタンを押します。Chromosome Number、Exon Number、Protein Length、Pseudogene、Familyを選択可能です。
キーワード	GPCR (G-protein coupled receptor)| Odorant/olfactory and gustatory receptors| Chemokines and chemotactic factors receptors| taste receptors| metabotropic glutamate and calcium receptors| Insect chemosensory receptors| Opsins| Nematode chemoreceptors| Serpentine receptor-like proteins| Sre proteins| Plant MLO receptors| slime mold receptors| Archaeal/bacterial/fungal opsin family| fungal receptors| BLAST|SOSUI|ALN|HMMER
ダウンロードデータ総量（Mbyte）| データ一括取得方法	7,000 | 現在一括ダウンロード不可。
使っている外部リソース	NCBI|KEGG|SWISS-Prot|Pfam
主な対象データ	タンパク質-配列 アミノ酸配列
生物種	全生物種
利用条件	表示-非営利-継承 アカウント不必要
データ更新頻度	1回/1～2年
最終更新日（調査日）	2008/07/18 (2008/12/16)
利用できるID	CBRC-species-2digit numbers-4digit numbers
IDを使った成果物の利用方法	http://sevens.cbrc.jp/detail.php?db=[生物種]&chr=[染色体番号]&l=1&r=1053750&twin=0&id=[ID]
外部リンク	ExPASy Proteomics Server(http://www.expasy.org/), NCBI(http://www.ncbi.nlm.nih.gov/), Pfam(http://pfam.sanger.ac.uk/)
論文等（PubMed ID）	PMID: 16126348, Automatic gene collection system for genome-scale overview of G-protein coupled receptors in eukaryotes. Ono Y, Fujibuchi W, Suwa M. Gene.

図 2.1.1-7 詳細画面（成果物）

プロジェクトの名称	科研費（208059）
分野	ゲノムインフォマティックス
目的	生物学的な知見を得るためGPCR遺伝子を全て保有し、それらの機能情報を網羅的に解析することが必要とされるが、現状の生化学実験ではコストと時間がかかり不可能である。バイオインフォマティクス手法で、実験を支援できるような研究成果を国際的にも、国内でも流通させるため国内でデータベース化することを目的としている。
紹介	本事業では、バイオインフォマティクス手法を応用し真核生物のゲノム配列からGPCR遺伝子を網羅的に同定し、生物種、染色体番号、ゲノム上の座標、タンパク質の機能情報を収納し、視覚的なインターフェースで表現した総合DBを作成している。　実験で発現が確認されずとも、ゲノム上に存在し発現可能性のある遺伝子を全て含むのが大きな特徴で、GPCRの総合的な理解やGPCR関連創薬に大きく貢献できる可能性が高い。　従来から構築しているSEVENSではデータの収集、評価およびDB化のための作成組織と、技術的方式は既に確立しているので今後、機能解析・予測情報をさらに充実させて、アクセス数を伸ばすことにより、世界的なDBとして定着させることを目的とする。
キーワード	GPCR | Genome|Function|annotation|
開始-終了年度	2007 - 2008
代表者	諏訪牧子
代表者所属組織	産業技術総合研究所　生命情報工学研究センター
予算（百万円）	5
代表委託機関	なし
参加機関	なし
論文等（PubMed ID）	PMID: 16126348, Automatic gene collection system for genome-scale overview of G-protein coupled receptors in eukaryotes. Ono Y, Fujibuchi W, Suwa M. Gene. 2005 Dec 30;364:63-73.
特許（日本、海外）	なし

図 2.1.1-8　詳細画面（プロジェクト）

- ダウンロードページ画面

ダウンロード画面（図 2.1.1-9）では、データベースやツールのアーカイブをダウンロードでき、2008 年度には 2 件のアーカイブを提供している。提供情報としては、データ（ソフトウエア）のダウンロードするためのリンク以外に、関連ドキュメント、詳細情報画面へのリンクが設置してある。

図 2.1.1-9　ダウンロード画面

2．1．2　アーカイブ対象データベースの選定

ポストゲノム以降の過去の経済産業省関連のプロジェクトの成果である各種データベースやソフトウエア、あるいは有用な測定データの多くは、プロジェクトが終わっても一般の研究者が簡単には取得できない状態にある。これらのデータを公開し、容易に取得できるようにするためのデータアーカイブが将来的に公開される。本年度はその準備として、過去のプロジェクトによって産生されたデータの調査、選定を行った。調査にあたっては、各プロジェクトの背景、測定データの詳細な内容等を、インタビュー等により調査した。対象となるデータはその利用価値を考慮して選定した。また、現在も継続的に維持管理されているデータベースの旧バージョンのデータも保管対象とした。

（1）アーカイブ対象の選定基準

各機関が産生した成果物のリスト（2.1.1 節）のうち、将来的にアーカイブから公開していく成果物を「アーカイブ対象」として選んだ。まず、アーカイブ対象、すなわちポータルサイトからダウンロードできるようにする成果物の対象としては、「オンラインで容易に取得することができないもの（独自なサイトを持たないために公開していない等）」とした。さらに、対象を絞る基準として、「利用価値の高いもの」という趣旨で以下の条件を決定した。

選定基準：以下の 2 項目の何れかに該当すること

1. 公開による利用を前提として出されたもの
2. 統一条件で収集される網羅的データ

選定するだけの情報が足りない成果物については、プロジェクト関係者や開発担当者へインタビュー等を行った。その際、各プロジェクトの背景、測定データの詳細な内容、世界に存在する類似データとの違い、データ品質の多様度、同時に取得したい関連知識、プロジェクト終了後の世界の動向とそれによるデータ価値の変容等について確認を行った。

この過程で、ライフサイエンスとしてはプロジェクトの趣旨に合わない成果や現在各組織で運営している成果に関してはアーカイブ候補から削除した。また、現在進行しているプロジェクトの成果に関してもアーカイブ候補から削除した。例えば、NEDO のプロジェクトである「開発成果標準化フォローアップ等標準化調査研究事業遺伝型-表現型データベースのデータ記述形式標準化事業」ではデータ記述形式定義書が成果物に該当する。これは標準化団体のサイト（http://www.openpml.org/）より配布等の運営が行われている為、アーカイブ候補から削除した。

これらの活動の結果、選定されたアーカイブ候補は表 2.1.2-1 に示す 63 件（35 プロジェクト）となった。

表 2.1.2-1 選定されたアーカイブリスト

NEDOプロジェクト		
プロジェクト項番	プロジェクト名	成果名
1	標準SNPs解析	解析工程支援システム コール作業支援システム 二次元散布図描画システム データ集計システム 公開用データチェックシステム アレル頻度集計システム 解析事業のアレル頻度データ
2	タンパク質機能解析・活用プロジェクト	ヒトの遺伝子全てに対するノックダウンsiRNA ライブラリーを作製し、ガンやHIV など、様々な疾患に関連する遺伝子、あるいは細胞分化等の様々な生物現象に関わる機能遺伝子を網羅的に同定するシステム BIOPREDsi off-target 排除システム Intris
3	ゲノムインフォマティクス技術開発	発現頻度情報収集用ソフトウェア 転写因子データベース(TFDB) 遺伝子同定支援ワークベンチシステム 遺伝子発現プロファイル計測実験支援システム
4	生体高分子構造情報利用技術開発	低分子処理のための支援ツール GamessViewer
5	生物機能を活用した生産プロセスの基盤技術開発	MGF 研究支援システム 代謝物質標準液の測定データベース 代謝物質標準液のMS/MS スペクトルデータベース 生命分子ネットワークシステム設計支援システムCADLIVE
6	植物の物質生産プロセス制御基盤技術開発	KATANA(ANNOME, Metabolic Pathway Search, Gene Ontology Search, マメ科植物糖の代謝化合物データ DAGViz KaPPa-View(Kazusa Pathway viewer) MapDrawer(ツール)
7	複合生物系等生物資源利用技術	転写因子破壊株から得られた遺伝子発現プロファイルからの遺伝子制御配列解析結果(CD-ROM) 77の遺伝子発現プロファイルデータ 出芽酵母を構成する生体高分子間の相互作用66、668件とタンパク質間相互作用11、102件を納めた再編成データベース
8	多種遺伝子導入技術開発	DNAアレイによる遺伝子発現プロファイルデータを公開するデータベースシステム 遺伝子発現プロファイルと代謝産物プロファイルの統合データベースとしての代謝経路データベース kazusa Pathway
9	医用化合物スクリーニング支援システム	化合物データベース 医用化合物スクリーニング支援システム
10	蛋白質発現・相互作用解析技術開発	組織別遺伝子発現データベース 癌組織遺伝子発現データベース
11	完全長cDNA構造解析	cDNA機能アノテーション統合データベースシステム(cDNA Integrated Database System) 完全長cDNA発現プロファイルデータベース
12	染色体解析技術開発	染色体異常関連データベースの構築 個別化医療実現のために必要な日本人ゲノム多型のデータ収集と情報データベース
14	定位的がん治療装置	治療計画データベース
15	微粒子利用型生体結合物質等創製技術	各種ライブラリデータ
16	グリコクラスター制御生体分子合成技術	バーチャルスクリーニングシステム
17	糖鎖合成関連遺伝子ライブラリーの構築	糖転移酵素データベース
18	エネルギー使用合理化生物触媒等技術開発	セルロース分解酵素、発現遺伝子のデータベース
19	環境中微生物の高精度・高感度モニタリング技術の開発	画像解析モジュールの開発Bacteria Automatic Counting System (BACS)
20	細胞内ネットワークのダイナミズム解析技術開発	リアルタイム・マルチカラー・デジタル録画システム
21	ゲノム情報に基づいた未知微生物遺伝資源ライブラリーの構築	RDP0Ⅱ
22	機能性糖鎖複合材料創製技術開発	糖鎖構造解析並びに機能予測システム
23	高機能簡易型有害性評価手法の開発	遺伝子情報、ラット毒性データ及び公開データを含めた統合データベース
24	化学物質総合リスク評価管理システムの開発	化学物質総合リスク評価管理システム
25	高精度・簡易有害性(ハザード)評価システムの開発	バイオインフォマティクスデータベースシステム
26	植物利用エネルギー使用合理化工業原料生産技術開発(植物の物質生産プロセス制御基盤技術開発)	遺伝子発現および代謝産物のプロファイルデータの統合データベース
27	細胞組織工学(ティッシュエンジニアリング)の研究開発	継代培養中に見られる種々の生物学的パラメータ(細胞接着、増殖、分化、細胞寿命のパラメータなど)を用いて整理したデータベース(観察システム)
28	タンパク質－汎用低分子医薬品相互作用の重点的解析による創薬研究のための基盤技術開発 [基盤技術研究促進事業(民間基盤技術研究支援制度)のー	相互作用解析用に開発した可視化ツールの相互作用データベースシステムとの連携機能 (汎用医薬品の構造、物性、薬理作用、適応疾患、副作用およびタンパク質の構造・機能の情報に関するデータベース)
29	糖鎖の極微量構造解析技術開発研究 [基盤技術研究促進事業(民間基盤技術研究支援制度)のー	構造単位データベース・検索インタフェース
30	自己集合性タンパク質に基づくバイオマテリアル創成基盤整備事業	塩濃度応答性自己集合ペプチドを最適化するためのペプチドライブラリーとデータベース
31	二本鎖RNA発現ベクターを用いた変異マウス作製法の開発 [基盤技術研究促進事業(民間基盤技術研究支援制度)のー	組織特異的に二本鎖RNA発現ベクター
32	siRNA発現ライブラリーを用いた迅速な標的探策と医薬品開	siRNA発現ライブラリー
33	遺伝子多様性モデル解析技術開発	SNP-system 集団遺伝学解析ソフト(AStat, POPGAN, Clark, LDConf, RIGHT) SNP解析による副作用予測システム 遺伝子発現解析による治療に対する感受性や転移の予測システム
経済産業省直轄プロジェクト		
プロジェクト項番	プロジェクト名	成果名
34	バイオインフォマティクス関連データベース整備	H-InvDB リリース1データセット
35	ゲノム情報統合	H-InvDB リリース2～5データセット

(3)アーカイブ前倒し公開

2008 年度の目標ではアーカイブの選定であるが、有用性がありアーカイブ公開の調整が可能になった 2 件については前倒しで公開を行った。今年度公開した 2 件は以下の通りである。

- myPresto　:医薬品開発支援用分子シミュレーション計算ソフトおよびデータベース集合体であり、国等から委託を受けて実施した「生体高分子立体構造情報解析」等の NEDO プロジェクトの中で開発した分子シミュレーションを構成するプログラム群等である。

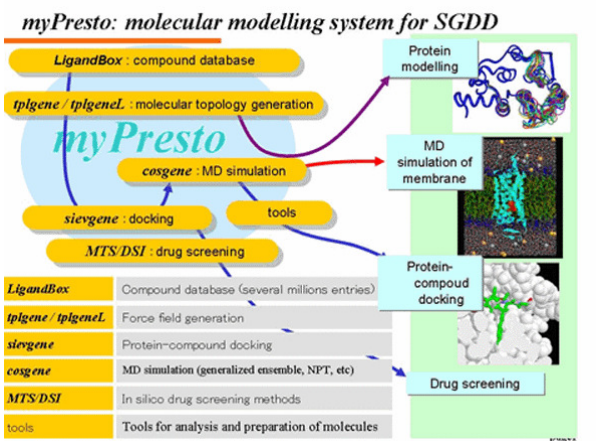


図 2.1.2-1　myPresto

- PubMedScan　:新規関連文献お知らせツールは、複数の文献と関連性の高い文献を従来のサービスより精度、感度が高く、定期的に収集してメールでお知らせが届くシステムである（詳細は、2.1.4 を参照）。

図 2.1.2-2　PubMedScan

2．1．3　サイト内検索サービスの開発

本ポータルサイトでどのようなデータが提供されているかを容易に知ることができるように、本ポータルサイト内のコンテンツを対象としたキーワード検索のサービスを作成し、10 月に公開した。本サービスでは、成果物やプロジェクト、当該サイトで配布可能な文書を検索できる仕様になっている。図 2.1.3-1 に、サイト内検索の概念図を示す。重要なキーワードに関しては、言い換えや別名などを準備するなどして、狭い特定のキーワードだけでなく、広く様々なキーワードでもヒットするよう工夫した。過去の経済産業省関連のプロジェクトや研究内容をまったく知らないユーザーが、利用可能で便利なデータを発掘する機会が増えるように、検索前後における情報の見せ方を工夫した。

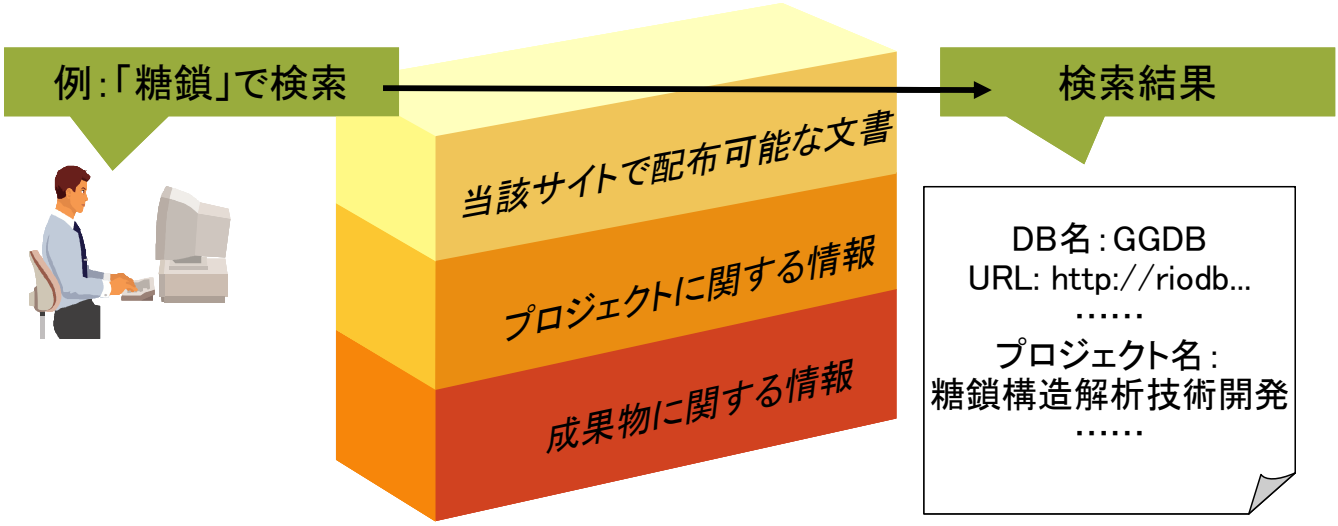


図 2.1.3-1　サイト内検索概念

（1）検索エンジンと検索式

サイト内検索では、内部の便覧及びその他の付記情報に対する全文検索を行い、たくさんの情報の中から、ユーザーの希望する特定語句を含む情報を探し出すことができる。検索エンジンとして HyperEstraier を採用した。この検索エンジンには、将来的拡張に対応しやすい次の特徴がある。

- インデックスを使った高速な検索ができる
- 大量の文書のインデックスを短時間で作成できる
- N-gram 方式による漏れのない検索ができる
- 形態素解析と N-gram のハイブリッド機構で検索精度を向上できる
- フレーズ検索や正規表現検索や属性検索や類似検索をサポートしている
- 世界各国の言語が扱える
- 対象文書の所在や形式に依存しない
- 高精度なウェブクローラが付属している
- ライブラリとして各種製品に組み込める
- P2P 連携機能をサポートしている

また、本エンジンは文部科学省ライフサイエンス統合データベースにおいても採用されており、今後の連携も容易に対応できると思われる。

表 2.1.3-1 検索式比較

検索	AND	OR	NOT	前方一致	後方一致
MEDALS	空白 &	|	!	文字＊	＊文字
Google	空白	|	-	文字＊	＊文字

利用できる検索式の内、利用される可能性の高い幾つかを表 2.1.3-1 に示す。広く利用されている検索エンジン Google との対比から分かるように、NOT 検索の記述方式が異なるが、大部分は同じ検索式を利用できる為、過去の経済産業省関連のプロジェクトや研究内容をまったく知らないユーザーでも利用可能である。

（2）様々なキーワードでもヒットする工夫

MEDALS では、便覧のキーワード項目を充実させることで、重要なキーワードに関しては、言い換えや別名など狭い特定のキーワードだけでなく、広く様々なキーワードでもヒットするよう工夫した。例えば、GPCR 網羅的データベース SEVENS では、18 個のキーワードを準備し、様々なキーワードでもヒットするようになっている為、G-protein と入力しても SEVENS がヒットする。

キーワード

1. GPCR (G-protein coupled receptor)
2. Odorant/olfactory and gustatory receptors
3. Chemokines and chemotactic factors receptors
4. taste receptors
5. metabotropic glutamate and calcium receptors
6. Insect chemosensory receptors
7. Opsins
8. Nematode chemoreceptors
9. Serpentine receptor-like proteins
10. Sre proteins
11. Plant MLO receptors
12. slime mold receptors
13. Archaeal/bacterial/fungal opsin family
14. fungal receptors

15. BLAST
16. SOSUI
17. ALN
18. HMMER

（3）サイト内検索の利用方法

サイト内検索を行うには、画面上段に設置している①簡易検索か、②検索ページ（トップ画面に設定してあるサイト内検索メニューからリンク先へ移動）を利用できる（図 2.1.3-2）。

MEDALS のサイト内検索を利用した方法を、例（GPCR）を使って説明する（図 2.1.3-3）。

（ア） 簡易検索ボックスに「GPCR」を入力し、検索ボタンをクリックする。

（イ） 「GPCR」が含まれるページの一覧を検索結果リストに表示される。

（ウ） 検索リストでは、検索したキーワードを含むページへのリンクと、ページのタイトル、検索した文字がヒットした前後の文章、検索したキーワードを含むページと類似性のあるページの抽出ボタンが表示される。（例では、「MEDALS　詳細情報　DB　SEVENS」がトップヒットしている。）

言語圏の文字コードを配慮して、英語版の検索サービス（簡易検索、検索ページ）では、日本語のページは検索対象外にしている。また、検索リストから詳細ページへ遷移した場合も、全て英語版を表示する。検索速度は高速であり、100 件の全文検索では 0.003 秒である。この速度はユーザーが実際に利用するに当っては、現実的に利用でき、且つフラストレーションを感じない速度である。

図 2.1.3-2 トップ画面でのサイト内検索機能の画面

図 2.1.3-3　サイト内検索例（GPCR を用いて）

2．1．4　研究ツールの開発

（1）背景

本事業では、製薬企業へのインタビューやインターネットによるアンケート調査を行った（詳細は 2.3.2 節を参照）。この調査から分かった「時間を最も短縮したい業務」は以下の通りであった。

1 位　　DB からのデータの収集

2 位　　実験

3 位　　論文の検索

そこで、「論文検索と情報収集の時間を短縮したい」や「キーワードでうまく絞れない場合に時間がかかる」などニーズの高い論文の検索に対して、バイオメディシナル情報研究センター（BIRC）で開発した文献収集ツールの提供を検討した。

一般的に文献を検索・収集する方法として「キーワードから探す」と、「文書から探す」の 2 種類がある。「キーワードから探す」方法では、分野や著者、雑誌名をキーワードに使い、逐次探す場合は、通常の PubMed や図書館、Google 等から探し出す。また、定期的に探し出す場合は、一般にサービス提供されている「myNCBI」が有名である。また、それらの 2 次利用したサービスもある。「文書から探す」方法では、幾つかの文献を持っていればキーワードを検討する必要は無い。逐次探す場合は、入力に 1 文献ずつ使って、Jane や eTBLAST で検索する。また、定期的に探し出す場合は、入力に複数文献を利用できる PURE がある。理想的には、ユーザーが興味を持っている複数の文献と関連性の高い文献を、キーワードで悩むことなく、定期的に収集するサービスが望ましい。しかし、PURE では文献に対する感度や精度が共に低いという問題があった。

そこで、バイオメディシナル情報研究センター（BIRC）では、文献収集ツールについて、文書間の類似度の指標を変えて改良した PubMedScan を開発していた。このシステムは、PubMed の提供する類似度を利用しており、複数の文献と関連性の高い文献を従来のサービスより精度、感度が高く、定期的に収集してメールでお知らせが届くシステムである。

本事業の調査から文献の検索はニーズが高いことが分かったため、チーム内で利用していた PubMedScan を配布可能なように修正し、機能追加を行い、MEDALS からユーザーに配布することにした。

（2）プログラムの稼動環境

MEDALS の「新規関連文献お知らせツール」メニューからプログラムをダウンロードできる。ダウンロードしたファイルをインストールする時に必要な稼働環境を以下に示す。現在、動作確認できている OS （Operation System）は以下の3つである。

- Cent OS4.6
- Redhat RL 3
- Mac OS 10.4

また、プログラムを起動するに当り、以下の環境が必要となる。

- ウェブサーバー： Apache
- データベースマネージメントシステム： MySQL
- 言語環境： Perl、PHP

- その他モジュール/ライブラリ：DBI&DBD::mysql、XML::DOM、LWP::Simple

（3）使い方

インストールされた PubMedScan の使い方を記載する。PubMedScan のサービスを利用する為にはユーザーの登録（E-mail アドレスとパスワード）が必要である（図 2.1.4-1）。また、メールによるお知らせのタイトルや送信数も設定できる。登録後には、設定したユーザーID とパスワードでログインし、サービスを利用できる。

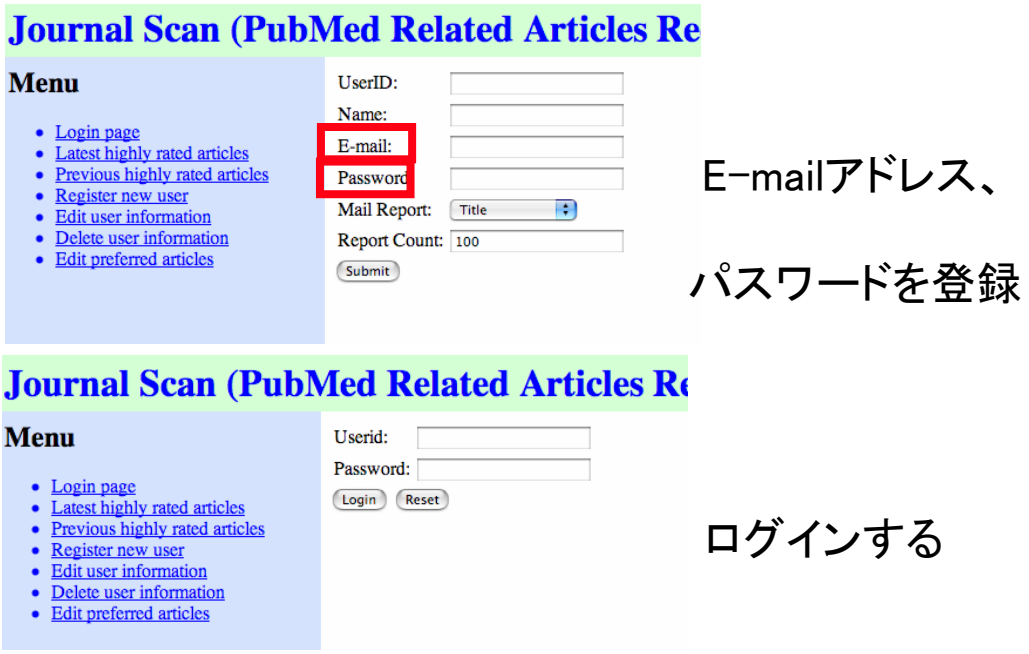


図 2.1.4-1　PubMedScan 利用方法（ユーザー登録～ログイン）

ログイン後に、興味のある文献の PubMed ID を入力し、入力した ID の文献タイトルを一覧で確認することができる（図 2.1.4-2）。このシステムでは、複数の文献を登録することができ、実際の文献を持っている必要も無く、ID だけをユーザーが持っているだけでよい。

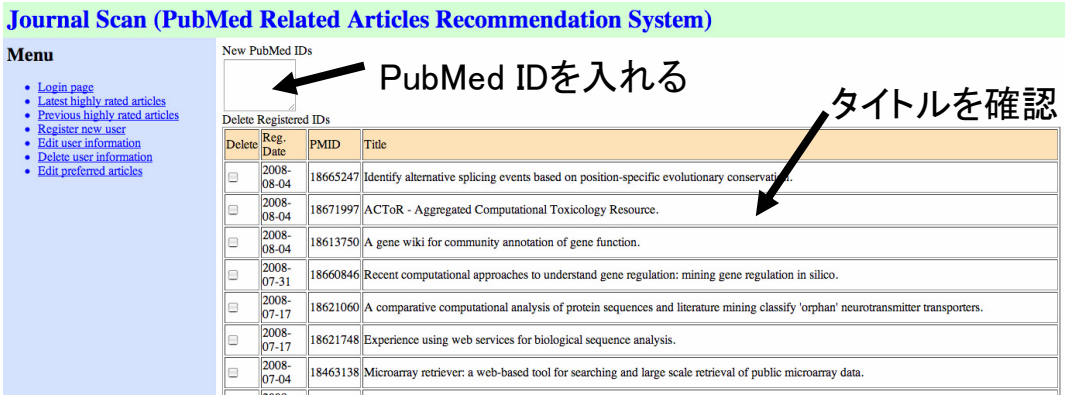


図 2.1.4-2　PubMedScan 利用方法（PubMed ID 登録）

興味のある PubMed ID をシステムに入力後、新規関連文献のお知らせがメールで届く。ユーザーは、届いたメール或いはシステムにログインして、抽出された新規の関連文献の PubMed ID や関連の強さを示す Score、文献のタイトルを確認できる（図 2.1.4-3）。PubMed ID には PubMed のアブストラクトへのリンクが付いているため、興味ある文献はシステムを通じて参照できる。

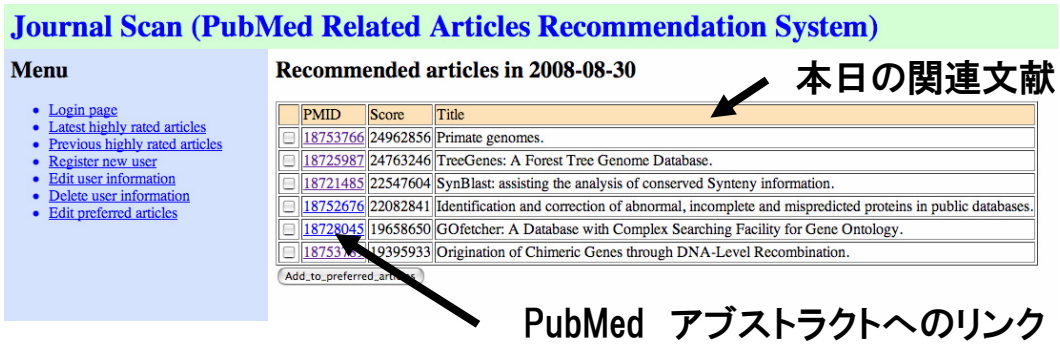


図 2.1.4-3 PubMedScan 利用方法（新規関連文献リスト結果）

（4）ユーザーの関心

本システムに対するユーザーの関心の高さは、MEDALS へのアクセスログを解析（調査期間 2008 年 10 月 28 日～2009 年 2 月 16 日[4 ヶ月]）して確認することができた。表 2.1.4-1 はポータルサイトでよくみられるページのうち上位 10 個を示した物である。PubMedScan のページは MEDALS 全ページの中でも 5 位であるが、ツール単独の紹介ページでは人気が 1 位であった。この結果から、本システムに対する関心が高いことが確認できた。また、情報提供ツールにおいて、事前のユーザーに対するアンケートやインタビューのニーズ分析が重要かと思われる。

また、広報活動の一環として実施した 2008 年分子生物学会のポスター発表では、発表時間（2 時間）に 30 人以上の聴講者が訪問し、研究者の関心が高いシステムであると思われた。

表 2.1.4-1 ポータルサイトでよくみられるページ

	ページ	ページビュー	割合（%）
1	トップ	3,808	36.6
2	データベース便覧（全リスト）	653	6.2
3	ツール便覧（全リスト）	417	4.0
4	データベース便覧（DNA・ゲノムリスト）	389	3.7
5	PubmedScan詳細	335	3.2
6	全ダウンロードリスト	327	3.1
7	ダウンロード	272	2.6
8	お問い合わせ	237	2.2
9	MEDALSツールリスト	237	2.2
10	データベース便覧（タンパク質リスト）	184	1.7

（5）システムの簡易評価

システムの性能（関連する文献の網羅性）の簡易評価を実施した。今回の性能の評価値には類似スコア（NCBI 版）を用いた。10 種の研究トピックス（各データセット内の論文数は 10～52）の文献でテストを行った（表 2.1.4-2）。単一文献の登録で残りの文献を抽出する場合(single query)と、複数文献の登録で残りの文献を抽出する場合(multiple query)のカバー率をテストした。その結果、複数文献で探す場合は、単一文献で探す場合に比べ、検索漏れが大きく減ることがわかった（図 2.1.4-4）。また、多くのトピックスでは 90%以上が網羅できた。ただし、トピックスによっては（set-7 参照）網羅性が低くなる場合もあり、今後の検討課題である。現段階では、知らなかった文献を知ることに関しては高いカバー率であるが、これだけでは網羅的とは言えない。その為、当面は他の手法との併用を行い、更にユーザーの検索要求を明確にする必要があると考えられる。

表 2.1.4-2 データセットのトピックス

Set ID	Topics	No. of prepared papers
set-1	Processed pseudogene	43
set-2	Gene annotation	52
set-3	Sensory rhodopsin transduction	10
set-4	Chimera fusion transcript	10
Set-5	Anthocyanidin	12
Set-6	HIV protease resistance	17
Set-7	Prostate cancer	13
Set-8	Translation initiation	15
Set-9	Gene finding	24
Set-10	Promoter motif	32

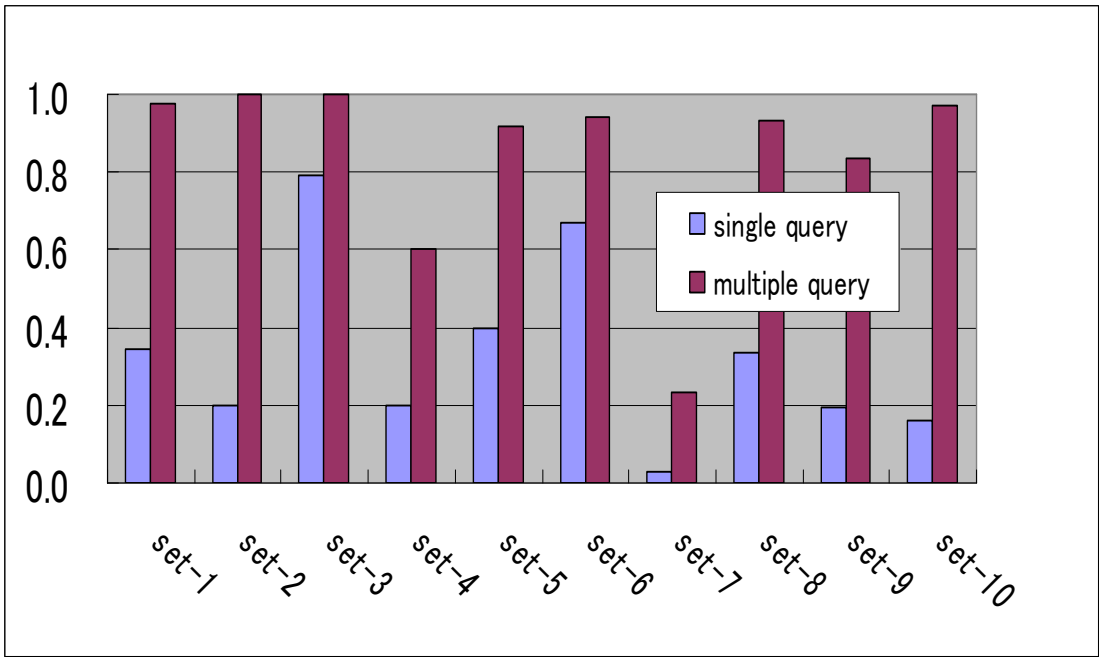


図 2.1.4-4　PubMedScan 性能簡易評価

（6）今後の課題

本システムは UNIX へのインストールが必要だが、一般のユーザーにとってその作業はスキルがないために出来ない、あるいは手間がかかるという課題がある。その為、MEDALS サイト上で IE 等のブラウザで利用できるサービスを開始する予定としている。また、PubMed ID のリストを準備する事が大変であるとの意見も出ている。これを受けて、PDF ファイルから ID リストを作成する別ソフトを提供する予定である（共に 2009 年 3 月予定）。

2．1．5　ポータルサイト運用実績

ポータルサイト運用にあたり、サイト更新、ログ解析、セキュリティ対策を行った。

（1）サイト更新

2008 年 10 月からポータルサイトを一部の情報を先行公開した。2009 年 3 月までの運用実績を以下にまとめる。

- MEDALS を公開[08.10.29]
- 新規関連文献お知らせツール(PubMedScan)を 0.81 へバージョンアップ[08.11.25]
- メンテナンス作業の為、サービスを一時停止。期間[2008 年 11 月 28 日（金）15 時 － 12 月 1 日（月）9 時][08.11.26]
- MEDALS データ更新で 8 件を新たに追加。[09.01.27]
  1. POODLE, 2. WoLF PSORT, 3. CellMontage, 4. GRIFFIN, 5. SAMURAI, 6. GENIUS II, 7. SEVENS, 8. LAST
- MEDALS データ更新で 20 件を新たに追加。また、英語版を公開。[09.02.06]

（2）ログ解析

MEDALS の公開により、当該プロジェクトの目標が達成されたか否かを評価する為には、MEDALS の利用状況を正確に把握することが肝要である。また一層の活用促進を図るために、利用ユーザーの視点からデータベースの使い勝手（ユーザービリティ）を改善する必要がある。特に、操作履歴等からユーザーが何を参照したのかを把握することは、主な利用ニーズと今後提供すべきコンテンツのニーズを把握する上で重要である。そこで、10 月の先行公開からのアクセスログについて解析を実施した。調査期間は、2008 年 10 月 28 日～2009 年 2 月 16 日である。アクセスの解析では、主に GoogleAnalytics などの一般的に解析に利用されるツールを利用した。その解析の結果について説明する。尚、本アクセスログ解析では、アクセス数は、ユーザー数と読み替え、また参照されたページの数をページビュー数と表記する。

- 公開から 4 ヶ月間のユーザー数

公開後の総ユーザー数は 2,190、週平均で 130 ユーザーが毎日アクセスしており、更新が行われる度にアクセス数が一時的に上昇した。特に 2 月 6 日の更新にともなった大規模な広報活動により急激なユーザー数の増加があった事実から、ユーザー数の獲得には大規模な広報活動が効果的であると推測する（図 2.1.5-1）。また、ユーザー数のうち一度だけ訪れたユーザー（ユニークユーザー）数は 1,523 で、ユーザーの大多数が一度の利用だけであることがわかった。更に、便覧に掲載されている成果物を見て、実際にその成果物があるサイトへ便覧を通じて訪問したユーザー数（誘導成功数）は 348 ユーザー（15.8%）で、15 人に 1 人の比率で実際の成果を参照している。このデータから、便覧が成果物の活用促進の足がかりになっていることが伺える。

表 2.1.5-1　サイトのユーザー数統計

項番	項目名	数値	備考
1	ユーザー数	2,190	ユニークユーザー数　1,523
2	誘導成功数	348（単位：ユーザー）	比率　15.8 %
3	ページビュー数	10,392	平均　4,75
4	平均サイト滞在時間	3 分 22 秒	

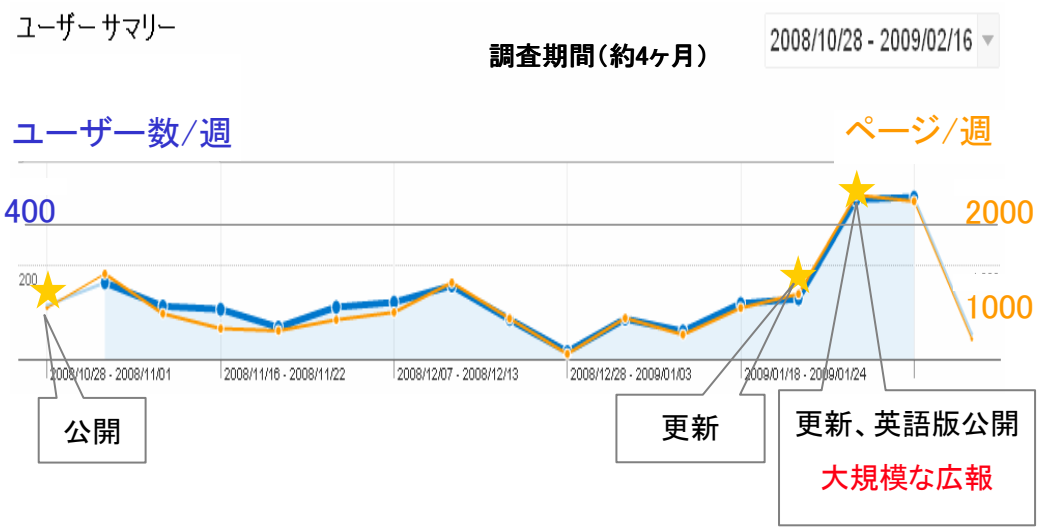


図 2.1.5-1　サイトの更新とアクセスの推移

- 外部からのリンクによるサイト訪問トップ 10

MEDALS では、プロジェクト目的のひとつである「プロジェクト成果の有効活用の促進」を達成する為に、サイトへの入り口である外部からのリンクを積極的に増やしている。その活動により、外部からのリンクを通じたユーザー数が増えつつある。表 2.1.5-2 に、外部からのリンクによるサイト訪問のベスト 10 をまとめた。関係機関である産業技術総合研究所（産総研）やバイオ産業情報化コンソーシアム（JBiC）だけではなく、連携機関である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のポータルサイトからのリンクを通じて訪問されている。

表 2.1.5-2 外部からのリンクによる訪問数トップ 10

	ページ
1	産業 技術 総合技術研究所
2	JBiC
3	千葉大亥鼻イノベーションプラザ
4	文科省統合データベースプロジェクト
5	Google
6	H-InvDB　オフィシャルウェブサイト
7	産業 技術 総合研究所ユニット紹介
8	H-InvDB　ニュースページ
9	ライフサイエンス統合データベースセンター
10	文部科学省

- 一般検索エンジンからのサイト訪問

また、Google による一般的な検索エンジンからもサイトへ訪問するユーザーが多く、全く MEDALS を知らない人が興味あるキーワードから MEDALS の成果を知る流れが構築できつつあることが伺える。例えば、Google からキーワード検索で辿ってくるユーザーの例では、「GPCR　真核生物」で検索した場合がある。このとき、CBRC の成果物である GPCR 網羅的データベースの SEVENS に関するポスター発表ページの次に当該サイトがヒットする（2009 年 3 月時点）（図 2.1.5-2）。このように、成果物のサイトへ直接辿りつけないユーザーを成果物へと誘導できている例をいくつか確認できた。

図 2.1.5-2　Google で検索した結果（例）

- サイト内でよく見られるページトップ 10

広報活動や外部からのリンクなどを通じて訪問したユーザーがよく見るページのトップ 10 をまとめた（表 2.1.5-3）。主に便覧でデータベースやツールの一覧を参照していることが分かる。この結果から、ユーザーが主に見るページに成果の活用へと導く仕組みをつけることが重要であると分かった。また、2.1.4 節で述べた様に、新規関連文献お知らせツールは、単独詳細ページとしてはトップであり、ユーザーの関連文献の調査の手間を省きたいというニーズの強さが把握できた。

表 2.1.5-3 よく見られるページトップ 10

	ページ	ページビュー	割合（%）
1	トップ	3,808	36.6
2	データベース便覧（全リスト）	653	6.2
3	ツール便覧（全リスト）	417	4.0
4	データベース便覧（DNA・ゲノムリスト）	389	3.7
5	PubmedScan詳細	335	3.2
6	全ダウンロードリスト	327	3.1
7	ダウンロード	272	2.6
8	お問い合わせ	237	2.2
9	MEDALSツールリスト	237	2.2
10	データベース便覧（タンパク質リスト）	184	1.7

- サイト内検索でよく使われるキーワードトップ 10

訪問したユーザーがサイト内検索でよく使うキーワードのトップ 10 をまとめた（表 2.1.5-4）。ユーザーは、分子種、データベース、あるいはツール名など多様なキーワードで検索していることが分かる。また、検索では AND 検索や OR 検索などの複雑な検索はほとんど利用されていない。この結果から、ユーザーが複合的な問い合わせを行わなくてもヒットするようにキーワードを充実させることはユーザービリティーの向上に寄与すると考える。

表 2.1.5-4 サイト内検索のキーワードトップ 10

順位	キーワード
1	RNA
2	ゲノム
3	H-ANGEL
4	griffin
5	異性
6	TM
7	rna
8	riodb
9	GMDB
10	GGDB

（3）セキュリティ対策

本プロジェクトで開発した MEDALS 及び 4 種のアプリケーション（H-InvDB のウェブサービス（Web サービス）、fRNA のウェブサービス、リンク自動管理システム、H-DBAS）は公共性が強く、インターネットを通じたユーザーに信頼性・可用性を担保する必要がある。しかし、現在のインターネットを取り巻く環境は非常に厳しく、脆弱性を悪用したサイトへの攻撃やウイルス感染によりシステムのサービス停止や情報の漏洩などの被害が発生することがある。その為、インターネットセキュリティの検査を行うことは、公共的に情報を提供するサイトには必要不可欠である。また、検査スキルに長けた第三者機関に検査を担当してもらう事で、結果に対する信憑性を確保し、検査漏れを防止し、一般性のあるセキュリティレベルを保持できる。

そこで、MEDALS 公開サーバーで動作しているウェブアプリケーション（Web アプリケーション）をはじめ、4 種のアプリケーションに対するセキュリティの検査を行い、セキュリティレベルを評価した（期間：2009 年 1 月 20 日～2009 年 1 月 27 日　平日 10 時～18 時）。

- 診断内容

実施したセキュリティ診断項目を表 2.1.5-5 に示す。

表 2.1.5-5 実施したセキュリティ診断項目

項番	名称
1	セッション管理
2	クッキー使用法対策
3	クロスサイトスクリプティング対策
4	バッファオーバーフロー対策
5	コマンドインジェクション対策
6	SQL インジェクション対策
7	ファイル改ざん、消去対策
8	エラー画面
9	強制ブラウジング
10	ファイルの誤った公開
11	パス名パラメータの未チェック
12	HTTP レスポンス分割
13	ディレクトリトラバーサル
14	セキュリティ上不適切な画面設計

- 総合評価

診断の結果総合評価は「C」であった（評価基準 A（最高）～E（最低）の 5 段階評価）。個別の脆弱箇所については、対策が必要な脆弱（2 件）は診断期間中に報告を受け、2009 年 1 月末までに修正を完了した。その他の指摘された箇所は、2009 年 3 月末までに修正を完了する。

2. 2　経済産業省関連機関から産生されるデータベース等の統合

2. 2. 1　データベース間を自動的に管理可能な相互リンクシステムの整備

ユーザーが効率よくデータを収集するためには、異なる機関で独立に運営されているデータベース間での密な連携が不可欠であり、そのためにはデータの対応関係を調べて相互リンクを設置することが効果的である。しかし、多数のデータベース間での相互リンクを維持管理するには、データベースの組み合わせが非常に多くなるために、運用コストが嵩む。例えば 6 個のデータベース間で相互リンクを設置する場合、30 通りのリンクが必要である（図 2.2.1-1）。リンクを正しく設定するためには、それぞれのデータベースが更新されるたびにデータの対応関係を調べてリンクの再設定作業が生じる。

リンク自動管理システムでは、主要なライフサイエンス分野のデータベースからデータ ID の一覧表を毎日ダウンロードし、それらの ID の対応表を自動作成しており、その対応表を元に、データ収集先間の全てのデータ ID の変換を実現している。（図 2.2.1-1）。

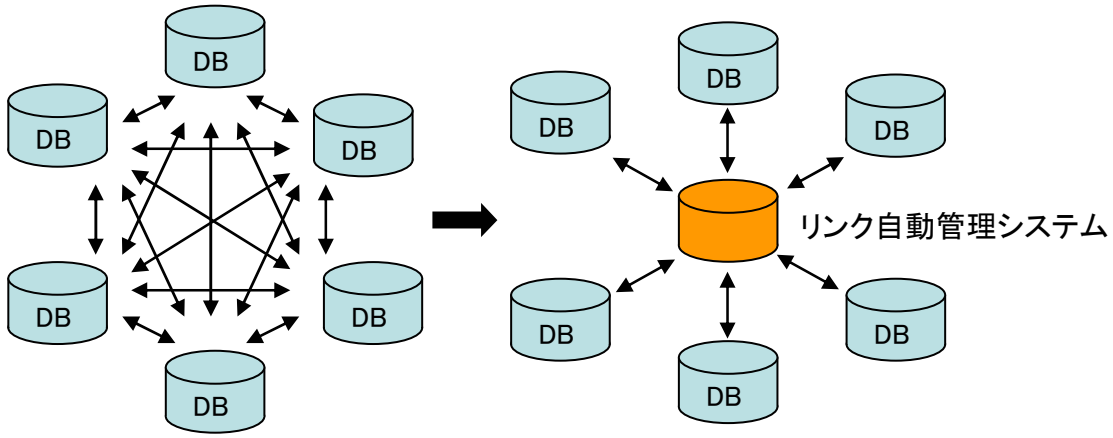


図 2.2.1-1　リンク自動管理システムによるデータ ID の管理

本システムはウェブサイト（Web サイト）（http://biodb.jp）で公開しており、データ ID 変換の仕組みを利用して自動維持されるリンク設定サービスを CGI で提供している。この CGI を使用してリンクを設定するためには、下記のように URL を設定する。

```
http://biodb.jp/hfs.cgi?id=[ID]&type=[ID Type]&db=[Database name]
```

[ID]はリンク元で使用しているデータ ID、[ID Type]はその ID の種類、[Database name]はリンク自動管理システムで登録しているリンク先のデータベース名を指定する。例えば下記のような URL をリンクに設定した場合、リンク自動管理システムが Accession Number BC053657 に対応する H-InvDB の Transcript view のデータ画面に転送する。

```
http://biodb.jp/hfs.cgi?id=BC053657&type=ACC_ID&db=TRANSCRIPTVIEW
```

また、リンクを自動で維持する仕組みを利用し、ウェブ(Web)上でデータ ID をリンクに変換するサービスやデータ ID を別のデータ ID に一括変換する ID 一括変換システムを公開している。

本プロジェクトでは、経済産業省関連のライフサイエンス分野のプロジェクトで産出されたデータベースをリンク自動管理システムで利用できるようにデータ ID の取得先の追加作業、ID 一括変換システム機能のウェブサービス(Web サービス)開発作業を行い、これらの機能を 2009 年 1 月に公開した。

● データ ID の取得先の追加

データ ID の取得先の追加作業では、機能性 RNA データベース(fRNAdb)、糖鎖関連遺伝子データベース(GGDB)から最新のデータID 取得し、既存のデータID との対応表を自動で作成できるモジュールをシステムに組み込んだ。さらに、経済産業省関連のプロジェクトで産出された Human Gene and Protein Database (HGPD)、ユーザーの利便性を向上させるため Ensembl からもデータID を取得しデータ対応表を自動で作成できるように改良を加えた。そして、それらのデータベースとの自動リンク管理システム経由での相互リンクを設定できるようにシステムの拡張を行った。これらの作業によりデータ ID の種類は 5 種類(fRNAdb ID、Ensembl Transcript ID、Ensembl Gene ID、cDNA cloneID、FLJ cDNA clones)追加され 12 から 17 種類となった(表 2.2.1-1)。

表 2.2.1-1 使用できるデータ ID の種類

データIDの種類
Accession Number
H-Inv transcript ID (HIT)
H-Inv cluster ID (HIX)
H-Inv protein ID (HIP)
UniProt ID
HUGO gene symbol
RefSeq
OMIM ID
Entrez Gene
dbSNP rs# (GemDBJ)
H-GOLD Marker ID
PDB ID
fRNAdb ID
Ensembl Transcript ID
Ensembl Gene ID
cDNA clone ID (HGPD)
FLJ cDNA clones (HGPD)

また、リンクを取得できるデータベースは 5 種類（GGDB、fRNAdb、Ensembl Transcript、Ensembl Gene、HGPD）追加され、12 から 17 種類となった（表 2.2.1-2）。

表 2.2.1-2 リンクを取得できるデータベース

データベース名
H-InvDB Transcript view
H-InvDB Locus view
H-InvDB G-integra
H-InvDB PPI view
NCBI Sequence Viewer
Entrez Gene
GeMDBJ
HUGO HGNC
MutationView
H-GOLD
PDBj
UniProt
GGDB
fRNAdb
Ensembl Transcript
Ensembl Gene
HGPD

● ID 一括変換機能のウェブサービス開発

リンク自動管理システム、ID 一括変換システムをブラウザ以外のアプリケーションで利用できるようにするために、ウェブサービスを開発した。出力のフォーマットはウェブサービスで一般的に使われる json 形式、テキスト形式を採用した。ウェブサービスは下記の形式で公開している。ウェブサービスで使われる文字列の解説（表 2.2.1-3）はシステムのヘルプ画面（後述）にも記載した。

http://biodb.jp/convert/[id_type]/[id_type or db_type]/[entry_id,entry_id2,...][format]/[output][count][format]/[offset,limit][format]

表 2.2.1-3 ウェブサービスで使用できる要素

要素	指定する文字列	概要
id_type	ex(acc_id)	データIDの種類指定する
db_type	ex(transcriptview)	データベースの種類を指定する
entry_id	ex(AB210043)	id_typeに該当するデータIDを指定する（100 IDまで指定可能）
output	id	ID対応表を返す
	url	リンクの対応表を返す
format	.txt	結果をテキスト形式で返す
	.json	結果をjson形式で返す
count	count	結果の件数を返す
offset,limit	0以上の整数 例(5,10)	結果の件数を指定する。例では5件目から14件目を取得できる(limitの上限は100)
method	id_type	登録しているデータIDのリストを返す
	db_type	登録しているデータベースのリストを返す
	update	登録しているデータベースの更新情報を返す
	url	各データIDからデータベースへのリンク情報を返す

■ウェブサービスの使用例（出力がテキストフォーマットの場合）

・クエリー

http://biodb.jp/convert/acc_id/locusview/AB210043.txt

・レスポンス

##query_id,result_id,state,url

AB210043,HIX0026954,1,http://www.h-invitational.jp/hinv/spsoup/locus_view?hix_id=HIX0026954

query_id はクエリーのデータ ID、result_id は取得できたデータ ID、url はクエリーのデータ ID の指定したデータベースへのリンク先を返す。state は 1 か 0 を返す。この値は本システムのデータ更新時にリンク先のサーバーにアクセスできたかどうかの情報である。それぞれアクセス可、不可を意味する。

■ウェブサービスの使用例（出力が json フォーマットの場合）

・クエリー

http://biodb.jp/convert/acc_id/locusview/AB210043.json

・レスポンス

```
{
   hfs:[
   "query":[
            "id_type":"acc_id",
            "id":"AB210043"
```

```
        },
    "result":{
        "id_type":"hix_id",
        "db_type":"locusview",
        "url":["http://www.h-invitational.jp/hinv/spsoup/locus_view?hix_id=HIX0026954"],
        "id":["HIX0026954"],
        "state":"1"
        }
    }
}
```


"id_type"はデータ ID の種類、"id"はデータ ID、"db_type"はデータベースの種類、"url"はクエリーのデータ ID に対応する"db_type"の画面へのリンクである。"result"の"id"、"url"は配列を返す。

このサービスを利用することにより、データベース開発者が自身のデータベースに格納していないデータを画面上に表示させることが可能になる。例えばユーザーがある画面にアクセスした場合、システムがウェブサービスを利用して下記のようなクエリーを要求する。このクエリーは H-Inv cluster ID HIX0001706 を使って HUGO Gene Nomenclature Committee（HGNC）への URL を取得したいことを意味する。

```
http://biodb.jp/convert/hix_id/hugo/HIX0001706/url
```

このリクエストを受け取ったウェブサービス提供側は下記のようなレスポンスを返す。これは、HIX0001706 に対応する HGNC でのデータ画面の URL を意味する。

```
http://www.genenames.org/data/hgnc_data.php?match=KCNK1
```

この結果を受け取ったシステムはこの URL を使って任意の画面にリンクを設定できる。よって画面上には常に最新のリンクが表示されていることになる。

データ ID 取得先の追加、ウェブサービスを開発し 2009 年 1 月 6 日にリンク自動管理システムの最新版を公開した。最新版では search ボタンを押すだけでユーザーがシステムの機能を理解できるように、予め画面上にサンプルデータを入力した（図 2.2.1-2）。

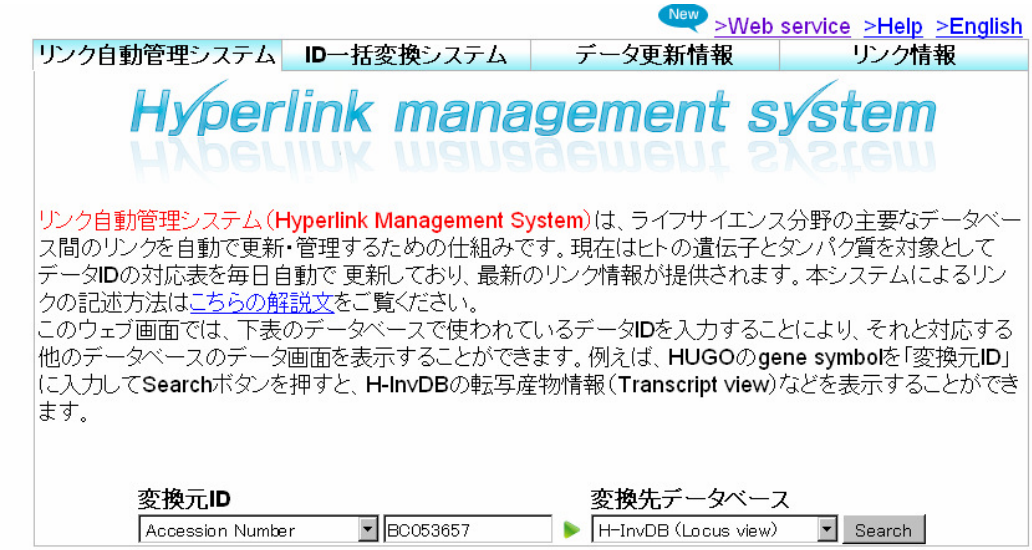


図 2.2.1-2 リンク自動管理システムのトップページ

また、ウェブサービスをより多くのユーザーに使って貰うために perl、java、ruby、python のサンプルコードをヘルプ画面(http://biodb.jp/help/ws_jp.html)から提供した(図 2.2.1-3)。

図 2.2.1-3 ウェブサービスのヘルプ画面の抜粋

2. 2. 2　ウェブサービスの開発と画面の改良によるデータ統合

ヒト全遺伝子のアノテーション統合データベース(H-Invitational DB: H-InvDB)を基礎として、糖鎖や機能性RNA 等の経済産業省関連機関から産生されるデータベース等を統合するために研究開発を行った。具体的には、ウェブサービス（Web サービス）を使ったバーチャルなデータベース連携のしくみを整備するため、連携の基盤となる H-InvDB のウェブサービスの改良・機能拡張を行うとともに、連携させるべき 2 種類のデータベース（糖鎖および機能性 RNA）に対してウェブサービスを新規に開発し公開した。

（1） 糖鎖関連遺伝子データベース(GlycoGene Database: GGDB)におけるウェブサービスの開発

GGDB には、産業技術総合研究所（産総研）・糖鎖医工学研究センターで独自に測定された糖転移反応に関する基質（糖鎖）、acceptor（糖鎖を受け取る分子）、生成物などの情報が示されている。この糖転移反応の情報はヒトの糖転移酵素の機能を正確に理解するために有用であり、H-InvDB に格納されている機能ドメイン、多型、遺伝子発現などの情報と連結することが望まれる。

そこで、GGDB の持つ糖転移反応の情報を取得するための糖鎖関連遺伝子データベース(GGDB)におけるウェブサービスの開発を行った。

具体的には、下記項目の開発を行った。

・XML ファイル設計

・データ検索、データ提供を行うウェブサービスプログラムの開発(SOAP および REST)

（データ項目）

① ID (gene name)

② acceptor

③ expression

④ PMID

⑤ ID 対応データ

⑥ 検索インデックス用データ

・ドキュメント、サンプルソースの提供

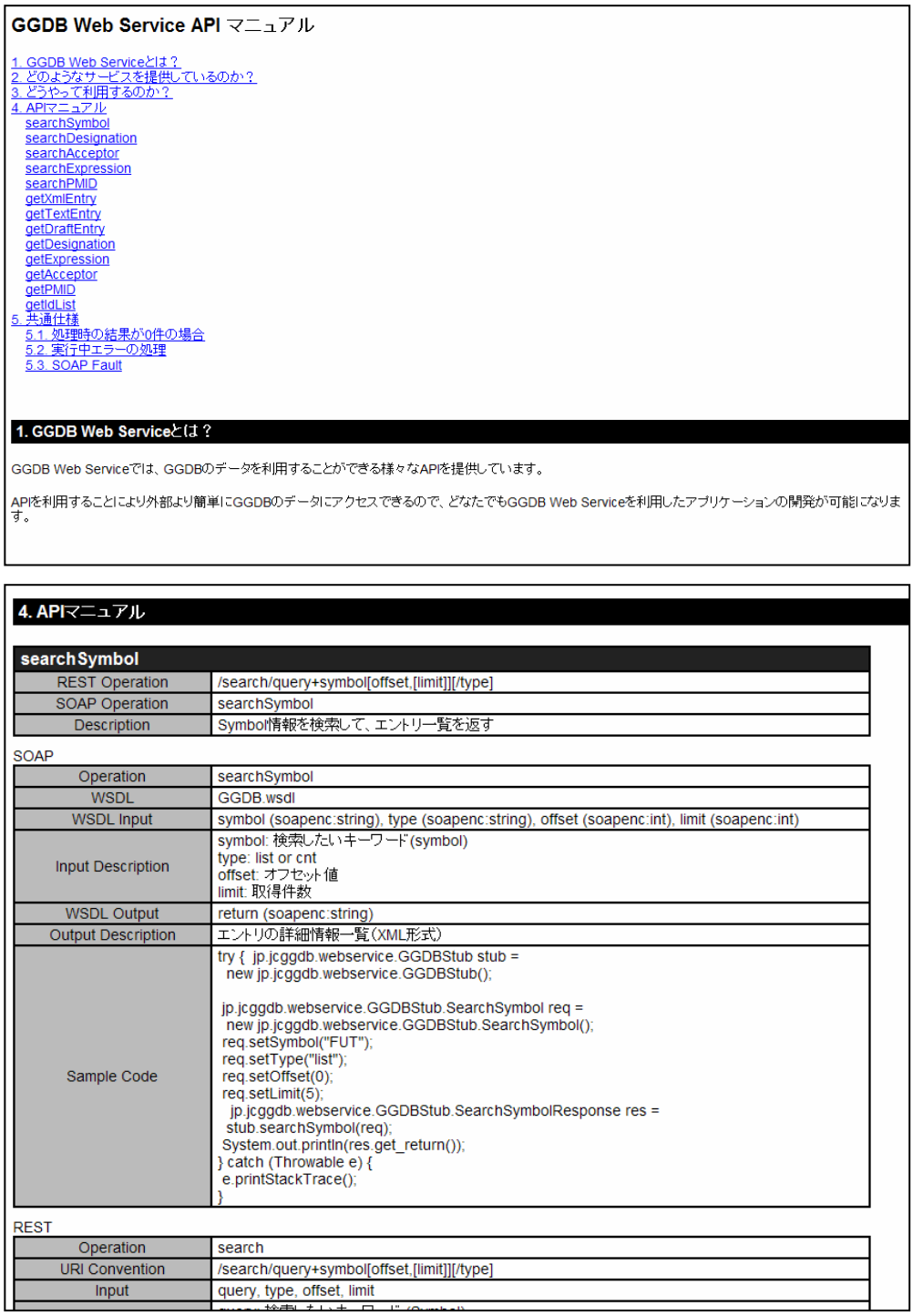
日本語および英語のドキュメント、および Perl, Phython, Ruby, Java のサンプルソースを提供した。（図 2.2.2-1 参照）

尚、GGDB ウェブサービスの設計は、平成 20 年度統合ウェブサービス マニフェスト(20081022 版)に準拠して行った。統合ウェブサービス マニフェストついては、後述の「 2.4.3 節　ウェブサービスの統合化」で詳しく報告する。

開発したウェブサービスは下記 URL から 3 月中に公開の予定である。

- GGDB ウェブサービス

  http://riodb.ibase.aist.go.jp/rcmg/ws/ggdb/



GGDB Web Service API マニュアル

1. GGDB Web Serviceとは？

GGDB Web Serviceでは、GGDBのデータを利用することができる様々なAPIを提供しています。

APIを利用することにより外部より簡単にGGDBのデータにアクセスできるので、どなたでもGGDB Web Serviceを利用したアプリケーションの開発が可能になります。

4. APIマニュアル

searchSymbol	
REST Operation	/search/query+symbol[offset,[limit]][/type]
SOAP Operation	searchSymbol
Description	Symbol情報を検索して、エントリー覧を返す

SOAP

Operation	searchSymbol
WSDL	GGDB.wsdl
WSDL Input	symbol (soapenc:string), type (soapenc:string), offset (soapenc:int), limit (soapenc:int)
Input Description	symbol: 検索したいキーワード(symbol) type: list or cnt offset: オフセット値 limit: 取得件数
WSDL Output	return (soapenc:string)
Output Description	エントリの詳細情報一覧(XML形式)

Sample Code	try { jp.jcggdb.webservice.GGDBStub stub =
	new jp.jcggdb.webservice.GGDBStub();
	
	jp.jcggdb.webservice.GGDBStub.SearchSymbol req =
	new jp.jcggdb.webservice.GGDBStub.SearchSymbol();
	req.setSymbol("FUT");
	req.setType("list");
	req.setOffset(0);
	req.setLimit(5);
	jp.jcggdb.webservice.GGDBStub.SearchSymbolResponse res =
	stub.searchSymbol(req);
	System.out.println(res.get_return());
	} catch (Throwable e) {
	e.printStackTrace();
	}

REST

Operation	search
URI Convention	/search/query+symbol[offset,[limit]][/type]
Input	query, type, offset, limit

図 2.2.2-1　糖鎖関連遺伝子データベース(GGDB)ウェブサービスドキュメント

(http://riodb.ibase.aist.go.jp/rcmg/ws/ggdb/doc/api_manual_jp/)

（2） 機能性 RNA データベース（fRNAdb）におけるウェブサービスの開発

NEDO「機能性 RNA プロジェクト」の成果である機能性 RNA データベース fRNAdb および機能性 RNA 用 UCSC ゲノムブラウザでは、公共データベースから集めた既知のヒト機能性 RNA の情報を網羅的に収集・整理して提供しており、さらに独自の手法で予測された機能性 RNA と、実験により検証された機能性 RNA の情報が追加される。一方、H-InvDB は転写が確認されたヒト遺伝子の情報を提供しており、両者のデータを対比するこ

とによって転写の証拠がある機能性 RNA の情報などの有益な情報が得られる。

そこで、fRNAdb および機能性 RNA 用ゲノムブラウザのデータを H-InvDB からプログラムによって入手できるようにするための機能性 RNA データベース（fRNAdb）におけるウェブサービスの開発と公開を行った。

具体的には、下記項目の開発を行い公開した。

・XML ファイル設計

・データ検索、データ提供を行うウェブサービスプログラムの開発(SOAP および REST)

（データ項目）

① ID

② description

③ 文献情報

④ 塩基配列長

⑤ ゲノム上位置情報

・ドキュメント、サンプルソースの提供

日本語および英語のドキュメント、および Perl, Phython, Ruby, Java のサンプルソースを提供した。（図 2.2.2-2 参照）

尚、fRNAdb ウェブサービスの設計は、平成 20 年度統合ウェブサービス マニフェスト(20081022 版)に準拠して行った。統合ウェブサービス マニフェストついては、後述の「2.4.3 節 ウェブサービスの統合化」で詳しく報告する。

開発したウェブサービスは下記 URL から公開した。

- fRNAdb ウェブサービス

http://www.ncrna.org/frnadb/api/

図 2.2.2-2　機能性 RNA データベース(fRNAdb)ウェブサービスドキュメント

(http://www.ncrna.org/frnadb/doc/api_manual_jp/)

（3） H-InvDB のウェブサービスの機能拡張と画面の改良による、ヒト遺伝子分子データの統合

H-InvDB ウェブサービスの機能拡張と画面改良により、ウェブサービスを使ったバーチャルなデータベース連携のしくみを整備してヒト遺伝子分子データの統合化を行った。

H-InvDB ウェブサービスの機能拡張

H-InvDB に格納されている情報には、機能ドメイン、多型、進化、発現、機能性 RNA の情報等がある。H-InvDB に含まれるデータを取得するためのウェブサービスを改良し、利用頻度の高い遺伝子発現パターンとモデル生物のオーソログ（オルソログ）等の情報を取得できるようにすることで、異なるデータベース間での相互データ提供による連携を実現することができる。

そこで、H-InvDB ウェブサービスの機能拡張について、具体的には下記項目の開発を行い公開した。

・データ検索、データ提供を行うウェブサービスプログラムの開発(SOAP および REST)

（データ項目）

① 分子進化（オーソログ）
② 遺伝子発現情報
③ 遺伝子名
④ ゲノム上の位置情報
⑤ 多型情報
⑥ ドメイン情報

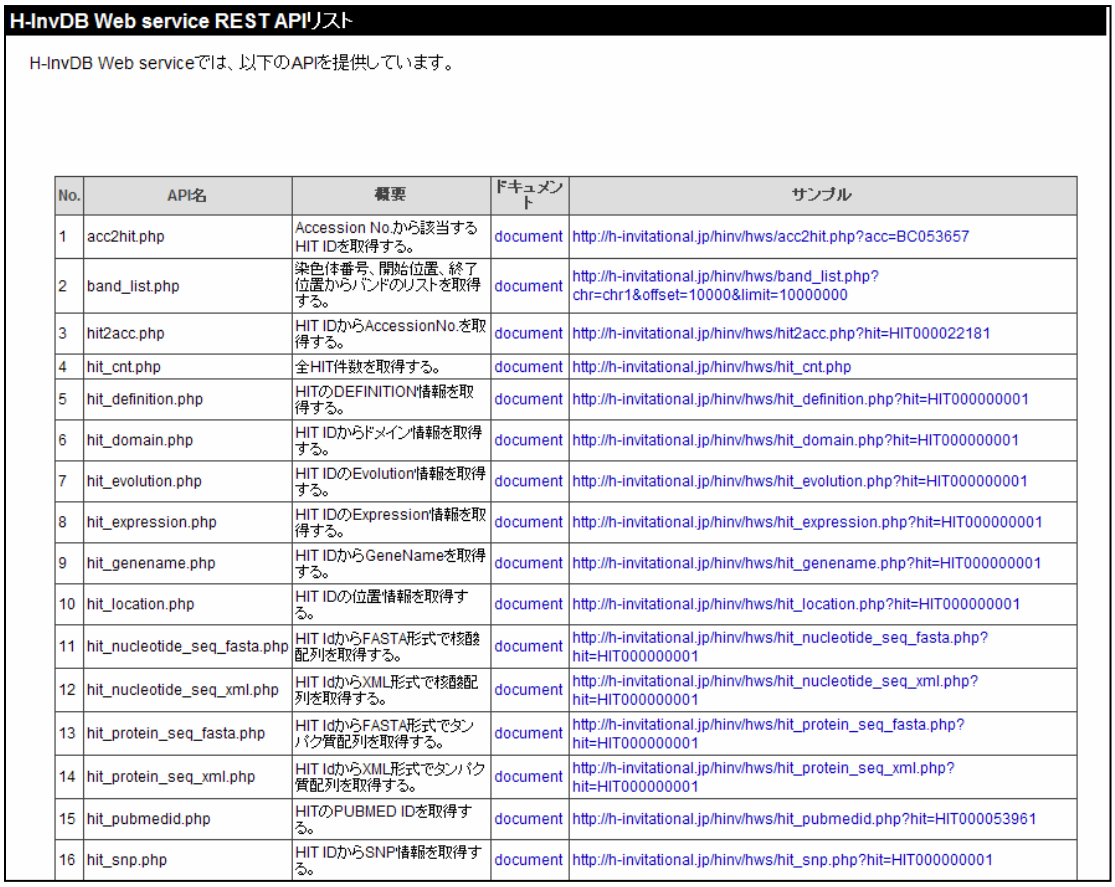
- 独立サーバーでの運用
- ドキュメント、サンプルコードの追加

日本語および英語のドキュメント、および Perl, PHP, Ruby, Java のサンプルソースを提供した。（図 2.2.2-3 参照）

開発したウェブサービスは下記 URL から公開した。

- H-InvDB ウェブサービス

  http://h-invitational.jp/hinv/hws/doc/index_jp.html



H-InvDB Web service REST APIリスト

H-InvDB Web serviceでは、以下のAPIを提供しています。

No.	API名	概要	ドキュメント	サンプル
1	acc2hit.php	Accession No.から該当するHIT IDを取得する。	document	http://h-invitational.jp/hinv/hws/acc2hit.php?acc=BC053657
2	band_list.php	染色体番号、開始位置、終了位置からバンドのリストを取得する。	document	http://h-invitational.jp/hinv/hws/band_list.php?chr=chr1&offset=10000&limit=10000000
3	hit2acc.php	HIT IDからAccessionNo.を取得する。	document	http://h-invitational.jp/hinv/hws/hit2acc.php?hit=HIT000022181
4	hit_cnt.php	全HIT件数を取得する。	document	http://h-invitational.jp/hinv/hws/hit_cnt.php
5	hit_definition.php	HITのDEFINITION情報を取得する。	document	http://h-invitational.jp/hinv/hws/hit_definition.php?hit=HIT000000001
6	hit_domain.php	HIT IDからドメイン情報を取得する。	document	http://h-invitational.jp/hinv/hws/hit_domain.php?hit=HIT000000001
7	hit_evolution.php	HIT IDのEvolution情報を取得する。	document	http://h-invitational.jp/hinv/hws/hit_evolution.php?hit=HIT000000001
8	hit_expression.php	HIT IDのExpression情報を取得する。	document	http://h-invitational.jp/hinv/hws/hit_expression.php?hit=HIT000000001
9	hit_genename.php	HIT IDからGeneNameを取得する。	document	http://h-invitational.jp/hinv/hws/hit_genename.php?hit=HIT000000001
10	hit_location.php	HIT IDの位置情報を取得する。	document	http://h-invitational.jp/hinv/hws/hit_location.php?hit=HIT000000001
11	hit_nucleotide_seq_fasta.php	HIT IdからFASTA形式で核酸配列を取得する。	document	http://h-invitational.jp/hinv/hws/hit_nucleotide_seq_fasta.php?hit=HIT000000001
12	hit_nucleotide_seq_xml.php	HIT IdからXML形式で核酸配列を取得する。	document	http://h-invitational.jp/hinv/hws/hit_nucleotide_seq_xml.php?hit=HIT000000001
13	hit_protein_seq_fasta.php	HIT IdからFASTA形式でタンパク質配列を取得する。	document	http://h-invitational.jp/hinv/hws/hit_protein_seq_fasta.php?hit=HIT000000001
14	hit_protein_seq_xml.php	HIT IdからXML形式でタンパク質配列を取得する。	document	http://h-invitational.jp/hinv/hws/hit_protein_seq_xml.php?hit=HIT000000001
15	hit_pubmedid.php	HITのPUBMED IDを取得する。	document	http://h-invitational.jp/hinv/hws/hit_pubmedid.php?hit=HIT000053961
16	hit_snp.php	HIT IDからSNP情報を取得する。	document	http://h-invitational.jp/hinv/hws/hit_snp.php?hit=HIT000000001

図 2.2.2-3　H-InvDB ウェブサービスドキュメント

（http://h-invitational.jp/hinv/hws/doc/index_jp.html）

H-InvDB の画面の改良

H-InvDB、GGDB、fRNAdb において開発・公開したウェブサービスの仕組みを通して、データベース間のリアルタイムでの相互のデータ提供が可能となった。また、2.2.1 で報告した自動リンク管理システムの整備により、データベース間の自動リンク設定が可能となった。

そこで、ウェブサービスの仕組みを通して H-InvDB、GGDB、fRNAdb で相互に取得したデータをリアルタイムで表示し、また、相互にリンクするための画面の改良を H-InvDB に対して実施した。

画面改良について、具体的には下記項目の開発を行い公開した。

- GGDB データの表示とリンク設置(対象:1画面)

  下記データ項目を H-InvDB 画面(Transcript view)にウェブサービスを介して表示し、リンク自動管理システムを介してリンクの設置するようシステム開発を行った。(図 2.2.2-4 参照)

  - ID (gene name)
  - expression
  - PMID

開発した画面の改良は、3 月中に公開の予定である。

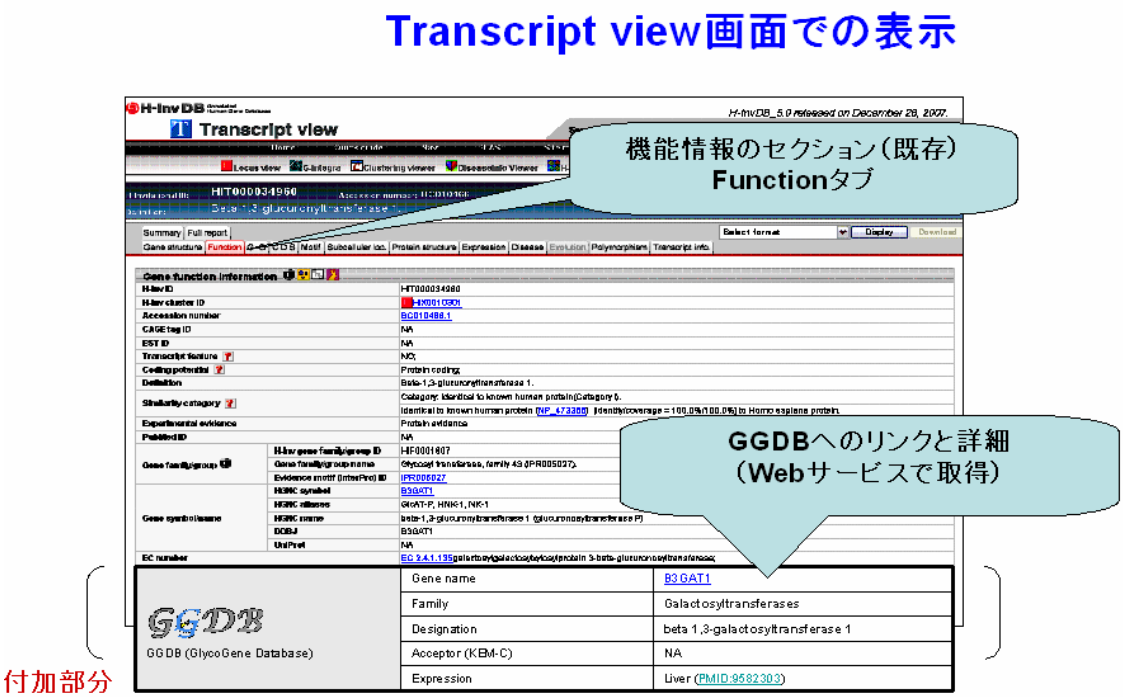


図 2.2.2-4 H-InvDB 画面への GGDB データ表示

- fRNAdb データの表示とリンク設置（対象：2 画面）

  下記データ項目を H-InvDB 画面（Transcript view および G-integra）にウェブサービスを介して表示し、リンク自動管理システムを介してリンクの設置を行った。（図 2.2.2-5 参照）

  - ID
  - description
  - 文献情報
  - ゲノム上位置情報

開発した画面の改良は、H-InvDB 6.0 より公開した。

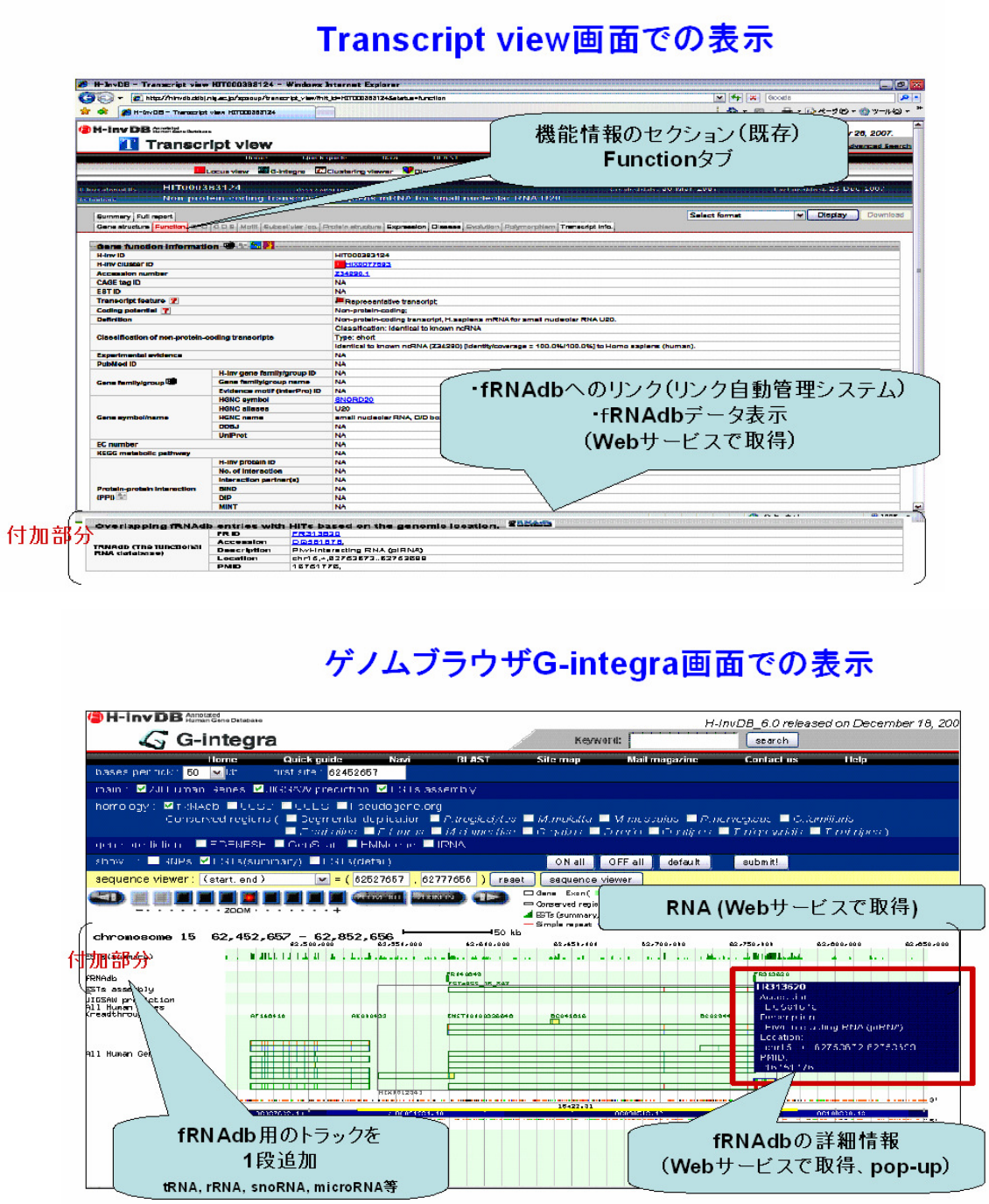


図 2.2.2-5　H-InvDB 画面への fRNAdb データ表示

- fRNAdb データ対応表のダウンロード機能の提供

H-InvDB は、タンパク質コード遺伝子についてのアノテーションを重点的に行っているヒト遺伝子データベースであり、一方、fRNAdb は機能性 RNA（非タンパク質コード遺伝子）の統合データベースである。ヒトの全ての遺伝子を正確に定義するためには、主に、タンパク質をコードしている遺伝子と非タンパク質コード遺伝子という機能・構造・性質などの大きく異なる遺伝子をアノテーションする必要があるが、両者を 100%厳密に区別する解析手法はまだ確立していない。

そこで、タンパク質コード遺伝子と非タンパク質コード遺伝子をより正確にアノテーションする為の情報を提供することを目的として、ゲノム上の位置の重なる H-InvDB と fRNAdb エントリーのデータ対応表の作成、データ対応表ダウンロード機能の開発および公開を行った。

具体的には、fRNAdb のエントリーと H-InvDB の転写産物エントリー（HIT）の対応付けを、ヒトゲノム上の位置情報の重なり（ゲノム上の両ストランドの exon 位置 1bp 以上）で判定した ID 対応表を作成し、新規ページ、「アノテーショントピック」より fRNAdb データ対応表をダウンロード可能にした（図 2.2.2-6 参照）。

URL:

http://www.h-invitational.jp/hinv/topic_annotation/frnadb_ja.cgi

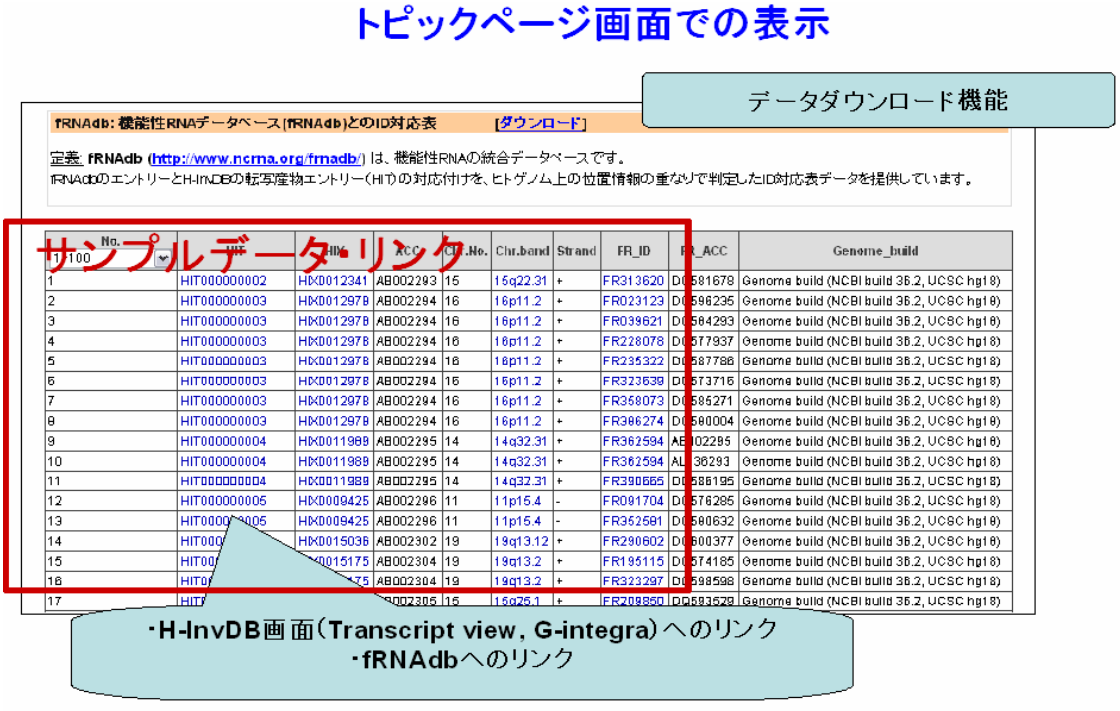


図 2.2.2-6 fRNAdb データ対応表ダウンロード画面

- fRNAdb データ検索機能の提供

前項で報告した、ゲノム上の位置の重なる H-InvDB と fRNAdb エントリーのデータ対応表に含まれるデータを検索するため fRNAdb データ検索機能の開発と公開を行った。

具体的には、fRNAdb データのうち FR ID および FR ACC# が検索できるよう、H-InvDB の新詳細検索画面の「機能性 RNA」コンテンツに項目を追加した（図 2.2.2-7 参照）。

URL:

http://www.h-invitational.jp/hinv/c-search/gotocrssearch.do?blockState=3

尚、H-InvDB の新詳細検索は、遺伝子のゲノム上位置、タンパク質としての機能、機能性 RNA 等の 16 の検索コンテンツをユーザーが任意に選択して複合的な検索を行うことができる検索システムである。

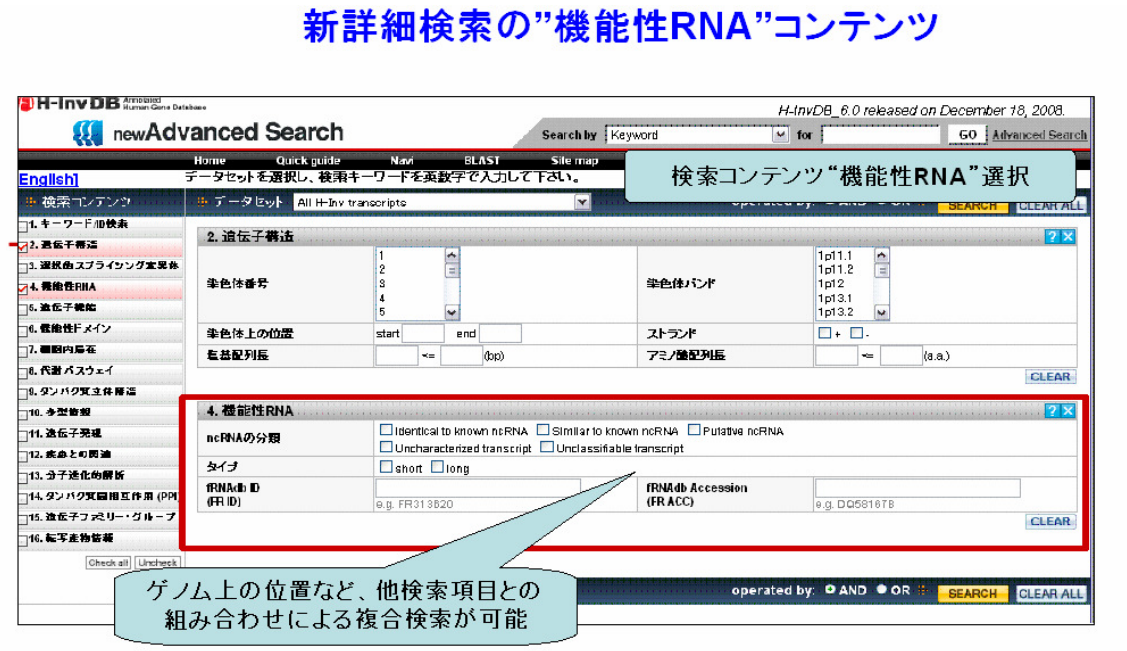


図 2.2.2-7 fRNAdb データ検索機能（新詳細検索画面）

2．2．3　ヒト選択的スプライシングバリアント・データベース（H-DBAS）のデータ更新

H-InvDB のヒト遺伝子情報に対して糖転移酵素（GGDB）と機能性 RNA（fRNAdb）の精査された情報を統合するにあたり、統合された情報の品質を保証することが何よりも重要である。一方で、ヒト遺伝子には多数の選択的スプライシングがあり、ヒトの転写産物やタンパク質の構造と機能に大きな影響を与えることが知られている。そのため、統合されたデータの品質を保つためには、H-InvDB と GGDB と fRNAdb の情報を統合する際に選択的スプライシングの影響を十分に考慮し、最新かつ高精度の情報を使うことが必要である。そこで、ヒトの選択的スプライシングに関する最新かつ高精度の情報を整備して H-InvDB に提供するため、「ゲノム情報統合プロジェクト」で構築されたヒトの選択的スプライシングのデータベース H-DBAS の情報を最新データに更新した。

選択的スプライシング（AS）とは、1つの遺伝子座から複数の転写物（バリアント）が生成される現象である。これにより、生物は遺伝子数以上に多様な機能を獲得することが可能となっている。それだけではなく、ガンなどの疾病にも関わっている重要な機構でもある。我々はこれまでに、ヒトの選択的スプライシングをゲノムワイドに同定し、選択的スプライシングパターンの分類・タンパク機能に影響を与える選択的スプライシング・ヒトとマウスで保存された選択的スプライシングなどを調べた。そして、これらの研究成果を公開するため、H-DBAS（URL: http://h-invitational.jp/h-dbas/）というヒトの選択的スプライシングデータベースを構築し、アップデートを繰り返してきた。

今回我々は、選択的スプライシングと同様に転写物の多様性に関して重要な機構である、選択的プロモーターおよび選択的ポリアデニレーションに注目した。そして、これらと密接に関係する選択的スプライシングを新たに同定し、H-DBAS へ登録することを目的とした。具体的には、選択的スプライシングのパターン分類をより厳密にし、5’末エクソンと 3’末エクソンにおける選択的スプライシングのアノテーションを実施した。特に、5’末エクソンの選択的スプライシングと転写開始点の変異が同時出現する場合や、3’末エクソンの選択的スプライシングと転写終結点の変異（または poly-A サイト変異）が同時出現する場合を正確に判定してアノテーションを行った。新しい選択的スプライシングパターンとして、5’末で選択的第一エクソン・選択的第一 3’スプライス、3’末で選択的最終エクソン・選択的最終 5’スプライスを同定し、それぞれ統計を取った（図 2.2.3-1）。以前から報告されているように、ヒトでは 5’末でより複雑な転写機構を備えていることを裏付けた。そして、これらを H-DBAS に登録し、トップページ（図 2.2.3-2）のリンクを経由して詳細検索画面（図 2.2.3-3）から検索し、Java アプレットで動作する AS ビューワ（図 2.2.3-4）上で表示できるようにした。他の更新内容として、転写物やマウスゲノムのバージョンアップ・AS エクソン判定の緩和・Materials&Methods や Operating manual ページの日本語化などを行った。これらを更新した H-DBAS を 2009 年 1 月 30 日に公開した。

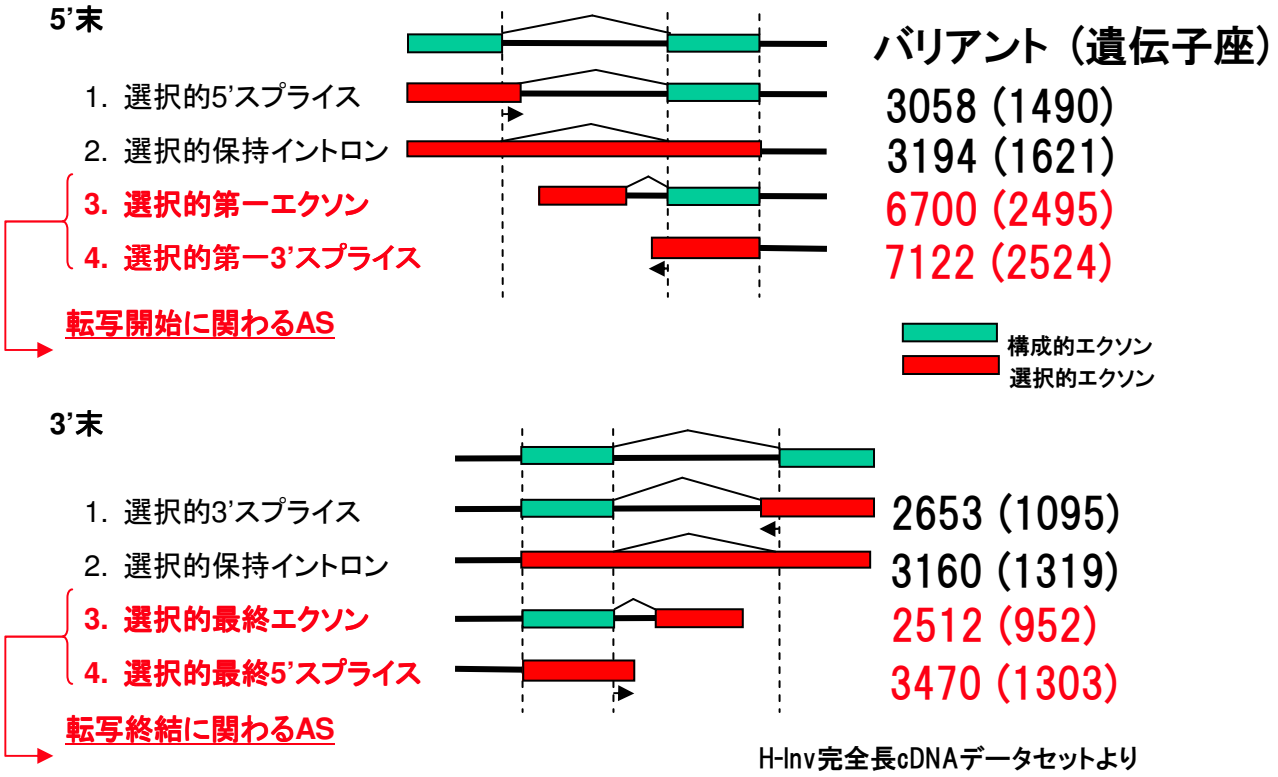


図 2.2.3-1　5'/3'末でアノテーションした転写開始・終結に関わる AS パターン（赤字）とそのバリアントおよび遺伝子座数

URL→http://h-invitational.jp/h-dbas/

図 2.2.3-2　H-DBAS のトップページ（日本語版）と URL

詳細検索(Advanced search)ページの一部

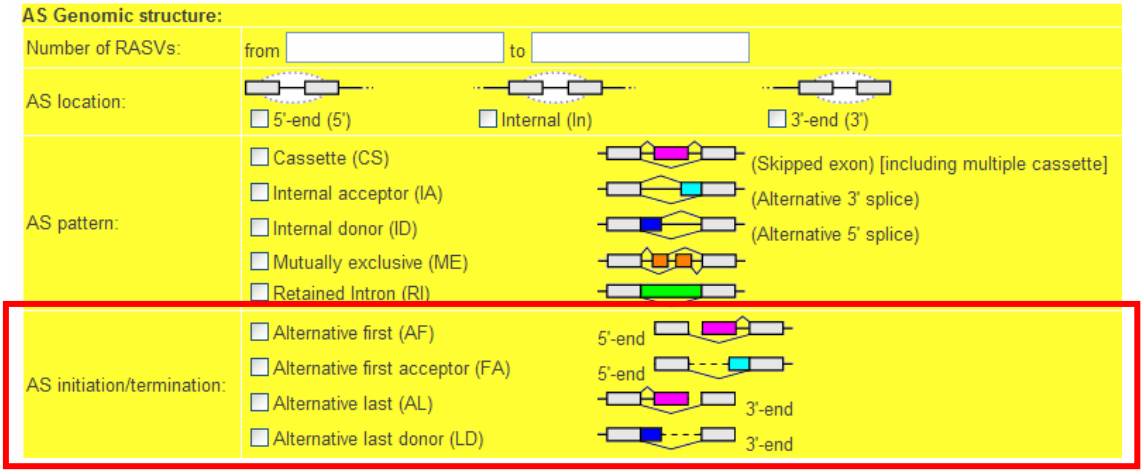


図 2.2.3-3　詳細検索ページに追加した新規 AS パターン

ASビューワ(Javaアプレット)

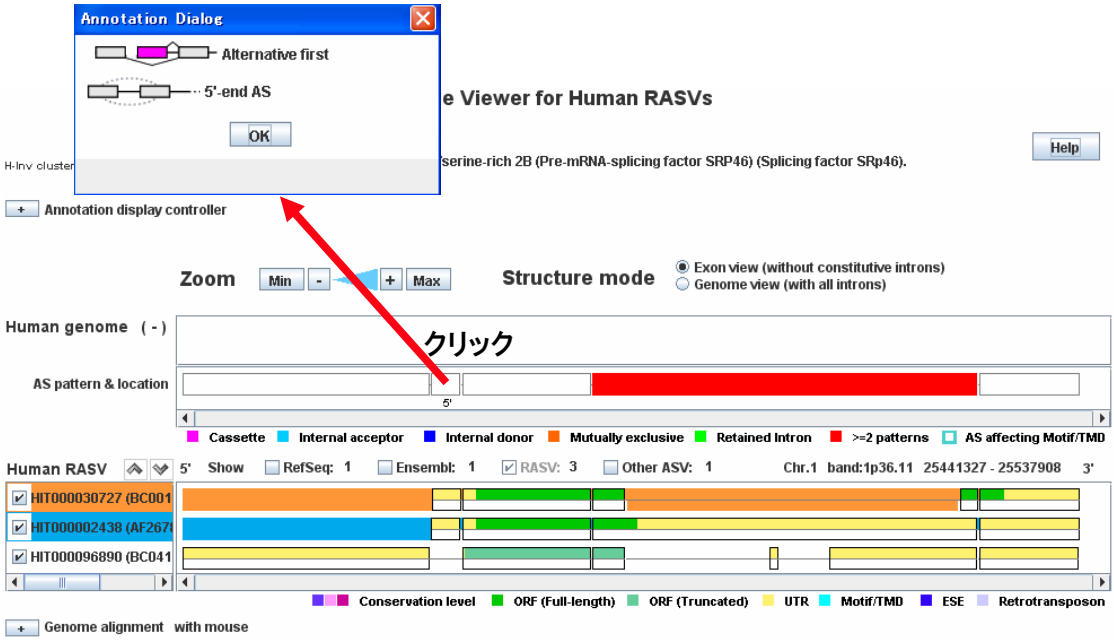


図 2.2.3-4　AS ビューワに追加した新規 AS パターン

2．3　国内外の最新動向とユーザーニーズ

2．3．1　国内外の最新動向調査と広報・普及活動

（1） 国内外の動向調査

経済産業省ライフサイエンスデータベース・ポータルサイト MEDALS に関する学会発表または聴講を目的として、国際学会に 1 人回、国内学会等に 17 人回参加した（表 2.3.1-1）。

表 2.3.1-1　参加した国際および国内学会等

日付	会議名	開催場所
2008 年 7 月 19-23 日	The 16th International Conference on Intelligent Systems for Molecular Biology [ISMB 2008]	トロント
2008 年 8 月 18-20 日	第 28 回日本糖質学会年会	つくば
2008 年 10 月 15-17 日	BioJapan 2008	横浜
2008 年 10 月 31 日	JBiC2008 プロジェクト研究成果報告会	品川
2008 年 12 月 9-12 日	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]	神戸
2009 年 1 月 26 日	クリエイティブコモンズシンポジウム	丸の内
2009 年 1 月 29-30 日	第 8 回産総研・産技連 LS-BT 合同研究発表会	つくば

（2） 学会発表およびその機会を利用した広報活動

学会等において MEDALS およびメイン統合データベースである H-InvDB に関連するポスター発表を行い、その機会を利用してデモンストレーションやパンフレット配布等の広報活動を実施した。また、JBiC 成果報告会にてプロジェクトの成果発表をおこなった。

○2008 年 7 月 19-23 日　The 16th International Conference on Intelligent Systems for Molecular Biology [ISMB 2008]（トロント）

ポータルサイトで提供する情報の収集を可能とするようなテキストマイニング技術の可能性を探るため、その動向について調査した。文献情報を扱う解析に関しては、文脈解析（談話解析：discourse analysis）に着目して前後の文章をつなげる意味を取り出すことで、1 文の場合に比べてより多くの情報が得られるという報告があった。文献から事実を取り出してデータベース化するために1つの進歩を示したといえる。

また、多くのデータベースで wet 実験のデータと文献データは PubMed ID によるリンクで連携されているが、今後はテキストマイニング等の技術を使って gene symbol や疾患名等のキーワードで連携できれば有用なデータを提供できると考えられる。

○2008 年 10 月 15-17 日　BioJapan 2008（横浜）

JBiC ブース内で、MEDALS および H-InvDB についてのデモンストレーションやパンフレットの配布を行い、MEDALS の特徴や H-InvDB の利用方法等を説明した。初日と最終日にデモンストレーションの時間を設定し、

特に MEDALS についてはサイト公開直前の時期であったため、広く紹介する良い機会となった。また、前年度にプロモーショングッズとして作製した H-InvDB オリジナルデザインのエコバッグもパンフレットと合わせて配布し好評を博した。

【配布物】

- MEDALS パンフレット：200 部
- 第 1 回講習会案内：100 部
- H-InvDB パンフレット：300 部（英語版 200 部、日本語版 100 部）
- エコバッグ：300 個

〇2008 年 10 月 31 日　JBiC2008 プロジェクト成果報告会（品川）

MEDALS および H-InvDB について、2 件の口頭発表と 10 件のポスター発表をおこなった。JBiC の研究開発に対する産業界等からの強い期待を反映し、口頭発表会場はほぼ満席でポスター会場も盛況であった。

また、ポスター発表時にはユーザーの意見や質問事項も集めることができた。

【意見等】

- データベースを作成中だが、リンク自動管理システムに追加してもらうことは可能か？
- 配列のバージョンの違いの影響はないのか？
- 新規関連文献お知らせツールを Windows から使いたい
- H-InvDB には、ヒト以外の生物種のデータはどのようなものがあるか？
- SuperTACT では、RNA のアノテーションパイプラインはあるのか？

〇2008 年 12 月 9-12 日　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]（神戸）

MEDALS および H-InvDB に関係する 9 件のポスターを連番で発表し、多くの方に研究成果を説明することができた。会場ではパンフレットを入れたエコバッグをポスター前に置いておくことで、事前のパンフレット配布等によるポスターへの誘導の手間が省けて、効率よく研究内容のアピールができた。

ポスター発表時には、事前に発表内容を知っていたわけではなく偶然目について興味を持った人が多い印象を受けた。MEDALS や H-InvDB ではまだ潜在的なユーザーが多い可能性があり、広報活動に力を入れる必要性を認識できた。また、講習会への関心も高く、初歩的な利用方法を知りたいというニーズが感じられた。

【配布物】

- MEDALS パンフレット：50 部
- 第 2 回講習会案内：100 部
- BMB2008 チーム内ポスター要旨集：150 部
- H-InvDB パンフレット：170 部
- エコバッグ：40 個

【意見等】

- 各種データベースに対して、ニーズはあるが常に使うほど慣れるまでには時間がかかる

- 業務で必要になったが、どんなデータベースがあるのかわからなかった
- 特許情報を無料で閲覧したい
- リンク自動管理システムのデータ ID 変換サービスをぜひ使ってみたい
- データID 変換サービスで、シノニムまで対応されていると gene symbol をキーにデータを検索することの多い wet 実験の研究者にとってさらに使いやすくなる

〇2009 年 1 月 29 日　第 8 回産総研・産技連 LS-BT 合同研究発表会（つくば）

MEDALS および H-InvDB に関連する 3 件のポスターを、生物機能とバイオインフォマティックスの区分で発表した。また、広報活動として、MEDALS と H-InvDB のパンフレット、および第 3 回講習会案内を配布した。

（3） MEDALS 更新時の一斉アナウンスの実施

2009 年 2 月 6 日の MEDALS 大規模更新の日程に合わせて、ニュース掲載やメール配信等各種媒体を用いた一斉アナウンスを実施した（表 2.3.1-2）。このような様式の案内は初の試みであったが、MEDALS へのアクセス数が著しく増加し大変効果があったことが確認できた。

表 2.3.1-2　MEDALS 更新に関する一斉アナウンス一覧

アナウンス形態	掲載・配信先/関連 URL
ニュース掲載	産業技術総合研究所公式サイト「主な研究成果」 http://www.aist.go.jp/aist_j/new_research/nr20090206/nr20090206.html
	JBiC 公式サイト「JBIC からのお知らせ」 http://www.jbic.or.jp/bio/c/sub_news/09/0206o.html
	バイオメディシナル情報研究センター公式サイト「News & Topics」 http://unit.aist.go.jp/birc2/
	分子システム情報統合チーム公式サイト「最新情報」 http://www.h-invitational.jp/idb/index.html
	H-InvDB 公式サイト日本語版「What's new」 http://h-invitational.jp/index_jp.html
掲示板記載	BIOWEB（バイオ研究者支援サイト）掲示板 http://www.bioweb.ne.jp/mid_board/stg_info/id_3640/
	Biotechnology Japan「皆のホームページ」（後日メーリングリストでも告知） http://biotech.nikkeibp.co.jp/100hpn/100hpdetail.jsp?id=4027&site=btj
メール配信	JBiC めるまが　第 242 号　090206
	H-InvDB　メールマガジン http://www.h-invitational.jp/hinv/mag/backnumber.jsp
	BIOWEB　メール情報 http://www.bioweb.ne.jp/mid_bnews/stg_info/id_2670/

（4）　リンクの設置

MEDALS　へのアクセスルートの拡大を目的として、関連するサイト（学術的ポータルサイト、大学や研究所のページ、解析サービスのサイト、または医学生物学関連の商業ウェブサイト（Web サイト）等）から MEDALS へのリンクの設置を実現した（表 2.3.1-3）。実際、外部から MEDALS への遷移を調べたアクセスログ解析結果でもこれらのサイト等から誘導されていることが確認できた。

表 2.3.1-3　MEDALS へのリンク設置一覧

サイト名/リンク元 URL
産総研公式サイト「産総研データベースポータル」 http://www.aist.go.jp/aist_j/database/portal/index.html
JBiC 公式サイト「研究開発成果」 http://www.jbic.or.jp/activity/r_d_pj_a.html
H-InvDB オフィシャルサイト「リンク集/H-InvDB 関連データベース＆ツール」 http://h-invitational.jp/links_jp.html
文部科学省統合データベースプロジェクト http://lifesciencedb.jp/
BIOWEB「バイオリンク」(情報収集サイト欄内) http://www.bioweb.ne.jp/mid_biolink/index.html#NEWS
千葉大亥鼻イノベーションプラザ「ニュース」 http://www.kanto-bi.smrj.go.jp/Ciip/archives/2008/11/medalsjp.html

（5） 講習会開催

MEDALS の公開後、MEDALS および H-InvDB に関する講習会を全国各地で 4 回開催し、内容の解説とデモンストレーションを実施した(表 2.3.1-4)。実際に PC を使用して体験していただくことにより、利用方法や機能をアピールした。MEDALS と H-InvDB だけでなく、H-InvDB のサブデータベースである遺伝子発現データベース H-ANGEL や比較ゲノムデータベース Evola についても解説し、また、H-InvDB とウェブサービスで連携を図った機能性 RNA データベース(fRNAdb)と糖鎖関連遺伝子データベース(GGDB)の開発機関の方も講師として招いてデータベースを紹介していただいた。

参加者からは多くの質問を受け、今後の改良項目について検討することができた。さらに、参加者へのアンケート調査も行い、MEDALS および H-InvDB への意見や講習会の感想を収集した。

表 2.3.1-4　今年度開催した講習会

日付	講習会名	開催場所	参加人数
2008 年 11 月 6 日	第 1 回講習会	大阪(千里ライフサイエンスセンター)	25 名
2008 年 12 月 19 日	第 2 回講習会	お台場(産総研・臨海副都心センター)	32 名
2009 年 2 月 6 日	第 3 回講習会	つくば(産総研・つくば中央 共用講堂)	21 名
2009 年 3 月 17 日	第 4 回講習会	福岡(九州大学病院キャンパス)	-

【参加者へのアンケート調査結果】

- 開催情報の入手は、H-InvDB メールマガジンや関係者からが多かった
- 参加の目的は、業務への活用と情報収集が同程度だった

- 講習会の満足度については、9 割程度の方から、期待に沿った内容で難易度も適当という回答が得られた

【意見等】

- 学生とのゼミ等で活用したい
- H-InvDB については、知らなかったので今後は使いたい
- ヒトを研究対象としていないため、H-InvDB は使う予定がない
- MEDALS に関しては、もっと広く宣伝するべきだ
- こんな場合にこう使うと便利、というストーリーの講習会にしてはどうか？

（6） 新聞報道

公式のプレス発表の機会はなかったが、今年度は MEDALS について 2 回の新聞報道があった。産業界から期待と注目が集まっているといえる。

- 2008 年 8 月 21 日　（化学工業日報）「経産関連ライフサイエンス DB　ポータルサイト構築へ」
- 2008 年 10 月 31 日　（化学工業日報）「経産のバイオプロ DB 統合、運用開始」

2. 3. 2　ユーザーのニーズ調査

（1） アンケート調査

産業界における MEDALS の潜在ニーズを調べることを主な目的として、専門会社に委託したウェブアンケート調査を 2008 年の 7 月から 8 月にかけて実施し、200 名までの結果を集計した。さらに今年度 H-InvDB との連携が実現した機能性 RNA データベース（fRNAdb）および糖鎖関連遺伝子データベース（GGDB）に関する項目も作成した。質問項目は自由記述回答も含む全 40 問で、閲覧頻度と感想に関しては回答しやすいようテーブル形式の選択肢を採用した。以下にアンケート調査結果をまとめる。

① 調査対象者について

今年度のニーズ調査は産業界を対象としていたため、アンケート案内の配信先を検討した結果、約 6 割の回答者の所属が公的研究機関と民間企業であった。

【アンケート案内の対象】

- JBiC プロジェクト参加企業
- JBiC めるまが購読者
- H-InvDB メールマガジン購読者
- 日本バイオインフォマティクス学会会員
- 首都圏バイオメルマガ購読者
- 日本糖質学会参加者
- BIOWEB 登録者のうちメールアドレスが企業ドメインの方

② ライフサイエンス分野のデータベースについて

ライフサイエンス分野のデータベースで情報収集をする際に利用している情報源を調べた結果、PubMed や DDBJ/EMBL/GenBank、Entrez は月に 1 回以上利用されている割合が高かった（図 2.3.2-1）。また、ライフサイエンス分野のデータベースで内容の充実が望まれている情報としては、疾患、遺伝子発現、タンパク質を選んだ回答者が多く（図 2.3.2-2）、より具体的なアノテーション情報が必要とされている傾向があった。

研究生活に役立つポータルサイト作成の手がかりを知るための質問項目として、時間を短縮したい業務を尋ねたところ、インフォマティクス関係が最も多く、実験関係と論文関係が次に多いという結果となった（図 2.3.2-3）。このデータを参考にして、MEDALS から新規関連文献お知らせツール（PubMedScan）を公開することになった。

以下に、データベースへの不満等についての自由記述回答の一部を紹介する。

【利用しているデータベースへの不満点等】

- それぞれのサイトに行かずに 1 つのサイトでまとめて検索したい
- 種類が多すぎてどれが良いのか良く分からない
- 操作方法が分かりづらいことがある
- 横断的な検索が煩雑で不便
- 思ったキーワードで検索できない

- cDNA クローンを分与・購入できるサイトへのリンクがあると良い
- クローン翻訳産物を認識する抗体とのリンクがあると良い
- 外部リソースや最新更新記録等が記載されていない場合がある

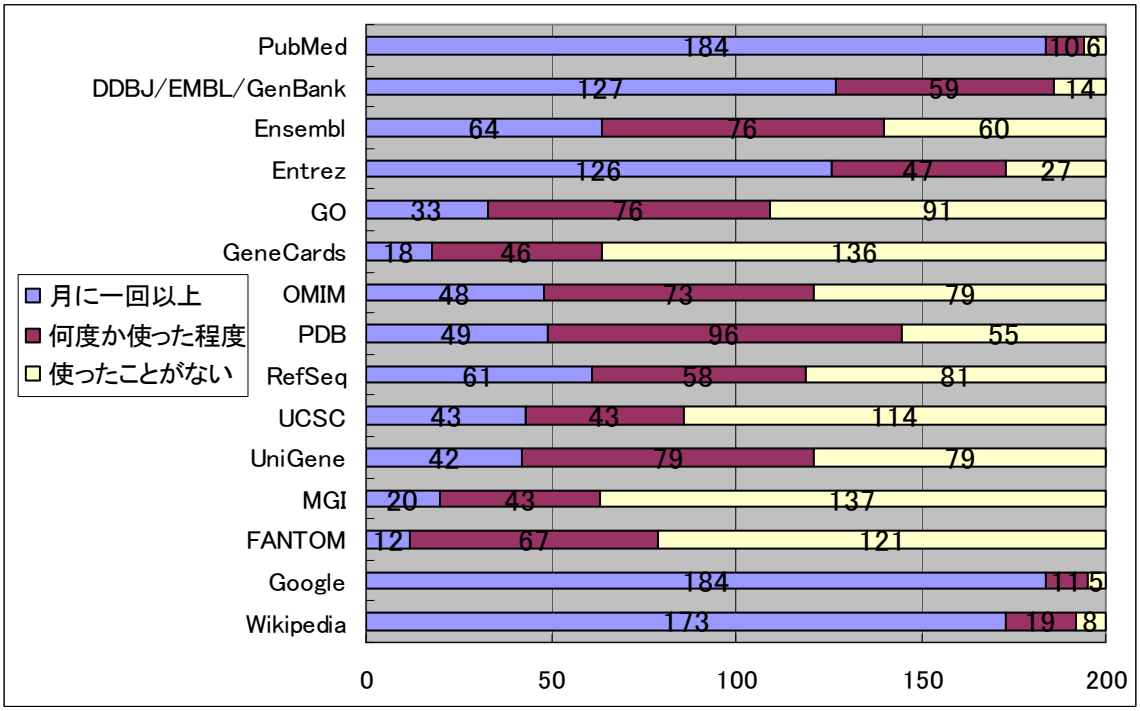


図 2.3.2-1 ライフサイエンス分野の情報収集で利用する情報源

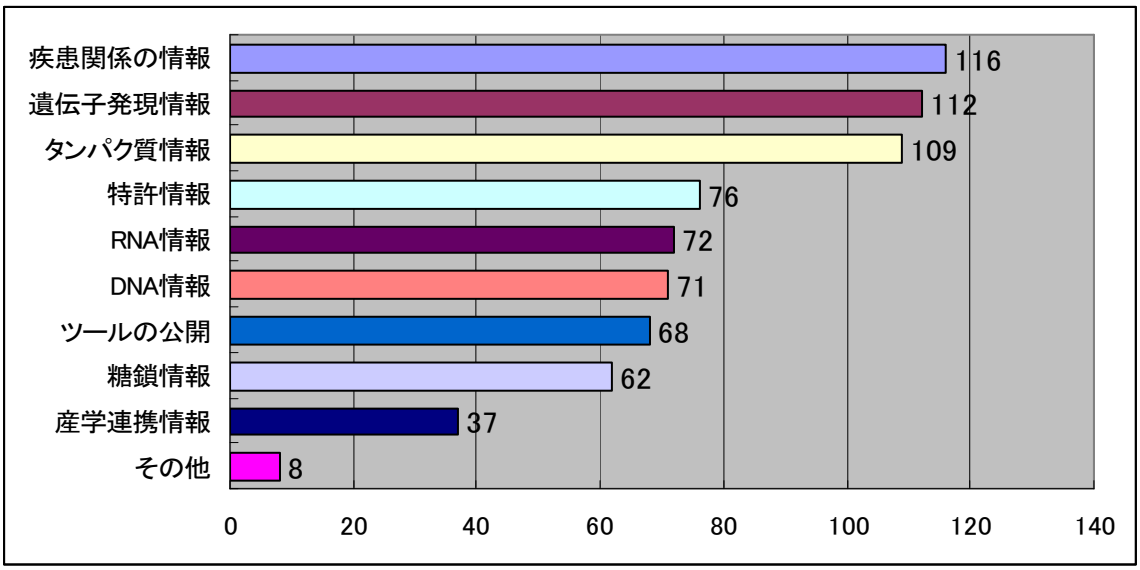


図 2.3.2-2 ライフサイエンス分野のデータベースで充実が望まれる情報

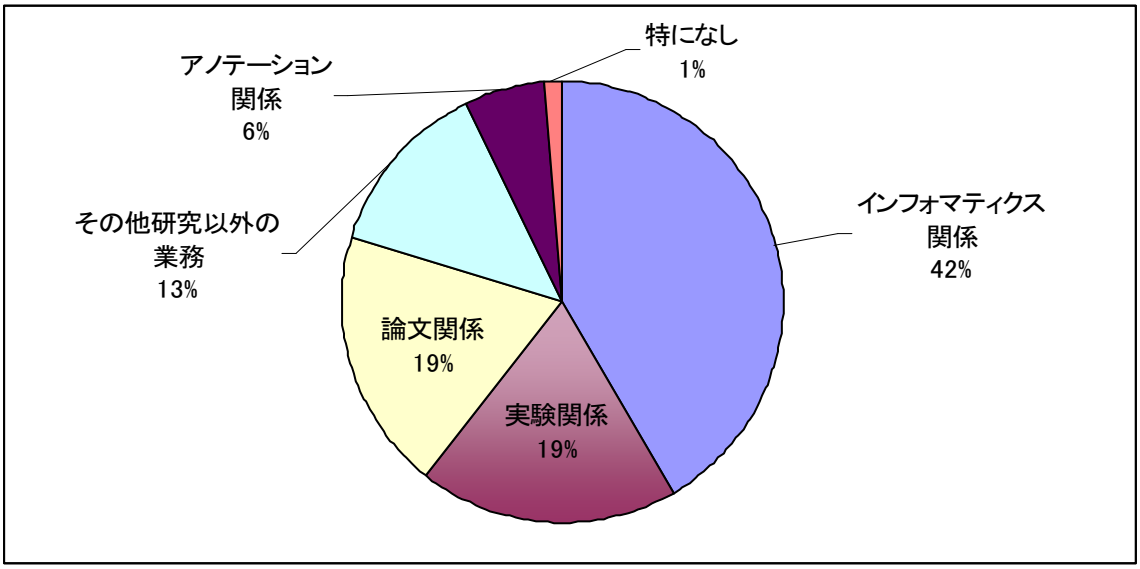


図 2.3.2-3 ライフサイエンス分野の研究生活で最も時間を短縮したい業務

③ 新規ポータルサイトについて

MEDALS の構築にあたり、ポータルサイトに関連するニーズについて自由記述の形式を中心に調査した。興味深いものとして選ばれたコンテンツ例は、論文サーチシステムやデータベースアーカイブ、データベース便覧、用語集等で(図 2.3.2-4)、上位 3 件は既に MEDALS から公開中である。また、ポータルサイトのコンセプトに関する調査では、利便性や信頼性といったキーワードが多くの回答者から選ばれており(図 2.3.2-5)、堅実な作りのポータルサイトが望まれているといえる。

以下に、ポータルサイトへの希望・要望についての自由記述回答の一部を紹介する。

【ポータルサイトへの希望・要望】

- ポータルというと単なる寄せ集めのイメージがあるが、横断検索や統合表示等をきちんと実現したものにしてほしい
- クロス検索の利便性を高めてほしい
- 横断検索はさせてほしいが、あまり時間はかけたくない
- 単なるデータベースのリンクではなくて、ツール等も含めた複合検索の利便性を上げてほしい
- ソース、網羅性等が既存のデータベースと比較できて、特徴がわかりやすければ使うかもしれない
- 異分野融合が流行りなので、そのような研究者同士が出会えるような、研究者検索リスト等があれば面白いと思う
- 初心者用とエキスパート用の 2 つの入り口が必要かと思う
- 便覧にアクセスできる窓口サイトを増やしてほしい
- ユーザーの意見を取り入れながら、バージョンアップを 2-3 年ごとに繰り返してほしい
- 経済産業省だけではなくて日本各地、世界各地のデータベースとの連携をしてほしい
- 標準化を最優先し、その次にデータのクオリティーを統一する登録方法を確立すべき
- 基本的なデータベースの使用方法についての講義を受ける機会がほしい

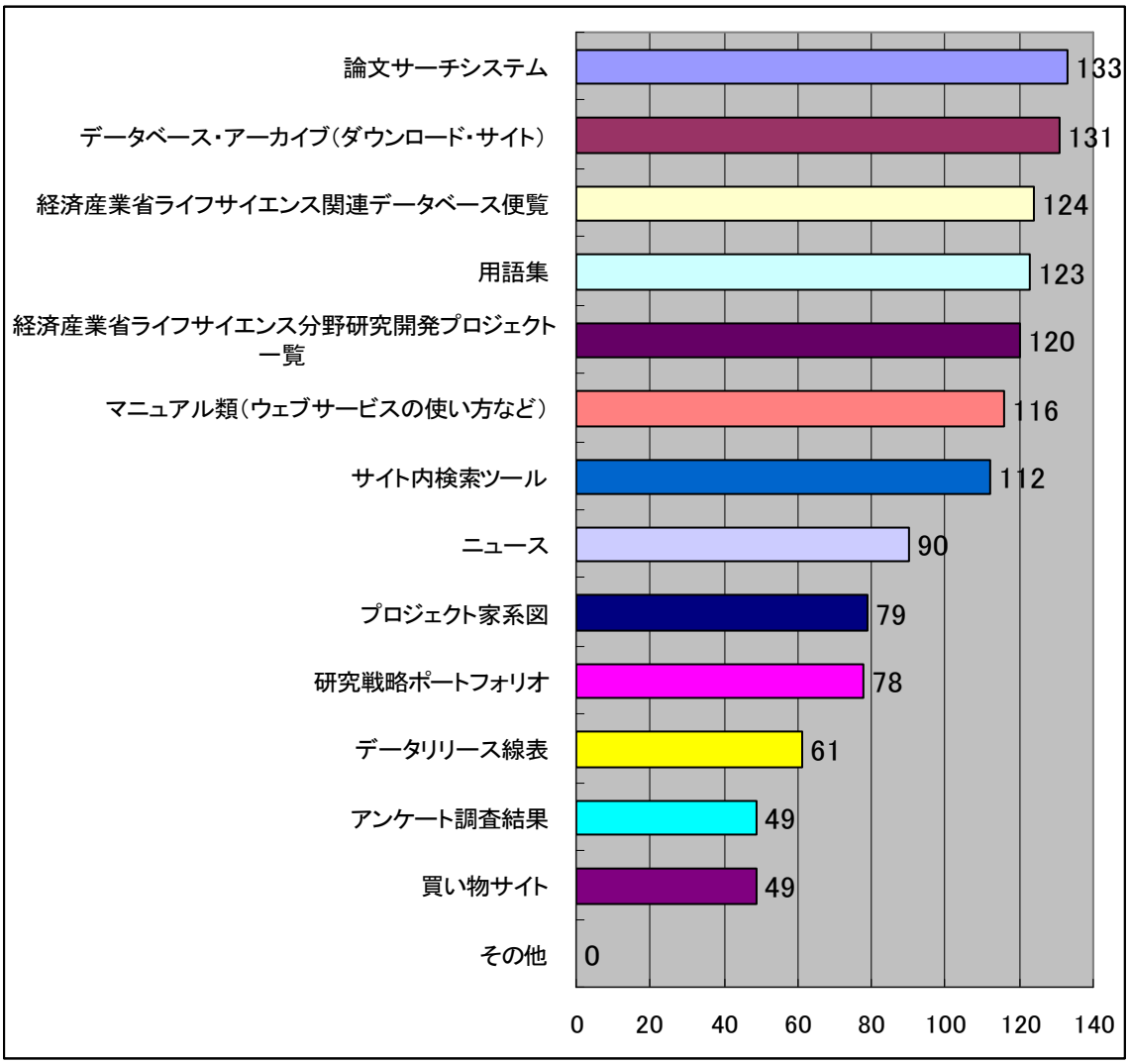


図 2.3.2-4　興味深いポータルサイトコンテンツ

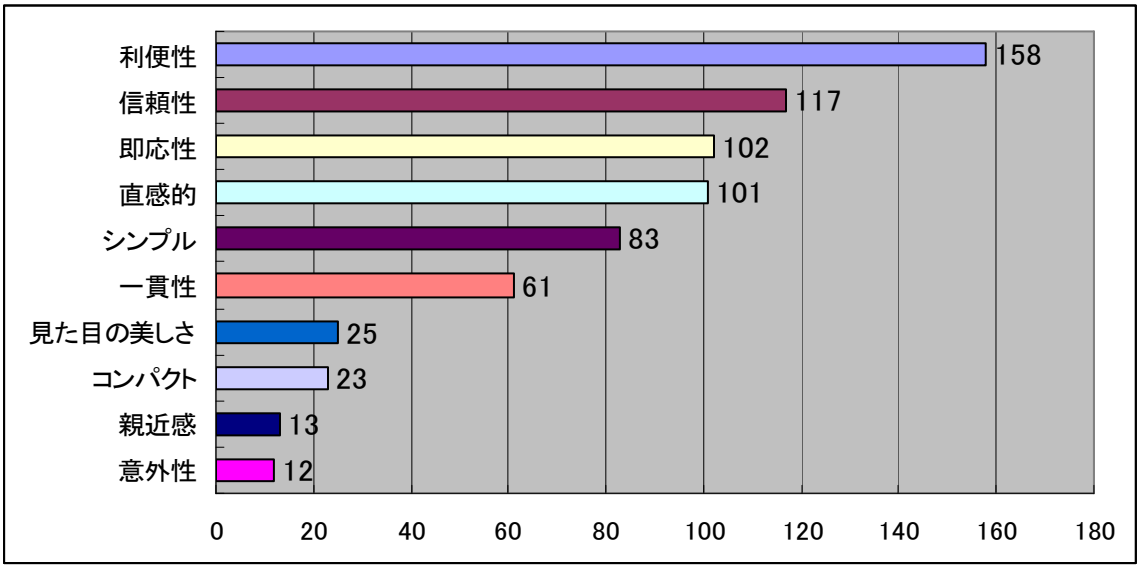


図 2.3.2-5　興味をひかれるポータルサイトのコンセプト

④ 広報活動に関連する調査結果について

情報収集によく利用しているデータベースと H-InvDB、機能性 RNA データベース、糖鎖関連遺伝子データベースについて、それらのサイトを知ったきっかけを調査した結果、研究者同士の情報交換で知る機会が多いことが示された（図 2.3.2-6）。そのため、MEDALS 公開後には学会等での広報活動や講習会の開催を積極的におこなった。また、H-InvDB および連携するデータベースに関しては、存在を知るきっかけとしてアンケートを選択している回答者も目立った。特に各連携データベースでは、本アンケート調査で初めて知ったという回答が約 3 割であり、宣伝活動にも貢献できたと考えられる。

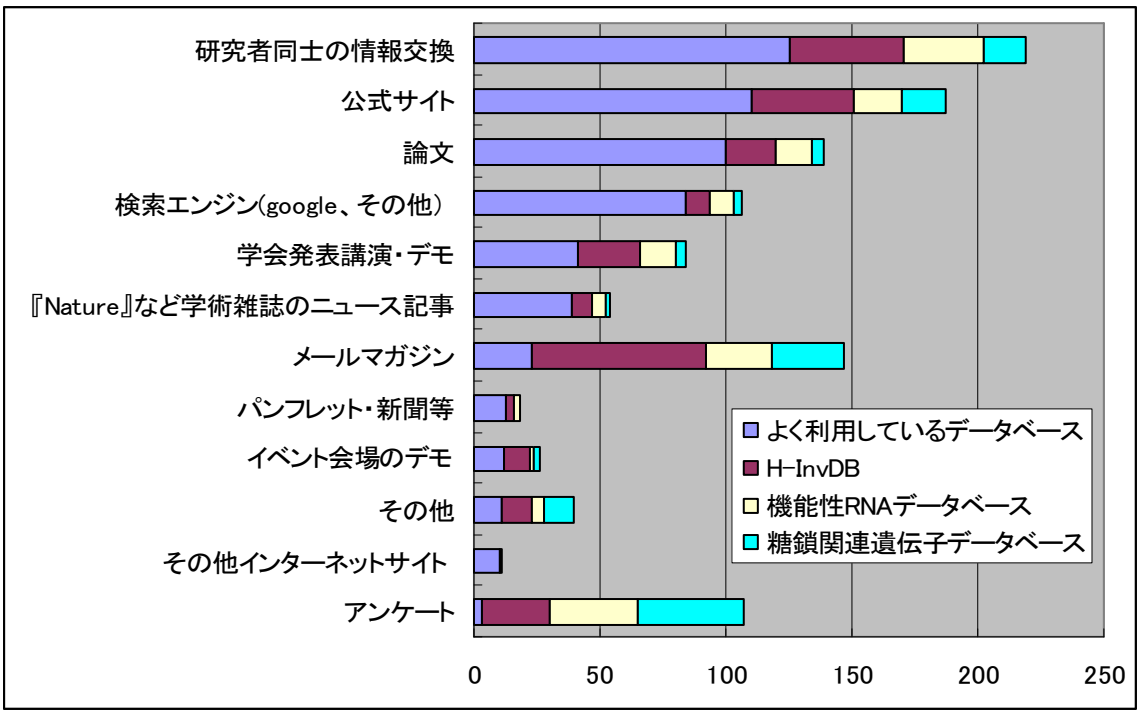


図 2.3.2-6 データベース等を知ったきっかけ

（2） ヒアリング調査

当初の計画には含まれていなかったが、現場の意見を集めることを目的として、ライフサイエンス分野の企業へのヒアリング調査を企画し、実施した。対象は、JBiC プロジェクト参画企業とし、事前に実施した簡単なアンケートの結果をもとに企業を絞り込んだ。さらに、モニター調査として、H-InvDB を利用している創薬ベンチャー企業での意見収集も行い、今年度は合計 5 社を訪問した（表 2.3.2-1）。

ヒアリング調査によって、ウェブアンケート調査だけでは調べきれない内容も聞くことができ、大変有意義であった。また、いただいたコメントはニーズ調査の質問項目の参考にもなった。

表 2.3.2-1 ヒアリング調査

日付	訪問先
2008 年 7 月 14 日	バイオ IT 企業
2008 年 8 月 4 日	製薬企業研究センター
2008 年 8 月 6 日	ゲノム創薬ベンチャー
2008 年 8 月 21 日	発酵系研究所
2008 年 12 月 25 日	創薬ベンチャー

① バイオ IT 企業

【データベース利用の実態】

- 顧客の要望等でデータ調査をおこなう際、少し間違えると迷路に迷ったように再び同じデータに辿り着けなくなることがある

【希望・要望】

- 企業の研究者がライフサイエンスデータベースについてどう思っているか、どれだけ使っているかといった、製薬企業等の意見を知りたい
- アクセス数だけでの評価は望ましくない
- H-InvDB をポータルとすれば良いのではないか

② 製薬企業研究センター

【データベース利用の実態】

- wet 実験とインフォマティクスで分業していて、バイオインフォマティクスの部署では wet 実験の研究者のデータ検索等の相談を受けている
- wet 実験の研究者自身で適切なキーワードを見つけられず、「調べ方がわからないので調べてほしい」という要望の事例が多い
- 文献以外の情報収集に関しては、サーチエンジン（主に Google）や NCBI を利用している

【希望・要望】

- ポータルサイトの入り口には、wet 実験の研究者向けのユーザーフレンドリーな対話形式と、インフォマティクスのヘビーユーザー向けのシンプルな形式の両方があると良い
- 検索式等での情報流出に関しては、公知のデータの扱いは特に制限していないが、化合物の構造式は権利化の可能性もあるため外部に漏れないようにしている

③ ゲノム創薬ベンチャー

【データベース利用の実態】

- バイオインフォマティクスに割く時間は 1 日に 30 分程度
- 日常業務では NCBI をよく閲覧している

【希望・要望】

- 新しいデータベース等の構築情報は、通知や案内がないと分からない
- データベースにも次の作業に役立つ機能が準備されていると便利だと思う

④ 発酵系研究所

【データベース利用の実態】

- 酵素名やタンパク質名を検索語として、関連情報（アミノ酸配列、遺伝子配列、細胞内局在、発現量、生理的な働き等）を調べることが多い
- GenBank 等で入手できる範囲のデータを利用している
- BLAST 等のツールで設定項目が多く、全てのオプションの詳細な説明を見つけられないこともあり、初期設定のまま使うことになる

【希望・要望】

- 組換え微生物での物質生産の取り組みで、工場スケールでの産業利用を試みる場合には経済産業省に届け出る必要があるため（実験室内の場合は文部科学省への届け出）、法的に問題がないことを確認できるような情報がほしい

⑤ 創薬ベンチャー

【データベース利用の実態】

- 抗体医薬開発において、H-InvDB を用いて特定のキーワードを含む膜タンパク質のデータを抽出し、各遺伝子のアノテーション情報を使用している
- 興味のある遺伝子について、H-DBAS を使用して選択的スプライシングバリアント（alternative splicing: AS）を確認している

【希望・要望】

- 発現データの比較でマーカーを探索しているので、H-ANGEL にノーマルの組織だけでなく癌組織等の疾患データも追加してほしい
- 抗体医薬開発のためには、マウスとヒトの結びつきをもっと示してほしい

２．４　他省庁との連携

本プロジェクトは、2005 年に総合科学技術会議の唱えた日本のライフサイエンス統合データベースを目指す取り組みの一つである。文部科学省でもこれを受けて、2006 年秋に「統合データベースプロジェクト」を開始し、4 省庁（経済産業省、文部科学省、厚生労働省、農林水産省）のデータベース統合のための中核団体として、情報･システム研究機構・ライフサイエンス統合データベースセンター（DBCLS）を設立した。そこで本プロジェクトでは、DBCLS と様々な点で連携をしてプロジェクトを実行した。本節では、その連携内容について述べる。

4 月、5 月、11 月に DBCLS にて連携のための会議を行い、連携内容の決定や進め方の議論を行った。第一回会議では、「全体の連携方針」について議論した。文部科学省統合 DB と経済産業省統合 DB の全体の連携の枠組みについて議論し、その後の定期的な会議や技術面の協力、情報交換を行う等の方針を確認した。特に、ポータルサイト構築、ウェブサービス標準化について連携することを決定した。第二回会議では、「ウェブサービスの連携」について、特にウェブサービス標準化について議論した。ここでは、1）文部科学省統合 DB ウェブサービス標準化コミュニティに参加し、共通ガイドライン作成等を連携して作成すること、2）本プロジェクトにおけるウェブサービスの仕様策定の際に、ガイドラインが決定していればなるべくそれに沿うように考慮すること、の 2 つが決定した。第三回会議では、「データベース便覧の共通化」や「著作権等の権利の表示方法」について議論した。ここで、本プロジェクトで調査した便覧内容を DBCLS へ提供した。さらに 2 月にはデータベース 27 件の便覧内容を DBCLS へ提供し、DBCLS の便覧に反映され、便覧内容の共通化をすすめた。

以下に、データベース便覧、ポータルサイトでの検索機能、ウェブサービスの 3 つにおける詳細を述べる。

２．４．１　便覧

データベース便覧内容については、項目と内容の共通化の議論を進め、以下の連携を行った。

（1） 説明項目の共通化

項目については、DBCLS 及び WingPro（独立行政法人・科学技術振興機構：JST のデータベース説明サイト）を参考にして項目候補を洗い出し、その後 DBCLS と協議の上決定した。協議の場では、新たな案としてデータベースとそれを産生したプロジェクトを分けることで一致した。さらに、DBCLS と便覧項目の統一可能性を検討した。すでに便覧作成や開発者への確認を始める関係上、当面は別々にすすめ、段階的に統一を進めることに決定した。2008 年度の下半期の時点で、DBCLS 側が便覧項目について根本的な見直しを行っている。

（2） 記述内容の共通化

本プロジェクトと DBCLS の双方で説明項目が一致する部分については、文言レベルで共通になるように、データの共有化を図った。具体的には、12 月にプロジェクトの詳細を、2 月にはデータベースの詳細を DBCLS へ提供した。この結果、データベースの詳細記述については、27 件のデータベースについて DBCLS のサイトに反映された。

2．4．2　検索

本ポータルサイトでは、「サイト内検索」の機能を持つ計画であった。一方、DBCLS では統合データベースサービスの 1 つとして、主要なデータベースの内容についてキーワードの横断検索を実現していた。そこで本ポータルサイトの「サイト内検索」では、DBCLS の横断検索と親和性を保てるように、同じ検索エンジン（HyperEstraier）を採用した。これにより、本ポータルサイトの内容を DBCLS のサービス上に載せる必要が生じても、容易に実現出来るようになっている。

また、DBCLS ポータルサイトでは、特許データベースを含めた横断検索ができるが、特許についてはライフサイエンス分野以外の特許もヒットする問題があった。そのため、検索対象をライフサイエンス分野の特許に絞り込む方法を検討していた。それに対して本プロジェクトでは、その検索結果に対しての絞り込み方法を検討し、国際特許分類によって絞ることを提案した。具体的には、バイオ系の区分としては C12、C07 に絞り、製薬系の区分としては A の区分に絞ることを提案した。

2．4．3　ウェブサービスの統合化

ウェブサービス（Web サービス）統合化については、下記のように実施した。

（1） 平成 20 年度統合ウェブサービス マニフェスト(20081022 版)の作成

文部科学省統合 DB ウェブサービス標準化コミュニティへ参加し、ウェブサービス共通ガイドライン作成に協力した。

ウェブサービス共通ガイドラインは、

・普遍的に ウェブサービスが守るべき必須事項

・推奨するガイドライン

の2段階の基準を作成する方針で、標準的なメソッド、設計、ドキュメント整備等を提案する。5 月および 9 月に DBCLS で開催された打合せに参加し、推奨ガイドラインのマニフェスト案を作成した。

最終的にはウェブを介して無料で一般に公開される予定であり一部は

http://togows.dbcls.jp/site/services.html から公開されている。

以下に平成 20 年度統合ウェブサービス マニフェスト(20081022 版)を示す。

[2008/10/22 版@統合 WS]

統合ウェブサービス マニフェスト ¶

（参考）http://togows.dbcls.jp/site/services.html

以下、ウェブサービスを提供するにあたり、容易に統合可能なサービスを開発するための指針をマニフェストとして提案します。

- データベース系のサービス提供におけるルール
    - REST での検索とエントリの取得を可能とすること（URL へのマッピングについては TogoWS のガイドラインに準拠）

    - クエリを統一するため、検索キーワードの書式は Common Query Language に準拠すること
    - 検索キーワードからヒット件数が取得できること
    - 検索キーワードからヒットしたエントリの ID のリストが取得できること
    - エントリの ID から実際のエントリが取得可能であること
    - エントリが無かった場合 HTTP 404 Not Found と空のエントリを返すこと
    - REST の URL にはクエリ変数（param1=value1&param2=value2....）は使用しない
    - リストを返す場合は、エントリ単位で改行し、エントリ内はタブ区切りとする
- SOAP サービス全般
    - メソッド毎にマニュアルを書くこと
- パラメータについて、ちゃんと動く値の例を示すこと
    - 少なくとも Java, Perl, Ruby, Python での動作確認を行うこと
- サンプルコードが提供されればベター
    - 非同期サービスはジョブ ID を発行する
    - 扱えないクライアントがあるため、クッキーは使わないこと
    - エラーが発生した場合は SOAP Fault を返す
- 戻り値の型を可能な限り統一すること
    - 配列 INSDC
    - 発現 MIAMI
    - 糖鎖 GLYDE-II
    - 配列アノテーション GFF
    - key – value ペア タブ区切り
- REST にするか SOAP にするか
    - データベースエントリの検索や取得については容易に URL にマッピング可能なので REST で提供することが望ましい。それ以外の解析サービスや実行時間のかかるサービス、さまざまな引数を必要とするサービスを提供する場合は SOAP を用いるとよい。

TogoWS ガイドライン ¶

REST サービス ¶

TogoWS では、以下のルールでエントリの取得や検索を URL にマッピングしました。

- http://togows.dbcls.jp/site/services.html
- http://togows.dbcls.jp/site/rest.html

エントリの取得 ¶

```
http://togows.dbcls.jp/entry/database/entry_id[,entry_id2,...]/field[.format]
```

URL のホスト部にデータベース名などの名前空間が含まれる場合:

- http://frnadb.example.org/entry/frnadb/id の代わりに http://frnadb.example.org/entry/id とする

データベースに組織等の名前空間やバージョン番号を含めたい場合:

- http://togows.dbcls.jp/entry/birc-hinv-5.0/HIT0012345

戻り値について:

基本的には、デフォルトはテキストフォーマットのエントリを返すが、フォーマットを指定したい場合は .xml など拡張子をつけることで指定する。

エントリの検索 ¶

```
http://togows.dbcls.jp/search/database/query+string[.format][/offset,limit[.format]]


http://togows.dbcls.jp/search/database/query+string/count
```

検索文字列 "query string" の書式は Common Query Language のサブセットを使用する。

データ形式や ID の変換 ¶

```
http://togows.dbcls.jp/convert/data_source.format
```

メタデータ ¶

- 対応 DB 一覧

```
http://togows.dbcls.jp/entry/
```

- 対応フィールド名一覧

```
http://togows.dbcls.jp/entry/DB 名/
```

- 対応フォーマット一覧(エントリ ID の一覧は取得できないので代わりに)

```
http://togows.dbcls.jp/entry/DB 名/formats
```

SOAP サービス ¶

- 非同期メソッドについて
    - 非同期メソッド名は Async で終わり、ジョブ ID を返す

- CheckAsyncStatus? メソッドにジョブ ID を渡し、RUNNING, COMPLETED, ERROR のいずれかの文字列を返す
- 非同期メソッドの実行結果を取得するためのメソッド名は「非同期メソッド名＋Result」とし、引数にジョブ ID を渡す

- 少なくとも以下の環境での動作確認を行う
  - Perl: Soap::Lite 0.69
  - Ruby: SOAP4R 1.5.5 (Ruby 1.8.6 bundle)
  - Python: SOAPpy 0.12.0
  - Java: Axis 1.4

（以上）

(2) 平成 20 年度統合ウェブサービス マニフェスト(20081022 版)に準拠した、fRNAdb および GGDB ウェブサービスの開発

2.2.2 で詳細を報告したとおり、機能性 RNA データベース（fRNAdb）および糖鎖関連遺伝子データベース（GGDB）ウェブサービスの開発を平成 20 年度統合ウェブサービス マニフェスト(20081022 版)に準拠して行った。具体的には、データ検索およびデータ取得ウェブサービスプログラム設計、ドキュメントやサンプルコード提供の点についてマニフェストに準拠した。

開発したウェブサービスの URL:

- fRNAdb ウェブサービス

  http://www.ncrna.org/frnadb/api/
- GGDB ウェブサービス

  http://riodb.ibase.aist.go.jp/rcmg/ws/ggdb/

2.５　運営委員会

経済産業省統合データベースプロジェクトの進捗状況を明らかにし、議論を通じて外部有識者（アドバイザー）らの意見を成果物に反映させることを目的として、年度内に 2 回、運営委員会を開催した。運営委員会では、まず実施テーマごとにプロジェクトの進捗報告がなされ、それを受けてアドバイザー等からさまざまな意見が述べられ、最後に総合討論が行われた。また、産業界におけるニーズ調査の結果や、ポータルサイト MEDALS へのユーザーのアクセス状況なども報告された。

開催日時

第一回　2008 年 9 月 12 日（金）9 時 30 分―12 時

第二回　2009 年 2 月 3 日（月）13 時―17 時

アドバイザー名

伊藤隆司（東京大学大学院新領域創成科学研究科）

長洲毅志（エーザイ株式会社 CEO オフィス）

3. 成果発表

（1） 論文

1. Imanishi T and Nakaoka H (2009) Hyperlink Management System and ID Converter System: enabling maintenance-free hyperlinks among major biological databases. Nucleic Acids Research (Web Server Issue) (submitted.)

（2） 発表

1. Murakami K, Sato Y, Gojobori T, and Imanishi T(2008) Searching PubMed articles queried by multiple articles. 16th Annual International Conference Intelligent Systems for Molecular Biology. (ISMB 2008/カナダ トロント)
2. 松矢 明宏、村上 勝彦、富所 布紗乃、中岡 源、五十嵐 由美子、今西 規、五條堀 孝(2008)「MEDALS：経済産業省ライフサイエンスデータベースのポータルサイト」JBIC2008 プロジェクト研究成果報告会（品川）
3. 中岡 源、今西 規(2008)「持続可能なデータベース統合のためのリンク自動管理システムとID 一括変換システムの開発」JBIC2008 プロジェクト研究成果報告会（品川）
4. 村上 勝彦、佐藤 慶治、今西 規、五條堀 孝(2008)「複数文献を入力とする PubMed サーチシステム」JBIC2008 プロジェクト研究成果報告会（品川）
5. 富所 布紗乃、山崎 千里、佐藤 慶治、中岡 源、羽原 拓哉、武田 淳一、五條堀 孝、今西 規(2008)「大量の転写産物の一括アノテーションのための自動解析システム SuperTACT」JBIC2008 プロジェクト研究成果報告会（品川）
6. 村上 勝彦(2008)「経済産業省関連ライフサイエンスデータベース・ポータルサイト(medals.jp)の構築」JBIC2008 プロジェクト研究成果報告会（品川）
7. 村上 勝彦(2008)「経産省統合データベースポータルサイト MEDALS の紹介」平成 20 年度 第 1 回講習会「ヒト遺伝子統合データベース H-InvDB の活用法」（大阪）
8. 松矢 明宏、村上 勝彦、富所 布紗乃、中岡 源、五十嵐 由美子、今西 規、五條堀 孝(2008)「MEDALS：経済産業省ライフサイエンスデータベースのポータルサイト」BMB2008（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/神戸）
9. 中岡 源、今西 規(2008)「持続可能なデータベース統合のためのリンク自動管理システムとID 一括変換システムの開発」BMB2008（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/神戸）
10. 村上 勝彦、佐藤 慶治、今西 規、五條堀 孝(2008)「複数文献を入力とする PubMed サーチシステム」BMB2008（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/神戸）
11. 富所 布紗乃、山崎 千里、佐藤 慶治、中岡 源、羽原 拓哉、武田 淳一、五條堀 孝、今西 規(2008)「大量の転写産物の一括アノテーションのための自動解析システム SuperTACT」BMB2008（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/神戸）
12. 村上 勝彦(2008)「経産省統合データベースポータルサイト MEDALS の紹介」平成 20 年度 第 2 回講習会「ヒト遺伝子統合データベース H-InvDB と経産省統合データベースポータルサイト MEDALS の活用法」

（台場）

13. 村上 勝彦、松矢 明宏、富所 布紗乃、中岡 源、五十嵐 由美子、今西 規、五條堀 孝(2009)「MEDALS：経済産業省ライフサイエンスデータベースのポータルサイト」平成 20 年度第 8 回産総研・産技連 LS－BT 合同研究発表会（つくば）
14. 今西 規、五條堀 孝、五十嵐 由美子(2009)「経済産業省ライフサイエンス関連プロジェクトのポータルサイト MEDALS の公開」産総研ホームページ
15. 村上 勝彦(2009)「経産省統合データベースポータルサイト MEDALS の紹介」平成 20 年度 第 3 回講習会「ヒト遺伝子統合データベース H-InvDB と経産省統合データベースポータルサイト MEDALS の活用法」（つくば）
16. 村上 勝彦(2009)「経産省統合データベースポータルサイト MEDALS の紹介」平成 20 年度 第 4 回講習会「ヒト遺伝子統合データベース H-InvDB と経産省統合データベースポータルサイト MEDALS の活用法」（福岡）

（3） データベース公開

2008 年

10 月 29 日	経済産業省ライフサイエンス関連プロジェクトのポータルサイトMEDALS 一部情報の先行公開
12 月 12 日	機能性 RNA 配列データベース fRNAdb ウェブサービス公開
12 月 25 日	ヒト遺伝子統合データベース H-InvDB ウェブサービス拡張版公開

2009 年

1 月 30 日	選択的スプライシング(alternative splicing: AS)バリアントデータベース H-DBAS リリース 4.0 更新
2 月 6 日	経済産業省ライフサイエンス関連プロジェクトのポータルサイトMEDALSの公開
3 月	糖鎖関連遺伝子データベース GGDB（GlycoGene DataBase） ウェブサービス公開予定

4. 謝辞

本プロジェクトの実施にあたり、多数の方のご協力、ご貢献を頂いた。ここに深く感謝申し上げます。(敬称略)

- 文部科学省統合データベースとの連携

  高木　利久　　文部科学省 ライフサイエンス統合データベースセンター

  西川　哲夫　　文部科学省 ライフサイエンス統合データベースセンター

- 便覧作成協力

  各データベース・解析ツール開発者の方々

- 民間企業でのデータベースニーズ調査

  社団法人 バイオ産業情報化コンソーシアム会員企業(5 社)

- プロジェクト運営の補助

  野田　彰子　　産業技術総合研究所　バイオメディシナル情報研究センター

  坂手　龍一　　産業技術総合研究所　バイオメディシナル情報研究センター

5. 参加者名簿

役職	氏名	所属機関
プロジェクトリーダー	五條堀　孝	産業技術総合研究所
		バイオメディシナル情報研究センター(併任)
サブリーダー	今西　規	産業技術総合研究所
		バイオメディシナル情報研究センター
	村上　勝彦	社団法人 バイオ産業情報化コンソーシアム
研究員	松矢　明宏	社団法人 バイオ産業情報化コンソーシアム(出向)
	富所　布紗乃	社団法人 バイオ産業情報化コンソーシアム(出向)
	中岡　源	社団法人 バイオ産業情報化コンソーシアム(出向)
	山崎　千里	産業技術総合研究所
		バイオメディシナル情報研究センター
	武田　淳一	産業技術総合研究所
		バイオメディシナル情報研究センター
	佐藤　慶治	産業技術総合研究所
		バイオメディシナル情報研究センター
技術員	羽原　拓哉	産業技術総合研究所
		バイオメディシナル情報研究センター
	遠藤　智宏	産業技術総合研究所
		バイオメディシナル情報研究センター)
	五十嵐　由美子	社団法人 バイオ産業情報化コンソーシアム

研究協力機関

独立行政法人 産業技術総合研究所 生命情報工学研究センター

特に経済産業省関連機関から産生されるデータベース等の統合において

浅井　潔

光山　統泰

独立行政法人 産業技術総合研究所 糖鎖医工学研究センター

特に経済産業省関連機関から産生されるデータベース等の統合において

成松　久

鹿内　俊秀

国立大学法人北海道大学大学院　情報科学研究科

学校法人 慶應義塾大学 医学部

国立大学法人東京大学大学院情報理工学研究科

大学共同利用機関法人 国立遺伝学研究所生命情報･DDBJ 研究センター

運営委員（アドバイザー）

伊藤　隆司	東京大学　大学院新領域創成科学研究科
長洲　毅志	エーザイ株式会社

＜参加企業＞

株式会社　シーズ･ラボ

株式会社　ダイナコム

日立公共システムエンジニアリング株式会社